Supplementary Material:

From Material Properties to Device Metrics: A Data-Driven Guide to Battery Design

Kevin W. Knehr[1], Joseph J. Kubal[1], Abhas Deva[1], Mohammed B. Effat[1,2], Shabbir Ahmed[1]

[1]Chemical Sciences and Engineering Division, Argonne National Laboratory, 9700 S. Cass Ave, Lemont, IL 60439, USA

[2]Department of Mechanical Power Engineering, Assiut University, Assiut Governorate, Egypt

Method for Generating Correlation Plots

Formulation

Correlation plots are used in this work to provide a quantitative way of explaining the degree to which achieving a given target cell metric (*e.g.*, Wh/kg, W/kg, or $/kWh) depends on a given material property, cell design, or cost parameter. Several standard correlation plots exist in the literature for statistical analysis of large datasets, notably the Pearson, Spearman, and Kendall plots.[1] The correlation values in the Pearson, Spearman, and Kendall plots quantify, respectively, the linearity, monotonic nature, and rank correlation of a dataset. These plots can be used to judge whether the inputs are positively or negatively correlated to outputs, but they provide minimal information on which inputs have the *greatest* relationship to the outputs. For instance, the correlation values depend on the variable ranges and sampling strategies used to generate the database. These correlations are also best suited for providing information on the database as a whole. They have limitations when attempting to determine finer-grained insights about how the importance of an input variable changes when attempting to achieve certain output values included within the database. These shortcomings motivated the development of the correlation plotting method described here. The goal is to provide a more quantitative correlation for all input parameters in the Monte Carlo database as a function of the target output metrics.

This process can be understood by first noting that the Monte-Carlo database was generated by sampling each input variable (V_i) using a uniform distribution, which is defined as follows:

$$V_i = U\left(V_{i,min}, V_{i,max}\right), \qquad \text{[S1]}$$

where U signifies a uniform distribution between $V_{i,min}$, and $V_{i,max}$, which are the minimum and maximum variable values listed in Tables 2 and 3 in the main text. The first step in the method involves normalizing the database for each variable, V_i, to make the formulation consistent with the statistics literature. The normalization is done by converting every value, v_j, of independent variable, V_i, into the normalized value, x_j, of the normalized independent variable, X_i, as follows:

$$X_i = \left[\, x_1, x_2, \dots, x_j\right], \qquad where \quad x_j = \frac{v_j \; - \; V_{i,min},}{V_{i,max} - V_{i,min}}. \qquad \text{[S2]}$$

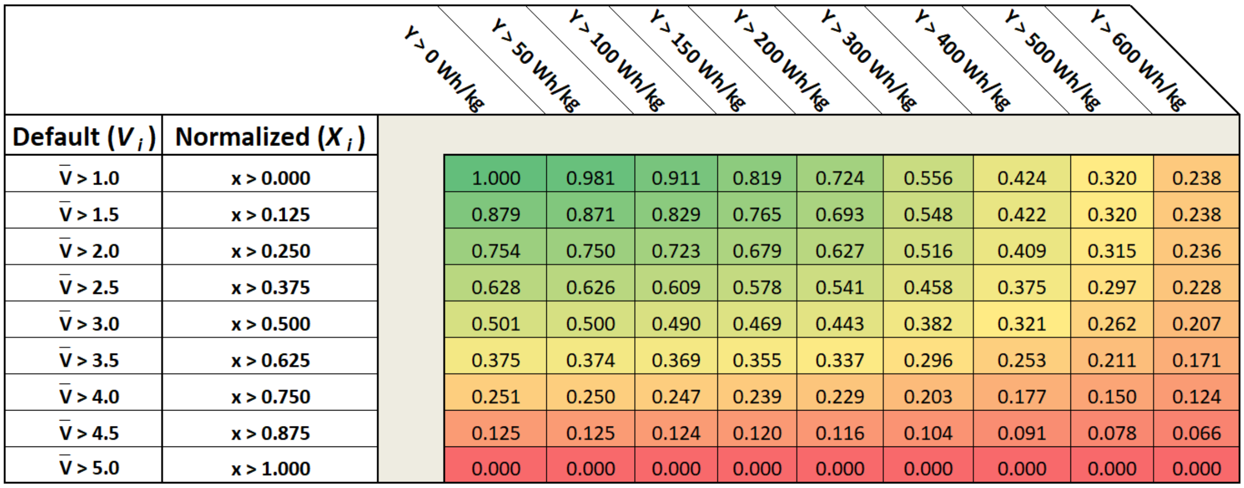
The second step requires binning the values of X_i based on their ability to achieve a target output metric (Y). Binning is achieved by evaluating the cumulative distribution function (CDF) defined in Eq. S3:

