LiBMap: a large-scale simulation-derived design–performance dataset for lithium-ion batteries

Anonymous authors
Paper under double-blind review

Abstract

Lithium-ion batteries (LIBs) play an important role in the development of new energy technologies, yet their design still relies heavily on costly trial-and-error experiments and simulations. Artificial intelligence has shown considerable potential for solving complex design problems, but its application to LIB design is constrained by the scarcity of design–performance data. To address this limitation, we construct a large-scale design–performance dataset for LIB design—the first of its kind, to the best of our knowledge. The dataset integrates cell format, geometric dimensions, material systems, kinetic parameters, and porous-electrode design parameters into a unified design space. Using material balance calculations and a coupled electrochemical–thermal model, we generate approximately 770,000 battery designs and corresponding performance data through large-scale parallel computing. Each sample contains over one hundred design parameters and 29 performance metrics covering typical operating scenarios in energy storage, electric vehicles, and consumer electronics, together with complete electrical and thermal process data for each metric. The dataset provides complete design information, comprehensive performance characterization, and fine-grained process data. Furthermore, we develop a retrieval–generation framework that uses k-nearest neighbor search to retrieve local data and fits the retrieved data using radial basis functions. The framework requires no large-scale neural network training and has low computational cost. For both LIB performance prediction and inverse design, a single task can be completed in tens of milliseconds; relative to simulation results, the mean absolute percentage error (MAPE) for both tasks is below 5%. We believe this work will provide a data foundation for LIB design and further advance LIB design capabilities.

1 Introduction

Lithium-ion batteries (LIBs) are widely used in energy storage, electric vehicles, and consumer electronics. Because different application scenarios impose different requirements on energy, power, and other performance metrics, cell designs must be tailored to specific applications (Kench et al., 2024; Wu et al., 2018). This work focuses on cell design and investigates how design choices affect performance metrics such as power, direct current resistance (DCR), and fast-charging time under different operating conditions. This problem involves a large number of design variables at three levels: cell format and geometric dimensions, electrode structure, and material systems together with their kinetic characteristics. These variables are interdependent and, together with operating conditions, determine cell performance across multiple performance dimensions, making cell design highly complex (Kench et al., 2024; Wu et al., 2018; Appiah et al., 2024; Ma et al., 2024).

Experiments and simulations are the primary means of evaluating cell designs. However, cell prototyping and testing are time-consuming and costly, while large-scale simulation-based searches require repeated solutions of complex physical models, resulting in substantial computational cost (Kench et al., 2024; Wu et al., 2018; Ma et al., 2024). Industrial research and development (R&D) expenditure also reflects the resource-intensive nature of battery development. For example, Contemporary Amperex Technology Co., Limited (CATL) reported RMB 5.845 billion in material costs for R&D in 2024, accounting for approximately 31.41% of its total R&D expenditure (CATL, 2025). Improving the efficiency of design screening and reducing repeated prototyping and simulation are therefore important for lowering development costs and resource consumption.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Advances in artificial intelligence (AI) for protein structure prediction (Jumper et al., 2021), materials discovery (Merchant et al., 2023), and chip placement (Mirhoseini et al., 2021) have demonstrated the potential of AI to solve complex scientific and engineering problems. In cell design, neural networks have been explored for performance prediction, and generative methods have been explored for electrode optimization (Kench et al., 2024; Wu et al., 2018). However, the applicability and generalization of models to cell design depend strongly on how well their training data cover the design space (Ma et al., 2024). Supporting broader cell-design research therefore requires data resources that adequately cover the design space and systematically characterize the design–performance relationships between design variables and cell performance.

For long-term battery performance issues such as cycle aging and lifetime prediction, a substantial body of data, methods, and tools has been developed, including BatteryLife (Tan et al., 2025), Discovery Learning (Zhang et al., 2026), and BatteryML (Zhang et al., 2024). In contrast, research on the design–performance relationships between cell designs and short-term performance under specified operating conditions remains relatively limited. Existing studies have explored the mapping from electrode design to energy and power (Kench et al., 2024; Wu et al., 2018), while others have used simulations to investigate the performance of specific cylindrical cells (Pang et al., 2024). However, these studies differ in their coverage of the design space and the performance metrics considered, and remain insufficient to support broader cell-design research. Cell design across different application scenarios therefore requires broader design-space coverage and a more systematic framework for characterizing short-term performance.

To address these limitations, we use large-scale parallel computing to construct LiBMap, a large-scale design–performance dataset for LIB design. LiBMap broadens both design-space coverage and the scope of performance characterization. On the design side, approximately 770,000 cell designs span multiple cell formats and representative geometric dimensions, different material systems and their representative kinetic parameters, and different porous-electrode design parameters. On the performance side, LiBMap establishes a performance characterization framework comprising 29 metrics for typical operating scenarios in energy storage, electric vehicles, and consumer electronics, enabling systematic comparison across different cell designs. In addition, each performance metric is associated with complete electrical and thermal process data under the corresponding operating conditions, allowing performance results to be analyzed in conjunction with the corresponding operating profiles.

With LiBMap as the primary contribution of this work, we further develop a retrieval–generation framework that requires no large-scale neural network training, and evaluate LiBMap's utility for cell design through two tasks: performance prediction and inverse design. In our evaluation, the average completion time is on the order of tens of milliseconds for both tasks, while the mean absolute percentage error (MAPE) relative to simulation results remains below 5% for both. These results demonstrate that LiBMap can provide a data foundation for efficient and accurate cell design.

2 Related Work

2.1 Lithium-Ion Battery Design Methods

Existing lithium-ion battery design methods can be broadly categorized into experimental, simulation-based, and data-driven approaches. Experimental approaches investigate the effects of design parameters on battery performance by fabricating and testing batteries with different designs. Zheng et al. (2012), Heubner et al. (2019), and Ogihara et al. (2015) investigated the effects of electrode thickness and porosity on performance metrics such as energy and power. Experiments provide direct measurements of actual battery performance, but sample fabrication and testing are time-consuming and costly, making it difficult to explore large design spaces.

Simulation-based approaches can reduce the need for extensive sample fabrication and testing while providing access to internal states that are difficult to measure experimentally. Previous studies have used porous-electrode models to investigate electrode-thickness effects (Appiah et al., 2024; Xu et al., 2019) and to optimize electrode thickness and porosity (Srinivasan & Newman, 2004) and particle-size-distribution models to investigate the effects of active-material particle size distributions on capacity and local current distributions (Röder et al., 2016). However, exploring design spaces

through simulation typically requires repeated solution of high-fidelity physical models, resulting in substantial computational cost (Pang et al., 2024).

Data-driven approaches learn design–performance relationships from experimental or simulation data, providing an alternative route for battery design. Previous studies have employed neural networks, generative models, and related methods for performance prediction, electrode-structure generation, and design optimization (Kench et al., 2024; Wu et al., 2018; Ma et al., 2024; Gao & Lu, 2020; Kim et al., 2020). For example, Kench et al. (2024) combined generative models with Bayesian optimization for electrode-structure design, while Kim et al. (2020) used a response surface model to optimize separator and electrode parameters. Once trained or constructed, these models can typically evaluate candidate designs rapidly at relatively low computational cost. However, their applicability depends strongly on how well the training data cover the design space and performance metrics of interest. Kench et al. (2024) restricted model application to the region covered by the training data, while Ma et al. (2024) also noted the limited extrapolation capability of neural networks beyond the training domain and emphasized the need to expand the range of design parameters. Therefore, extending data-driven approaches to broader battery design problems still requires design–performance data with wider design-space coverage and more comprehensive performance characterization.

2.2 Lithium-Ion Battery Data Resources

Existing lithium-ion battery data resources can be broadly divided into three categories: material and electrode data, operational and aging data, and battery design–performance data. The first category primarily describes material compositions and properties as well as electrode structures (Kench et al., 2024; Ward et al., 2022). For example, Dave et al. (2020) released formulations of 251 aqueous electrolytes together with properties such as ionic conductivity and pH, while Usseglio-Viretta et al. (2018) released three-dimensional structural data for 14 electrode samples. These data provide an important foundation for battery material and electrode-structure research, but they primarily characterize materials and electrode structures rather than the cross-scale battery design problem considered here.

The second category primarily records current, voltage, temperature, and related measurements from batteries or battery packs during operation and aging. BatteryLife, BatteryML, and ENPOLITE aggregate or provide access to multi-source battery aging data (Tan et al., 2025; Zhang et al., 2024; Dechent et al., 2021); Dubarry & Beck (2020) generated synthetic aging curves using mechanistic models; and Liu et al. (2025) released three years of operational data from 300 electric vehicles. These resources encompass experimental, synthetic, and real-world operational data and primarily characterize battery or battery-pack operation and aging. They can support tasks such as state estimation and lifetime prediction, but generally lack complete battery design information and are therefore difficult to use directly for battery design research.

The third category describes design–performance relationships between battery design parameters and performance and is most closely related to the problem considered in this work. Pang et al. (2024) used models of two 18650 batteries to generate voltage and temperature data under different thermophysical parameters and operating conditions. However, existing data of this type are generally limited to specific batteries and a restricted set of design variables. Overall, these data resources still lack design–performance data that simultaneously cover cross-scale design variables and systematically characterize battery performance.

3 LiBMap Dataset

3.1 Data Generation

Designing lithium-ion batteries requires specifying coupled design parameters across the material, porous-electrode, and cell scales to meet performance requirements under multiple constraints (Ma et al., 2024; Richardson et al., 2022). LiBMap targets three common applications: energy storage, electric vehicles (EVs), and consumer electronics (3C). To balance design-space coverage and computational cost, we construct a representative design space from key variables at each scale. High-throughput material-balance calculations followed by coupled P2D–lumped thermal simulations yield application-specific performance metrics and the corresponding electrical and thermal process data (Figure 1(a)).

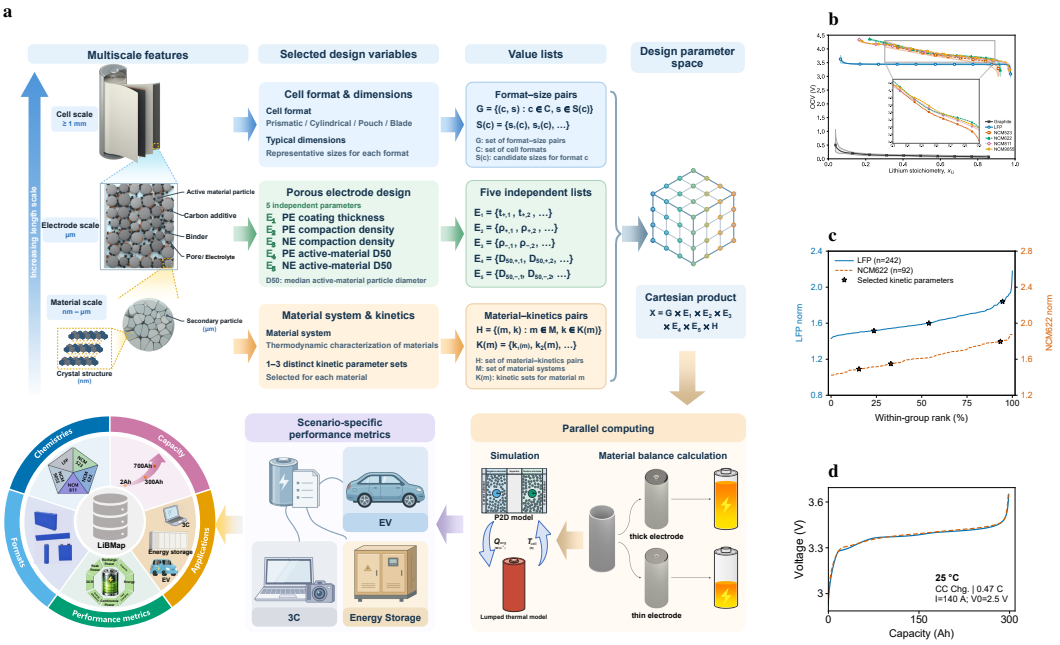


Figure 1: Data generation workflow. PE and NE denote the positive and negative electrodes, respectively; subscripts + and − identify the corresponding electrodes in (a). (a) Materials with distinct kinetics are selected within each material system, and representative dimensions are selected for each cell format. For each material–cell-model combination, the Cartesian product of the sampled value sets of the five porous-electrode design parameters defines the design space. High-throughput material-balance calculations followed by coupled P2D–lumped thermal simulations yield detailed design specifications, performance metrics, and electrical, thermal, and polarization process data, with metrics and test conditions specified for energy-storage, EV, and 3C applications. The schematics depict NCM positive electrode secondary particles and crystal structure, and illustrate how different electrode designs yield different capacities at fixed external cell dimensions. "1–3 sets" denotes the initial selection; LFP-4 was subsequently added to the LFP–graphite system. (b) For the 13 material sets, positive electrode open-circuit potential curves are grouped by the five materials, while all negative electrode curves form one graphite group. Within the degree-of-lithiation range shared by all curves in each group, piecewise linear interpolation gives potentials at common lithiation values. Lines show pointwise arithmetic means, and shading shows pointwise minimum–maximum ranges, not confidence intervals. (c) Kinetic-score rankings within LFP and NCM622 are shown, with stars marking the initially selected material sets. (d) Experimental (solid blue) and simulated (dashed orange) capacity–voltage curves are compared for LFP-1 during constant-current charging at 25 °C and approximately 0.47C.

Design space across scales. We select design variables at three levels: material properties, porous-electrode structure, and cell geometry. The material level includes five material systems, with LFP, NCM523, NCM622, NCM811, and NCM9055 as the positive electrode materials and graphite as the negative electrode material in all cases, covering mainstream commercial material systems. To capture kinetic differences among materials within each system, we select 4, 3, 3, 2, and 1 material sets with distinct kinetic characteristics from the five systems, respectively, yielding 13 sets in total. The NCM523 sets include both single-crystal and polycrystalline materials. Figure 1(b) shows the open-circuit potential curves for the 13 material sets, with positive electrode curves grouped by the five materials and all negative electrode curves forming one graphite group. Curves within each group are generally similar but not identical. Over the degree-of-lithiation range shared by all curves within each group, lines show pointwise means and shaded bands show minimum–maximum ranges. The LFP and NCM622 systems have relatively large pools of candidate materials. From 242 and 92 material sets with different kinetic characteristics, respectively, we select three sets per system based jointly on their within-system kinetic scores and the agreement between experimental results and simulations using the corresponding material parameters, covering different reaction and transport characteristics (Figure 1(c)). For the remaining systems, all available material sets are included be-

cause the candidate pools are small. Initially selected material sets are numbered in ascending order of their kinetic scores within each system; the subsequently added LFP set is denoted LFP-4. The kinetic score is constructed from seven reaction and transport parameters and is used only for ranking and selection within the same material system. The parameters and scoring method are detailed in Appendix A.4.

Porous electrodes link material properties to cell performance. We select five design variables: positive electrode coating thickness, the compaction densities of the positive and negative electrodes, and the active-material particle sizes D50 of the positive and negative electrodes. Their ranges depend on the material system and whether the positive electrode material is single-crystal or polycrystalline (Table 2). We extend these ranges beyond conventional design ranges to improve coverage and explore extreme designs. At the cell-geometry level, a survey of commercial cells informs the selection of 33 widely used cell models (Table 1), spanning four formats—prismatic, blade, cylindrical, and pouch—with at least five models per format. Energy-storage applications place stringent demands on cycle life and cost. LFP's advantages in safety and cycle life make it an important choice for energy storage (Hesse et al., 2017). We therefore restrict the 10 energy-storage cell models to the LFP material system, while the remaining 23 EV and 3C cell models can be paired with all five material systems. Material system and kinetics, as well as cell format and dimensions, are interdependent: kinetic characteristics must be specified within a particular material system, and dimensions must be defined for a given cell format.

Table 1: Representative cell dimensions and reference capacities for different applications.

ES		EV		EV and 3C	
Dimensions (mm)	**Capacity (Ah)**	**Dimensions (mm)**	**Capacity (Ah)**	**Dimensions (mm)**	**Capacity (Ah)**
Prismatic		**Prismatic**		**Cylindrical**	
174 × 207 × 54	176–230	**148 × 101 × 27**	**15–55**	18 × 65	1–4
173 × 207 × 71	**280 or 314**	135 × 220 × 30	72–80	21 × 70	2.5–6
275 × 215 × 74	530–587	130 × 200 × 36	90–105	**46 × 80**	**30**
352 × 207 × 72	**560–720**	160 × 118 × 50	100–102	**46 × 95**	**35**
182 × 470 × 54	625	200 × 173 × 33	125	46 × 135	30
Blade		**194 × 113 × 50**	**130–210**	**Pouch**	
300 × 112 × 50	214	173 × 170 × 48	120–152	199 × 140 × 11	52
412 × 145 × 57	270	200 × 207 × 33	150	227 × 161 × 10	30–55
684 × 239 × 34	680	**Blade**		202 × 133 × 10	15–30
580 × 215 × 19	317	960 × 90 × 13.5	135 or 202	272 × 181 × 16	75
506 × 215 × 26	350	961.6 × 122 × 27.4	145	301 × 100 × 12	50
—		574 × 118 × 22	198	—	
—		905 × 118 × 13.5	202	—	
—		510 × 120 × 16	124	—	

Notes: Dimensions are given as width × height × thickness (W × H × T) for non-cylindrical cells and diameter × height (D × H) for cylindrical cells, all in mm; capacities are in Ah. EV, 3C, and ES denote electric-vehicle, consumer-electronics, and energy-storage cells, respectively. The tabulated capacities are reference values; the thermodynamic capacity of each design in the dataset is determined by material-balance calculations. Bold entries identify cell models used for dense sampling. The six material–cell-model combinations are LFP-3 prismatic energy-storage, 173 × 207 × 71; LFP-1 prismatic energy-storage, 352 × 207 × 72; NCM811-1 cylindrical, 46 × 80; LFP-4 prismatic EV, 194 × 113 × 50; NCM9055-1 cylindrical, 46 × 95; and NCM622-3 prismatic EV, 148 × 101 × 27.

Parameter sampling and dataset size. For each material–cell-model combination, we construct Cartesian-product grids over the five electrode design variables using equally spaced values including both range endpoints. The sparse and dense grids contain $5 \times 5 \times 5 \times 3 \times 3 = 1{,}125$ and $20 \times 20 \times 20 \times 3 \times 3 = 72{,}000$ points, respectively. Positive electrode coating thickness and the compaction densities of both electrodes affect electrode thickness, changing active-material loading within a fixed-size casing and hence thermodynamic capacity; these variables therefore receive denser sampling. Active-material particle sizes D50 of both electrodes do not affect thermodynamic capacity in the current material-balance calculation and retain three sampled values each. Low-rate parameter sweeps in Appendix B show small, smooth capacity changes with particle size, supporting this allocation. Sparse sampling provides broad coverage, while dense sampling further resolves the design-performance relationships of representative combinations. Dense combinations pair material systems whose thermodynamic and kinetic parameters permit faster model solution and better convergence with representative cell models further selected from the 33 models.

The sparse grid yields 344,250 records: 310,500 from the initial 12 material sets paired with 23 EV and 3C cell models, and 33,750 from LFP-1 through LFP-3 paired with 10 energy-storage models. The six dense combinations listed in the notes to Table 1 contribute 72,000 records each, totaling 432,000; LFP-4 is used only in one dense combination. Removing duplicate designs at grid endpoints leaves 775,890 records.

Material balance and performance simulation. Each design first undergoes material-balance calculations using cell dimensions, material formulations, processing parameters, and assembly constraints to determine electrode geometry, available thermodynamic capacity, cell mass, and other quantities. The resulting detailed design specification contains more than 100 parameters describing cell structure and material composition (Table 3). Thermodynamic capacity and effective electrode

area are used to convert C-rate into current and areal current density for subsequent coupled pseudo-two-dimensional (P2D)–lumped thermal simulations.

Table 2: Ranges and sampling counts for porous-electrode design parameters.

Design variable	Range		Samples per variable	
	LFP–graphite	NCM–graphite	Dense	Sparse
Positive electrode single-side coating thickness (µm)	[60, 80]	[30, 65]	20	5
Positive electrode compaction density (g cm^{-3})	[2.3, 2.8]	[3.3, 3.7]	20	5
Negative electrode compaction density (g cm^{-3})	[1.3, 1.75]	[1.3, 1.75]	20	5
Positive electrode particle diameter, D_{50} (µm)	[0.6, 1.5]	[3, 8] / [0.2, 0.8]	3	3
Negative electrode particle diameter, D_{50} (µm)	[8, 16]	[8, 16]	3	3
Total combinations	—	—	72 000	1 125

Notes: LFP and NCM denote the positive electrode materials lithium iron phosphate and lithium nickel–cobalt–manganese oxides, respectively. D50 is the median active-material particle size. Positive electrode coating thickness is the coating thickness on one side of the current collector. For NCM positive electrode particle sizes, ranges before and after the slash apply to single-crystal and polycrystalline materials, respectively; their other parameter ranges are identical. Each of the five variables is sampled at equally spaced values including both range endpoints. The Cartesian product of the five sampled value sets gives 1,125 sparse-grid and 72,000 dense-grid design points.

Table 3: Categories and counts of material-balance parameters for four cell formats.

Cell format	Shared parameter categories	Format-specific parameter categories	Parameter count
Blade	Material formulations and properties Coating structure and processing states Electrode capacity balance Separator, electrolyte and current collectors Material quantities and cell properties Capacity, energy and energy densities	Stacking and assembly geometry Tabs, case and end caps	142
Cylindrical		Winding and assembly geometry Full-tab design, case and end caps	145
Prismatic		Winding and assembly geometry Tab geometry and layout, case and top cap	162
Pouch		Stacking and assembly geometry Tabs and pouch-film packaging	140

Notes: Parameter counts include both input parameters and calculated results. The capacities and energies in the table are obtained from material-balance calculations and are distinct from simulation results under specific operating conditions.

The P2D model captures diffusion within particles, porous-electrode transport, and interfacial reactions; the lumped thermal model describes mean cell temperature and its effects on electrochemical processes (Richardson et al., 2022; Bernardi et al., 1985; Gu & Wang, 2000). This coupled electrochemical–thermal model balances the representation of key mechanisms against computational cost while capturing design-performance relationships at the material and porous-electrode levels. The effects of cell geometry are represented primarily through material-balance calculations, whose capacity, area, mass, and other results influence the coupled model's performance response. Experimental benchmarking of all 13 material parameter sets indicates high model accuracy. Figure 1(d) shows a representative comparison, and Appendix E collects 228 sets of experimental and simulated curves. During porous-electrode and cell-geometry sweeps, each material set's thermodynamic and kinetic parameterization remains fixed; design differences enter through the five porous-electrode design variables and material-balance results. Model and parameterization details are given in Appendices A.1–A.3.

For the different applications, we define eight categories of performance metrics: discharge capacity, discharge energy, DC resistance (DCR), peak discharge power, continuous discharge power, Recharge Power, fast-charging time, and energy efficiency. We specify conditions at different temperatures, states of charge (SOC), or C-rates according to the characteristics of each metric category, yielding 29 performance metrics in total. EV and 3C applications use 28 metrics across the seven categories other than energy efficiency. Energy-storage applications use 11 metrics across four categories: discharge capacity, discharge energy, DCR, and energy efficiency (Table 4). Each record corresponds to a complete battery design, including design parameters across scales, detailed material-balance results, and performance metrics together with electrical, thermal, and polarization process data under different operating conditions. For a very small number of operating conditions, simulations fail to converge or the associated data are missing.

3.2 Design-Performance Relationships and Process Responses

Figure 2(a,b) illustrates statistical design trends obtained by selecting the valid records with the best target performance in each material–cell-model group and summarizing the within-group normalized positions of their five electrode design parameters (Appendix C.2). Designs maximizing 60 s continuous discharge power at 25 °C and 50% SOC (hereafter, power; Figure 2(a)) tend to favor thinner positive electrode coatings and lower positive electrode compaction density. Designs maximizing discharge capacity at 25 °C and 1/3C (hereafter, capacity; Figure 2(b)) instead favor greater positive electrode coating thickness and higher positive electrode compaction density. Both favor higher negative electrode compaction density and smaller active-material particle sizes D50 in both electrodes.

Table 4: Cell performance metrics and simulation conditions for different applications.

Metric	Application	Temperature (°C)	SOC (%)	Simulation protocol
Discharge capacity	EV, 3C, ES	25, 45	—	1/3C, 1C CC discharge.
Discharge energy	EV, 3C, ES	25, 45	—	1/3C, 1C CC discharge.
DC resistance	EV, 3C, ES	25	80 50	10 s CC discharge pulses: 3C at 80% SOC; 2C at 50% SOC.
Continuous discharge power	EV, 3C	25, 45	20, 50, 80	Maximum CP discharge power for 60 s.
Peak discharge power	EV, 3C	25, 45	20, 50, 80	Maximum CP discharge power for 10 s.
Recharge Power	EV, 3C	−10, 25, 45	90	Maximum CP charging power for 5 s at 90% SOC.
Fast-charging time	EV, 3C	−10, 25, 45	20 → 80	12 CC stages, 5% SOC each; maximize each stage current without Li plating.
Energy efficiency	ES	25	—	Fully charge at 0.5P, rest for 30 min, then discharge at 0.5P.

Notes: EV, 3C, and ES denote electric-vehicle, consumer-electronics, and energy-storage applications; SOC, CC, CP, and DCR denote state of charge, constant current, constant power, and DC resistance. The application label 3C differs from a 3C current rate. "—" means SOC is not separately specified; 20→80 denotes charging from 20% to 80% SOC. Each row uses all combinations of its listed temperatures, SOC values, and C-rates, one metric per combination, except DCR, which uses only (80% SOC, 3C) and (50% SOC, 2C) at 25 °C. EV and 3C each use 28 metrics; ES uses 11, covering 29 distinct metrics overall.
Capacity and energy are obtained from the same CC discharge simulation. DCR is the terminal-voltage drop from immediately before the pulse to 10 s into discharge divided by the discharge pulse current. Continuous and peak discharge power are the maximum constant discharge powers that can be maintained for 60 s and 10 s, respectively, under applicable constraints. Recharge Power is the maximum constant charging power that can be maintained for 5 s at 90% SOC while satisfying the model's lithium-plating constraint. Fast charging divides 20%–80% SOC into 12 equal stages of five percentage points each; the maximum constant charging current satisfying the model's lithium-plating constraint is found separately for each stage. For 0.5P, $P_{\mathrm{set}} = 0.5 Q_{\mathrm{MB}} U_{\mathrm{nom}}/(1\,\mathrm{h})$, where Q_{MB} is the material-balance capacity and U_{nom} is the nominal voltage. Energy efficiency is discharged energy divided by charged energy for the sequence of 0.5P CP charging, a 30 min rest, and 0.5P CP discharging at 25 °C.
Temperature, voltage, and charging-related lithium-plating cutoffs are applied according to operating conditions. A simulation terminates when any applicable cutoff is reached: the temperature reaches 80 °C; the voltage reaches the upper limit during charging or the lower limit during discharging (LFP: 2.5–3.65 V; NCM: 3.0–4.2 V); or, during charging, the negative electrode potential versus Li/Li$^+$ at the separator–negative electrode interface falls to 0 V or below, triggering the lithium-plating-risk cutoff.

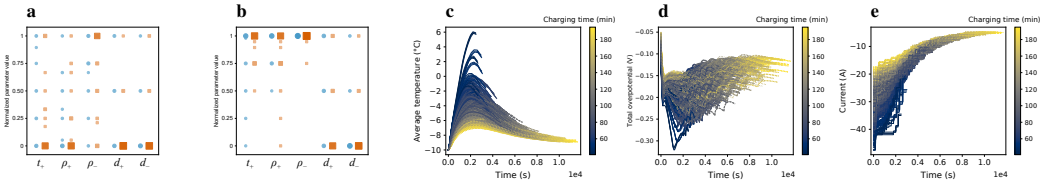


Figure 2: Distributions of porous-electrode parameters in designs maximizing continuous discharge power or discharge capacity. Panels (a,b) summarize the five electrode design parameters of designs maximizing the corresponding target metric within each material–cell-model group: (a) 60 s continuous discharge power at 25 °C and 50% SOC (276 valid groups); (b) discharge capacity at 25 °C and 1/3C (307 valid groups). In (a,b), t_+ denotes positive electrode coating thickness; ρ_+ and ρ_- denote positive and negative electrode compaction densities; and d_+ and d_- denote the corresponding active-material median particle sizes (D_{50}). Each parameter is normalized by its observed within-group range, with 0 and 1 denoting the lower and upper bounds. Blue circles and orange squares denote LFP and NCM; marker size and color intensity indicate occurrence counts for each parameter value across optimal designs. Parameters are summarized separately, without retaining their combinations within individual designs. Panels (c–e) show mean temperature, total overpotential, and charging current, respectively, for the LFP-3 prismatic EV cell (W × H × T = 148 × 101 × 27 mm) during multistage constant-current charging at −10 °C. Charging from 20% to 80% SOC uses 12 equal stages of five percentage points each, with the maximum constant charging current found under the model's lithium-plating constraint for each stage. Both total overpotential and current follow the convention that charging is negative and discharging is positive. Colors from dark blue to yellow indicate increasing total charging time; solid, dashed, and dotted lines denote short-, intermediate-, and long-duration groups defined by the tertile thresholds of 91.71 and 124.05 min.