$$F_{XY}(x,y) = \;\; P(X \geq x, Y \epsilon\, y), where\, \epsilon = \begin{cases} \leq, & e.g.\;\; \$/kWh \\ \geq, & e.g.\;\; Wh/kg \end{cases}, \qquad \text{[S3]}$$

where P is the probability that the normalized independent variable, X, is greater than or equal to a certain normalized value, x, *and* the output metric, Y, is greater (or less) than or equal to the target output value, y. The probability is calculated by counting the number of simulations in the Monte Carlo database that meet these criteria and dividing by the total number of simulations. The condition on Y can be greater than or less than y depending on the metric. For instance, the cost metric, \$/kWh, is met if its value is *less than* or equal to y, while the specific energy metric, Wh/kg, is met if its value is *greater than* or equal to y. In this work, the values of x used to evaluate F_{XY} are selected to evenly segment the database among 15 values. Note that different numbers of segments were tested, and consistent results were observed between 5 and 25 segments.

An example CDF using nine x values is shown in Table S1, where the average open circuit voltage ($\bar{V}$) is the independent variable, V_i, and the specific capacity is the target output metric, Y. Results are shown for nine v_i values from 1.0 to 5.0 V and nine target specific capacities, y, from 0 to 600 Wh/kg. The colored values are the results from F_{XY}. Note that the F_{XY} value in the upper left corner is equal to one, indicating every value in the database meets these criteria (*i.e.*, $\bar{V} > 1.0$ V and $Y > 0$ Wh/kg).

Table S1. Example results for the cumulative distribution function, F_{XY}, with the average open circuit voltage, $\bar{V}$, as the variable, V_i, and the specific capacity, Wh/kg, as the output metric, Y. Each value in the table denotes the probability, P, that $\bar{V} > v_j$ (*i.e.*, $X_i > x_j$) and $Y > y$ for all simulations in the Monte Carlo database.



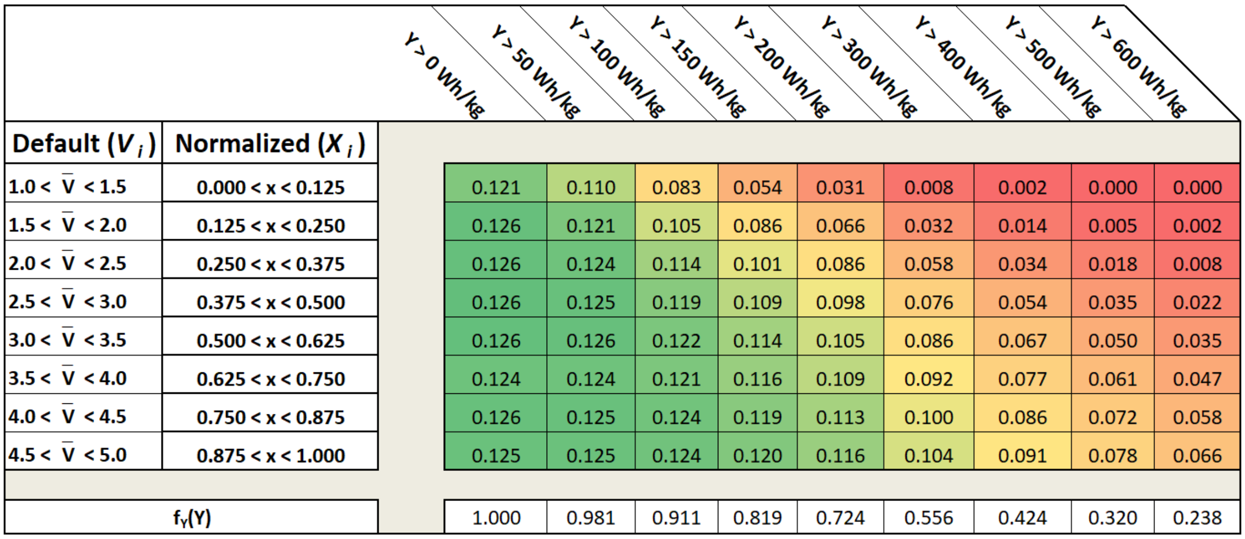
Default (V_i)	Normalized (X_i)	Y > 0 Wh/kg	Y > 50 Wh/kg	Y > 100 Wh/kg	Y > 150 Wh/kg	Y > 200 Wh/kg	Y > 300 Wh/kg	Y > 400 Wh/kg	Y > 500 Wh/kg	Y > 600 Wh/kg
$\bar{V}$ > 1.0	x > 0.000	1.000	0.981	0.911	0.819	0.724	0.556	0.424	0.320	0.238
$\bar{V}$ > 1.5	x > 0.125	0.879	0.871	0.829	0.765	0.693	0.548	0.422	0.320	0.238
$\bar{V}$ > 2.0	x > 0.250	0.754	0.750	0.723	0.679	0.627	0.516	0.409	0.315	0.236
$\bar{V}$ > 2.5	x > 0.375	0.628	0.626	0.609	0.578	0.541	0.458	0.375	0.297	0.228
$\bar{V}$ > 3.0	x > 0.500	0.501	0.500	0.490	0.469	0.443	0.382	0.321	0.262	0.207
$\bar{V}$ > 3.5	x > 0.625	0.375	0.374	0.369	0.355	0.337	0.296	0.253	0.211	0.171
$\bar{V}$ > 4.0	x > 0.750	0.251	0.250	0.247	0.239	0.229	0.203	0.177	0.150	0.124
$\bar{V}$ > 4.5	x > 0.875	0.125	0.125	0.124	0.120	0.116	0.104	0.091	0.078	0.066
$\bar{V}$ > 5.0	x > 1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The third step converts the CDF into a probability density function (PDF) using the following equation:

$$f_{XY}(x,y) = P(x_j \geq X \geq x_{j+1}, Y \epsilon y) = F_{XY}(x_{j+1}, y) - F_{XY}(x_j, y). \qquad [S4]$$

This equation provides the probability that a simulation has a value of X from x_j to x_{j+1} *and* a value of Y that meets the criteria y. Table S2 provides the PDF for the CDF in Table S1. Note that the use of nine v_j/x_j values in Table S1 corresponds to eight bins in Table S2. Also note that, for the condition where $Y > 0$ Wh/kg, all bins have probabilities of ~1/8, indicating the database is divided equally among the eight bins. In this case, the values of $\bar{V}$ have no impact on whether $Y > 0$ Wh/kg because all results have $Y > 0$ Wh/kg. Slight deviations from 1/8 in this column correspond to slightly uneven sampling in the Monte Carlo simulations.

Table S2. Example results for probability density function, f_{XY}. Each colored value denotes the probability, P, that $v_j < \bar{V} < v_{j+1}$ (*i.e.*, $x_j < X_i < x_{j+1}$) and $Y > y$ for all simulations in the Monte Carlo database.