These statistical trends reflect a trade-off between thermodynamic capacity and kinetic processes. At low C-rates, greater active-material loading within a fixed-size casing benefits capacity (Appendix B). Thinner positive electrodes and higher porosity help alleviate ionic transport limitations but may reduce active-material loading (Heubner et al., 2019). The opposing preferences for positive electrode coating thickness and compaction density in Figure 2(a,b) are consistent with these mechanisms and reflect a capacity–power trade-off, in line with previous studies of the energy–power trade-off (Zheng et al., 2012; Heubner et al., 2019). Smaller active-material particles can shorten solid-state diffusion paths (Zheng et al., 2012; Richardson et al., 2022) without changing thermodynamic capacity in the current material-balance calculation (Appendix B), which can explain the shared preference for smaller particles. These shared preferences suggest prioritizing investigation of higher negative electrode compaction density and smaller particles in both electrodes when seeking to improve both capacity and power. Such statistical trends can narrow the design search space. Although single-metric preferences are relatively clear, practical design remains a complex opti-

mization problem involving multiple performance objectives and constraints (Ma et al., 2024) and coupled design variables across scales (Richardson et al., 2022).

The dataset also provides electrical and thermal process data. Figure 2(c–e) shows mean temperature, total overpotential, and charging current, respectively, during multistage constant-current fast charging of an LFP-3 prismatic EV cell ($148 \times 101 \times 27$ mm) from 20% to 80% SOC at $-10\,^{\circ}$C. Across electrode designs, shorter fast-charging times are typically associated with larger charging-current and total-overpotential magnitudes and higher mean temperatures. These curves show the evolution of current, temperature, and polarization under the present model and charging strategy. Together with the operating potentials, heat generation, and polarization decomposition data in Appendix C.1, they support comparisons of designs in terms of mechanisms and processes.

These examples illustrate LiBMap's design-performance relationships and detailed electrical and thermal process data; further results are provided in Appendix C. We have released the 72,000-point dense subset for the NCM811-1 cylindrical 46800 cell and plan to expand the public release incrementally to support AI-based battery design. Data and code access instructions are provided in Appendix F.

4 Lithium-Ion Battery Design Tasks

Lithium-ion battery design seeks designs satisfying performance requirements and design constraints.

4.1 Forward Performance Prediction

For design x, forward prediction yields $y = \mathcal{F}(x; u)$, where u specifies operating conditions and the initial state. Outputs may be scalar metrics or spatiotemporal fields. Partially specified designs require specifying the remaining parameters' values, ranges, or distributions and may yield nonunique predictions. This task evaluates existing designs against requirements without generating new ones.

4.2 Inverse Design

Inverse design seeks designs satisfying performance requirements and constraints:

$$x \in \mathcal{X}, \qquad y = \mathcal{F}(x; u), \qquad h(x, y) = 0, \qquad g(x, y) \leq 0.$$

Here, $\mathcal{X}$ is the design-variable feasible domain; h and g encode equality and inequality constraints, respectively, on designs and scalar metrics or fields. Interval bounds become inequalities. Inverse design generally balances competing objectives through constrained multiobjective optimization; feasible designs may be absent or nonunique.

5 The Retrieval–Generation Method

Retrieval–generation (RG) fits local surrogates to retrieved neighbors for prediction or optimization (Chang et al., 2025; Figure 3(a)). Simplifying the tasks in Section 4, we fix the material system and cell model; five scalar porous electrode design parameters map to scalar performance metrics under specified operating conditions, excluding response trajectories and fields.

Retrieved designs satisfying requirements are returned directly. Otherwise, retrieval using the Canberra distance in normalized design or performance space supplies forward or inverse neighborhoods, respectively, for local rescaling, standardization, and fitting. All reported RG experiments include local fitting. Using expanded data, we compare sampling, dataset size, neighborhood size, surrogate models and their parameters, and baseline candidates (Appendix D). Balancing accuracy, computational cost, and data generation cost, we select UG500 ($5 \times 5 \times 5 \times 2 \times 2 = 500$ records), $K = 200$, and a radial basis function (RBF) surrogate with a linear radial kernel and 10^{-6} smoothing. CatBoost and a standard multilayer perceptron (MLP) are selected as tree-based and neural global baselines, respectively.

On 1,000 common LFP-3 test designs, RG achieves mean absolute percentage errors (MAPEs) of **0.054%–0.926%** across four metrics on UG500 (Figure 3(b–d); Appendix D). Capacity, energy,

and energy efficiency MAPEs are comparable to MLP's; direct-current resistance (DCR) MAPE is higher (0.926% versus 0.513%), but both methods remain below 1% on all metrics. Without global pretraining, RG averages **3.52 ms** per forward query in independent serial timing (Figure 3(h)).

Forward accuracy supports inverse design with the same configuration. L-BFGS-B starts from the retrieved design closest to target performance and minimizes the maximum relative target deviation across predicted metrics within neighborhood design-variable bounds. Generated designs are independently re-evaluated with the coupled P2D–lumped thermal model; performance deviations from targets quantify inverse accuracy.

Three distinct cell models using LFP-3, NCM622-3, and NCM811-1 materials have separate UG500 databases, with 1,000 cases per task and cell (Figure 3(e–g)). All per-metric MAPEs are below 5%. Requiring every metric's error below 5% gives forward and inverse pass rates of **94.6%–100%** and **96.5%–100%**, respectively. Inverse queries average **40.47 ms** for LFP-3 in independent serial timing (Figure 3(i)).

RG requires no global retraining after data updates. After filtering by material system and cell model, adding or deleting records from other cell types has no effect if the target subset remains unchanged. These results illustrate an application of LiBMap: a small UG500 dataset for a fixed material–cell combination enables accurate millisecond predictions and rapid candidate design generation at low data generation cost. However, RG remains a local surrogate constrained by data coverage, making it difficult to learn a general design-performance relationship.

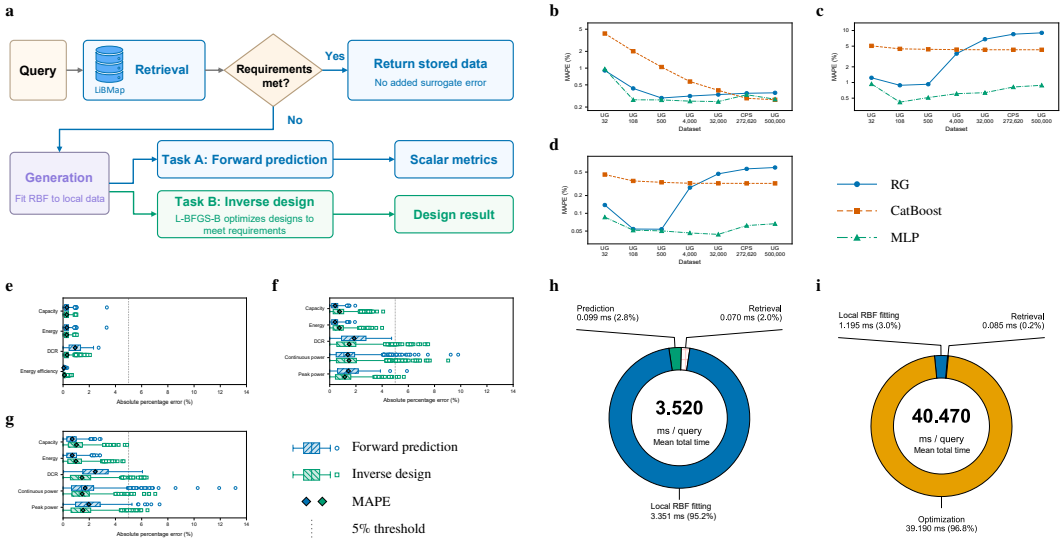


Figure 3: Retrieval–generation: workflow, accuracy, and efficiency. (a) Workflow: return matching designs directly; otherwise fit a local surrogate for forward prediction or inverse design. (b–d) Dataset/model comparison for the LFP-3 cell in (e): (b) energy; (c) DCR; (d) energy efficiency. Six uniform-grid (UG) and one change-point-sampled (CPS) dataset supply retrieval for RG and training for both baselines; all share 1,000 test designs. Lines connect categorical datasets; MAPE axes are logarithmic. RG uses $\min(200, N)$ neighbors (N: dataset size). Conditions at 25 °C: (b) 1/3C discharge; (c) 50% SOC, 2C, 10 s discharge; (d) 0.5P charge/discharge. Capacity and implementation details: Appendix D.1–D.3. (e–g) Forward/inverse errors: (e) LFP-3 prismatic, 173 × 207 × 71 mm; (f) NCM622-3 prismatic, 148 × 101 × 27 mm (W × H × T); (g) NCM811-1 cylindrical, 46 × 80 mm (D × H). Each cell: UG500, 1,000 cases/task; inverse performance independently recomputed with the coupled P2D–lumped thermal model. Boxes show quartiles and medians; whiskers extend to the most extreme observations within 1.5 interquartile ranges of the box, with all outliers shown. Diamonds denote MAPE; dotted lines mark 5%. (h,i) Serial query time: (h) forward prediction; (i) inverse design. LFP-3, UG500, $K = 200$; i7-14700HX, single process; numerical libraries and FAISS single-threaded. After warm-up: 1,000 cases × 5 rounds. Excludes database/index preparation and physical re-evaluation.

6 Conclusion

We introduced LiBMap, a large-scale design–performance dataset for lithium-ion battery design. Building on LiBMap, we proposed retrieval–generation (RG) for millisecond-scale forward prediction and inverse design, with MAPEs below 5% against simulation results, demonstrating LiBMap's utility for efficient battery design. LiBMap is limited by its design space, performance characterization, and physics-based data-generation models, while RG depends on LiBMap's coverage. Overall, LiBMap provides a data foundation for data-driven battery design.

AI Use Statement

Generative AI tools were used to assist with language editing, translation, and improving the readability and organization of the manuscript. All scientific content, methodology, experiments, results, and conclusions were developed and verified by the authors. The authors reviewed all AI-assisted content and take full responsibility for the final manuscript.

REFERENCES

Williams Agyei Appiah, Dohwan Kim, Yong Min Lee, Juan Maria Garcia-Lastra, and Ivano E. Castelli. Continuum insight into the effects of electrode design parameters on the electrochemical performance of lithium-ion batteries. *Electrochimica Acta*, 494:144470, 2024. doi: 10.1016/j.electacta.2024.144470.

D. Bernardi, E. Pawlikowski, and J. Newman. A General Energy Balance for Battery Systems. *Journal of The Electrochemical Society*, 132(1):5–12, 1985. doi: 10.1149/1.2113792. URL https://doi.org/10.1149/1.2113792.

Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995. doi: 10.1137/0916069. URL https://doi.org/10.1137/0916069.

Liming Chang, Zhixuan Wu, Hongkai Zhao, Weinan Zhou, and Zhe Li. Efficient design method for batteries based on data retrieval and generation. In Ronghai Qu, Zhengxiang Song, Zhiming Ding, Gang Mu, Rui Xiong, and Li Han (eds.), *Proceedings of the 1st Electrical Artificial Intelligence Conference, Volume 2*, volume 1395 of *Lecture Notes in Electrical Engineering*, pp. 483–494, Singapore, 2025. Springer Nature Singapore. doi: 10.1007/978-981-96-4063-8_47.

Contemporary Amperex Technology Co., Limited. 2024 Annual Report, 2025. URL https://static.cninfo.com.cn/finalpage/2025-03-15/1222806982.PDF.

Adarsh Dave, Jared Mitchell, Kirthevasan Kandasamy, Han Wang, Sven Burke, Biswajit Paria, Barnabás Póczos, Jay Whitacre, and Venkatasubramanian Viswanathan. Autonomous discovery of battery electrolytes with robotic experimentation and machine learning. *Cell Reports Physical Science*, 1(12):100264, 2020. doi: 10.1016/j.xcrp.2020.100264.

Philipp Dechent, Alexander Epp, Dominik Jöst, Yuliya Preger, Peter M. Attia, Weihan Li, and Dirk Uwe Sauer. ENPOLITE: Comparing lithium-ion cells across energy, power, lifetime, and temperature. *ACS Energy Letters*, 6(6):2351–2355, 2021. doi: 10.1021/acsenergylett.1c00743.

Marc Doyle, Thomas F. Fuller, and John Newman. Modeling of Galvanostatic Charge and Discharge of the Lithium/Polymer/Insertion Cell. *Journal of The Electrochemical Society*, 140(6):1526–1533, 1993. doi: 10.1149/1.2221597. URL https://doi.org/10.1149/1.2221597.

Matthieu Dubarry and David Beck. Big data training data for artificial intelligence-based Li-ion diagnosis and prognosis. *Journal of Power Sources*, 479:228806, 2020. doi: 10.1016/j.jpowsour.2020.228806.

Thomas F. Fuller, Marc Doyle, and John Newman. Simulation and Optimization of the Dual Lithium Ion Insertion Cell. *Journal of The Electrochemical Society*, 141(1):1–10, 1994. doi: 10.1149/1.2054684. URL https://doi.org/10.1149/1.2054684.

Tianhan Gao and Wei Lu. Physical model and machine learning enabled electrolyte channel design for fast charging. *Journal of The Electrochemical Society*, 167(11):110519, 2020. doi: 10.1149/1945-7111/aba096.

W. B. Gu and C. Y. Wang. Thermal-Electrochemical Modeling of Battery Systems. *Journal of The Electrochemical Society*, 147(8):2910–2922, 2000. doi: 10.1149/1.1393625. URL https://doi.org/10.1149/1.1393625.

Holger C. Hesse, Michael Schimpe, Daniel Kucevic, and Andreas Jossen. Lithium-ion battery storage for the grid—a review of stationary battery storage system design tailored for applications in modern power grids. *Energies*, 10(12):2107, 2017. doi: 10.3390/en10122107.

Christian Heubner, Alexander Nickol, Jann Seeba, Sebastian Reuber, Nils Junker, Mareike Wolter, Michael Schneider, and Alexander Michaelis. Understanding thickness and porosity effects on the electrochemical performance of $LiNi_{0.6}Co_{0.2}Mn_{0.2}O_2$-based cathodes for high energy Li-ion batteries. *Journal of Power Sources*, 419:119–126, 2019. doi: 10.1016/j.jpowsour.2019.02.060.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

Steve Kench, Isaac Squires, Amir Dahari, Ferran Brosa Planella, Scott A. Roberts, and Samuel J. Cooper. Li-ion battery design through microstructural optimization using generative AI. *Matter*, 7(12):4260–4269, 2024. doi: 10.1016/j.matt.2024.08.014.

Ji-San Kim, Dong-Chan Lee, Jeong-Joo Lee, and Chang-Wan Kim. Optimization for maximum specific energy density of a lithium-ion battery using progressive quadratic response surface method and design of experiments. *Scientific Reports*, 10(1):15586, 2020. doi: 10.1038/s41598-020-72442-4.

Hongao Liu, Chang Li, Xiaosong Hu, Jinwen Li, Kai Zhang, Yang Xie, Ranglei Wu, and Ziyou Song. Multi-modal framework for battery state of health evaluation using open-source electric vehicle data. *Nature Communications*, 16(1):1137, 2025. doi: 10.1038/s41467-025-56485-7.

Xiao-Ying Ma, Wen-Ke Zhang, Ying Yin, Kailong Liu, and Xiao-Guang Yang. Multi-objective optimization of lithium-ion battery designs considering the dilemma between energy density and rate capability. *Energy and AI*, 18:100416, 2024. doi: 10.1016/j.egyai.2024.100416.

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023. doi: 10.1038/s41586-023-06735-9.

Azalia Mirhoseini, Anna Goldie, Mustafa Yazgan, Joe Wenjie Jiang, Ebrahim Songhori, Shen Wang, Young-Joon Lee, Eric Johnson, Omkar Pathak, Azade Nova, Jiwoo Pak, Andy Tong, Kavya Srinivasa, William Hang, Emre Tuncer, Quoc V. Le, James Laudon, Richard Ho, Roger Carpenter, and Jeff Dean. A graph placement methodology for fast chip design. *Nature*, 594(7862):207–212, 2021. doi: 10.1038/s41586-021-03544-w.

Nobuhiro Ogihara, Yuichi Itou, Tsuyoshi Sasaki, and Yoji Takeuchi. Impedance spectroscopy characterization of porous electrodes under different electrode thickness using a symmetric cell for high-performance lithium-ion batteries. *The Journal of Physical Chemistry C*, 119(9):4612–4619, 2015. doi: 10.1021/jp512564f.

Yiheng Pang, Yun Wang, and Zhiqiang Niu. Deep learning from three-dimensional lithium-ion battery multiphysics model part I: Data development. *Energy and AI*, 18:100428, 2024. doi: 10.1016/j.egyai.2024.100428.

G. W. Richardson, J. M. Foster, R. Ranom, C. P. Please, and A. M. Ramos. Charge transport modelling of Lithium-ion batteries. *European Journal of Applied Mathematics*, 33(6):983–1031, 2022. doi: 10.1017/S0956792521000292. URL `https://doi.org/10.1017/S0956792521000292`.

Fridolin Röder, Sören Sonntag, Daniel Schröder, and Ulrike Krewer. Simulating the impact of particle size distribution on the performance of graphite electrodes in lithium-ion batteries. *Energy Technology*, 4(12):1588–1597, 2016. doi: 10.1002/ente.201600232.

Venkat Srinivasan and John Newman. Design and optimization of a natural graphite/iron phosphate lithium-ion cell. *Journal of The Electrochemical Society*, 151(10):A1530–A1538, 2004. doi: 10.1149/1.1785013.

Ruifeng Tan, Weixiang Hong, Jiayue Tang, Xibin Lu, Ruijun Ma, Xiang Zheng, Jia Li, Jiaqiang Huang, and Tong-Yi Zhang. BatteryLife: A comprehensive dataset and benchmark for battery life prediction. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 5789–5800. Association for Computing Machinery, 2025. doi: 10.1145/3711896.3737372.

Francois L. E. Usseglio-Viretta, Andrew Colclasure, Aashutosh N. Mistry, Koffi Pierre Yao Claver, Fezzeh Pouraghajan, Donal P. Finegan, Thomas M. M. Heenan, Daniel Abraham, Partha P. Mukherjee, Dean Wheeler, Paul Shearing, Samuel J. Cooper, and Kandler Smith. Resolving the discrepancy in tortuosity factor estimation for Li-Ion battery electrodes through micro-macro modeling and experiment. *Journal of The Electrochemical Society*, 165(14):A3403–A3426, 2018. doi: 10.1149/2.0731814jes.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2. URL `https://doi.org/10.1038/s41592-019-0686-2`.

Logan Ward, Susan Babinec, Eric J. Dufek, David A. Howey, Venkatasubramanian Viswanathan, Muratahan Aykol, David A. C. Beck, Benjamin Blaiszik, Bor-Rong Chen, George Crabtree, Simon Clark, Valerio De Angelis, Philipp Dechent, Matthieu Dubarry, Erica E. Eggleton, Donal P. Finegan, Ian Foster, Chirranjeevi Balaji Gopal, Patrick K. Herring, Victor W. Hu, Noah H. Paulson, Yuliya Preger, Dirk Uwe-Sauer, Kandler Smith, Seth W. Snyder, Shashank Sripad, Tanvir R. Tanim, and Linnette Teo. Principles of the Battery Data Genome. *Joule*, 6(10):2253–2271, 2022. doi: 10.1016/j.joule.2022.08.008.

Bin Wu, Sangwoo Han, Kang G. Shin, and Wei Lu. Application of artificial neural networks in design of lithium-ion batteries. *Journal of Power Sources*, 395:128–136, 2018. doi: 10.1016/j.jpowsour.2018.05.040.

Meng Xu, Benjamin Reichman, and Xia Wang. Modeling the effect of electrode thickness on the performance of lithium-ion batteries with experimental validation. *Energy*, 186:115864, 2019. doi: 10.1016/j.energy.2019.115864.

Han Zhang, Xiaofan Gui, Shun Zheng, Ziheng Lu, Yuqi Li, and Jiang Bian. BatteryML: An open-source platform for machine learning on battery degradation. In *International Conference on Learning Representations*, 2024. URL `https://arxiv.org/abs/2310.14714`.

Jiawei Zhang, Yifei Zhang, Baozhao Yi, Yao Ren, Qi Jiao, Hanyu Bai, Weiran Jiang, and Ziyou Song. Discovery Learning predicts battery cycle life from minimal experiments. *Nature*, 650 (8100):110–115, 2026. doi: 10.1038/s41586-025-09951-7.

Honghe Zheng, Jing Li, Xiangyun Song, Gao Liu, and Vincent S. Battaglia. A comprehensive understanding of electrode thickness effects on the electrochemical performances of Li-ion battery cathodes. *Electrochimica Acta*, 71:258–265, 2012. doi: 10.1016/j.electacta.2012.03.161.

A Coupled P2D–Lumped Thermal Model and Its Parameterization

A.1 Overview of the Coupled P2D–Lumped Thermal Model

The pseudo-two-dimensional (P2D) model describes charge and ion transport in porous electrodes along the electrode-thickness coordinate x, together with solid-phase lithium diffusion along the radial coordinate r of active-material particles. By solving these processes in a coupled manner, the model determines the solid-phase lithium concentration, electrolyte concentration, solid-phase and electrolyte potentials, interfacial reaction current, and cell voltage (Doyle et al., 1993; Fuller et al., 1994). In the data-generation workflow of this study, the P2D model maps material properties, electrode structures, and operating conditions to voltage, capacity, and the evolution of internal states.

The P2D model is further coupled with a lumped thermal model. Heat generated by the electrochemical processes determines the cell temperature, which in turn affects solid-phase diffusion, electrolyte transport, interfacial reaction kinetics, and equilibrium potentials, thereby forming a bidirectional electrochemical–thermal coupling (Bernardi et al., 1985; Gu & Wang, 2000).

The model assumes that the electrodes and separator are homogeneous in the in-plane directions and that the active material can be represented by spherical particles. The electrolyte is assumed to be locally electroneutral, and convection is neglected. Only single-electron lithium intercalation/deintercalation reactions are considered. Double-layer effects, side reactions, aging, mechanical deformation, and phase separation are neglected. The thermal model adopts a lumped-temperature assumption, under which the temperature is spatially uniform throughout the cell and internal temperature gradients are neglected.

A.2 Governing Equations and Boundary Conditions

The electrode-thickness coordinate is defined from the negative-electrode current collector toward the positive-electrode current collector. Discharge is defined as $I > 0$, with $i_{\mathrm{app}} = I/A$. A positive interfacial current density $j_k > 0$ denotes delithiation/oxidation. For $k \in \{n, p\}$, the specific surface area of spherical particles is

$$a_k = \frac{3\varepsilon_{s,k}}{R_k},$$

and the volumetric reaction current density is $\mathcal{J} = a_k j_k$, with $\mathcal{J} = 0$ in the separator.

The local lithium stoichiometry within an active-material particle is defined as

$$\theta_k(r, x, t) = \frac{c_{s,k}(r, x, t)}{c_{s,k}^{\max}},$$

and the solid-phase diffusion coefficient is evaluated as

$$D_{s,k} = D_{s,k}(\theta_k, T).$$

Solid-phase lithium diffusion within the active-material particles is governed by

$$\frac{\partial c_{s,k}}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2 D_{s,k}(\theta_k, T)\frac{\partial c_{s,k}}{\partial r}\right],$$

subject to the boundary conditions

$$\left.\frac{\partial c_{s,k}}{\partial r}\right|_{r=0} = 0, \qquad -D_{s,k}(\theta_k, T)\left.\frac{\partial c_{s,k}}{\partial r}\right|_{r=R_k} = \frac{j_k}{F}.$$

Neglecting double-layer effects, charge conservation in the solid and electrolyte phases and Ohm's law in the solid phase are given by

$$\frac{\partial i_{s,k}}{\partial x} = -a_k j_k, \qquad \frac{\partial i_e}{\partial x} = \mathcal{J}, \qquad i_s + i_e = i_{\mathrm{app}},$$

$$i_{s,k} = -\sigma_k^{\mathrm{eff}} \frac{\partial \phi_{s,k}}{\partial x}.$$

The effective electronic conductivity σ_k^{eff} is provided directly as a model parameter for each electrode.

Using a lithium-reference potential convention, the ionic flux and species conservation in the electrolyte are written as

$$N_+ = -D_e^{\mathrm{eff}} \frac{\partial c_e}{\partial x} + \frac{t_+^0 i_e}{F}, \qquad \frac{\partial(\varepsilon_e c_e)}{\partial t} = -\frac{\partial N_+}{\partial x} + \frac{\mathcal{J}}{F},$$

while the electrolyte current under concentrated-solution theory is

$$i_e = -\kappa^{\mathrm{eff}} \frac{\partial \phi_e}{\partial x} + \frac{2RT\kappa^{\mathrm{eff}}}{F}(1 - t_+^0)\chi \frac{\partial \ln c_e}{\partial x},$$

where

$$\chi = 1 + \left(\frac{\partial \ln f_\pm}{\partial \ln c_e} \right)_T$$

(Richardson et al., 2022, see Eqs. (2.61) and (4.15), pp. 1000 and 1017). The effective diffusivity and ionic conductivity are obtained from the corresponding intrinsic transport properties through the Bruggeman relation (Richardson et al., 2022, Secs. 4.2.2 and 4.3, pp. 1016–1017). In the negative electrode, separator, and positive electrode,

$$D_{e,\ell}^{\mathrm{eff}} = D_e \varepsilon_{e,\ell}^{b_\ell}, \qquad \kappa_\ell^{\mathrm{eff}} = \kappa \varepsilon_{e,\ell}^{b_\ell}, \qquad \ell \in \{n, s, p\},$$

where b_ℓ denotes the Bruggeman coefficient of the corresponding region. The separator Bruggeman coefficient b_s is treated as a fixed model parameter and is not included among the seven kinetic features used for kinetic parameter-set selection.

The surface lithium stoichiometry of the active-material particles is defined as

$$\theta_k^{\mathrm{surf}} = \frac{c_{s,k}^{\mathrm{surf}}}{c_{s,k}^{\mathrm{max}}}.$$

The interfacial overpotential and symmetric Butler–Volmer kinetics are then given by

$$\eta_k = \phi_{s,k} - \phi_e - U_k(\theta_k^{\mathrm{surf}}, T),$$

$$j_k = 2 j_{0,k} \sinh\left(\frac{F\eta_k}{2RT} \right),$$

where the exchange current density used in the data-generation model is

$$j_{0,k} = F k_k^{\mathrm{cfg}}(\theta_k^{\mathrm{surf}}, T) \sqrt{c_e \, c_{s,k}^{\mathrm{surf}} \left(1 - \frac{c_{s,k}^{\mathrm{surf}}}{c_{s,k}^{\mathrm{max}}} \right)}.$$

Here, $k_k^{\mathrm{cfg}}(\theta_k^{\mathrm{surf}}, T)$ is the reaction rate constant provided by the kinetic dataset, with units of $\mathrm{m\,s^{-1}}$. Its dependence on surface lithium stoichiometry and temperature is obtained directly from the corresponding tabulated data.

For the boundary conditions, the solid-phase current density at the negative- and positive-electrode current-collector boundaries is set to the corresponding applied current density and vanishes at the electrode/separator interfaces. At the external electrolyte boundaries, $i_e = N_+ = 0$. Across internal interfaces containing the same electrolyte, c_e, N_+, ϕ_e, and i_e are continuous. Taking

$$\phi_{s,n}(0,t) = 0,$$

the voltage predicted directly by the P2D model is

$$V_{\mathrm{P2D}}(t) = \phi_{s,p}(L,t) - \phi_{s,n}(0,t).$$

V_{P2D} does not include the resistive voltage drop across the tabs, connection plates, or other electrical interconnections. The terminal voltage used for comparison with experimental measurements is therefore

$$V_{\mathrm{term}} = V_{\mathrm{P2D}} - IR_{\mathrm{conn}},$$

where R_{conn} denotes the equivalent resistance of these electrical interconnections.

The initial solid-phase lithium concentration, electrolyte concentration, and temperature are determined by the initial state of charge, the initial electrolyte state, and the environmental or experimental conditions, respectively.

The lumped thermal model follows a Bernardi-type energy balance (Bernardi et al., 1985; Gu & Wang, 2000),

$$C_{\mathrm{th}}\frac{\mathrm{d}T}{\mathrm{d}t} = Q_{\mathrm{irr}} + Q_{\mathrm{rev}} + Q_{\mathrm{conn}} - hA_{\mathrm{cool}}(T - T_\infty),$$

where the electrochemical irreversible heat-generation rate is

$$Q_{\mathrm{irr}} = I\left(U_{\mathrm{OCV}} - V_{\mathrm{P2D}}\right),$$

the reversible heat-generation rate is

$$Q_{\mathrm{rev}} = -IT\frac{\partial U_{\mathrm{OCV}}}{\partial T},$$

and the ohmic heat-generation rate in the tabs, connection plates, and other electrical interconnections is

$$Q_{\mathrm{conn}} = I^2 R_{\mathrm{conn}}.$$

For the Bernardi-type thermal model, the volume-averaged solid-phase lithium concentration over the active material of electrode k is defined as

$$\bar{c}_{s,k}(t) = \frac{\displaystyle\int_{\Omega_k} \varepsilon_{s,k}(x)\left[\frac{3}{R_k^3}\int_0^{R_k} c_{s,k}(x,r,t)r^2\,\mathrm{d}r\right]\mathrm{d}x}{\displaystyle\int_{\Omega_k}\varepsilon_{s,k}(x)\,\mathrm{d}x}.$$

The corresponding average lithium stoichiometry is

$$\bar{\theta}_k = \frac{\bar{c}_{s,k}}{c_{s,k}^{\max}}.$$

The open-circuit voltage is evaluated from the average lithium stoichiometries of the two electrodes as

$$U_{\mathrm{OCV}} = U_p(\bar{\theta}_p, T) - U_n(\bar{\theta}_n, T),$$

with

$$\frac{\partial U_{\mathrm{OCV}}}{\partial T} = S_p(\bar{\theta}_p) - S_n(\bar{\theta}_n),$$

where

$$S_k(\theta) = \left(\frac{\partial U_k}{\partial T}\right)_{\theta}.$$

Because V_{P2D} excludes the resistive voltage drop across the electrical interconnections, Q_{conn} is included separately and is not double-counted in Q_{irr}. Equivalently,

$$Q_{\mathrm{irr}} + Q_{\mathrm{conn}} = I\left(U_{\mathrm{OCV}} - V_{\mathrm{term}}\right).$$

The present implementation therefore accounts for heat generation through electrochemical irreversible heat, reversible heat, and ohmic heat from the electrical interconnections, without separately evaluating local reaction- or transport-heat source terms.