Default (V_i)	Normalized (X_i)	$Y > 0$ Wh/kg	$Y > 50$ Wh/kg	$Y > 100$ Wh/kg	$Y > 150$ Wh/kg	$Y > 200$ Wh/kg	$Y > 300$ Wh/kg	$Y > 400$ Wh/kg	$Y > 500$ Wh/kg	$Y > 600$ Wh/kg
$1.0 < \bar{V} < 1.5$	$0.000 < x < 0.125$	0.121	0.110	0.083	0.054	0.031	0.008	0.002	0.000	0.000
$1.5 < \bar{V} < 2.0$	$0.125 < x < 0.250$	0.126	0.121	0.105	0.086	0.066	0.032	0.014	0.005	0.002
$2.0 < \bar{V} < 2.5$	$0.250 < x < 0.375$	0.126	0.124	0.114	0.101	0.086	0.058	0.034	0.018	0.008
$2.5 < \bar{V} < 3.0$	$0.375 < x < 0.500$	0.126	0.125	0.119	0.109	0.098	0.076	0.054	0.035	0.022
$3.0 < \bar{V} < 3.5$	$0.500 < x < 0.625$	0.126	0.126	0.122	0.114	0.105	0.086	0.067	0.050	0.035
$3.5 < \bar{V} < 4.0$	$0.625 < x < 0.750$	0.124	0.124	0.121	0.116	0.109	0.092	0.077	0.061	0.047
$4.0 < \bar{V} < 4.5$	$0.750 < x < 0.875$	0.126	0.125	0.124	0.119	0.113	0.100	0.086	0.072	0.058
$4.5 < \bar{V} < 5.0$	$0.875 < x < 1.000$	0.125	0.125	0.124	0.120	0.116	0.104	0.091	0.078	0.066
$f_Y(Y)$		1.000	0.981	0.911	0.819	0.724	0.556	0.424	0.320	0.238

The value $f_Y(y)$ at the bottom of Table S2 is the marginal probability that a simulation meets the criteria $Y > y$. It is calculated as follows:

$$f_Y(y) = \sum_k f_{XY}(\bar{x}_k, y)\,, \qquad \text{[S5]}$$

where $\bar{x}_k$ is the average value of x in each bin. It represents the conversion of nine normalized values, x_j, into eight bins with an average normalized value of $\bar{x}_k$. The marginal probability is used to convert the PDF (*i.e.*, f_{XY}) into a conditional probability as follows:

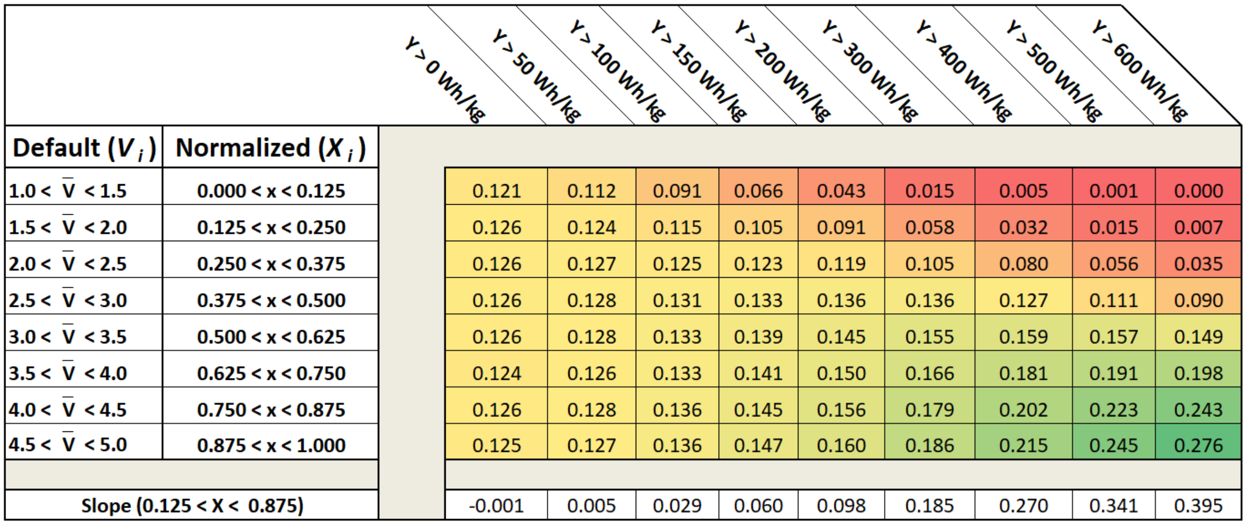
$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}\,, \qquad \text{[S6]}$$

which is the probability that a simulation has a value of X from x_j to x_{j+1} given it has a value of Y that meets the criteria y. The conditional probability values for the PDF values in Table S2 are given in Table S3.