Near the reference temperature, the equilibrium potential is written as

$$U_k(\theta, T) \approx U_{k,\mathrm{ref}}(\theta) + (T - T_{\mathrm{ref}})S_k(\theta).$$

The tabulated entropic coefficients are provided in $\mathrm{mV\,K^{-1}}$ and are converted to $\mathrm{V\,K^{-1}}$ before use in the model.

A.3 Model Parameterization

The model parameterization procedure consists of four interrelated steps. First, the half-cell thermodynamic parameters, kinetic parameters, and electrode-structure parameters are mapped to the corresponding governing equations, with consistent notation and units. Second, the lithium-stoichiometry ranges of the positive and negative electrodes are determined from their electrode-potential curves as functions of lithium stoichiometry together with the full-cell state-of-charge (SOC)–open-circuit-voltage (OCV) curve. The coupled model is then solved under prescribed current and ambient-temperature conditions to obtain terminal-voltage responses that can be compared with experimental measurements. Finally, only parameters that are difficult to measure directly and are identifiable are estimated. This procedure specializes the general coupled P2D–lumped thermal model to a specific material–cell system.

The parameterization of NCM811-1 is used below as an example. Its kinetic parameters include positive- and negative-electrode Bruggeman coefficients of 1.6887 and 2.7292, respectively. $D_{s,p}$, $D_{s,n}$, k_p^{cfg}, and k_n^{cfg} are each defined on a 16×21 temperature–lithium-stoichiometry grid covering -50 to $100\,^{\circ}\mathrm{C}$ and $\theta_k \in [0, 1]$. D_e is defined on a 16×31 temperature–concentration grid covering the same temperature range and $c_e \in [0, 3000]\,\mathrm{mol\,m^{-3}}$. During computation, $D_{s,k}$, k_k^{cfg}, and D_e are obtained by interpolation of the corresponding tabulated data within the ranges specified above, and temperatures in degrees Celsius are converted to kelvin.

The thermodynamic data include U_p, U_n, S_p, and S_n, with lithium-stoichiometry ranges of

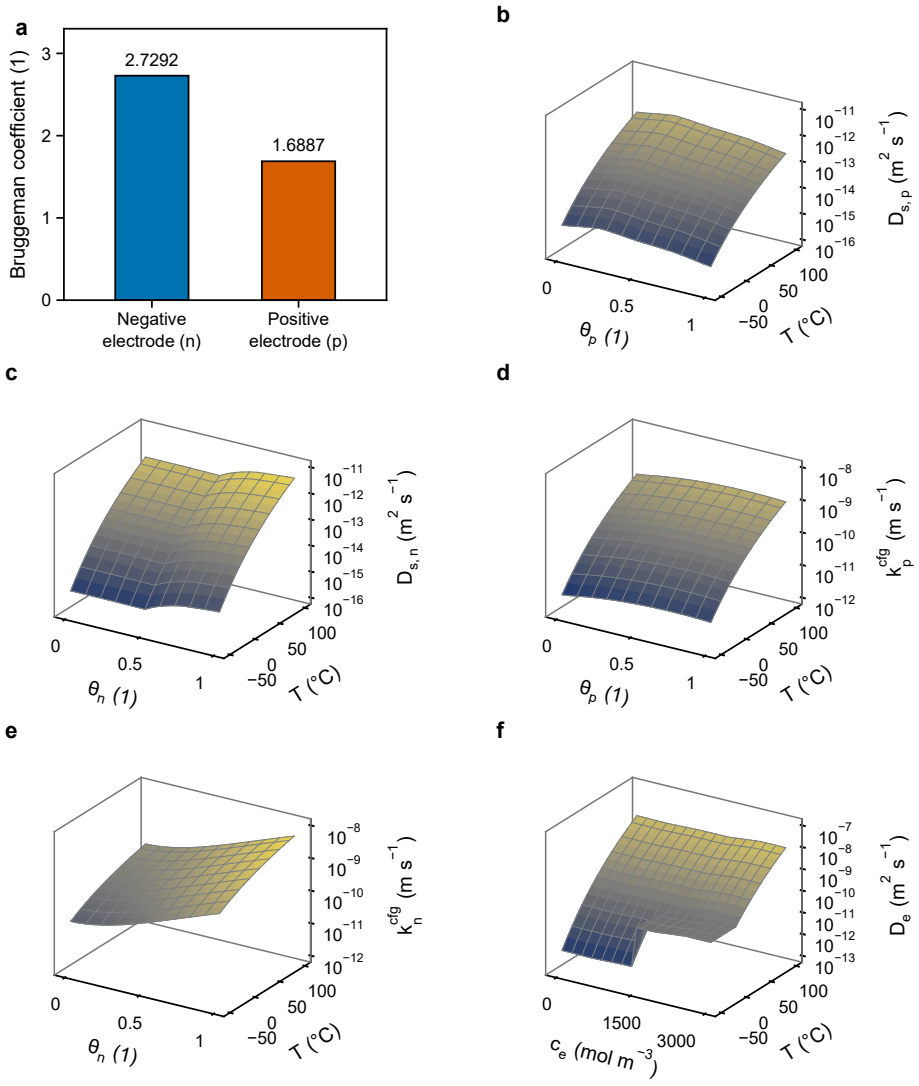


Figure A.1: **Kinetic parameters of NCM811-1.** (a) Positive- and negative-electrode Bruggeman coefficients; (b, c) positive- and negative-electrode solid-phase diffusion coefficients; (d, e) positive- and negative-electrode reaction rate constants; (f) electrolyte diffusion coefficient.

$$[0.1546, 0.9694], \quad [0.0100, 0.9500], \quad [0.2204, 0.9832], \quad [0.0050, 0.9000],$$

respectively.

Within the tabulated lithium-stoichiometry ranges, the thermodynamic quantities are obtained by interpolation. Outside a tabulated range, the nearest boundary value is retained. In particular, when the negative-electrode entropic coefficient is required above its upper tabulated limit,

$$S_n(\theta_n) = S_n(0.9000), \qquad \theta_n > 0.9000.$$

For model initialization, the initial electrode lithium stoichiometry is determined from the initial cell SOC z according to

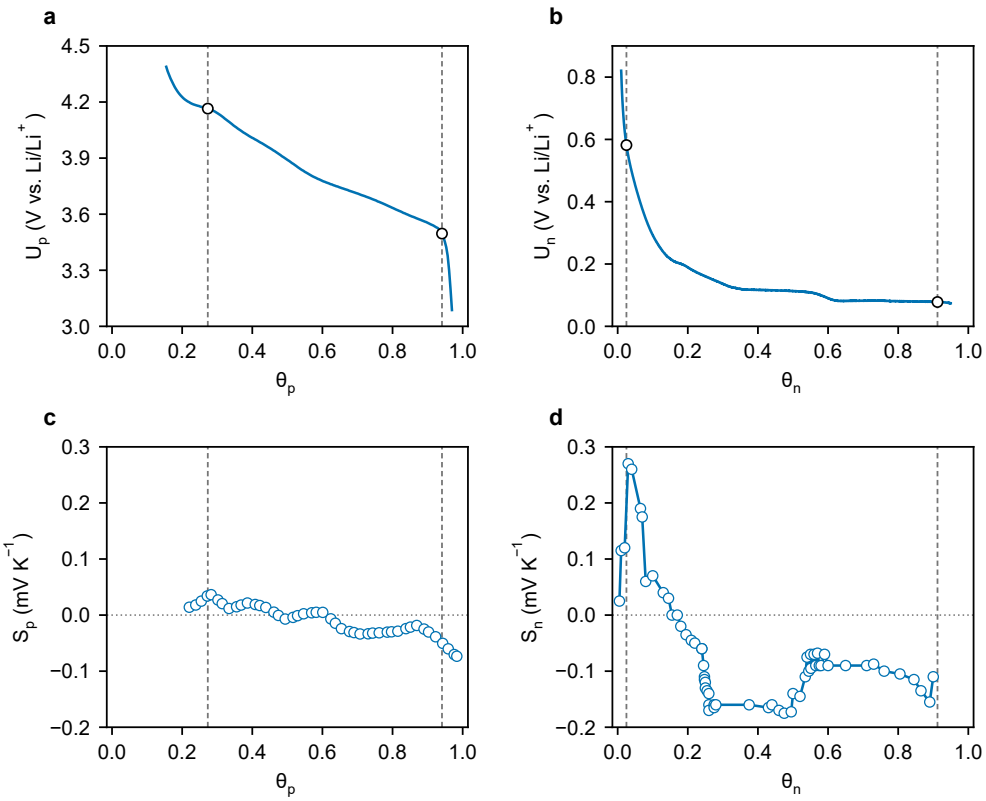


Figure A.2: **Thermodynamic data of NCM811-1.** (a, b) Open-circuit potentials of the positive and negative electrodes versus $\mathrm{Li}/\mathrm{Li}^+$; (c, d) positive- and negative-electrode entropic coefficient curves. The vertical dashed lines indicate the determined lithium-stoichiometry ranges of the positive and negative electrodes.

$$\theta_{k,0}(z) = (1 - z)\theta_k(0) + z\theta_k(1),$$

where

$$\theta_p(0) = 0.9410, \qquad \theta_p(1) = 0.2731,$$

and

$$\theta_n(0) = 0.0250, \qquad \theta_n(1) = 0.9121.$$

This SOC–lithium-stoichiometry mapping is used only to determine the initial solid-phase lithium concentration and is distinct from the local lithium stoichiometry $\theta_k(r, x, t)$ used to evaluate $D_{s,k}$ and the surface lithium stoichiometry θ_k^{surf} used in the interfacial kinetics.

The reference temperature is $25\,^{\circ}\mathrm{C}$ (298.15 K). The specified specific heat capacity is $1000\,\mathrm{J\,kg^{-1}\,K^{-1}}$, which, together with the cell mass, determines the total thermal capacity C_{th}.

A.4 Kinetic Parameter Set Selection Method

To characterize the differences among kinetic parameter sets within the same material system and select representative sets, we consider seven parameters with relatively strong effects on battery performance to construct a unified feature representation: the negative-electrode reaction rate constant,

Table A.1: Main Parameters Used for NCM811-1 Model Parameterization

Input	Value	Unit	Role
b_p, b_n	1.6887, 2.7292	1	Porous-electrode transport correction
$D_{s,p}, D_{s,n}$	16×21; $T \in [-50, 100]\,^\circ\mathrm{C}$, $\theta_k \in [0, 1]$	$\mathrm{m^2\,s^{-1}}$	Solid-phase diffusion
$k_p^{\mathrm{cfg}}, k_n^{\mathrm{cfg}}$	16×21; same as above	$\mathrm{m\,s^{-1}}$	Interfacial reaction kinetics
D_e	16×31; $T \in [-50, 100]\,^\circ\mathrm{C}$, $c_e \in [0, 3000]\,\mathrm{mol\,m^{-3}}$	$\mathrm{m^2\,s^{-1}}$	Electrolyte diffusion
U_p, U_n	201 points, $[0.1546, 0.9694]$; 2502 points, $[0.0100, 0.9500]$	V	Equilibrium-potential and overpotential calculations
S_p, S_n	45 points, $[0.2204, 0.9832]$; 60 points, $[0.0050, 0.9000]$	$\mathrm{mV\,K^{-1}}$	Equilibrium-potential temperature correction and reversible-heat calculation
$\theta_p(0), \theta_p(1)$	0.9410, 0.2731	1	Positive-electrode SOC–lithium-stoichiometry mapping
$\theta_n(0), \theta_n(1)$	0.0250, 0.9121	1	Negative-electrode SOC–lithium-stoichiometry mapping
T_{ref}	25	$^\circ\mathrm{C}$	Reference temperature for equilibrium-potential correction
c_p	1000 at 25 $^\circ\mathrm{C}$	$\mathrm{J\,kg^{-1}\,K^{-1}}$	Determines the total thermal capacity together with the cell mass

positive-electrode reaction rate constant, electrolyte diffusion coefficient, negative-electrode solid-phase lithium-ion diffusion coefficient, positive-electrode solid-phase lithium-ion diffusion coefficient, and the Bruggeman coefficients of the positive and negative electrodes. Here, the term kinetic parameter set is used in a broad sense to refer to parameters governing interfacial reaction and transport processes, including reaction rate constants, solid-phase and electrolyte diffusion coefficients, and Bruggeman coefficients.

The reaction rate constants and solid-phase lithium-ion diffusion coefficients of both electrodes are functions of lithium stoichiometry and temperature. The lithium stoichiometry ranges from 0 to 1 with a step size of 0.05, including both endpoints, while the temperature ranges from -50 to $100\,^\circ\mathrm{C}$ with a step size of $10\,^\circ\mathrm{C}$, also including both endpoints. These two variables form a regular two-dimensional sampling grid. The electrolyte diffusion coefficient is a function of electrolyte concentration and temperature. The temperature is sampled in the same manner, while the electrolyte concentration ranges from 0 to $3000\ \mathrm{mol\,m^{-3}}$ with a step size of $100\ \mathrm{mol\,m^{-3}}$.

For the above five parameters that vary with state variables, each two-dimensional parameter surface is converted into a scalar feature for consistent comparison by taking the mean of the logarithmically transformed valid samples. Let $\mathcal{G}_{ij}$ denote the set of valid data points for the j-th parameter in the i-th kinetic parameter set. We define

$$x_{ij} = \frac{1}{|\mathcal{G}_{ij}|} \sum_{q \in \mathcal{G}_{ij}} \log_{10} p_{ijq}, \qquad j = 1, \ldots, 5,$$

where p_{ijq} denotes the value of the j-th kinetic parameter in the i-th parameter set at sampling point q. This transformation characterizes each parameter by its mean logarithmic value over the valid

sampling domain while reducing the influence of variations across multiple orders of magnitude on subsequent comparisons.

For the negative- and positive-electrode Bruggeman coefficients, we define

$$x_{i6} = \frac{1}{b_{n,i}}, \qquad x_{i7} = \frac{1}{b_{p,i}}.$$

The reciprocals are used to ensure a consistent direction across all features. For the first five parameters, larger values generally correspond to faster reaction or transport processes, whereas a larger Bruggeman coefficient typically indicates stronger transport limitations in porous media. After the above transformation, larger values of all seven features therefore consistently correspond to enhanced reaction or transport capability.

To eliminate differences in numerical scale among the features, each feature is independently normalized within each material system using min–max normalization:

$$z_{ij}^{(g)} = \frac{x_{ij} - \min_{r \in g} x_{rj}}{\max_{r \in g} x_{rj} - \min_{r \in g} x_{rj}}, \qquad j = 1, \ldots, 7,$$

where g denotes the material system and r indexes all kinetic parameter sets within that system. Each kinetic parameter set can thus be represented by a seven-dimensional normalized feature vector,

$$\mathbf{z}_i = \left[z_{i1}, z_{i2}, \ldots, z_{i7} \right]^{\mathrm{T}}.$$

We further define the Euclidean distance between this feature vector and the origin in the seven-dimensional space as

$$d_i = \left(\sum_{j=1}^{7} z_{ij}^2 \right)^{1/2}.$$

Because all features have been directionally aligned and normalized within each material system, d_i is used as a ranking metric for the overall kinetic level of a parameter set. A larger d_i indicates that the parameter set has, overall, higher normalized values across the seven major kinetic features and can therefore be regarded as approximately representing a higher overall kinetic level. It should be emphasized that this metric is used only for ranking and selecting parameter sets within the same material system and should not be interpreted as a kinetic performance metric with strict physical meaning.

For each material system, d_i is calculated for all kinetic parameter sets and sorted in ascending order. Three parameter sets with relatively low, intermediate, and high d_i values are then selected to cover the major variation among the candidate kinetic parameters.

In addition to the seven parameters described above, each complete kinetic parameter set contains other parameters. These parameters are retained in the subsequent model calculations according to their corresponding parameter sets but are not included in the ranking and selection procedure described in this section. All kinetic parameter data used in this work were provided by a data provider whose identity is withheld for anonymous review.

As shown in Fig. A.3, each box represents the normalized distribution of the corresponding kinetic feature across all parameter sets within a material system, while the circles, squares, and triangles denote the three selected kinetic parameter sets. The selected sets occupy different positions in the distributions of the seven major kinetic features, indicating that the proposed selection method can capture the major differences among the candidate kinetic parameters using only a limited number of parameter sets.

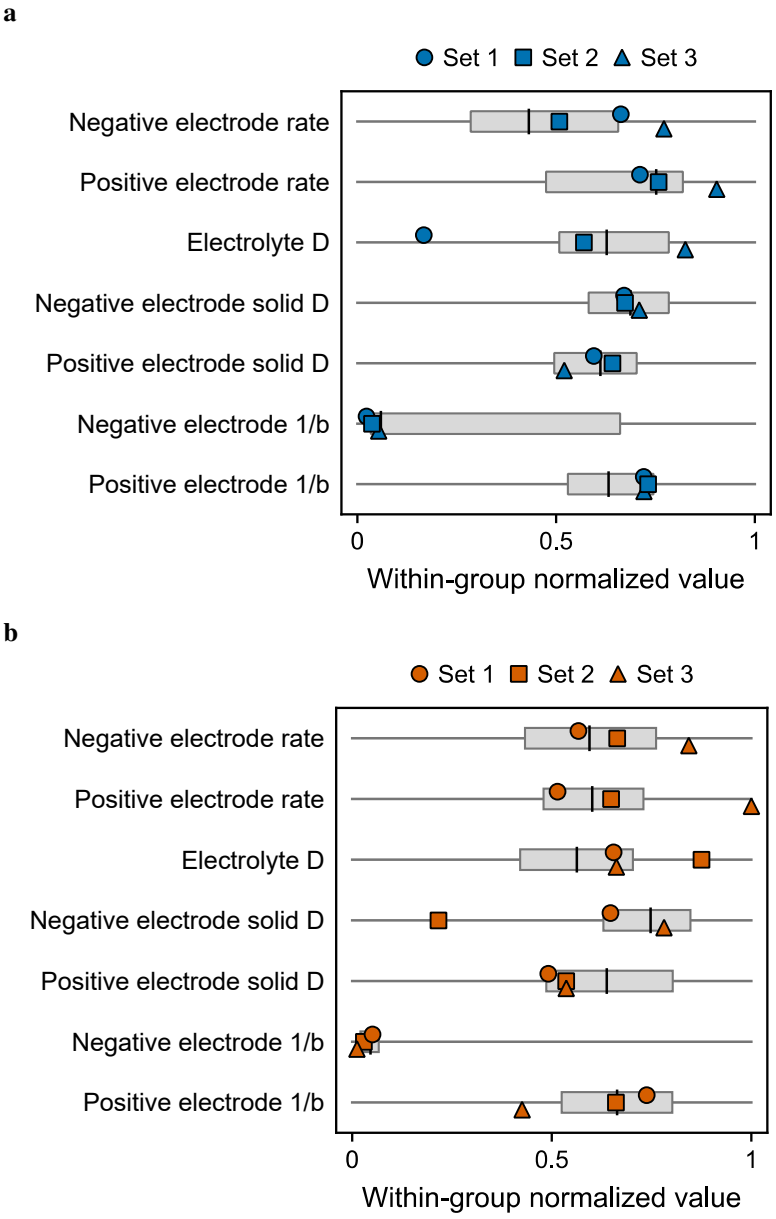


Figure A.3: **Distribution of the selected kinetic parameter sets among all candidate parameter sets for LFP and NCM622.** (a) LFP; (b) NCM622. The first five rows show the mean $\log_{10}$ values of the negative-electrode reaction rate constant, positive-electrode reaction rate constant, electrolyte diffusion coefficient, negative-electrode solid-phase lithium-ion diffusion coefficient, and positive-electrode solid-phase lithium-ion diffusion coefficient over the valid data points, respectively. The last two rows show the reciprocals of the negative- and positive-electrode Bruggeman coefficients, respectively. Each feature is independently min–max normalized within the corresponding material system. The horizontal lines, gray boxes, and black lines indicate the minimum-to-maximum range, interquartile range, and median, respectively. Circles, squares, and triangles denote the three selected kinetic parameter sets; the vertical offsets of the markers are introduced only to avoid overlap.

B Rationale for the Sampling Densities of Porous-Electrode Design Variables

Data generation involves five porous-electrode design variables: positive-electrode coating thickness, positive-electrode compaction density, negative-electrode compaction density, positive-electrode active-material particle size D_{50}, and negative-electrode active-material particle size D_{50}. Because these variables have different levels of influence on the model outputs, higher sampling densities are used for the first three variables, whereas relatively lower sampling densities are used for the two D_{50} variables, so as to balance design-space coverage and the ability to characterize design–performance relationships under a limited data budget.

To illustrate this choice, Fig. B.1 presents a representative example using an NCM811-1 46800 cylindrical cell (46 mm in diameter and 80 mm in height), in which the five variables are examined through separate one-factor-at-a-time sweeps. In each sweep, only one variable is varied while the other four are fixed at their baseline values. Each variable is sampled at 50 equally spaced points over the range used for data generation. The baseline values of the five variables are 54.5 μm, 3.6 $\mathrm{g\,cm^{-3}}$, 1.6 $\mathrm{g\,cm^{-3}}$, 6.5 μm, and 13.6 μm, respectively. For each design point, a material-balance calculation is first performed, followed by calculation of the discharge capacity at 25 $^{\circ}$C and 0.333 C using the coupled P2D–lumped thermal model.

The positive-electrode coating thickness and the positive-electrode and negative-electrode compaction densities affect the electrode thickness, total active-material mass, and the number of winding turns that can be accommodated within a fixed-size cell casing, and therefore have a pronounced effect on capacity. From the lower to the upper bound of their respective sweep ranges, the net changes in material-balance capacity are approximately 3.18, 1.55, and 4.03 Ah, respectively, while the corresponding net changes in simulated discharge capacity are approximately 2.78, 1.41, and 3.57 Ah. Because the cell casing dimensions are fixed, the number of winding turns that can be accommodated is geometrically constrained. As the electrode thickness increases, the capacity generally increases; however, once the thickness crosses a geometric threshold and the number of winding turns decreases abruptly, the total active-material mass drops abruptly, resulting in a local decrease in capacity. The responses therefore exhibit an overall increasing trend with superimposed step-like or sawtooth-like discontinuities. Under the current low-rate condition of 0.333 C, kinetic limitations are relatively weak, and the simulated discharge capacity generally follows a trend similar to that of the material-balance capacity. Because these three variables induce large capacity variations and pronounced local non-smoothness, they are sampled more densely.

In contrast, positive-electrode and negative-electrode D_{50} are not included in the current material-balance calculation, and therefore their variation does not affect the material-balance capacity. Although increasing particle size increases the solid-phase lithium-ion diffusion distance and thereby strengthens kinetic limitations, its effect is relatively small under the current low-rate condition. Across their respective sweep ranges, increasing positive-electrode and negative-electrode D_{50} reduces the simulated discharge capacity by only approximately 0.17 and 0.28 Ah, respectively, and the corresponding responses are smooth and nearly linear. Therefore, relatively lower sampling densities are used for the two D_{50} variables.

In summary, more sampling points are allocated to the positive-electrode coating thickness and the positive-electrode and negative-electrode compaction densities, which produce larger capacity variations and exhibit local non-smoothness, whereas relatively sparse sampling is used for positive-electrode and negative-electrode D_{50}, which produce smaller and smoother responses. Fig. B.1 provides a representative basis for this sampling strategy rather than an exhaustive sensitivity analysis across all cell formats, material systems, and operating conditions.

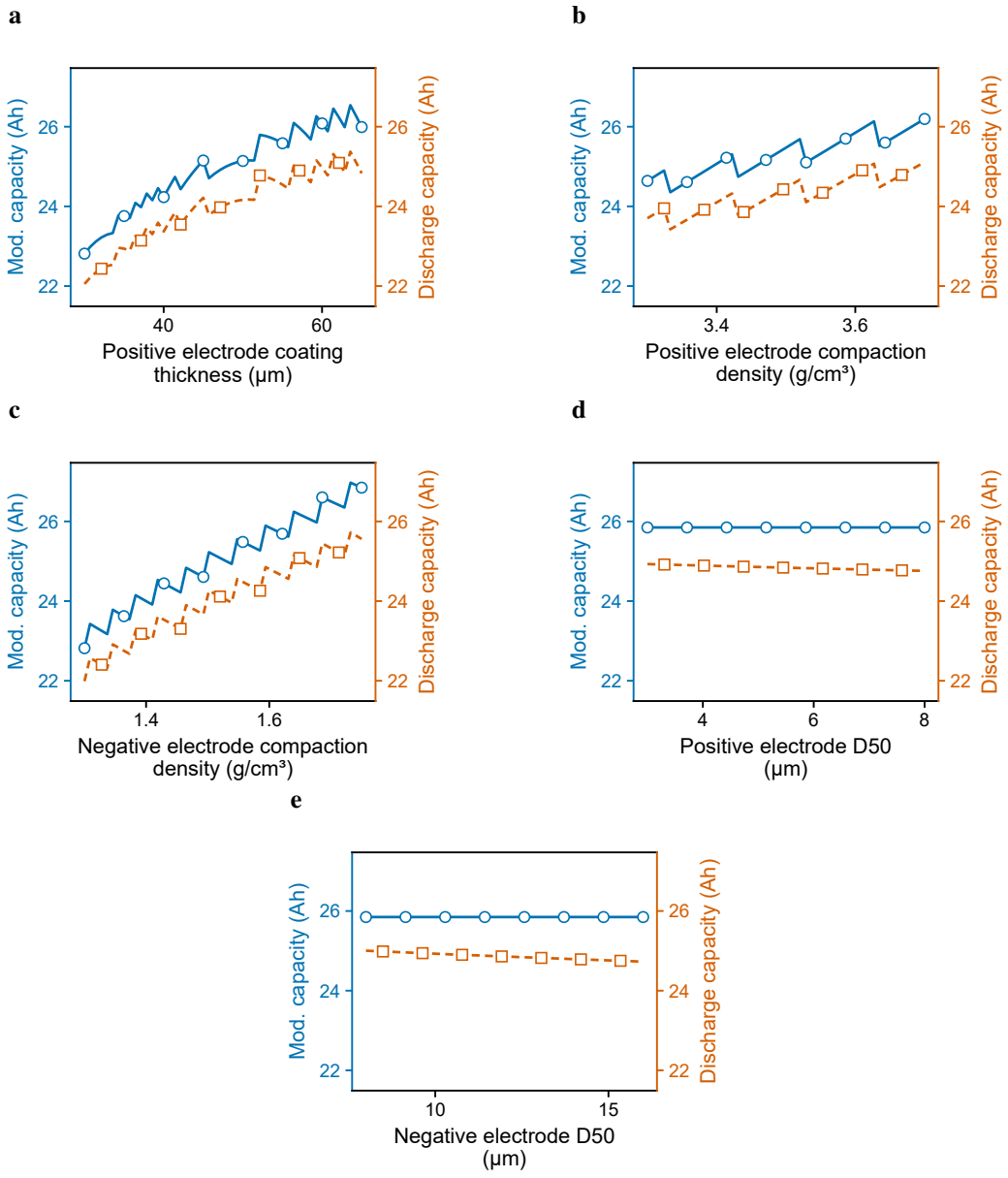


Figure B.1: **One-factor-at-a-time sweeps of porous-electrode design variables.** Using the NCM811-1 46800 cylindrical cell as an example, the following variables are swept individually: (a) positive-electrode coating thickness, 30–65 μm; (b) positive-electrode compaction density, 3.3–3.7 $\mathrm{g\,cm^{-3}}$; (c) negative-electrode compaction density, 1.30–1.75 $\mathrm{g\,cm^{-3}}$; (d) positive-electrode active-material D_{50}, 3–8 μm; and (e) negative-electrode active-material D_{50}, 8–16 μm. When one variable is swept, the other four are fixed at their corresponding baseline values. Blue solid curves with open-circle markers denote the material-balance capacity (left axis), while orange dashed curves with open-square markers denote the simulated discharge capacity at 25 °C and 0.333 C (right axis). Each curve contains 50 equally spaced sampling points. The first three variables produce larger and locally non-smooth capacity responses, whereas the two D_{50} variables produce only small and smooth changes in simulated capacity; accordingly, higher and relatively lower sampling densities are used for the two groups, respectively.

C Electrothermal Process Data and Statistical Analysis of the Dataset

The dataset encompasses multiple cell models, porous-electrode designs, and material systems, together with multiple classes of performance metrics. Complete electrothermal process data are retained for every performance metric evaluated under each design. To illustrate these data, we selected a representative prismatic power cell using the LFP-3 material system, with dimensions of 148 mm $\times$ 101 mm $\times$ 27 mm. Among the performance metrics evaluated for this cell, we selected the multistage constant-current fast-charging time from 20% to 80% SOC at $-10\,^{\circ}\mathrm{C}$ and visualized the corresponding electrothermal process data. The target SOC range is divided into twelve 5-percentage-point windows. Within each window, a constant current is applied, and the maximum charging current for that stage is determined subject to the constraint that the lithium-plating reaction overpotential remains no lower than 0 V. This procedure produces a design-specific multistage constant-current fast-charging profile.

In addition to visualizing the electrothermal process data, we statistically analyzed the complete database from three perspectives: correlations among performance metrics, design-performance relationships, and design-performance sensitivity.

First, we analyzed correlations among performance metrics of the same class under different temperatures, SOCs, or C-rates. Metrics within the same class have similar physical meanings and differ primarily in their test conditions. Most metrics within the same class exhibited strong Pearson linear correlations and Spearman rank correlations. This finding provides a statistical basis for estimating performance under other operating conditions from performance obtained under a representative condition. However, correlation alone does not guarantee predictive accuracy. Such inference still requires an explicit quantitative mapping and validation of the associated prediction error. The strong within-class correlations also provide a basis for dimensionality reduction in subsequent analyses. One representative metric can therefore be selected from each performance class for focused analysis. This avoids the simultaneous analysis of numerous correlated metrics and reduces redundancy while retaining the principal information.

Second, we analyzed the design-performance relationships between representative performance metrics and porous-electrode design parameters. For each cell model–material system combination in the database, we selected the design that optimized the corresponding performance metric from the generated data. Capacity, energy, power, and energy efficiency were maximized, whereas fast-charging time and direct-current resistance were minimized. We then examined the distributions of positive-electrode coating thickness, positive-electrode compaction density, negative-electrode compaction density, and the D_{50} particle sizes of the positive- and negative-electrode active materials. These distributions reveal qualitative design trends associated with different performance metrics.

The optimum for a single performance metric generally exhibited a clear preference for specific parameter values. For example, designs that maximized capacity or energy favored a greater positive-electrode coating thickness, higher positive- and negative-electrode compaction densities, and smaller D_{50} particle sizes of the positive- and negative-electrode active materials. These trends provide valuable prior information for screening candidate designs. However, the optimal porous-electrode design differed across performance metrics. Parameter combinations that increased capacity or energy did not necessarily benefit power, direct-current resistance, fast-charging capability, or energy efficiency. Practical cell design is therefore not a simple superposition of independent single-objective problems, but a trade-off among multiple competing objectives. Such conflicts are already evident when only five porous-electrode parameters are considered. When hundreds of design variables across the cell, electrode, and material scales are further introduced, the design space and its coupled relationships become substantially more complex, making multi-objective optimization considerably more challenging.

Finally, we performed a sensitivity analysis based on permutation feature importance to quantify the sensitivity of different performance metrics to the five porous-electrode design parameters. The design-performance analysis identifies the regions of the design space in which optimal designs for individual objectives are located. The sensitivity analysis further evaluates the relative importance of the different porous-electrode design parameters. These results provide a basis for balancing parameters across multiple performance objectives and for reducing the design space.

Permutation feature importance measures the dependence of model predictions on each input parameter under the current data distribution. It does not represent the physical derivative or causal effect of changing that parameter.

Overall, the database not only provides numerical performance metrics but also supports process-level mechanistic analysis, extraction of design trends, and high-dimensional multi-objective battery optimization.

C.1 Low-Temperature Multistage Fast-Charging Process Data

This section presents the low-temperature multistage fast-charging process from three complementary perspectives: electrical response, thermal response, and polarization. The current and voltage data include the stage-wise charging current, terminal voltage, and operating potentials of the positive and negative electrodes. These data characterize the fast-charging profiles and the voltage responses of the cell and electrodes to the current applied during each stage. The temperature and heat-generation data include the average cell temperature, the total volumetric heat-generation rate of the jelly roll, its reversible and irreversible components, and the heat-generation power of the mechanical components. These data quantify the temperature rise, thermal load, and contributions of different heat sources during fast charging. The polarization data further decompose the contributions associated with electrode reactions, electrolyte-phase ion transport, solid-phase ion transport, electronic conduction, and ion transport through the separator. The lithium-plating reaction overpotential is reported separately. These quantities enable comparisons among the different polarization components and help identify performance-limiting processes. Together, the three types of data describe the progression from the externally applied fast-charging current to the voltage response, thermal behavior, and internal polarization. They provide a basis for comparing the low-temperature fast-charging characteristics of different porous-electrode designs.

Discharge current is defined as positive, whereas charging current is defined as negative. The curve color progresses from dark blue to yellow as the total charge time increases. Based on the two tertile boundaries of total charge time, 91.71 and 124.05 min, the charging profiles are divided into short-, medium-, and long-duration groups. These groups are represented by solid, dashed, and dotted lines, respectively. The resulting plots show the fast-charging profiles obtained with different porous-electrode designs and provide a basis for comparing their fast-charging capabilities.

C.1.1 Current and Voltage Data

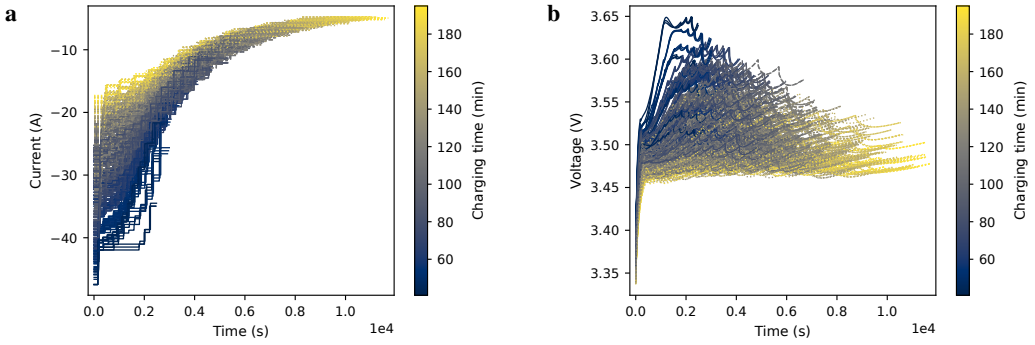


Figure C.1: Current and voltage during multistage constant-current fast charging from 20% to 80% SOC at $-10\,^{\circ}$C. (a) Charging current; (b) terminal voltage;

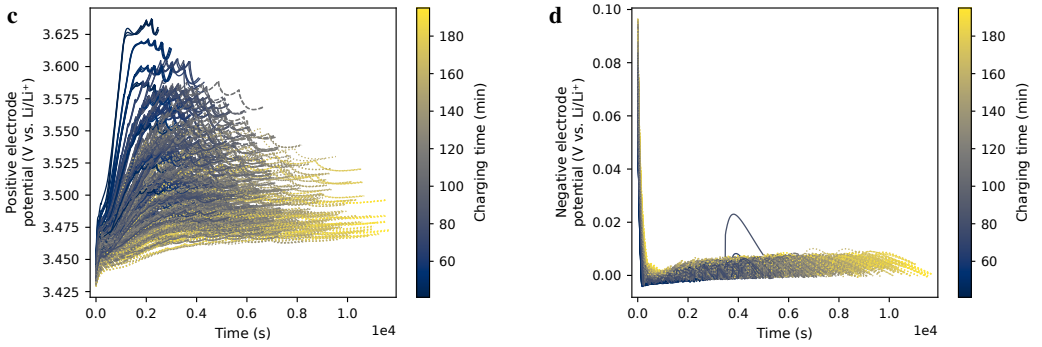


Figure C.1: (c) positive-electrode operating potential versus Li/Li$^+$; (d) negative-electrode operating potential versus Li/Li$^+$.

C.1.2 Average Temperature and Heat-Generation Data

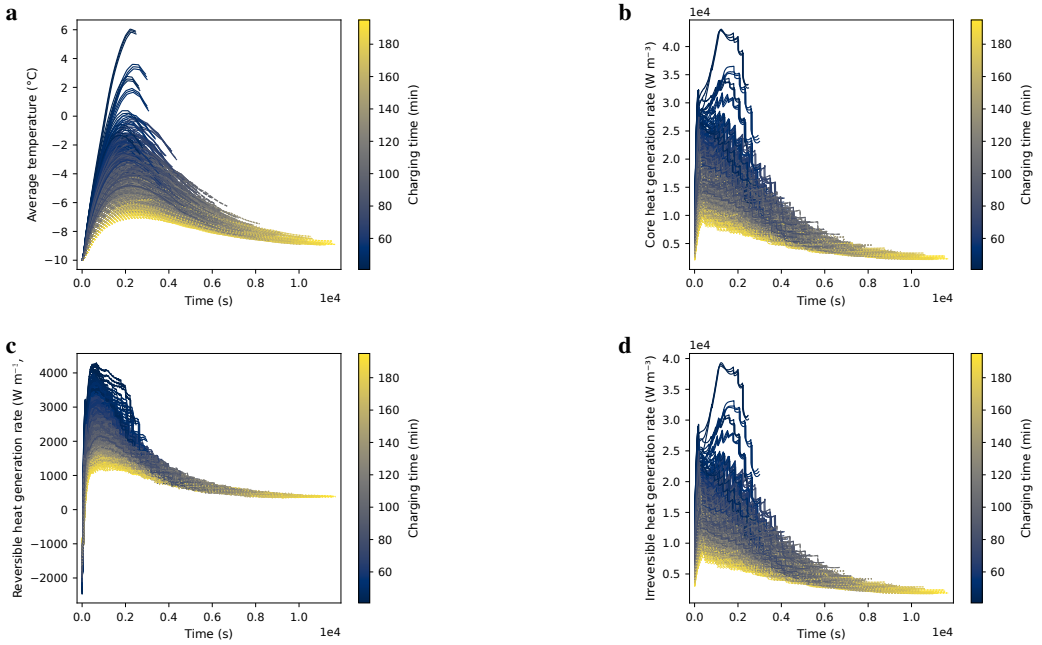


Figure C.2: Average temperature and heat generation during multistage constant-current fast charging from 20% to 80% SOC at $-10\,^{\circ}$C. (a) Average cell temperature; (b) total volumetric heat-generation rate of the jelly roll; (c) reversible volumetric heat-generation rate of the jelly roll; (d) irreversible volumetric heat-generation rate of the jelly roll;

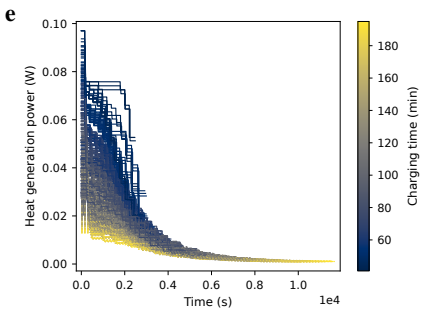


Figure C.2: (e) heat-generation power of the mechanical components.

C.1.3 POLARIZATION DATA

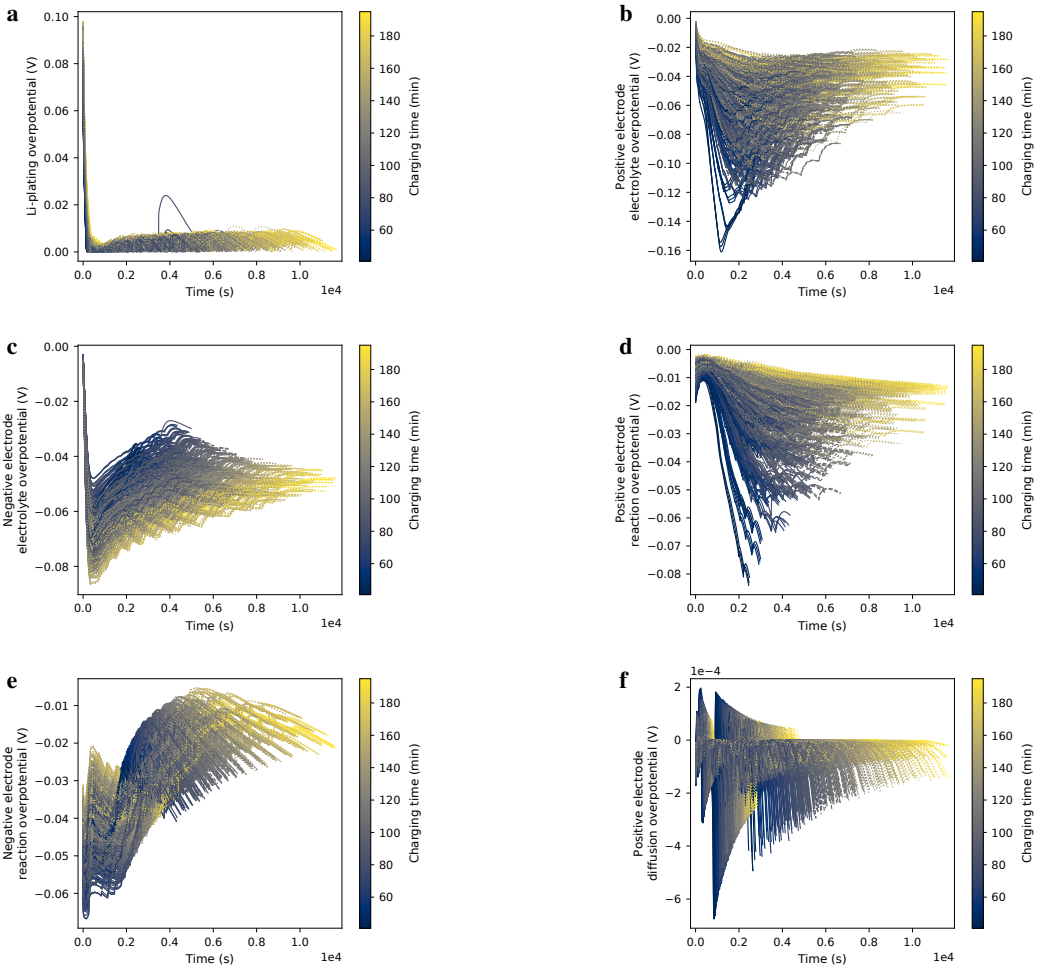


Figure C.3: Polarization decomposition during multistage constant-current fast charging from 20% to 80% SOC at $-10\,^{\circ}\mathrm{C}$. (a) Lithium-plating reaction overpotential; (b) positive-electrode electrolyte-phase ion-transport overpotential; (c) negative-electrode electrolyte-phase ion-transport overpotential; (d) positive-electrode reaction overpotential; (e) negative-electrode reaction overpotential; (f) positive-electrode solid-phase ion-transport overpotential;

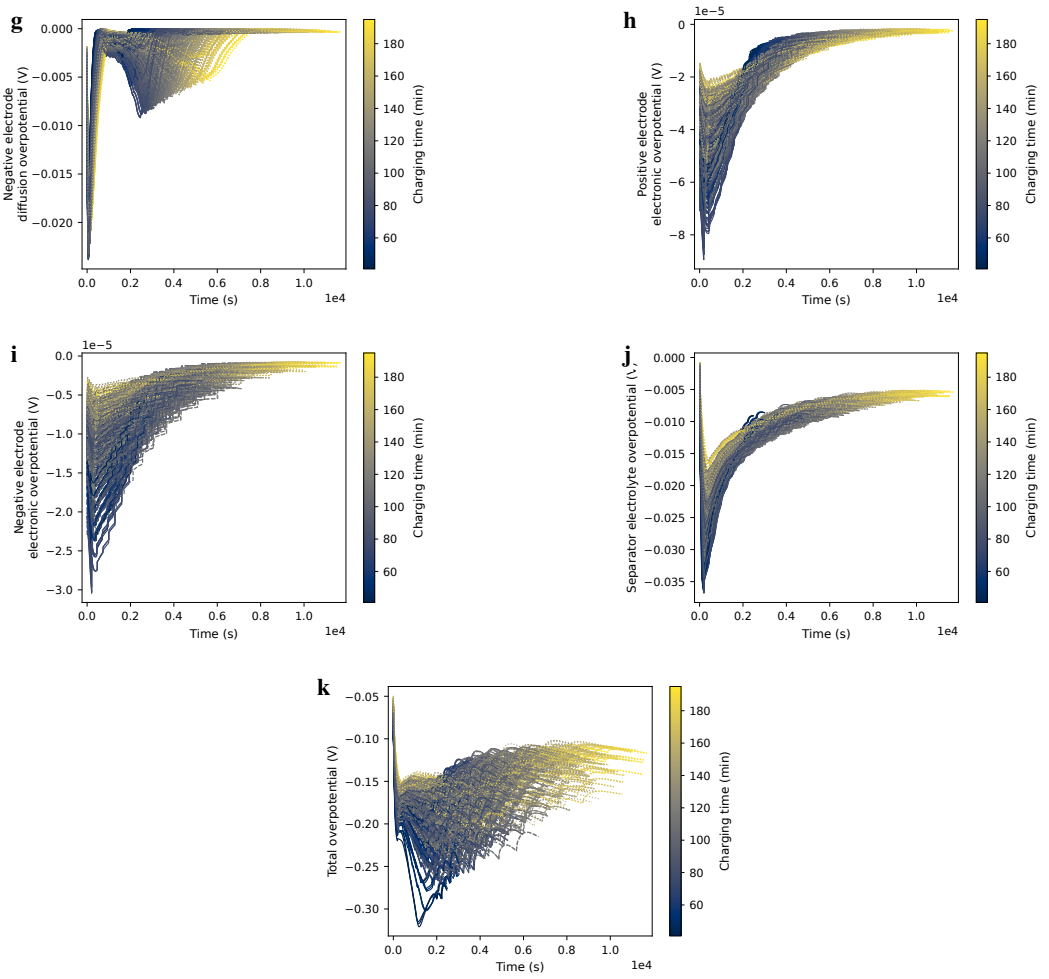


Figure C.3: (g) negative-electrode solid-phase ion-transport overpotential; (h) positive-electrode solid-phase electronic-conduction overpotential; (i) negative-electrode solid-phase electronic-conduction overpotential; (j) separator electrolyte-phase ion-transport overpotential; (k) signed sum of the nine overpotential components. Panel (a) is not included in the sum shown in panel (k).

C.2 Statistical Analysis of the Generated Dataset

Using the cleaned dataset, we conducted statistical analyses at three levels: correlations among performance metrics of the same class, design-performance relationships between performance metrics and porous-electrode design parameters, and the sensitivity of performance predictions to porous-electrode design parameters. The correlation analysis evaluates the consistency of metrics within the same class across different operating conditions. The design-performance analysis extracts the parameter distributions of performance-optimal records. The within-group permutation feature importance analysis measures the dependence of model predictions on each design parameter.

Table C.1: Abbreviations and test conditions for the 29 performance metrics

Abbreviation	Performance metric	Test condition	Unit
Q1	Discharge capacity	25 °C, 1/3C	Ah
Q2	Discharge capacity	45 °C, 1/3C	Ah
Q3	Discharge capacity	25 °C, 1C	Ah
Q4	Discharge capacity	45 °C, 1C	Ah
E1	Discharge energy	25 °C, 1/3C	Wh
E2	Discharge energy	45 °C, 1/3C	Wh
E3	Discharge energy	25 °C, 1C	Wh
E4	Discharge energy	45 °C, 1C	Wh
Ppk1	Peak discharge power	25 °C, 20% SOC, 10 s	W
Ppk2	Peak discharge power	25 °C, 50% SOC, 10 s	W
Ppk3	Peak discharge power	25 °C, 80% SOC, 10 s	W
Ppk4	Peak discharge power	45 °C, 20% SOC, 10 s	W
Ppk5	Peak discharge power	45 °C, 50% SOC, 10 s	W
Ppk6	Peak discharge power	45 °C, 80% SOC, 10 s	W
Pct1	Continuous discharge power	25 °C, 20% SOC, 60 s	W
Pct2	Continuous discharge power	25 °C, 50% SOC, 60 s	W
Pct3	Continuous discharge power	25 °C, 80% SOC, 60 s	W
Pct4	Continuous discharge power	45 °C, 20% SOC, 60 s	W
Pct5	Continuous discharge power	45 °C, 50% SOC, 60 s	W
Pct6	Continuous discharge power	45 °C, 80% SOC, 60 s	W
R1	Direct-current resistance	25 °C, 80% SOC, 3C	mΩ
R2	Direct-current resistance	25 °C, 50% SOC, 2C	mΩ
Prc1	Recharge Power	−10 °C, 90% SOC, 5 s	W
Prc2	Recharge Power	25 °C, 90% SOC, 5 s	W
Prc3	Recharge Power	45 °C, 90% SOC, 5 s	W
t1	Fast-charging time	−10 °C, 20%–80% SOC	min
t2	Fast-charging time	25 °C, 20%–80% SOC	min
t3	Fast-charging time	45 °C, 20%–80% SOC	min
η	Energy efficiency	25 °C, 0.5P	Dimensionless

Note. DCR denotes direct-current resistance, and SOC denotes state of charge. Energy efficiency is dimensionless and is reported as a fraction. Here, η denotes the 0.5P energy-efficiency metric in the statistical-analysis figures and this table; polarization overpotentials are defined separately in Section C.1.3.

C.2.1 Correlations among Performance Metrics of the Same Class

For each pair of performance metrics in the dataset, we calculated Pearson's correlation coefficient r and Spearman's rank correlation coefficient ρ, with tied values assigned average ranks. The upper triangle of each matrix reports r, whereas the lower triangle reports ρ.

Capacity and energy exhibited extremely strong correlations across operating conditions. The minimum Pearson and Spearman coefficients were approximately 0.997836 and 0.997806, respectively. Among the 15 pairs of peak-power metrics, 12 had both coefficients above 0.90. For continuous power, 13 pairs met this criterion. For all metric pairs in both power categories, both coefficients exceeded 0.85. The Pearson and Spearman coefficients between the two DCR conditions were 0.995903 and 0.997645, respectively.

The influence of temperature was more pronounced for charging-related metrics. For Recharge Power, r and ρ between $-10\,^{\circ}$C and 45 °C were 0.825694 and 0.896364, respectively. These values were lower than those obtained for the other two comparisons involving 25 °C. Fast-charging time remained strongly correlated only between 25 °C and 45 °C, with coefficients of 0.928592 and 0.944362, respectively. In contrast, both coefficients were below 0.8 when $-10\,^{\circ}$C was compared with either of the other temperatures.

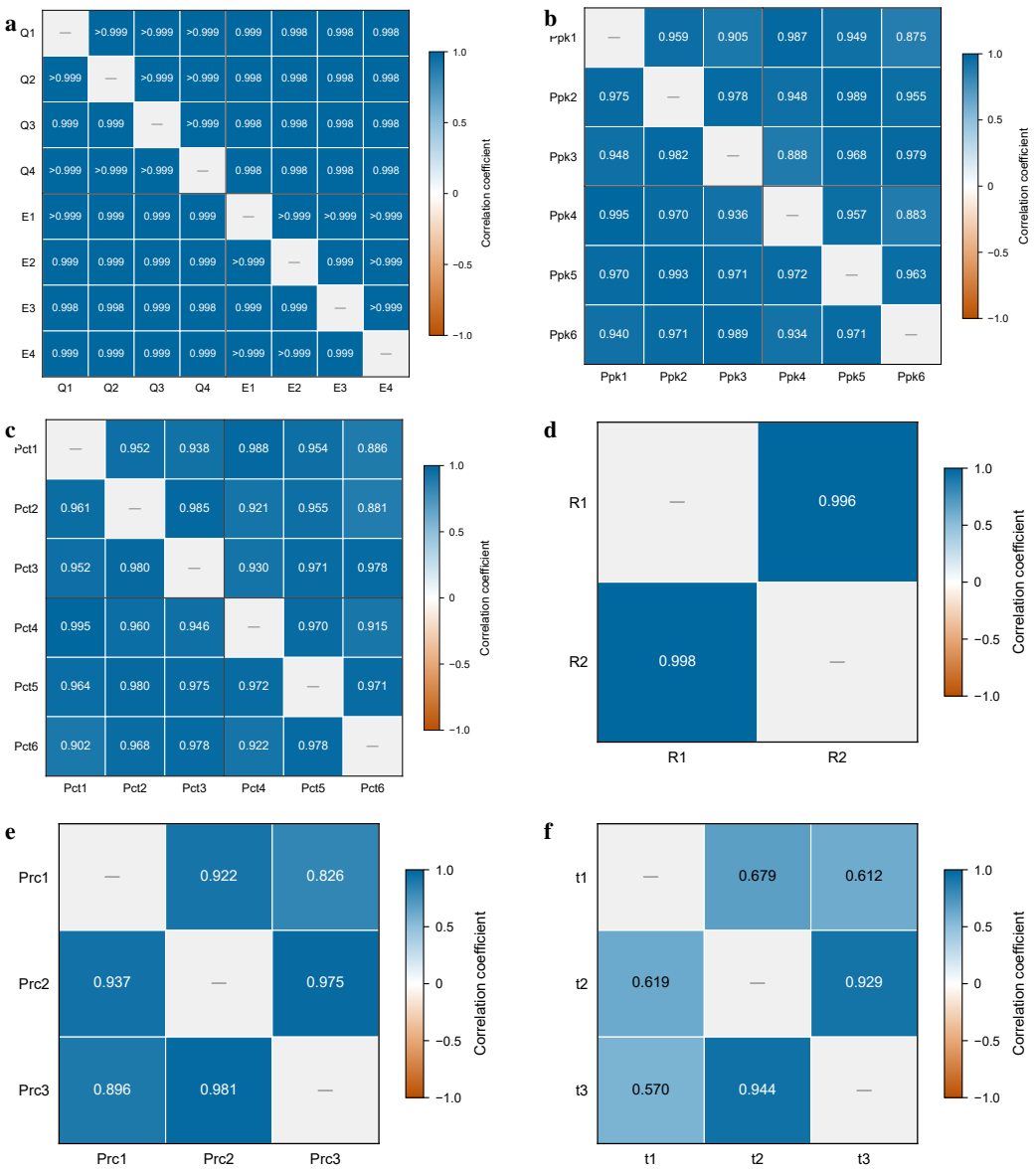


Figure C.4: Correlations among performance metrics of the same class. (a) Capacity and energy; (b) peak discharge power; (c) continuous discharge power; (d) direct-current resistance; (e) Recharge Power; (f) fast-charging time. Each matrix cell was calculated using only data for which both metrics were valid. The upper triangle reports Pearson's r, the lower triangle reports Spearman's ρ, and the diagonal was not calculated.

C.2.2 Porous-Electrode Design Parameter Distributions at Performance Optima

To examine the porous-electrode design parameter distributions associated with optimal values of different performance metrics, we first divided the data into subgroups defined by material system and cell model. Within each subgroup, we extracted the porous-electrode design parameters associated with maximum capacity, power, and energy efficiency, or with minimum direct-current resistance and fast-charging time. Each of the five porous-electrode design parameters was then min–max normalized using its range within the corresponding subgroup. Accordingly, ordinate values of 0 and 1 represent only the lower and upper bounds of the parameter range within that

subgroup. They do not represent absolute parameter values shared across different material systems or cell models.

Marker size and color intensity jointly represent the raw occurrence frequency of each normalized parameter value rather than a normalized proportion. More frequent values are represented by larger and darker markers with lower transparency, and vice versa. These frequencies describe only the marginal frequency distributions of parameter values among the selected optimal records. They cannot be interpreted as the independent effect of any individual parameter. Because the LFP and NCM datasets cover different numbers of subgroups, their relative degrees of concentration cannot be compared directly using the raw frequencies. In addition, the figure does not preserve the correspondence among different parameters belonging to the same design. Consequently, the most frequent values on different parameter axes cannot be combined to construct an actual "optimal design." Tied optima occurred in 84 subgroups for Recharge Power and 93 subgroups for fast-charging time, further limiting interpretation in terms of a unique optimum.

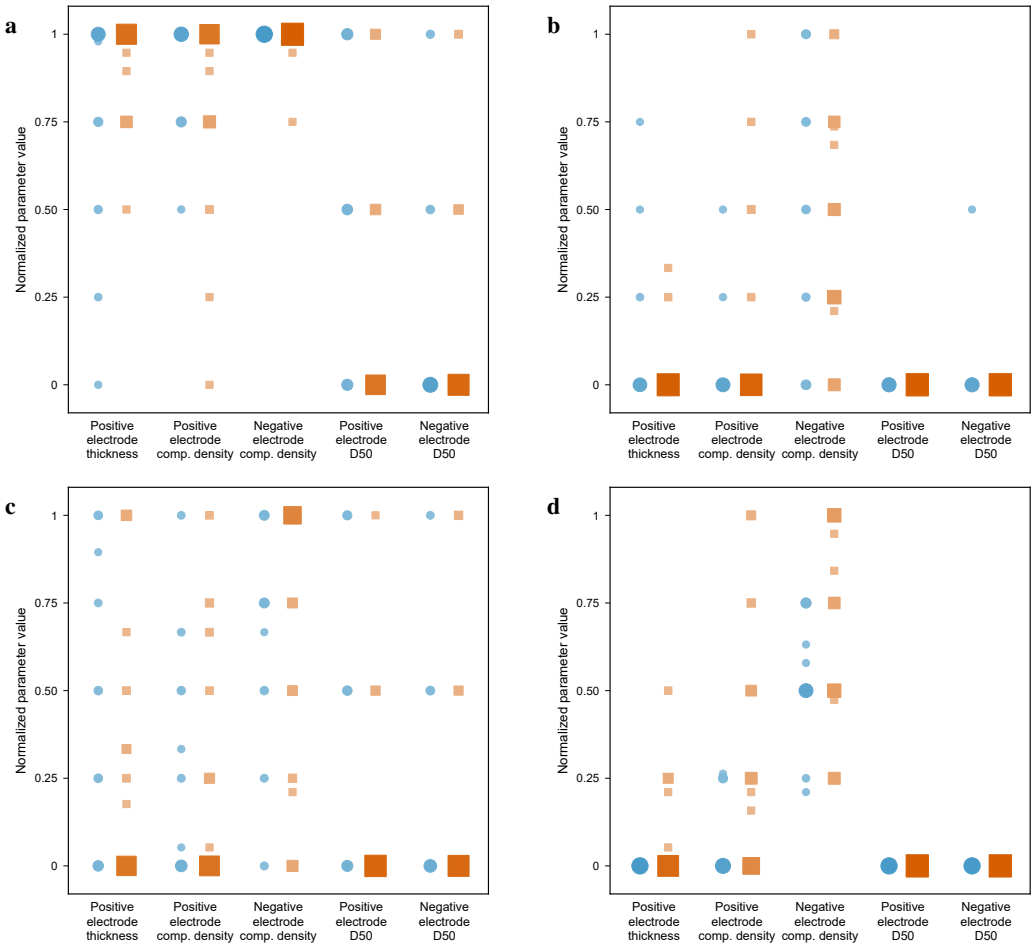


Figure C.5: Porous-electrode design parameter distributions at performance optima. (a) Discharge capacity at 25 °C and 1/3C; (b) 10-s peak discharge power at 25 °C and 50% SOC; (c) 60-s continuous discharge power at 25 °C and 50% SOC; (d) direct-current resistance at 25 °C, 50% SOC, and 2C;

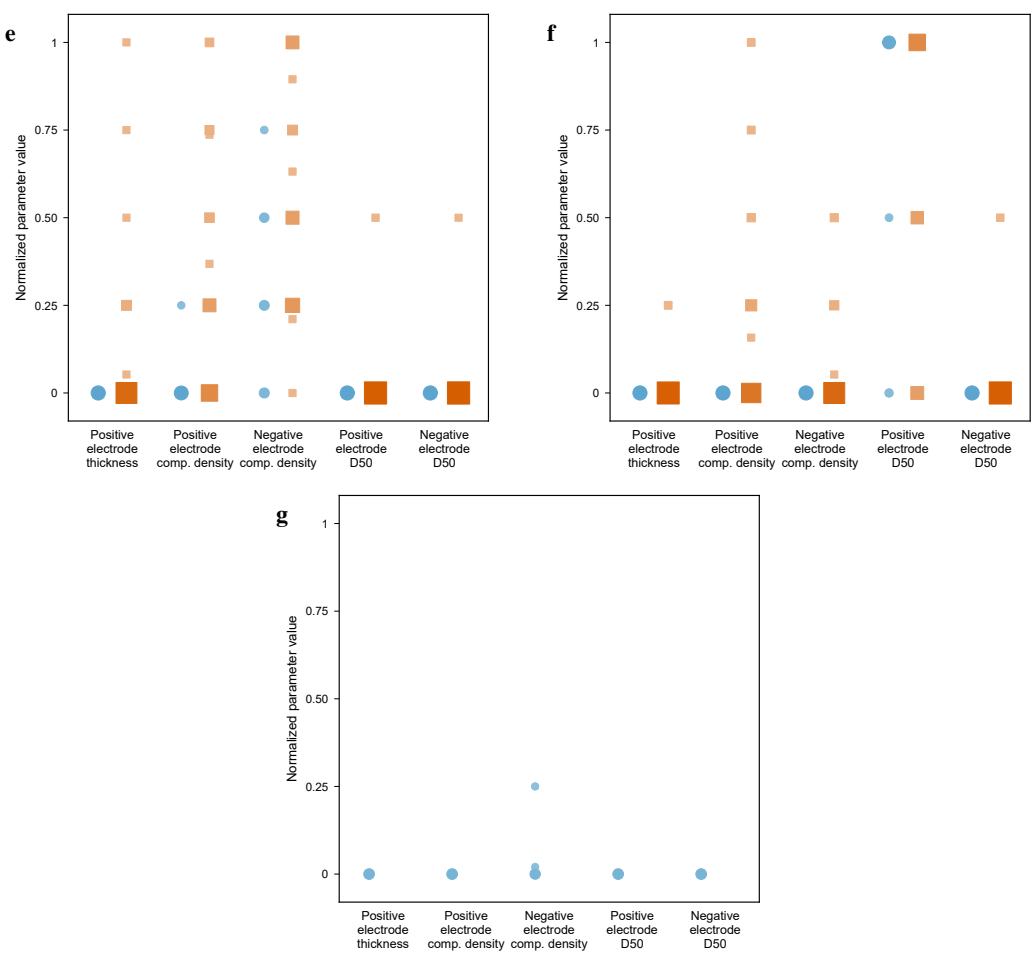


Figure C.5: (e) 5-s Recharge Power at 25 °C and 90% SOC; (f) fast-charging time from 20% to 80% SOC at 25 °C; (g) energy efficiency at 25 °C and 0.5P. The ordinate represents the normalized position of each parameter within the observed range of the corresponding subgroup. Blue circles and orange squares denote LFP and NCM, respectively. Marker size and color intensity indicate the raw occurrence count. Panels (a)–(g) contain optimal records from 307, 277, 276, 307, 277, 277, and 30 subgroups, respectively. Panel (g) contains only LFP.

C.2.3 Sensitivity Analysis of Porous-Electrode Design Parameters Based on Permutation Feature Importance

We trained a separate LightGBM prediction model for each of the 29 performance metrics. The five porous-electrode design parameters were used as numerical inputs, whereas material system and cell model were included as categorical inputs. Across the 29 models, R^2 on the independent test sets ranged from 0.99178 to 0.99997.

Parameter importance was estimated by permutation within each material system–cell model group. During each permutation, only one parameter was randomly rearranged. We defined ΔR^2 as the increase in weighted prediction error caused by the permutation, normalized by the total sum of squares over the complete test set. Each parameter was permuted independently three times, and the resulting values were averaged. All 145 importance values were positive and ranged from approximately 1.543801×10^{-7} to 0.7760984. The colors are mapped on a logarithmic scale, whereas the text in each cell reports the original ΔR^2.

Permutation feature importance describes the dependence of a fixed model on its input information under the current test distribution. It does not provide the positive or negative direction of a parameter's effect and is not equivalent to a local derivative, a Sobol index, statistical significance, or a

causal effect. The parameter importance values are not required to sum to one. Comparisons should primarily be made among parameters within the same performance row. Cross-performance differences in ΔR^2 cannot be used to determine which type of physical performance is "more sensitive."

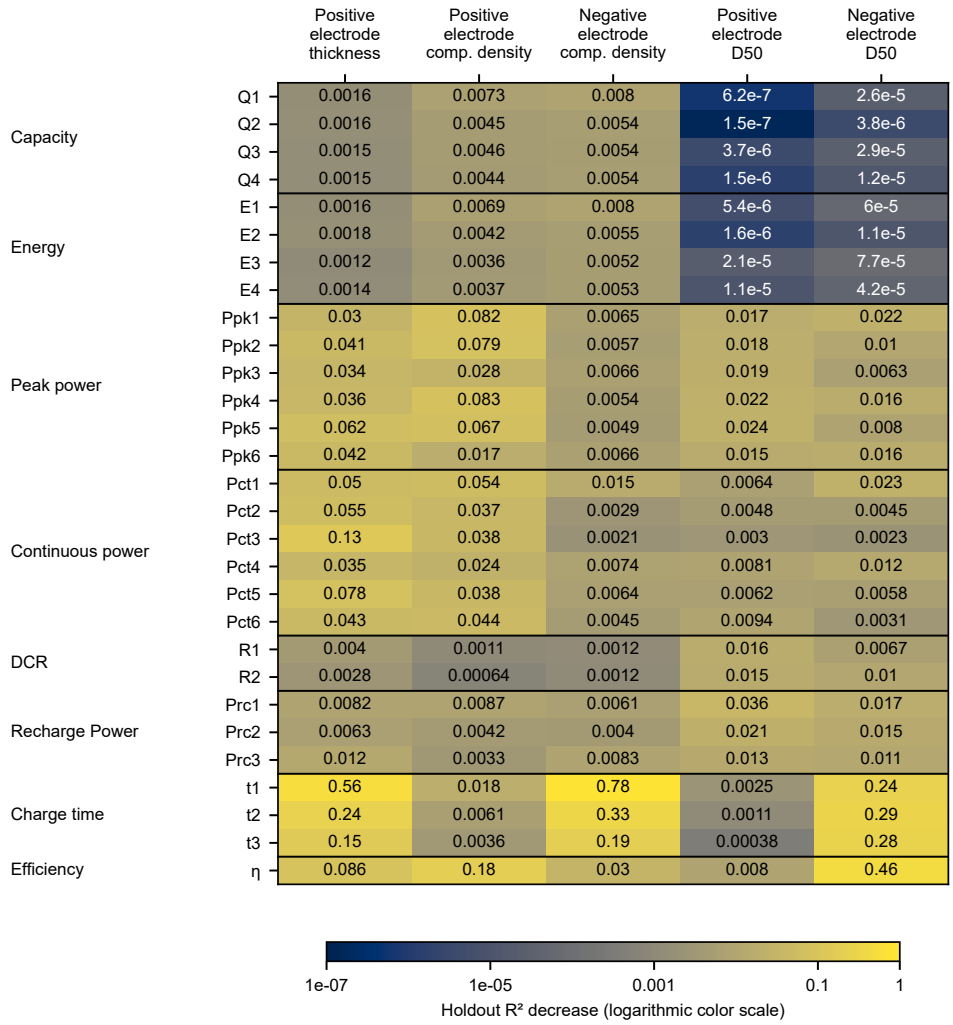


Figure C.6: Sensitivity analysis of porous-electrode design parameters based on permutation feature importance. Rows represent the test conditions for capacity, energy, peak discharge power, continuous discharge power, DCR, Recharge Power, fast-charging time, and energy efficiency. Columns represent positive-electrode coating thickness, positive-electrode compaction density, negative-electrode compaction density, positive-electrode active-material particle size D_{50}, and negative-electrode active-material particle size D_{50}. Cell text reports the original ΔR^2, whereas color is mapped on a logarithmic scale. Larger values indicate a stronger dependence of model predictions on the corresponding parameter, but do not indicate effect direction, local effects, statistical significance, or causality.

D Data Sampling Strategies, Global Regression Baselines, and the Retrieval–Generation Method

This appendix provides implementation details of the retrieval–generation (RG) method, describes the configurations and training settings of the tree-based and neural network supervised learning baselines, and explains how the datasets are constructed using uniform grid sampling and change-point sampling.

D.1 Data Sampling Strategies

D.1.1 Uniform Grid Sampling

For a given combination of cell model and material system, let the five porous electrode design parameters be

$$\mathbf{x} = (x_1, x_2, x_3, x_4, x_5).$$

Let the range of parameter x_j be $[\ell_j, u_j]$. Selecting n_j equally spaced values, including both endpoints, gives the following list of values:

$$\mathcal{X}_j = \left\{ \ell_j + \frac{k}{n_j - 1}(u_j - \ell_j) \;\middle|\; k = 0, \ldots, n_j - 1 \right\}.$$

When a parameter takes only the two endpoints of its range, $n_j = 2$. Once the value lists for all five parameters have been specified, their full Cartesian product defines the design parameter space for that combination:

$$\mathcal{G} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{X}_4 \times \mathcal{X}_5, \qquad |\mathcal{G}| = \prod_{j=1}^{5} n_j.$$

Thus, each design point in the uniform grid corresponds to a specific combination of the five porous electrode design parameters.

For each design point in $\mathcal{G}$, material balance calculations are first performed using the corresponding cell model, material system, and porous electrode design parameters. The performance metrics are then obtained by simulation with the coupled P2D–lumped thermal model. The design parameters, the calculated material balance results, and the performance metrics together form one data record. The results for all design points constitute the corresponding uniform grid dataset. Different cell models and material systems may use their own parameter definitions and ranges while following the same dataset construction procedure.

For a prismatic energy-storage cell with the LFP-3 material system and dimensions of $173 \times 207 \times 71$ mm (W $\times$ H $\times$ T), we construct six uniform grid datasets with different sampling densities, denoted by UG32, UG108, UG500, UG4,000, UG32,000, and UG500,000.

These datasets use five porous electrode design parameters: positive electrode coating thickness t_+, compaction density of the positive electrode ρ_+, compaction density of the negative electrode ρ_-, particle size of the positive electrode active material d_+, and particle size of the negative electrode active material d_-. Their ranges and sampling counts are listed below.

For a given sampling count s, the positive electrode coating thickness and the compaction densities of the positive and negative electrodes each take s equally spaced values including both endpoints. Each of the two particle-size parameters takes the two endpoints of its design range. The resulting full grid is

$$\mathcal{G}_s = \mathcal{T}_s \times \mathcal{R}_{+,s} \times \mathcal{R}_{-,s} \times \mathcal{D}_+ \times \mathcal{D}_-,$$

with the number of design points given by

$$|\mathcal{G}_s| = s^3 \times 2 \times 2 = 4s^3.$$

The sizes of the six uniform grid datasets are listed below.

All six datasets share the same cell configuration, geometric dimensions, material system, design parameter ranges, and computational procedure. They differ only in the sampling counts along

Table D.1: Ranges and sampling counts of porous electrode design parameters

Design parameter	Symbol	Range	Sampling count
Positive electrode coating thickness (μm)	t_+	$[60, 80]$	s equally spaced values including both endpoints
Compaction density of the positive electrode ($\mathrm{g\,cm^{-3}}$)	ρ_+	$[2.3, 2.8]$	s equally spaced values including both endpoints
Compaction density of the negative electrode ($\mathrm{g\,cm^{-3}}$)	ρ_-	$[1.3, 1.75]$	s equally spaced values including both endpoints
Particle size of the positive electrode active material (μm)	d_+	$[0.6, 1.5]$	2
Particle size of the negative electrode active material (μm)	d_-	$[8, 16]$	2

Table D.2: Sampling counts and sizes of the six uniform grid datasets

Dataset	s	Sampling counts for the five parameters	Number of design points
UG32	2	$2 \times 2 \times 2 \times 2 \times 2$	32
UG108	3	$3 \times 3 \times 3 \times 2 \times 2$	108
UG500	5	$5 \times 5 \times 5 \times 2 \times 2$	500
UG4,000	10	$10 \times 10 \times 10 \times 2 \times 2$	4,000
UG32,000	20	$20 \times 20 \times 20 \times 2 \times 2$	32,000
UG500,000	50	$50 \times 50 \times 50 \times 2 \times 2$	500,000

the positive electrode thickness and the positive and negative electrode compaction-density directions. As s increases, the grid spacing decreases along these three continuous parameter directions, whereas each particle-size parameter continues to take only the two endpoints of its range.

D.1.2 Change-Point Sampling

Change-point sampling first constructs a dense regular grid through material balance calculations. These calculations involve only algebraic operations and do not require solving partial differential equations, making them substantially less expensive than the subsequent solution of the coupled P2D–lumped thermal model. A large number of material balance results can therefore be obtained at relatively low computational cost. The grid is constructed in the same way as in the uniform grid sampling procedure in Section D.1.1: equally spaced values including both endpoints are selected within each parameter range, and the full Cartesian product of the corresponding value lists yields a regular grid of material balance results.

Under the material balance relationships used in this study, the porous electrode design parameters that substantially affect the calculated capacity are the positive electrode coating thickness and the compaction densities of the positive and negative electrodes. The active-material particle sizes D50 of the positive and negative electrodes do not affect this capacity. Change-point identification is therefore performed only on the three-dimensional regular grid formed by the former three parameters; the two particle-size parameters are excluded. In this subsection, the complete three-dimensional candidate grid is denoted by $\mathcal{G}$, and each design point is represented by $\mathbf{x} = (x_1, x_2, x_3)$, where x_1, x_2, and x_3 denote the positive electrode coating thickness and the compaction densities of the positive and negative electrodes, respectively.

After obtaining the capacity calculated from material balance at each design point in the dense regular grid, we examine its first-order local variation along the positive electrode coating thickness and the positive and negative electrode compaction-density directions, retaining design points with large capacity changes. After threshold filtering, points on the boundary of the design domain are added. The retained three-parameter combinations are then combined with all prescribed combinations of the positive and negative electrode active-material particle sizes D50 through a Cartesian product, recovering the complete five-dimensional porous electrode design parameters. These design points are then evaluated using the coupled P2D–lumped thermal model.

Here, a "change point" refers to a location where the capacity calculated from material balance varies substantially along a design-variable direction. The term does not refer to a statistical change-point test or a significance test.

To calculate the local capacity variation along a design-variable direction, we fix the other two design variables and vary only the variable corresponding to direction j, extracting a one-dimensional grid slice from the three-dimensional candidate grid. Let the grid points on this slice be ordered by increasing coordinate in direction j:

$$x_{j,0} < x_{j,1} < \cdots < x_{j,n_j-1},$$

The corresponding capacities calculated from material balance are $Q_0, Q_1, \ldots, Q_{n_j-1}$. The absolute slope between adjacent grid points is defined as

$$\delta_{j,i} = \frac{|Q_{i+1} - Q_i|}{|x_{j,i+1} - x_{j,i}|}, \qquad i = 0, \ldots, n_j - 2.$$

To quantify the local variation in capacity near a grid point along direction j, we define the directional change indicator $s_{j,i}$ as

$$s_{j,i} = \begin{cases} \delta_{j,0}, & i = 0, \\ \max\left(\delta_{j,i-1}, \delta_{j,i}\right), & 1 \le i \le n_j - 2, \\ \delta_{j,n_j-2}, & i = n_j - 1. \end{cases}$$

At an interior grid point, the directional change indicator is the larger of the absolute slopes on its two adjacent intervals. At either endpoint of a slice, it is the absolute slope on the single adjacent interval. This indicator describes the first-order local variation in the capacity calculated from material balance along a design-variable direction.

For each design-variable direction, we introduce a dimensionless threshold parameter $u_j \in [0, 1]$ and determine the actual filtering threshold from the minimum and maximum values of the directional change indicator:

$$\tau_j = s_j^{\min} + u_j \left(s_j^{\max} - s_j^{\min}\right),$$

Here, $s_j^{\min}$ and $s_j^{\max}$ are the minimum and maximum directional change indicators in direction j across all candidate grid points. A candidate design point passes threshold filtering if its change indicator is strictly greater than the corresponding threshold in at least one direction. All such points form the initially retained set $\mathcal{S}_{\boldsymbol{\tau}}$:

$$\mathcal{S}_{\boldsymbol{\tau}} = \{\mathbf{x} \in \mathcal{G} \mid \exists j,\ s_j(\mathbf{x}) > \tau_j\}.$$

Threshold selection considers two aspects: reducing the number of retained design points and controlling the error in reconstructing the capacity calculated from material balance using those points. The maximum slice reconstruction error in each design-variable direction and the fraction of candidate design points retained jointly form the objectives for multi-objective threshold selection.

After threshold filtering, all candidate design points on the boundary of the design domain are added to the initially retained set. Denoting the set of boundary points in the candidate grid by $\mathcal{B}$, the final retained set is

$$\mathcal{S}_{\text{final}} = \mathcal{S}_{\boldsymbol{\tau}} \cup \mathcal{B}.$$

Using the LFP-3 cell described in Section D.1.1 as an example, we perform change-point sampling in the three-dimensional space spanned by three continuous design parameters: the positive electrode coating thickness and the compaction densities of the positive and negative electrodes. The points to retain are determined by local variations in the capacity calculated from material balance along each parameter direction.

Each of the three parameters takes 50 equally spaced values including both endpoints. Their full Cartesian product contains

$$50^3 = 125{,}000$$

three-parameter combinations. Material balance calculations are performed for all 125,000 combinations to obtain their corresponding capacities Q, forming the dense regular capacity grid used for change-point identification.

Table D.3: Design parameter ranges and sampling counts of the candidate grid for change-point identification

Design parameter used for change-point identification	Symbol	Range	Sampling count
Positive electrode coating thickness (μm)	t_+	$[60, 80]$	50
Compaction density of the positive electrode ($\mathrm{g\,cm^{-3}}$)	ρ_+	$[2.3, 2.8]$	50
Compaction density of the negative electrode ($\mathrm{g\,cm^{-3}}$)	ρ_-	$[1.3, 1.75]$	50

When constructing this three-dimensional capacity grid, the particle size D50 of the positive electrode active material and the particle size D50 of the negative electrode active material are fixed at $0.6\ \mu$m and $8\ \mu$m, respectively. The directional change indicator has units of $\mathrm{Ah}\,\mu\mathrm{m}^{-1}$ along the positive electrode coating thickness direction and $\mathrm{Ah\,cm^3\,g^{-1}}$ along both electrode compaction-density directions.

For a one-dimensional slice g in direction j with at least three retained points, we construct a piecewise linear interpolation function $\widetilde{Q}_{j,g}$ from those points and use linear extrapolation beyond the first and last retained points. We then calculate the mean squared relative error over all 50 original grid points on the slice:

$$E_{j,g} = \min\left\{ \frac{1}{n_g} \sum_{i \in g} \left[\frac{Q_i - \widetilde{Q}_{j,g}(x_{j,i})}{Q_i + 10^{-9}} \right]^2 , 999000 \right\}.$$

where:

- j denotes the design-variable direction; in this example, $j = 1, 2, 3$ correspond to the positive electrode coating thickness and the compaction densities of the positive and negative electrodes, respectively.
- g denotes a one-dimensional grid slice in direction j, obtained by fixing the other two design variables and varying only the variable corresponding to direction j.
- i denotes an original grid point on slice g, and $\sum_{i \in g}$ sums over all original grid points on that slice.
- n_g is the number of original grid points on slice g. Each slice in this example contains 50 original grid points, so $n_g = 50$.
- $x_{j,i}$ is the value of the design parameter in direction j at grid point i.
- Q_i is the capacity calculated from material balance at original grid point i.
- $\widetilde{Q}_{j,g}(x_{j,i})$ is the capacity estimate at original grid point i, obtained by interpolation or extrapolation from the design points on slice g that pass threshold filtering.
- The term 10^{-9} in the denominator provides numerical stability by preventing division by zero.
- $E_{j,g}$ is the mean squared relative error of slice g in direction j. It is dimensionless and is not multiplied by 100, so it is not a percentage.
- The outer $\min\{\cdot, 999000\}$ caps the MSRE of an individual slice at 999000.

For each design-variable direction, the maximum MSRE over all evaluable slices in that direction is used as the corresponding optimization objective:

$$F_j = \max_{g \in \mathcal{G}_j^\star} E_{j,g}, \qquad j = 1, 2, 3.$$

where:

- $\mathcal{G}_j^\star$ is the set of all evaluable one-dimensional grid slices in direction j.
- We include in $\mathcal{G}_j^\star$ only slices with at least three design points that pass threshold filtering and evaluate them using piecewise linear interpolation and extrapolation.

- F_j is the maximum MSRE over all evaluable slices in direction j, i.e., the worst-case slice reconstruction error in that direction.

In addition to the maximum slice reconstruction errors in the three directions, the fraction of candidate design points retained is included as a fourth objective:

$$F_4 = \frac{|\mathcal{S}_{\boldsymbol{\tau}}|}{|\mathcal{G}|}.$$

where:

- $\mathcal{S}_{\boldsymbol{\tau}}$ is the set of design points initially retained by directional threshold filtering.
- $\mathcal{G}$ is the complete three-dimensional candidate grid.
- $|\mathcal{S}_{\boldsymbol{\tau}}|$ and $|\mathcal{G}|$ denote the numbers of design points in the respective sets.
- F_4 is the fraction of all candidate design points that pass threshold filtering.

The multi-objective search therefore jointly minimizes F_1, F_2, F_3, and F_4. The first three objectives control the worst-case slice reconstruction errors along the three design-variable directions, and the fourth controls the fraction of design points retained.

If a direction has no evaluable slice with at least three retained points, its maximum slice reconstruction error cannot be calculated, and the corresponding objective F_j is set to 10^6. In this example, each design-variable direction has $50 \times 50 = 2{,}500$ one-dimensional slices because each of the other two fixed design variables takes 50 values. All 2,500 slices are checked individually; no random sampling is used.

Thresholds are searched using the NSGA-III implementation in pymoo, optimizing all four objectives simultaneously. To maintain diversity among Pareto candidate solutions in the four-objective space, reference directions are generated using the Das–Dennis simplex lattice method with partition parameter $H = 10$. For four objectives, this produces

$$\binom{10+4-1}{4-1} = 286$$

reference directions, so the NSGA-III population size is set to 286. The search uses simulated binary crossover with a probability of 0.9 and a distribution index of 15, polynomial mutation with a distribution index of 30, and duplicate-candidate elimination. Optimization runs for 50 generations with random seed 1.

The final thresholds are selected manually from the resulting Pareto candidate solutions. The selected normalized threshold parameters and their corresponding actual thresholds are listed below.

Table D.4: Normalized and actual thresholds for change-point identification

Design-variable direction (units of $\boldsymbol{\tau_j}$)	$\boldsymbol{u_j}$	$\boldsymbol{\tau_j}$
Positive electrode coating thickness ($\mathrm{Ah}\ \mu\mathrm{m}^{-1}$)	0.997986	5.6637208931
Compaction density of the positive electrode ($\mathrm{Ah\ cm^3\ g^{-1}}$)	0.974590	256.1451519445
Compaction density of the negative electrode ($\mathrm{Ah\ cm^3\ g^{-1}}$)	0.626056	264.1262302639

The slice evaluation results for the selected thresholds, before adding boundary points, are listed below.

After threshold filtering, all design points on the boundary of the three-dimensional candidate grid are added to the initially retained set. For direction $j \in \{1, 2, 3\}$, $x_j^{\min}$ and $x_j^{\max}$ denote the minimum and maximum values of the corresponding design parameter in the candidate grid. The boundary-point set is defined as

$$\mathcal{B} = \left\{ \mathbf{x} \in \mathcal{G} \;\middle|\; \exists j \in \{1,2,3\}, \quad x_j \in \left\{ x_j^{\min}, x_j^{\max} \right\} \right\}.$$

Table D.5: Slice evaluation results before adding boundary points

Design-variable direction	Slices included in MSRE calculation	Slices with no retained points	Slices with only 1–2 retained points	Maximum slice MSRE
Positive electrode coating thickness	2,500	0	0	7.85553×10^{-5}
Compaction density of the positive electrode	2,500	0	0	2.09036×10^{-4}
Compaction density of the negative electrode	2,490	9	1	4.03657×10^{-3}

Under this definition, a candidate design point belongs to $\mathcal{B}$ if any of its parameters equals that parameter's minimum or maximum value. Each of the three design parameters has a minimum and a maximum boundary, giving six boundary faces of the three-dimensional candidate grid. Intersections and vertices shared by multiple faces are counted only once in $\mathcal{B}$.

The candidate grid in this example has 50 sampled values along each direction, giving $50^3 = 125{,}000$ design points in total. Removing the outer boundary leaves 48 sampled values along each direction of the interior grid. The total number of boundary points is therefore

$$|\mathcal{B}| = 50^3 - 48^3 = 14{,}408.$$

Let $\mathcal{S}_{\boldsymbol{\tau}}$ denote the design points initially retained by directional threshold filtering. The threshold rule retains 59,388 three-parameter combinations, corresponding to the following fraction of all candidate design points:

$$\frac{59{,}388}{125{,}000} = 47.5104\%.$$

Of the 14,408 boundary points, 5,641 already pass threshold filtering and need not be added again. The remaining

$$14{,}408 - 5{,}641 = 8{,}767$$

boundary points are added to the initially retained set.

The final number of retained three-parameter combinations is

$$\begin{aligned} |\mathcal{S}_{\text{final}}| &= |\mathcal{S}_{\boldsymbol{\tau}}| + |\mathcal{B} \setminus \mathcal{S}_{\boldsymbol{\tau}}| \\ &= 59{,}388 + 8{,}767 \\ &= 68{,}155. \end{aligned}$$

These 68,155 three-parameter combinations account for the following fraction of the complete candidate grid:

$$\frac{68{,}155}{125{,}000} = 54.524\%.$$

The maximum slice MSRE values reported above are calculated using the initially retained set $\mathcal{S}_{\boldsymbol{\tau}}$, before adding boundary points. These objectives are not recalculated using the final retained set $\mathcal{S}_{\text{final}}$. They therefore describe only the capacity reconstruction error at the threshold-filtering stage and cannot be interpreted as the reconstruction error of the final retained set after boundary points are added.

The active-material particle sizes D50 of the positive and negative electrodes are not used to calculate the directional change indicators of the capacity obtained from material balance. After three-dimensional change-point filtering and the addition of boundary points, each of the 68,155 retained three-parameter combinations is combined with all prescribed values of the positive and negative electrode active-material particle sizes D50. Let their respective value sets be

$$\mathcal{D}_+ = \{0.6, 1.5\}\ \mu\text{m}, \qquad \mathcal{D}_- = \{8, 16\}\ \mu\text{m}.$$

Here, $\mathcal{D}_+$ contains two particle-size values for the positive electrode active material, and $\mathcal{D}_-$ contains two for the negative electrode active material. Their Cartesian product gives $2 \times 2 = 4$ combinations of positive and negative electrode particle sizes. The final five-dimensional design parameter

set is

$$\mathcal{G}_{\mathrm{CPS}} = \mathcal{S}_{\mathrm{final}} \times \mathcal{D}_{+} \times \mathcal{D}_{-}.$$

The total number of five-dimensional design points is therefore

$$\begin{aligned} |\mathcal{G}_{\mathrm{CPS}}| &= |\mathcal{S}_{\mathrm{final}}| \times |\mathcal{D}_{+}| \times |\mathcal{D}_{-}| \\ &= 68{,}155 \times 2 \times 2 \\ &= 272{,}620. \end{aligned}$$

Each five-dimensional design point specifies the positive electrode coating thickness, the compaction densities of the positive and negative electrodes, the particle size D50 of the positive electrode active material, and the particle size D50 of the negative electrode active material. The same material balance calculations and coupled P2D–lumped thermal model used for the uniform grid datasets are applied to these design points to obtain the corresponding cell performance metrics, forming the CPS272,620 dataset.

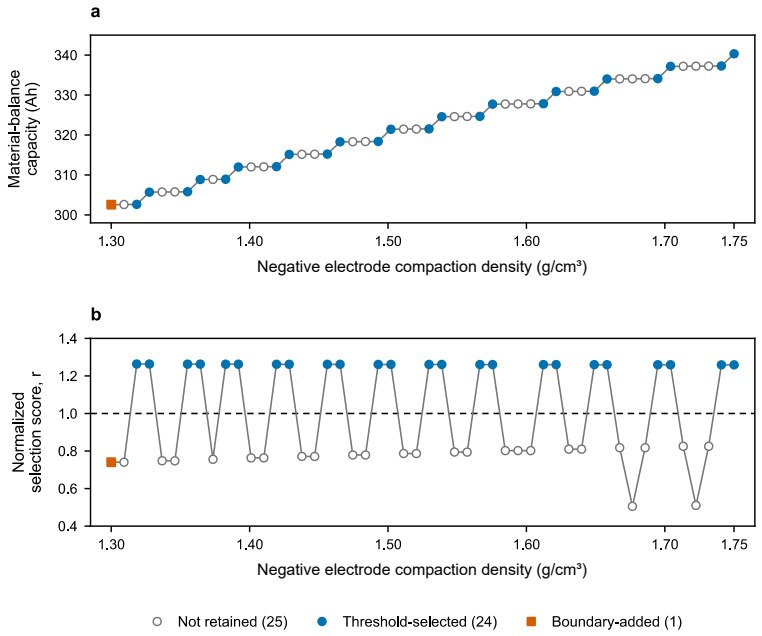


Figure D.1: Example of change-point sampling based on variations in the capacity calculated from material balance. (a) Capacities at the 50 candidate design points along the negative electrode compaction-density direction, together with their retention status after threshold filtering and the addition of boundary points. (b) Maximum normalized change indicator across the three design-variable directions, $r = \max_j(s_j/\tau_j)$; the horizontal dashed line marks the filtering boundary $r = 1$. Filled blue circles indicate the 24 design points retained by the strict threshold rule, the orange square indicates the 1 additional point that fails threshold filtering but is retained by the boundary-point rule, and open gray circles indicate the 25 points not retained. The positive electrode coating thickness and compaction density are fixed at $69.795918\ \mu\mathrm{m}$ and $2.544898\ \mathrm{g\,cm^{-3}}$, respectively. The particle size D50 of the positive electrode active material and the particle size D50 of the negative electrode active material are fixed at $0.6\ \mu\mathrm{m}$ and $8\ \mu\mathrm{m}$, respectively.

Figure D.1 shows a one-dimensional grid slice along the negative electrode compaction-density direction. The slice contains 50 candidate design points. The compaction density of the negative electrode varies across its range, while the positive electrode coating thickness and the compaction density of the positive electrode are fixed, respectively, at

$$69.795918\ \mu\mathrm{m} \qquad \text{and} \qquad 2.544898\ \mathrm{g\,cm^{-3}}.$$

When constructing this slice of capacity calculated from material balance, the particle size D50 of the positive electrode active material and the particle size D50 of the negative electrode active material are fixed at $0.6\ \mu\mathrm{m}$ and $8\ \mu\mathrm{m}$, respectively.

The change indicators along the three design-variable directions have different numerical ranges and units. To determine on a common scale whether a candidate design point exceeds the threshold in any direction, we define the maximum normalized change indicator as

$$r(\mathbf{x}) = \max_{j\in\{1,2,3\}} \frac{s_j(\mathbf{x})}{\tau_j},$$

Here, $s_j(\mathbf{x})$ is the directional change indicator at design point $\mathbf{x}$ in direction j, and τ_j is the corresponding filtering threshold. The ratio $s_j(\mathbf{x})/\tau_j$ expresses the change indicator in direction j relative to its threshold, and $r(\mathbf{x})$ is the maximum of these ratios across the three directions.

When $r(\mathbf{x}) > 1$, the design point has a change indicator strictly greater than its threshold in at least one direction and therefore passes threshold filtering. When $r(\mathbf{x}) \leq 1$, it does not satisfy the filtering condition. Thus, the strict threshold conditions across the three directions can be expressed jointly as

$$r(\mathbf{x}) > 1.$$

In the slice shown in Figure D.1, 24 candidate design points pass threshold filtering. All of these points have change indicators in the negative electrode compaction-density direction that exceed the corresponding threshold. The lower endpoint has a maximum normalized change indicator of $r = 0.740270$ and thus does not pass threshold filtering. However, because it lies on the boundary of the design domain, it is subsequently added to the final retained set by the boundary-point rule. The upper endpoint already passes threshold filtering and need not be added again.

After threshold filtering and the addition of boundary points, 25 of the 50 three-parameter combinations on this slice are retained and the other 25 are not. Pairing each retained combination with the four combinations of positive and negative electrode active-material particle sizes D50 means that these 25 three-parameter combinations correspond to

$$25 \times 4 = 100$$

five-dimensional porous electrode design points.

This slice is selected from three prespecified central-axis slices to illustrate the effects of both threshold filtering and boundary-point inclusion in a single example. In addition, CPS272,620 differs from the uniform grid datasets in both sampling strategy and dataset size.

D.2 Details of the Global Regression Baselines

This section evaluates two categories of global regression baselines, tree-based models and neural networks, for cell performance prediction and specifies the models used in subsequent experiments. We compare XGBoost, LightGBM, CatBoost, and Random Forest among the tree-based models. For neural networks, we compare a multilayer perceptron (MLP) and a residual multilayer perceptron (Residual MLP), with a linear model containing no hidden layers or nonlinear activation functions as a control. Here, "global" means that a model is trained on the entire given training dataset and used to predict all test samples; RG instead retrieves a local neighborhood and fits a local surrogate model for each test sample. The two model categories are trained and evaluated separately. CatBoost and MLP are used as the tree-based and neural network baselines, respectively, in subsequent experiments.

All experiments in this section use the LFP-3 cell described in Section D.1.1, with the five porous electrode design parameters as model inputs. The four prediction targets are discharge capacity at $25\,^{\circ}\mathrm{C}$ and $1/3\,\mathrm{C}$; discharge energy at $25\,^{\circ}\mathrm{C}$ and $1/3\,\mathrm{C}$; discharge DCR at $25\,^{\circ}\mathrm{C}$, 50% SOC, 2C, and 10 s; and charge–discharge energy efficiency at $25\,^{\circ}\mathrm{C}$ and 0.5P. Each tree-based model is trained independently for each of the four outputs, whereas the linear model, MLP, and Residual MLP each predict all four performance metrics jointly. All models use the same LFP-3 test set, containing 1,000 porous electrode designs and their corresponding high-fidelity simulation results. Mean absolute percentage error (MAPE) is used to evaluate each performance metric on a common basis.

The mean absolute percentage error for output j is defined as

$$\mathrm{MAPE}_j = \frac{100}{n} \sum_{i=1}^{n} \frac{|\hat{y}_{ij} - y_{ij}|}{|y_{ij}| + \varepsilon},$$

Here, $n = 1{,}000$, and y_{ij} and $\hat{y}_{ij}$ are the high-fidelity simulation value and model prediction, respectively, for output j of porous electrode design i in the test set. The denominator term ε prevents division by zero. In the tree-based model experiments, ε is set to 10^{-12}; in the neural network and linear model experiments, ε is set to 10^{-8}. We also report Macro-3 MAPE, defined as the arithmetic mean of the MAPEs for capacity, energy, and DCR. Energy efficiency is excluded from this average, and its MAPE is reported separately.

D.2.1 Selection of the Tree-Based Baseline

XGBoost, LightGBM, CatBoost, and Random Forest all use the UG500,000 dataset for the LFP-3 cell. Raw input values are used directly, without input normalization or target standardization.

Hyperparameters are searched independently for each output of each of the four tree-based models. Each model–output combination is configured for 200 Optuna trials, using RMSE from shuffled five-fold cross-validation as the selection metric. The fold-splitting seed is 42, and at most 10,000 validation samples are used in each fold. The search ranges are listed below.

Table D.6: Hyperparameter search ranges for the four tree-based models

Model	Parameter	Candidate range or values
XGBoost	`max_depth`	$\{3, \ldots, 20\}$
	`learning_rate`	$[10^{-3}, 0.3]_{\log}$
	`min_child_weight`	$[1, 20]$
	`subsample`	$[0.5, 1]$
	`colsample_bytree`	$[0.5, 1]$
	`gamma`	$[0, 10]$
	`reg_alpha`	$[10^{-8}, 10]_{\log}$
	`reg_lambda`	$[10^{-8}, 10]_{\log}$
LightGBM	`max_depth`	$\{1, \ldots, 20\}$
	`num_leaves`	$\{31, \ldots, U(d)\}$
	`learning_rate`	$[10^{-3}, 0.3]_{\log}$
	`min_child_samples`	$\{5, \ldots, 100\}$
	`feature_fraction`	$[0.5, 1]$
	`bagging_fraction`	$[0.5, 1]$
	`lambda_l1`	$[10^{-8}, 10]_{\log}$
	`lambda_l2`	$[10^{-8}, 10]_{\log}$
CatBoost	`depth`	$\{4, \ldots, 10\}$
	`learning_rate`	$[10^{-3}, 0.3]_{\log}$
	`l2_leaf_reg`	$[10^{-3}, 100]_{\log}$
	`bagging_temperature`	$[0, 10]$
	`random_strength`	$[0, 10]$
	`min_data_in_leaf`	$\{1, \ldots, 100\}$
	`border_count`	$\{32, \ldots, 255\}$
	`rsm`	$[0.5, 1]$
Random Forest	`n_estimators`	$\{200, \ldots, 2000\}$
	`max_depth`	$\{5, \ldots, 50\}$
	`min_samples_leaf`	$\{1, \ldots, 20\}$
	`max_features`	$\{\texttt{sqrt}, \texttt{log2}, 0.5, 1.0\}$

Note. $\{a, \ldots, b\}$: discrete integer candidates; $[a, b]$: a continuous interval on a linear scale; $[a, b]_{\log}$: a continuous interval on a logarithmic scale. For LightGBM, $d = \texttt{max_depth}$ and $U(d) = \max(31, \min(1023, 2^d))$. The initial iteration limits are 3,000 for XGBoost and LightGBM and 5,000 for CatBoost, with an early-stopping patience of 80 for all three. The three boosted-tree models further compare tree counts at 0.9, 1.0, and 1.1 times the base tree count. Random Forest uses at most 500 trees during cross-validation.

The main results are summarized below.

LightGBM, CatBoost, and Random Forest have similar prediction errors, whereas XGBoost has substantially higher DCR prediction error. Random Forest achieves slightly lower errors for capacity, energy, and DCR, but its four model files total 135.357 GiB, substantially exceeding the 0.299

Table D.7: Test-set prediction errors and model file sizes of the four tree-based models

Model	Capacity MAPE (%)	Energy MAPE (%)	DCR MAPE (%)	Energy efficiency MAPE (%)	Macro-3 MAPE (%)	Total model file size (GiB)
XGBoost	0.329	0.330	11.677	0.730	4.112	1.162
LightGBM	0.239	0.282	4.256	**0.322**	1.592	**0.299**
CatBoost	0.253	0.275	4.257	0.322	1.595	0.336
Random Forest	**0.237**	**0.266**	**4.253**	0.322	**1.585**	135.357

Note. The capacity, energy, DCR, energy efficiency, and Macro-3 columns report MAPE in %. Capacity and energy refer, respectively, to discharge capacity and discharge energy, both at 25 °C and 1/3 C. DCR denotes discharge direct-current resistance at 25 °C, 50% SOC, 2C, and 10 s. Energy efficiency refers to charge–discharge energy efficiency at 25 °C and 0.5P. Macro-3 is the arithmetic mean of the MAPEs for capacity, energy, and DCR, excluding energy efficiency. Total model file size is the combined disk space occupied by the four trained model files corresponding to the four outputs, where 1 GiB $= 2^{30}$ bytes. Bold values indicate the lowest value in each numeric column.

GiB of LightGBM and 0.336 GiB of CatBoost. LightGBM and CatBoost differ only slightly in prediction error and model file size. Considering prediction accuracy, model storage requirements, and CatBoost's support for categorical features, we select CatBoost as the tree-based baseline for subsequent experiments.

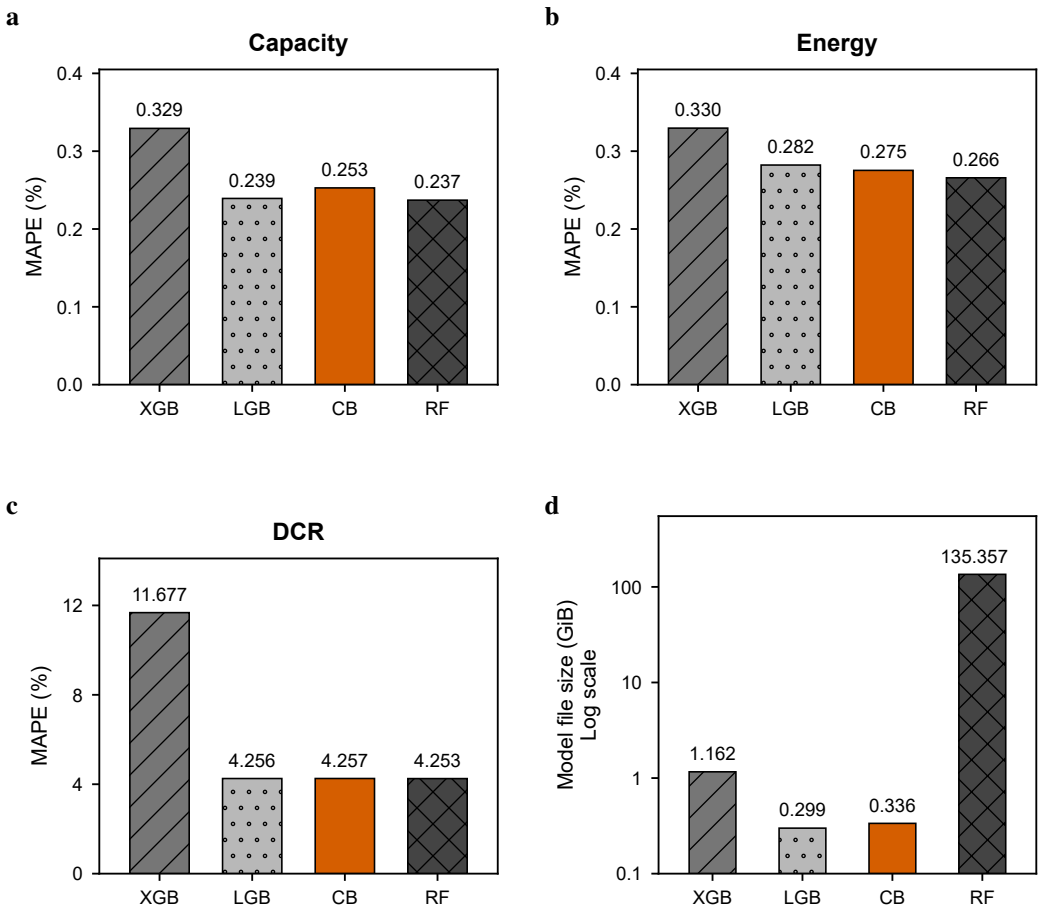


Figure D.2: Prediction errors and model file sizes of the four tree-based models trained on the LFP-3 UG500,000 dataset. (a) Capacity MAPE; (b) energy MAPE; (c) DCR MAPE; (d) combined disk space occupied by the trained model files, shown on a logarithmic vertical axis. XGB, LGB, CB, and RF denote XGBoost, LightGBM, CatBoost, and Random Forest, respectively.

D.2.2 CatBoost Hyperparameter Selection and Training on Seven Datasets

Hyperparameters are selected independently for each of the seven datasets, which differ in density and sampling strategy. The cross-validation and search settings are adapted to the number of samples.

Table D.8: CatBoost hyperparameter search settings for the seven datasets

Dataset	Five-fold repetitions	Optuna trials per output	Depth search range	Maximum training iterations during search	Early-stopping patience (iterations)
UG32	5	100	1–5	2,000	15
UG108	3	100	1–6	3,000	20
UG500	2	100	2–8	5,000	30
UG4,000	1	150	3–10	8,000	50
UG32,000	1	200	4–12	10,000	80
CPS272,620	1	200	4–12	12,000	200
UG500,000	1	200	4–12	15,000	200

Note. Five-fold repetitions specify how many times the full five-fold cross-validation procedure is run. For example, UG32 uses 5 repetitions, yielding 25 training–validation splits; a repetition count of 1 means that five-fold cross-validation is run once. All cross-validation splits use random seed 42. The maximum number of training iterations and early-stopping patience apply only to hyperparameter selection.

In addition to the depth ranges listed in the table, CatBoost uses the following common search domains across the seven datasets: a learning rate of 0.001–0.3 and `l2_leaf_reg` of 0.001–100, both on logarithmic scales; `bagging_temperature` and `random_strength` of 0–10; `border_count` of 32–255; and `rsm` of 0.5–1.0. The lower bound of `min_data_in_leaf` is 1 for all datasets, and its upper bounds for UG32, UG108, UG500, UG4,000, UG32,000, CPS272,620, and UG500,000 are 3, 5, 10, 30, 50, 100, and 100, respectively. All models use RMSE loss, Bayesian bootstrap, SymmetricTree, and random seed 42. `min_data_in_leaf` is a recorded search field; its presence does not establish that it imposes an effective constraint under SymmetricTree.

For UG4,000 and smaller datasets, all samples in each validation fold are used. For UG32,000, each validation fold uses at most 5,000 samples; for CPS272,620 and UG500,000, the limit is 10,000 samples. Because the number of cross-validation repetitions, Optuna trials, and maximum iterations varies with dataset size, the search budgets differ across datasets.

After hyperparameter search, the median of the best iteration counts across validation folds is taken as the base tree count. UG32, UG108, UG500, and UG4,000 use this count directly. For UG32,000, CPS272,620, and UG500,000, three-fold cross-validation further compares 0.9, 1.0, and 1.1 times the base count to determine the final tree count. Thus, the final tree counts for these three larger datasets may exceed the maximum number of training iterations allowed during search, as listed in the table. The final models are refitted on all training data in the corresponding dataset using the selected tree counts, without early stopping.

The UG500,000 CatBoost model in Section D.2.1 and the UG500,000 CatBoost model in this subsection use different hyperparameter search and tree-count selection procedures and belong to separate experiments.

D.2.3 Selection of the Neural Network Baseline

We train MLP and Residual MLP on the LFP-3 UG500,000 dataset and compare their performance, with a linear model as a control. All three models take the five porous electrode design parameters as inputs and jointly predict the four performance metrics specified at the beginning of Section D.2.

The linear model consists of a single $5 \to 4$ fully connected layer with bias and has 24 trainable parameters. It is optimized numerically using AdamW rather than solved by closed-form least squares. Here, "linear model" refers to an affine mapping from inputs to outputs.

The MLP has five hidden layers, each with width 128. Each hidden layer consists of a fully connected layer, a GELU activation, and Dropout in sequence. A final $128 \to 4$ output layer predicts the four performance metrics, giving a total of 67,332 trainable parameters. Residual MLP first maps the inputs to a 64-dimensional feature space through a $5 \to 64$ fully connected layer, followed by four residual blocks. Each block expands the feature dimension from 64 to 128, applies SiLU

activation and Dropout, maps back to 64 dimensions, adds the block input, and applies SiLU again. A final $64 \rightarrow 4$ fully connected layer produces the predictions, giving a total of 66,948 trainable parameters.

All three models scale the input features to $[-1, 1]$ and standardize the four outputs separately. Training uses AdamW with cosine learning-rate decay. All final models are trained in FP32, without BatchNorm or LayerNorm. Each model uses its own selected loss function, batch size, and number of training epochs, so the results below reflect overall performance under the respective training settings.

The final training settings of the three models are listed below.

Table D.9: Final training settings of Linear, MLP, and Residual MLP

Model	Loss function	Batch size	Final training epochs	Hyperparameter search
Linear	MSE	4,096	100	Not performed
MLP	MSE	4,096	200	200 attempted; 95 completed, 105 pruned
Residual MLP	Smooth L1	1,024	200	150 attempted; 43 completed, 107 pruned

Model	Initial learning rate	Weight decay coefficient	Dropout
Linear	1.000×10^{-3}	1.000×10^{-5}	0
MLP	1.858×10^{-3}	7.542×10^{-6}	2.664×10^{-5}
Residual MLP	4.46061×10^{-4}	2.72889×10^{-6}	0.092466

Note. The initial learning rate and weight decay coefficient are AdamW parameters, and Dropout denotes the configured probability. Linear uses a fixed configuration. For MLP and Residual MLP, "completed/pruned" refers to the recorded Optuna trial states. Only one final training result is saved for each configuration.

Table D.10: Prediction errors and trainable parameter counts of the linear model, MLP, and Residual MLP

Model	Trainable parameters	Capacity MAPE (%)	Energy MAPE (%)	DCR MAPE (%)	Energy efficiency MAPE (%)	Macro-3 MAPE (%)
Linear	**24**	0.323	0.322	0.756	0.061	0.467
MLP	67,332	0.272	0.276	0.884	0.067	0.477
Residual MLP	66,948	**0.255**	**0.253**	**0.733**	**0.050**	**0.414**

Note. All models are trained on the LFP-3 UG500,000 dataset and evaluated on the same test set of 1,000 LFP-3 cell designs. Macro-3 MAPE is the arithmetic mean of the MAPEs for capacity, energy, and DCR, excluding energy efficiency. Bold values indicate the lowest value in each numeric column.

Residual MLP achieves the lowest MAPE for all four performance metrics and the lowest Macro-3 MAPE. MLP has lower capacity and energy prediction errors than the linear model, but higher DCR and energy efficiency errors, resulting in a slightly higher Macro-3 MAPE. The three models differ in loss function, batch size, hyperparameter search settings, and number of training updates. This comparison therefore does not constitute a controlled ablation in which only the network architecture changes. The subsequent experiments on the seven datasets use a standard MLP as the neural network baseline, with its depth, width, and training configuration adjusted separately for each dataset, as detailed in Section D.2.4.

D.2.4 MLP Hyperparameter Selection and Training on Seven Datasets

Hyperparameters are selected independently for each of the seven datasets, which differ in data density and sampling strategy. The training–validation split ratios and hyperparameter search ranges are set according to dataset size. The search variables include the number and width of hidden layers, activation function, normalization method, loss function, batch size, learning rate, weight decay coefficient, and Dropout probability. The main selected network and training configurations are listed below.

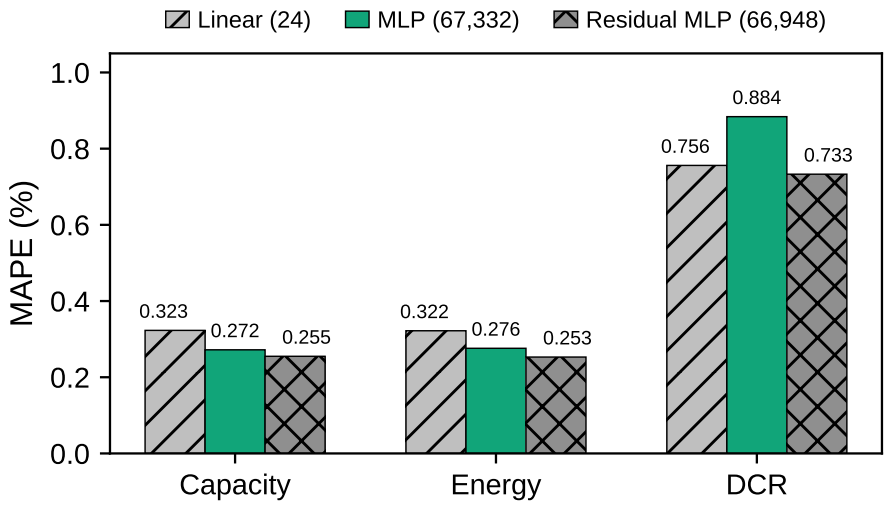


Figure D.3: Performance prediction errors of the linear model, MLP, and Residual MLP. The number in parentheses after each model name is its trainable parameter count. Each bar shows the model's MAPE (%) for the corresponding performance metric.

Table D.11: Final MLP architectures and training configurations for the seven datasets

Dataset	Hidden-layer configuration (layers × width)	Trainable parameters	Activation function	Loss function	Batch size	Final training epochs
UG32	1×32	324	GELU	MSE	2	607
UG108	1×64	644	GELU	MSE	8	537
UG500	1×128	1,284	GELU	Smooth L1	16	380
UG4,000	2×128	17,796	SiLU	MSE	256	230
UG32,000	2×256	68,356	GELU	Smooth L1	1,024	227
CPS272,620	4×128	50,820	GELU	Smooth L1	1,024	195
UG500,000	5×128	67,332	GELU	MSE	4,096	200

Note. Batch size is the number of samples used for each parameter update. None of the seven final configurations uses BatchNorm or LayerNorm.

The preceding MLP configuration table specifies the final architectures, loss functions, batch sizes, and training epochs. The corresponding AdamW initial learning rates, weight decay coefficients, and Dropout probabilities are listed below.

Table D.12: Learning rates and regularization settings of the final MLPs for the seven datasets

Dataset	Initial learning rate	Weight decay coefficient	Dropout
UG32	7.819×10^{-3}	6.480×10^{-2}	6.509×10^{-5}
UG108	4.146×10^{-3}	1.260×10^{-3}	7.167×10^{-2}
UG500	4.176×10^{-3}	3.916×10^{-5}	5.315×10^{-2}
UG4,000	5.723×10^{-4}	1.971×10^{-5}	3.297×10^{-2}
UG32,000	1.421×10^{-3}	1.685×10^{-5}	2.698×10^{-2}
CPS272,620	1.971×10^{-4}	1.272×10^{-3}	7.493×10^{-5}
UG500,000	1.858×10^{-3}	7.542×10^{-6}	2.664×10^{-5}

Note. Learning rate denotes the initial AdamW learning rate, which subsequently decays according to a cosine schedule. The weight decay coefficient is the decoupled weight decay parameter of AdamW. Dropout denotes the configured probability. Values are rounded. The architectures, loss functions, batch sizes, and final training epochs are given in the adjacent table of final MLP configurations.

For UG32, UG108, and UG500, 30% of the data are randomly held out for validation; for each of the other four datasets, 20% are held out. All splits use random seed 42. Optuna performs at most 200 trials per dataset, with MedianPruner terminating some unpromising configurations early. The recorded number of successfully completed trials ranges from 60 to 121 across datasets.

The network architecture, batch size, training limits, and regularization search ranges used for the seven datasets are listed below.

Table D.13: MLP hyperparameter search ranges and training limits for the seven datasets

(a) Network architecture and batch size

Dataset	Number of hidden layers	Hidden-layer width candidates	Batch size candidates
UG32	$\{1, 2\}$	$\{8, 16, 32\}$	$\{2, 4\}$
UG108	$\{1, 2, 3\}$	$\{16, 32, 64, 128\}$	$\{4, 8, 16\}$
UG500	$\{1, \ldots, 4\}$	$\{32, 64, 128, 256\}$	$\{16, 32, 64, 128\}$
UG4,000	$\{2, 3, 4\}$	$\{128, 256, 512\}$	$\{256, 512, 1024\}$
UG32,000	$\{2, 3, 4\}$	$\{256, 512\}$	$\{512, 1024, 2048\}$
CPS272,620	$\{3, \ldots, 10\}$	$\{64, 128, 256, 512, 1024, 2048\}$	$\{1024, 2048, 4096, 8192\}$
UG500,000	$\{3, \ldots, 10\}$	$\{64, 128, 256, 512, 1024, 2048\}$	$\{1024, 2048, 4096, 8192\}$

(b) Training limits and regularization ranges

Dataset	Maximum training epochs	Early-stopping patience	Dropout range	Weight decay coefficient range
UG32	1,000	80	$[0, 0.7]$	$[10^{-3}, 10^{-1}]_{\log}$
UG108	800	50	$[0, 0.5]$	$[10^{-4}, 10^{-1}]_{\log}$
UG500	500	30	$[0, 0.3]$	$[10^{-5}, 10^{-1}]_{\log}$
UG4,000	300	20	$[0, 0.2]$	$[10^{-6}, 10^{-2}]_{\log}$
UG32,000	250	20	$[0, 0.15]$	$[10^{-6}, 10^{-2}]_{\log}$
CPS272,620	200	15	$[0, 0.1]$	$[10^{-6}, 10^{-2}]_{\log}$
UG500,000	200	15	$[0, 0.1]$	$[10^{-6}, 10^{-2}]_{\log}$

(c) Candidate ranges shared by all datasets

Parameter	Candidate range or values
Normalization method	None, `BatchNorm1d`, `LayerNorm`
Activation function	GELU, SiLU, ReLU
Loss function	MSE, Smooth L1
Initial learning rate	$[10^{-4}, 10^{-2}]_{\log}$

Note. $\{a, \ldots, b\}$: discrete integer candidates; $[a, b]$: a continuous interval on a linear scale; $[a, b]_{\log}$: a continuous interval on a logarithmic scale. Validation is performed every five training epochs. Early-stopping patience counts consecutive validation checks without a lower validation RMSE. These settings apply only to hyperparameter search.

Automatic mixed precision (AMP) is used for training during hyperparameter search.

During model training, MSE or Smooth L1 loss is calculated on the four standardized outputs. To evaluate a hyperparameter configuration, predictions and reference values are first transformed back to their original physical scales. The errors of all four outputs over the validation set are then combined to calculate the root mean squared error:

$$R = \sqrt{\frac{1}{4n_{\mathrm{val}}} \sum_{i=1}^{n_{\mathrm{val}}} \sum_{j=1}^{4} (y_{ij} - \hat{y}_{ij})^2},$$

Here, n_{val} is the number of validation samples, and y_{ij} and $\hat{y}_{ij}$ are the reference value and prediction, respectively, for output j of validation sample i. For datasets without a model-size penalty, Optuna selects candidate configurations directly using R.

For UG32, UG4,000, and UG32,000, hyperparameter search instead uses a selection metric penalized by the parameter count:

$$R_{\mathrm{sel}} = R \left[1 + 0.15 \max \left(0, \log_{10} \frac{P}{N_{\mathrm{train}}} \right) \right],$$

Here, P is the number of trainable parameters, and N_{train} is the number of training samples actually used during hyperparameter search. A penalty is applied only when $P > N_{\mathrm{train}}$. The metric R_{sel}

is used for Optuna trial pruning and ranking candidate configurations. It is not added to the training loss and thus is not a regularization term in model fitting.

For each dataset, all successfully completed trials are first ranked. Trials for UG108, UG500, CPS272,620, and UG500,000 are ranked in ascending order of R, whereas those for UG32, UG4,000, and UG32,000 are ranked in ascending order of the parameter-count-penalized metric R_{sel}. Lower values are preferred in both cases. The top ten trials are then selected. Regardless of whether trials are ranked by R or R_{sel}, the best training epoch of each trial is defined as the epoch at which the combined validation RMSE R is minimized. The median of these ten epoch counts is used as the number of epochs for retraining the model on the full dataset.

After selecting the network architecture and training configuration, the final model is retrained in FP32 on all data in the corresponding dataset, without a validation split or early stopping.

Because R directly combines errors from four outputs with different physical units and numerical scales, outputs with larger numerical scales have greater influence on this metric. Therefore, R and R_{sel} are used only for hyperparameter selection and should not be interpreted as equally weighted aggregate measures of the four performance metrics. Final prediction performance is compared using per-metric MAPE and Macro-3 MAPE.

D.3 Details of the Retrieval–Generation Method

We define a combination of one cell model and one material system as configuration c, with dataset

$$\mathcal{D}_c = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N},$$

Here, $\mathbf{x}_i \in \mathbb{R}^5$ contains, in order, the positive electrode coating thickness, the compaction densities of the positive and negative electrodes, and the active-material particle sizes D_{50} of the positive and negative electrodes. The corresponding performance metrics are represented by the vector $\mathbf{y}_i \in \mathbb{R}^{m_c}$, where m_c is the number of performance metrics for configuration c.

For forward prediction, the porous electrode design to be evaluated is represented as a 5-dimensional query vector, and the K records with design parameters closest to that vector are retrieved from $\mathcal{D}_c$. For inverse design, the target performance metrics form the query vector, and the K records whose performance metrics are closest to the targets are retrieved from $\mathcal{D}_c$.

Before K-nearest-neighbor retrieval, the minimum and maximum of each variable in $\mathcal{D}_c$ are used to apply min–max normalization to both the query vector and the variables in $\mathcal{D}_c$ used to calculate the Canberra distance. Forward prediction normalizes the five porous electrode design variables, whereas inverse design normalizes the performance metrics. Canberra distances between the query vector and the records in $\mathcal{D}_c$ are then calculated in the corresponding normalized space, and the K records with the smallest distances are selected. If the total dataset size N is smaller than K, all N records are returned, giving

$$K_{\mathrm{eff}} = \min(K, N).$$

The K_{eff} retrieved records form the retrieval neighborhood for the current query, referred to below as the neighborhood, and are used to construct the corresponding local surrogate model. Letting $\mathbf{q}$ denote the query vector, the neighborhood is

$$\mathcal{N}(\mathbf{q}) = \{(\mathbf{x}_i, \mathbf{y}_i) : i \in \mathcal{I}_{\mathrm{nbr}}(\mathbf{q})\}, \qquad |\mathcal{N}(\mathbf{q})| = K_{\mathrm{eff}},$$

Here, $\mathcal{I}_{\mathrm{nbr}}(\mathbf{q})$ is the set of retrieved record indices for query vector $\mathbf{q}$. The Canberra distance is used only to identify the retrieval neighborhood and serves a different purpose from the Euclidean distance used internally by the RBF surrogate model.

For the K_{eff} retrieved records, the five porous electrode design variables are rescaled within the neighborhood to $[-1, 1]$, and the performance metrics are standardized using the neighborhood means and population standard deviations.

The retrieval–generation method uses SciPy (Virtanen et al., 2020), specifically its `RBFInterpolator`, whose linear radial kernel is

$$\phi(r) = -r,$$

Here, r is the Euclidean distance between two input vectors after local scaling, i.e., the radial distance in the RBF. For an input vector $\widetilde{\mathbf{x}}$ and a kernel center $\widetilde{\mathbf{x}}_i$,

$$r = \|\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}_i\|_2 .$$

The smoothing parameter is set to $\lambda = 10^{-6}$, and a constant polynomial term is included. Let $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ denote the locally scaled inputs and standardized outputs, respectively. The entries of the kernel matrix $\mathbf{\Phi}$ are

$$\Phi_{ij} = \phi\big(\|\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j\|_2\big) .$$

The RBF coefficients satisfy

$$(\mathbf{\Phi} + \lambda \mathbf{I})\mathbf{A} + \mathbf{1}\mathbf{b}^\top = \widetilde{\mathbf{Y}}, \qquad \mathbf{1}^\top \mathbf{A} = 0,$$

Here, $\mathbf{I}$ is the identity matrix of order K_{eff}, $\mathbf{1}$ is a vector of ones of length K_{eff}, and $\mathbf{A}$ and $\mathbf{b}$ are the coefficients to be determined.

For any locally scaled input $\widetilde{\mathbf{x}}$, the standardized output of the RBF surrogate model is

$$\widehat{\widetilde{\mathbf{y}}}(\widetilde{\mathbf{x}}) = \mathbf{b} + \sum_{i=1}^{K_{\text{eff}}} \mathbf{a}_i \phi(\|\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}_i\|_2) ,$$

Here, $\mathbf{a}_i$ is the coefficient vector corresponding to row i of matrix $\mathbf{A}$.

Because $\lambda > 0$, the model is a lightly regularized local RBF fit rather than an exact interpolant through all neighborhood data points.

For forward prediction, each component of the query vector is also clipped to the minimum and maximum values in the retrieved neighborhood.

For inverse design, we first retrieve the record closest to the target performance vector in performance space and use its design parameters to initialize the optimization. L-BFGS-B (Byrd et al., 1995) then optimizes within the ranges of the design variables in the neighborhood, minimizing the maximum relative error between the RBF surrogate predictions and the targets. The mathematical formulation is given below. The bounds of the design space determined by the neighborhood data are

$$\mathcal{B}_{\text{nbr}} = \prod_{\ell=1}^{5} \left[\min_{i \in \mathcal{I}_{\text{nbr}}} x_{i\ell},\ \max_{i \in \mathcal{I}_{\text{nbr}}} x_{i\ell} \right] .$$

Let

$$\widehat{\mathbf{y}}(\mathbf{x}) = [\widehat{y}_1(\mathbf{x}), \ldots, \widehat{y}_{m_c}(\mathbf{x})]^\top$$

denote the RBF surrogate prediction vector after reversing the output standardization, and let

$$\mathbf{y}^{\text{tar}} = \left[y_1^{\text{tar}}, \ldots, y_{m_c}^{\text{tar}}\right]^\top$$

denote the target performance vector. Inverse design then solves

$$\mathbf{x}^\star = \arg\min_{\mathbf{x} \in \mathcal{B}_{\text{nbr}}} \max_{1 \le j \le m_c} \frac{\left|\widehat{y}_j(\mathbf{x}) - y_j^{\text{tar}}\right|}{\left|y_j^{\text{tar}}\right| + 10^{-9}}.$$

D.3.1 Selection of Dataset Density and Retrieval Neighborhood Size K

This subsection uses the LFP-3 datasets constructed in Section D.1 to examine the effects of dataset density and retrieval neighborhood size K. We compare the six uniform grid datasets in Table D.2 and the change-point sampling dataset CPS272,620 constructed in Section D.1.2.

The prediction targets are the four performance metrics specified at the beginning of Section D.2. All combinations of dataset density and retrieval neighborhood size K use Canberra retrieval, an RBF model with a linear kernel, and $\lambda = 10^{-6}$. Candidate K values are first generated at equal

intervals within three ranges: 10 values from 5–800, 5 from 800–1500, and 5 from 1500–6000. The values are then rounded down and deduplicated, yielding 18 distinct integers. The actual number of retrieved records, K_{eff}, follows the definition in Section D.3. The error for each metric is

$$\text{MAPE} = \frac{1}{n_{\text{valid}}} \sum_{i=1}^{n_{\text{valid}}} 100 \frac{|\hat{y}_i - y_i|}{|y_i| + 10^{-9}}.$$

Here, y_i is the high-fidelity simulation value of the corresponding performance metric for porous electrode design i in the test set, $\hat{y}_i$ is the prediction from the retrieval–generation method, and n_{valid} is the number of valid prediction cases completed for that dataset–K combination.

Unless otherwise stated, this subsection and all subsequent experiments use the LFP-3 test set described in Section D.2. For each dataset–K combination, predictions are made on this test set, and MAPE is calculated from the valid predictions.

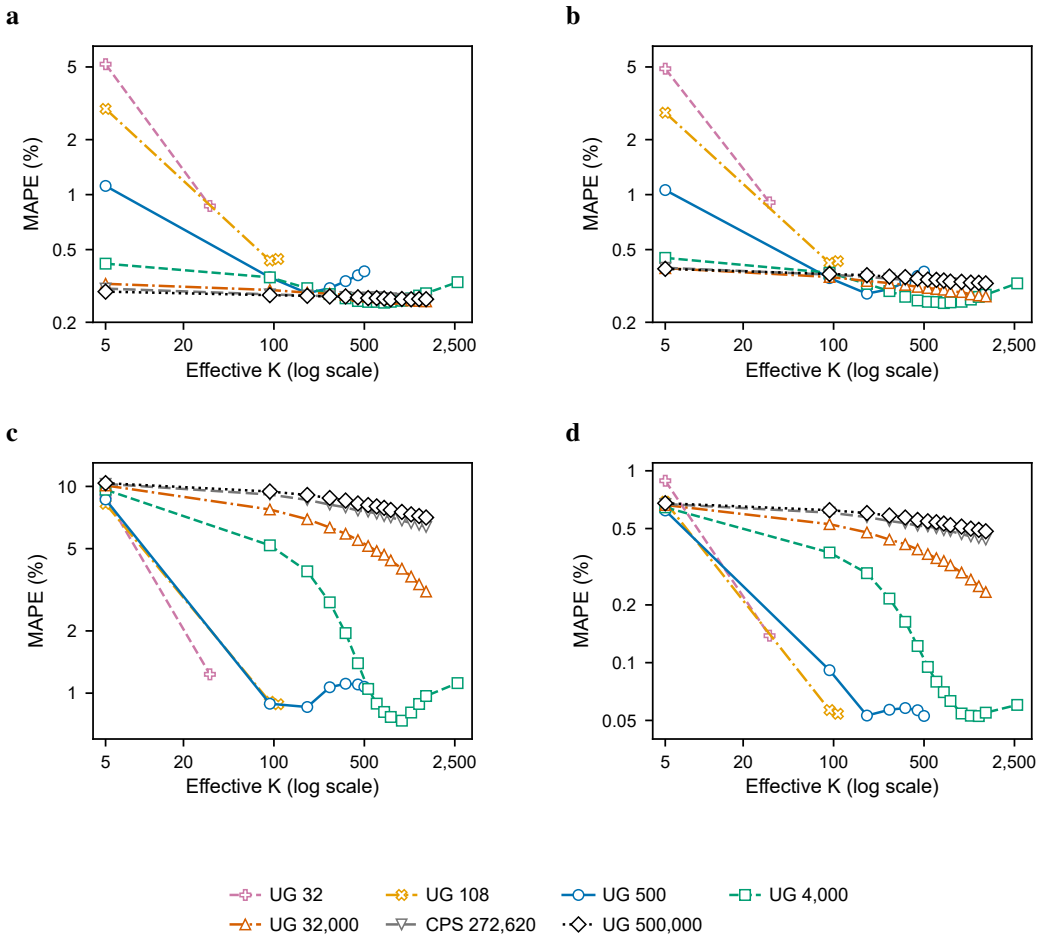


Figure D.4: Effects of dataset density and local neighborhood size on forward prediction errors. (a) Discharge capacity at 25 °C and 1/3 C; (b) discharge energy at 25 °C and 1/3 C; (c) discharge DCR at 25 °C, 50% SOC, 2C, and 10 s; (d) charge–discharge energy efficiency at 25 °C and 0.5P. All dataset–K combinations use Canberra retrieval and an RBF model with a linear kernel ($\lambda = 10^{-6}$). The horizontal axis is $K_{\text{eff}} = \min(K, N)$, and both axes use logarithmic scales. CPS changes both the number and the locations of samples.

The results show that overly sparse datasets reduce prediction accuracy, while further increases in dataset density or K do not necessarily improve it. CPS272,620, constructed using local variations in the capacity calculated from material balance along the design-variable directions, yields only a limited improvement in capacity prediction and no clear advantage for energy, DCR, or energy

efficiency. Considering data generation cost and surrogate prediction accuracy, we use UG500 and set the number of retrieved records K to 200.

D.3.2 Selection of Local Surrogate Models and Kernel Functions

Using the UG500 dataset selected in the preceding subsection, we compare two types of local surrogate models, radial basis function (RBF) models and Gaussian process regression (GPR), in terms of prediction accuracy and computational efficiency across kernel functions and retrieval neighborhood sizes K. The RBF models use linear, multiquadric, and Gaussian kernels, while GPR uses Matérn-1.5, Matérn-2.5, RBF, and rational-quadratic kernels. Figure D.5 presents the results for these 7 model–kernel combinations at the following 6 values of K:

$$K \in \{5, 93, 181, 270, 358, 446\}.$$

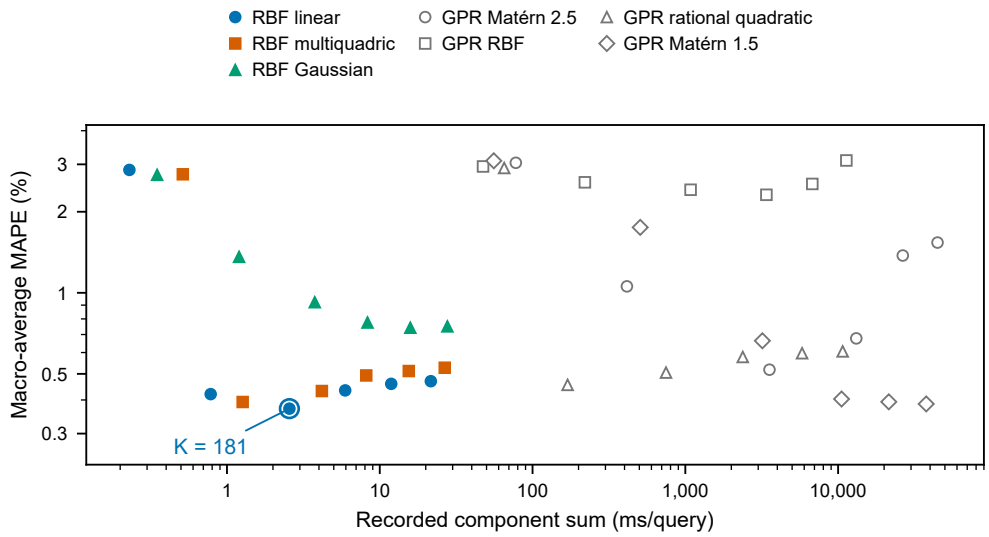


Figure D.5: Comparison of local surrogate models, kernel functions, and neighborhood size K. Each point represents a model–kernel–K configuration; filled and open markers denote RBF and GPR, respectively. The vertical axis is the equally weighted mean of the MAPEs for capacity, energy, DCR, and energy efficiency. The horizontal axis is the sum of the mean times for the three computational stages per performance prediction task. Both axes use logarithmic scales. The annotated point is the linear RBF at $K = 181$.

Figure D.5 therefore shows $7 \times 6 = 42$ configurations. The RBF smoothing parameter λ and the diagonal regularization coefficient α of the GPR kernel matrix are both set to 10^{-6}. The shape parameters of the multiquadric and Gaussian RBFs are set to $1/\bar{d}$ and $0.5/\bar{d}$, respectively, where $\bar{d}$ is the mean nearest-neighbor distance among the normalized local samples. The vertical axis is the equally weighted mean of the MAPEs for capacity, energy, DCR, and energy efficiency:

$$E_{\text{macro}} = \frac{1}{4} \sum_{q=1}^{4} \text{MAPE}_q,$$

Each MAPE is calculated as in Section D.3.1. The horizontal axis is the sum of the mean times for three computational stages:

$$T_{\text{sum}} = \overline{T}_{\text{retrieval}} + \overline{T}_{\text{fit}} + \overline{T}_{\text{predict}},$$

Here, $\overline{T}_{\text{retrieval}}$, $\overline{T}_{\text{fit}}$, and $\overline{T}_{\text{predict}}$ are the mean times for K-nearest-neighbor retrieval, local surrogate fitting, and performance prediction, respectively. The sum T_{sum} measures the computational

cost of different configurations and excludes data loading, normalization, index construction, and task scheduling. The term $\overline{T}_{\mathrm{retrieval}}$ is the total batch retrieval time averaged over all tasks; $\overline{T}_{\mathrm{fit}}$ is the mean time to construct a local surrogate model for a task; and $\overline{T}_{\mathrm{predict}}$ is the mean time for the local surrogate model to make predictions. The experiment was conducted on an HP Z8 G4 workstation running Ubuntu 22.04.5 LTS (Linux 6.8.0-117), equipped with two Intel Xeon Gold 6242R processors (40 physical cores, 80 logical threads, 3.10 GHz) and 377 GB of DDR4 ECC memory. Computation used CPUs only, with numerical libraries restricted to a single thread. Among the 42 configurations shown in Figure D.5, the linear RBF at $K = 181$ achieves the lowest macro-average MAPE of 0.371%, with a recorded time of $2.562\ \mathrm{ms/query}$. The lowest macro-average MAPEs of the multiquadric and Gaussian RBFs are 0.393% and 0.742%, respectively. The lowest GPR error is 0.387%, achieved by Matérn-1.5 at $K = 446$, but its recorded time reaches $37.7\ \mathrm{s/query}$. Based on these results, the final retrieval–generation method uses a linear RBF with neighborhood size $K = 200$.

D.3.3 Comparison of RG and Global Regression Baselines Using Data from a Single Cell Type or Multiple Cell Types

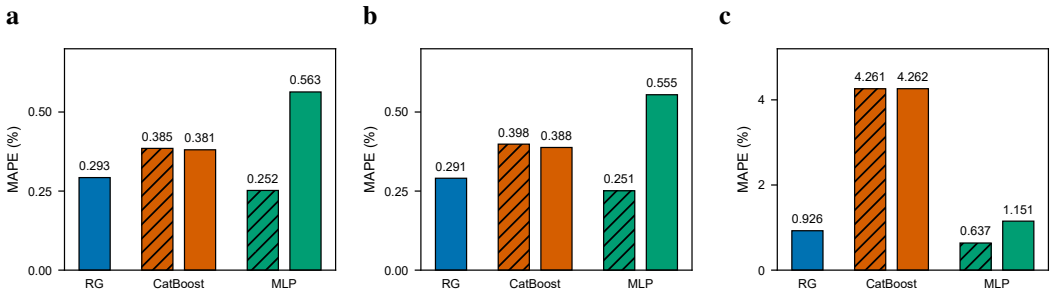
Using the same LFP-3 test set described above, we compare the prediction accuracy of RG and the two global regression baselines, CatBoost and MLP, when the dataset is expanded from a single cell type to multiple cell types. The single-type baselines are trained only on the LFP-3 UG32,000 dataset. The multi-type setting additionally includes the UG32,000 dataset of the NCM622-3 prismatic power cell ($148\times101\times27$ mm) and the UG32,000 dataset of the NCM811-1 46800 cylindrical cell, increasing the total training set size from 32,000 to 96,000 records. With data from multiple cell types, CatBoost treats cell type as a categorical feature and trains a separate model for each performance metric. MLP is trained jointly on data from all three cell types, encoding cell type as a two-dimensional vector that is updated during training and concatenating it with the five design parameters as input to a shared backbone. Each cell type has its own residual branch and output layer. Both methods also predict additional metrics specific to each cell type, but this subsection compares only capacity, energy, and DCR, which are shared by all three. For brevity, we refer to the models under the two settings as "single-type CatBoost" and "multi-type CatBoost," and "single-type MLP" and "multi-type MLP," respectively.

RG uses the UG500 dataset for the LFP-3 cell, retrieves $K = 200$ nearest-neighbor records, and predicts with a local RBF surrogate model using a linear kernel. Because RG can filter records by cell model and material system before retrieval, the additional NCM622-3 and NCM811-1 records do not enter nearest-neighbor retrieval or local fitting as long as the target cell's data subset remains unchanged. The RG predictions on the LFP-3 test set therefore remain unchanged.

The MAPEs of RG for capacity, energy, and DCR are 0.293%, 0.291%, and 0.926%, respectively. When CatBoost training is expanded from a single cell type to multiple cell types, these MAPEs change from 0.385%, 0.398%, and 4.261% to 0.381%, 0.388%, and 4.262%, respectively. Capacity and energy errors decrease slightly, while DCR error remains essentially unchanged. For MLP, the three MAPEs increase from 0.252%, 0.251%, and 0.637% to 0.563%, 0.555%, and 1.151%, respectively.

All model configurations above are selected by searching their respective predefined hyperparameter spaces. For capacity, energy, and DCR, the final CatBoost tree counts increase from approximately 11,000 in the single-type setting to approximately 13,200 in the multi-type setting, while the corresponding tree depths change from 9, 12, and 9 to 10, 12, and 11, respectively. The single-type MLP has 68,356 parameters, whereas the multi-type MLP has 98,996.

Thus, under the training and hyperparameter selection procedures used here, multi-type CatBoost maintains approximately the same prediction accuracy for LFP-3 while increasing model complexity and computational cost. The multi-type MLP is larger, takes longer to train, and has higher prediction errors on the LFP-3 test set. Because the single-type and multi-type settings differ in model architecture, hyperparameter search space, and training constraints, these results do not establish that adding data from multiple cell types necessarily reduces the prediction accuracy of CatBoost or MLP. Although type-based filtering makes RG insensitive to changes in data from other cell types, it also prevents RG from using those data to learn relationships shared across cell types.



For CatBoost and MLP, hatched bars indicate training on data from a single cell type, whereas solid bars indicate training on data from multiple cell types.

Figure D.6: Prediction errors of RG, single-type baselines, and multi-type baselines for the LFP-3 cell. (a) Discharge capacity at $25\,^{\circ}\mathrm{C}$ and $1/3\,\mathrm{C}$; (b) discharge energy at $25\,^{\circ}\mathrm{C}$ and $1/3\,\mathrm{C}$; (c) discharge DCR at $25\,^{\circ}\mathrm{C}$, 50% SOC, 2C, and 10 s. All results use the same 1,000 LFP-3 test samples. RG retrieves $K = 200$ nearest neighbors from the LFP-3 UG500 dataset and uses a local RBF surrogate with a linear kernel. Because NCM622-3 and NCM811-1 records are filtered out before retrieval, these two datasets do not affect its predictions. Single-type CatBoost and MLP use only the LFP-3 UG32,000 dataset, whereas multi-type CatBoost and MLP use the UG32,000 dataset for each of the three cell types.

E Experimental Validation of the Coupled P2D–Lumped Thermal Model Used for Data Generation

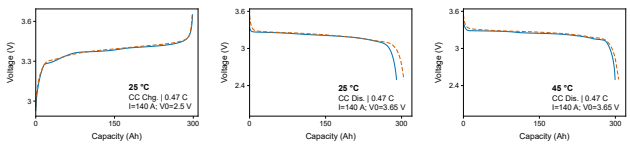
This section summarizes comparisons between experimental and simulated capacity–voltage curves for the coupled P2D–lumped thermal model used for data generation across five material systems, providing a direct assessment of the model calibration accuracy. The positive electrode material systems include LFP, NCM523, NCM622, NCM811, and NCM9055, with graphite used as the negative electrode material in all cases. Each material system contains 1–4 materials with distinct kinetic characteristics, namely LFP-1 to LFP-4, NCM523-1 to NCM523-3, NCM622-1 to NCM622-3, NCM811-1 to NCM811-2, and NCM9055-1, yielding a total of 228 experimental–simulation curve comparisons. The corresponding results are presented below in sequence according to the positive electrode material system and material identifier.

All figures in this section use a consistent visual encoding: blue solid lines represent the experimental results, while orange-yellow dashed lines represent the simulation results of the calibrated model. The legend is defined here once for all figures. The agreement between the experimental and simulated curves provides a direct indication of the model calibration accuracy.



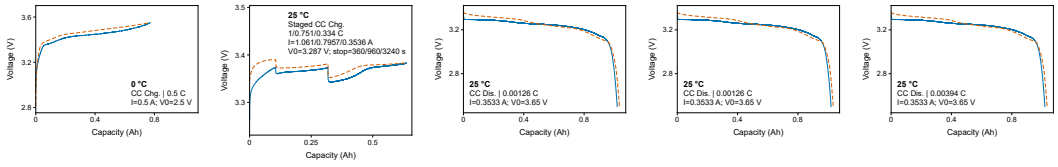
LFP–graphite

LFP-1

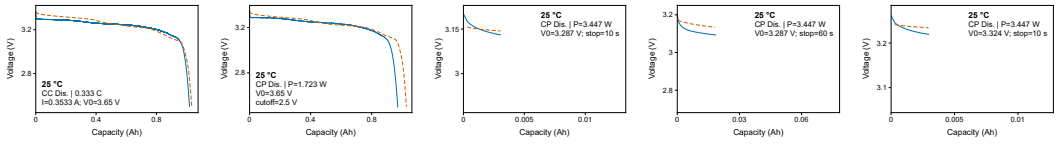


LFP-1, C001 LFP-1, C002 LFP-1, C003

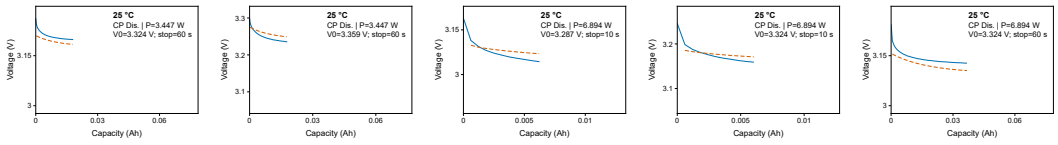
LFP-2



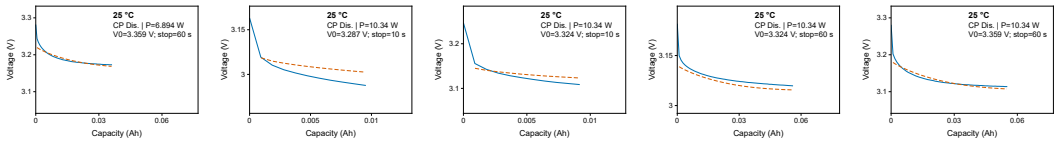
LFP-2, C001 LFP-2, C002 LFP-2, C003 LFP-2, C004 LFP-2, C005



LFP-2, C006 LFP-2, C007 LFP-2, C008 LFP-2, C009 LFP-2, C010

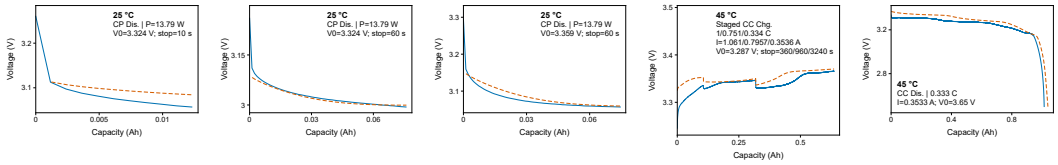


LFP-2, C011 LFP-2, C012 LFP-2, C013 LFP-2, C014 LFP-2, C015

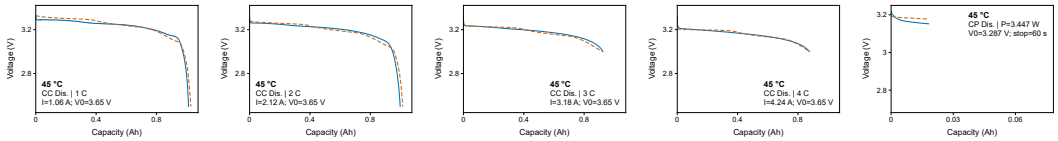


LFP-2, C016 LFP-2, C017 LFP-2, C018 LFP-2, C019 LFP-2, C020

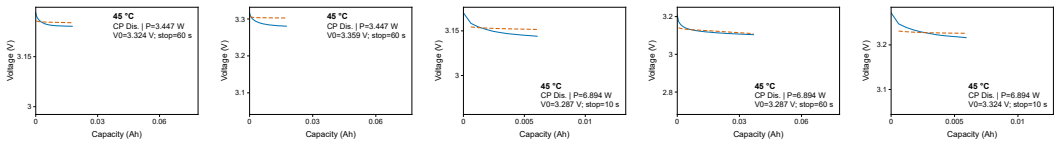
LFP-2



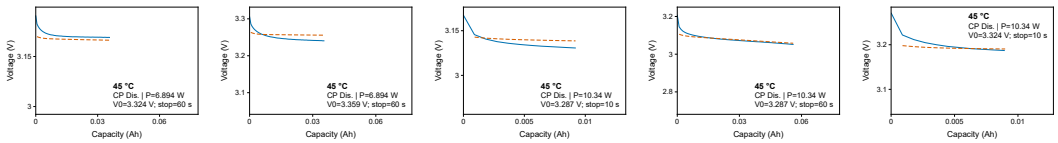
LFP-2, C021 LFP-2, C022 LFP-2, C023 LFP-2, C024 LFP-2, C025



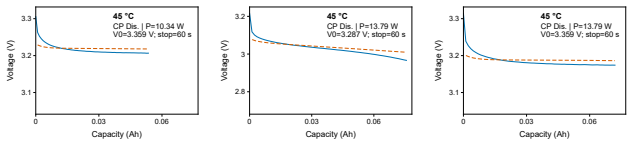
LFP-2, C026 LFP-2, C027 LFP-2, C028 LFP-2, C029 LFP-2, C030



LFP-2, C031 LFP-2, C032 LFP-2, C033 LFP-2, C034 LFP-2, C035

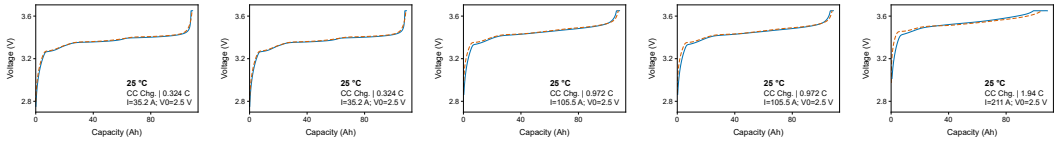


LFP-2, C036 LFP-2, C037 LFP-2, C038 LFP-2, C039 LFP-2, C040

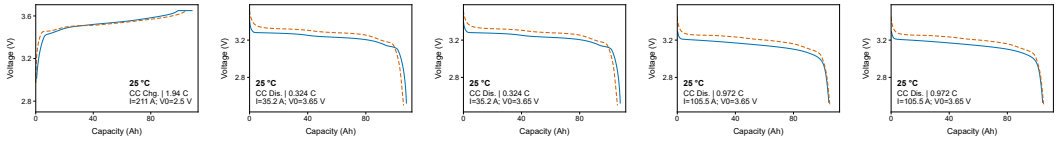


LFP-2, C041 LFP-2, C042 LFP-2, C043

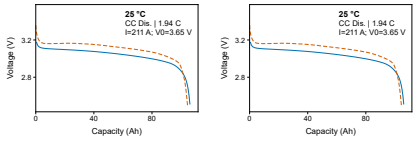
LFP-3



LFP-3, C001 LFP-3, C002 LFP-3, C003 LFP-3, C004 LFP-3, C005

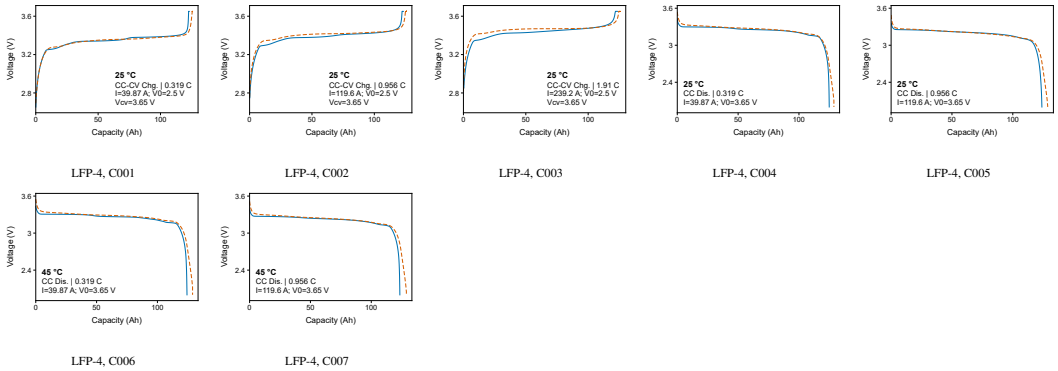


LFP-3, C006 LFP-3, C007 LFP-3, C008 LFP-3, C009 LFP-3, C010



LFP-3, C011 LFP-3, C012

LFP-4



LFP-4, C001

LFP-4, C002

LFP-4, C003

LFP-4, C004

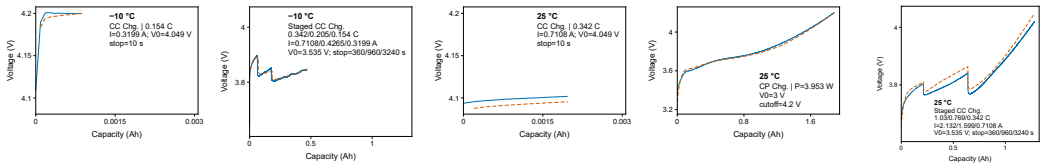
LFP-4, C005

LFP-4, C006

LFP-4, C007

NCM523–graphite

NCM523-1



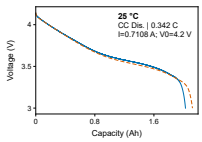
NCM523-1, C001

NCM523-1, C002

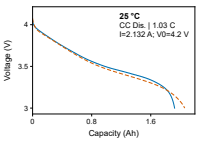
NCM523-1, C003

NCM523-1, C004

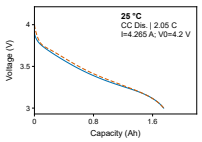
NCM523-1, C005



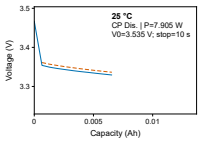
NCM523-1, C006



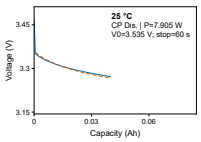
NCM523-1, C007



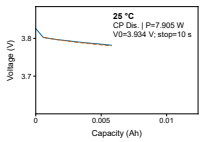
NCM523-1, C008



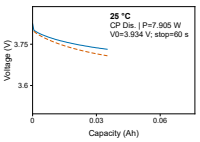
NCM523-1, C009



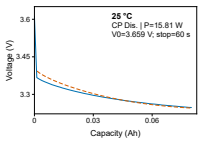
NCM523-1, C010



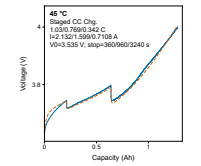
NCM523-1, C011



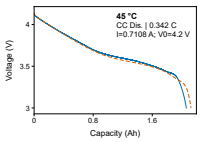
NCM523-1, C012



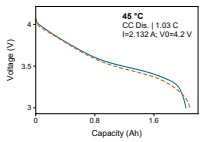
NCM523-1, C013



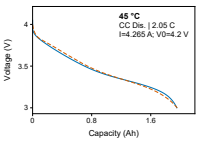
NCM523-1, C014



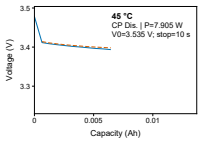
NCM523-1, C015



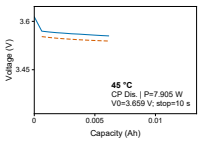
NCM523-1, C016



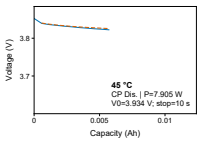
NCM523-1, C017



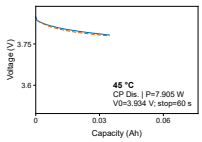
NCM523-1, C018



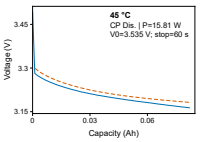
NCM523-1, C019



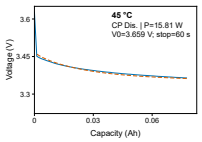
NCM523-1, C020



NCM523-1, C021

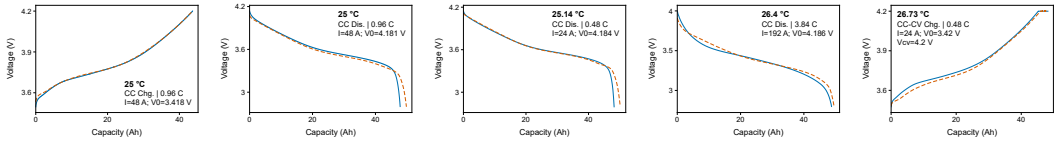


NCM523-1, C022



NCM523-1, C023

NCM523-2



NCM523-2, C001

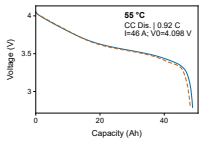
NCM523-2, C002

NCM523-2, C003

NCM523-2, C004

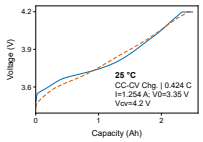
NCM523-2, C005

NCM523-2

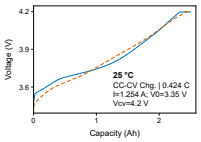


NCM523-2, C006

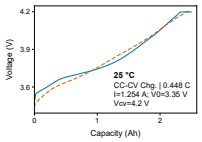
NCM523-3



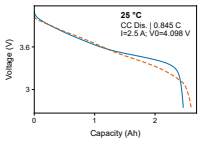
NCM523-3, C001



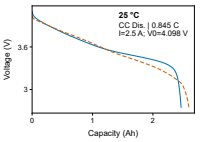
NCM523-3, C002



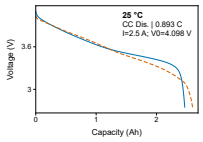
NCM523-3, C003



NCM523-3, C004



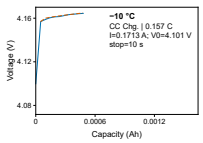
NCM523-3, C005



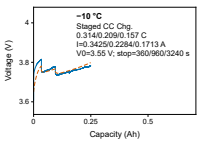
NCM523-3, C006

NCM622–graphite

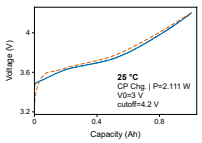
NCM622-1



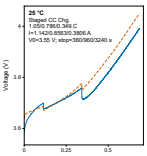
NCM622-1, C001



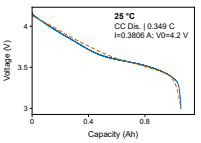
NCM622-1, C002



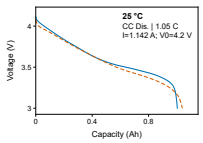
NCM622-1, C003



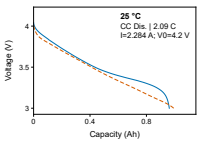
NCM622-1, C004



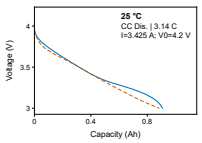
NCM622-1, C005



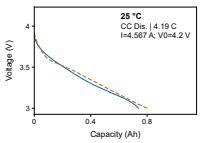
NCM622-1, C006



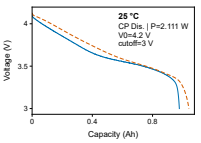
NCM622-1, C007



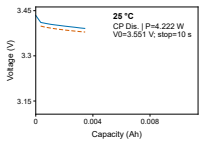
NCM622-1, C008



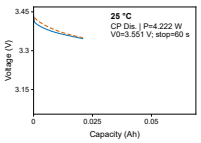
NCM622-1, C009



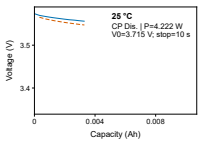
NCM622-1, C010



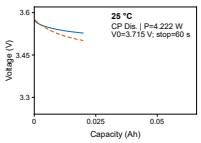
NCM622-1, C011



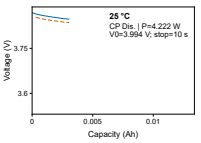
NCM622-1, C012



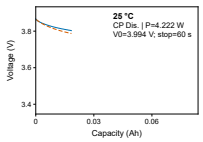
NCM622-1, C013



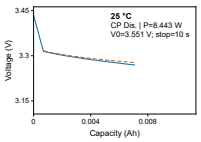
NCM622-1, C014



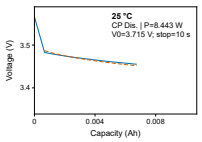
NCM622-1, C015



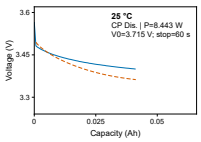
NCM622-1, C016



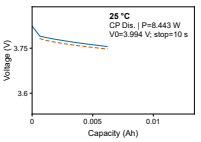
NCM622-1, C017



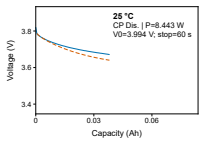
NCM622-1, C018



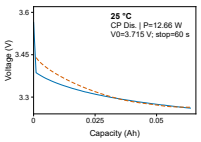
NCM622-1, C019



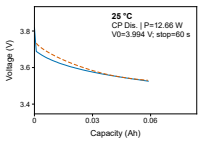
NCM622-1, C020



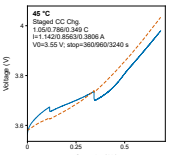
NCM622-1, C021



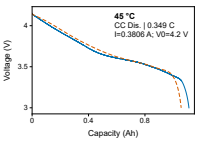
NCM622-1, C022



NCM622-1, C023

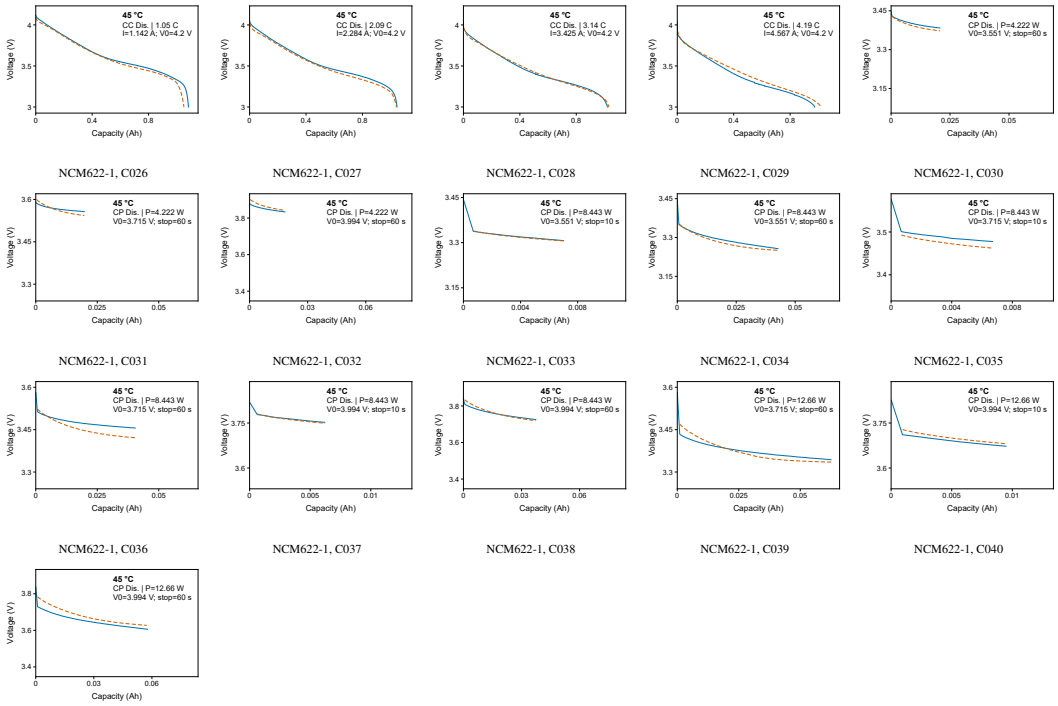


NCM622-1, C024



NCM622-1, C025

NCM622-1



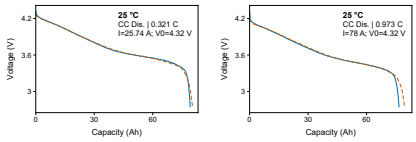
NCM622-1, C026 NCM622-1, C027 NCM622-1, C028 NCM622-1, C029 NCM622-1, C030

NCM622-1, C031 NCM622-1, C032 NCM622-1, C033 NCM622-1, C034 NCM622-1, C035

NCM622-1, C036 NCM622-1, C037 NCM622-1, C038 NCM622-1, C039 NCM622-1, C040

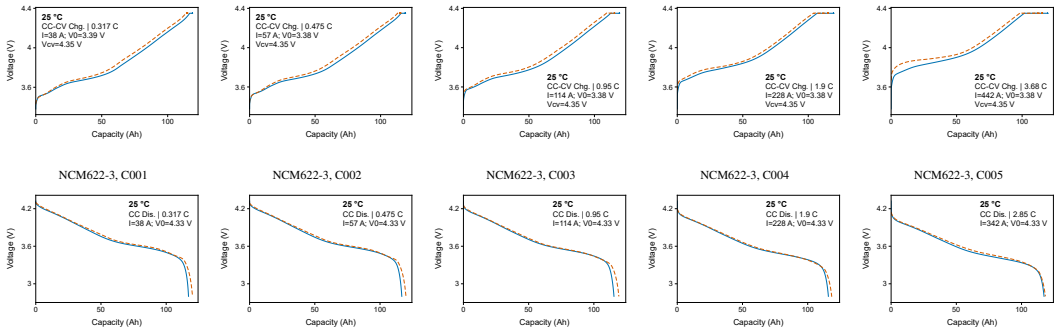
NCM622-1, C041

NCM622-2



NCM622-2, C001 NCM622-2, C002

NCM622-3

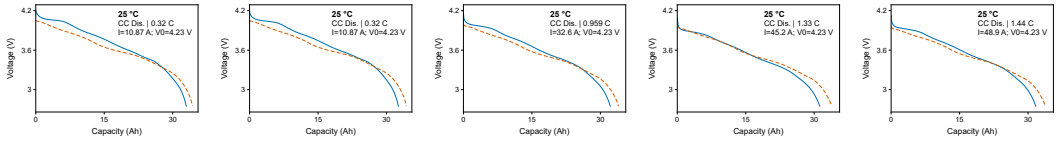


NCM622-3, C001 NCM622-3, C002 NCM622-3, C003 NCM622-3, C004 NCM622-3, C005

NCM622-3, C006 NCM622-3, C007 NCM622-3, C008 NCM622-3, C009 NCM622-3, C010

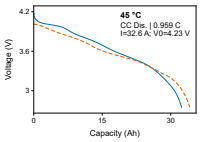
NCM811–graphite

NCM811-1

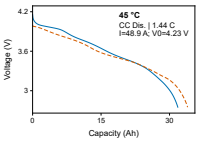


NCM811-1, C001 NCM811-1, C002 NCM811-1, C003 NCM811-1, C004 NCM811-1, C005

NCM811-1

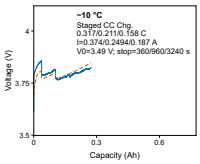


NCM811-1, C006

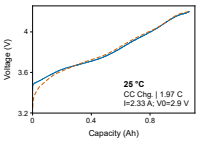


NCM811-1, C007

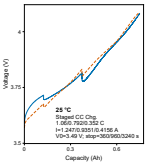
NCM811-2



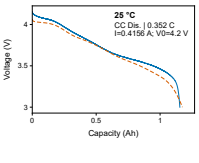
NCM811-2, C001



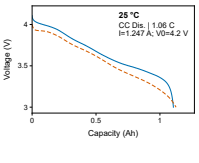
NCM811-2, C002



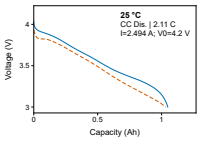
NCM811-2, C003



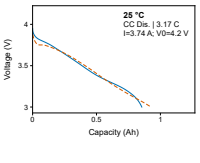
NCM811-2, C004



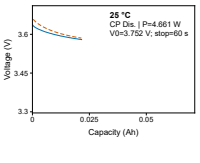
NCM811-2, C005



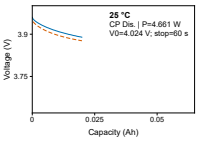
NCM811-2, C006



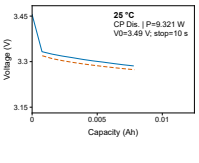
NCM811-2, C007



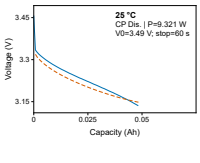
NCM811-2, C008



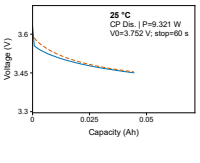
NCM811-2, C009



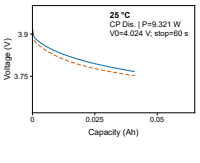
NCM811-2, C010



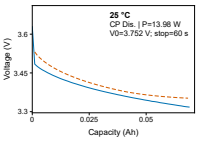
NCM811-2, C011



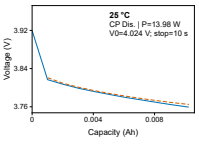
NCM811-2, C012



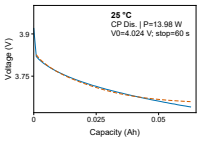
NCM811-2, C013



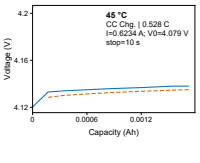
NCM811-2, C014



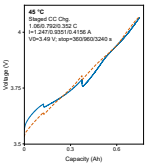
NCM811-2, C015



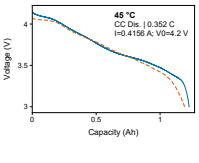
NCM811-2, C016



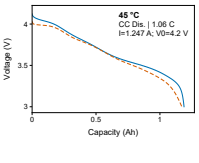
NCM811-2, C017



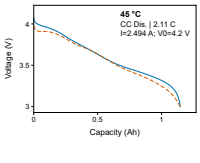
NCM811-2, C018



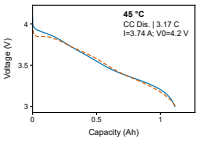
NCM811-2, C019



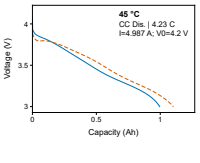
NCM811-2, C020



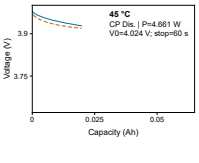
NCM811-2, C021



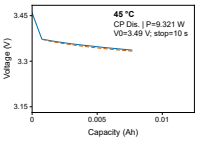
NCM811-2, C022



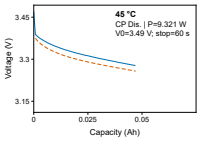
NCM811-2, C023



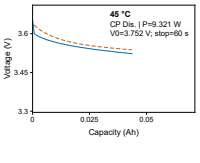
NCM811-2, C024



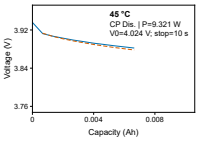
NCM811-2, C025



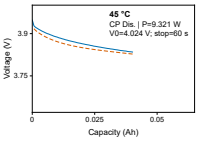
NCM811-2, C026



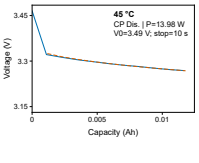
NCM811-2, C027



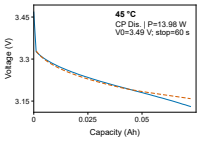
NCM811-2, C028



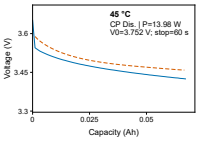
NCM811-2, C029



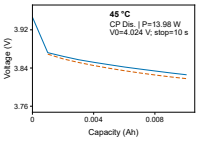
NCM811-2, C030



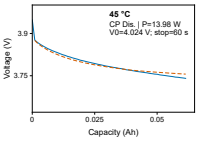
NCM811-2, C031



NCM811-2, C032



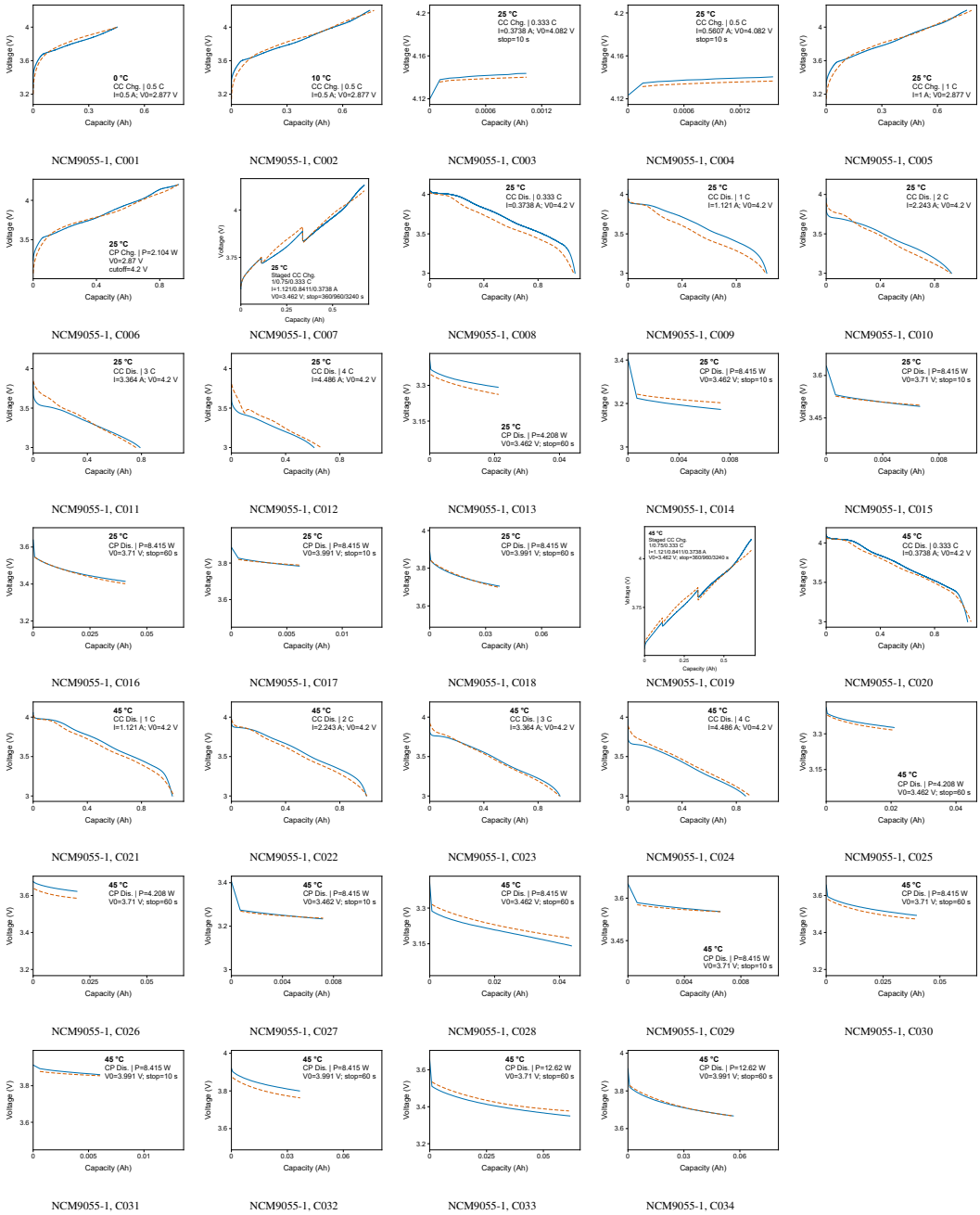
NCM811-2, C033



NCM811-2, C034

NCM9055–graphite

NCM9055-1



F CODE AND DATA AVAILABILITY

To comply with ICLR's double-blind review requirements, we provide anonymized code and data at the links below.

The shared LiBMap subset provides data for 72,000 cell designs obtained by dense sampling of porous-electrode design parameters for cylindrical 46800 cells with the NCM811-1–graphite material system. We plan to progressively release additional LiBMap data and evaluate the feasibility of releasing the full dataset.

- **Code related to retrieval generation:** Anonymous repository
- **Data, Part 1:** Zenodo review access
- **Data, Part 2:** Zenodo review access
- **Data, Part 3:** Zenodo review access

All three data parts are required to obtain the complete shared subset.