The correlation values for V_i with respect to the target are calculated by taking the slope of the conditional probability, $f_{X|Y}$, with respect to the normalized value, $\bar{x}_k$, using a linear regression. The slope ($\partial f_{X/Y}/\partial \bar{x}_k$,) is determined for all bins where $0.05 < x < 0.95$ to remove the occasional noise observed at the extrema. Note that an empirical distribution function was also investigated, but it was found that the binning technique used here provided a cleaner average of slopes.

For the example in Table S3, the correlation of V_i (*i.e.*, $\bar{V}$ or X_i) to each criterion $Y > y$ is calculated from the slope of the middle six bins. The first and last bins are excluded from the linear regression because they contain x-values that are <0.05 and >0.95, respectively. The resulting slopes from the linear regressions are shown at the bottom of Table S3. The criterion $Y > 0$ Wh/kg has a slope ~0, indicating $\bar{V}$ is not correlated with this target. The average open circuit voltage does not influence whether a simulation has Y > 0 Wh/kg because all simulations have Y > 0 Wh/kg. The slope progressively increases with increasing y, indicating the correlation between $\bar{V}$ and Y increases as y increases. This trend indicates that maximizing $\bar{V}$ is increasingly important when trying to achieve high Wh/kg.

Table S3. Example results for the conditional probability function, $f_{X|Y}$. Each colored value denotes the probability, P, that $v_j < \bar{V} < v_{j+1}$ (*i.e.*, $x_j < X_i < x_{j+1}$) for a given $Y > y$ for all simulations in the Monte Carlo database.



Default (V_i)	Normalized (X_i)	Y > 0 Wh/kg	Y > 50 Wh/kg	Y > 100 Wh/kg	Y > 150 Wh/kg	Y > 200 Wh/kg	Y > 300 Wh/kg	Y > 400 Wh/kg	Y > 500 Wh/kg	Y > 600 Wh/kg
$1.0 < \bar{V} < 1.5$	$0.000 < x < 0.125$	0.121	0.112	0.091	0.066	0.043	0.015	0.005	0.001	0.000
$1.5 < \bar{V} < 2.0$	$0.125 < x < 0.250$	0.126	0.124	0.115	0.105	0.091	0.058	0.032	0.015	0.007
$2.0 < \bar{V} < 2.5$	$0.250 < x < 0.375$	0.126	0.127	0.125	0.123	0.119	0.105	0.080	0.056	0.035
$2.5 < \bar{V} < 3.0$	$0.375 < x < 0.500$	0.126	0.128	0.131	0.133	0.136	0.136	0.127	0.111	0.090
$3.0 < \bar{V} < 3.5$	$0.500 < x < 0.625$	0.126	0.128	0.133	0.139	0.145	0.155	0.159	0.157	0.149
$3.5 < \bar{V} < 4.0$	$0.625 < x < 0.750$	0.124	0.126	0.133	0.141	0.150	0.166	0.181	0.191	0.198
$4.0 < \bar{V} < 4.5$	$0.750 < x < 0.875$	0.126	0.128	0.136	0.145	0.156	0.179	0.202	0.223	0.243
$4.5 < \bar{V} < 5.0$	$0.875 < x < 1.000$	0.125	0.127	0.136	0.147	0.160	0.186	0.215	0.245	0.276
Slope ($0.125 < X < 0.875$)		-0.001	0.005	0.029	0.060	0.098	0.185	0.270	0.341	0.395

Example Correlation Plot

Figure S1a provides an example correlation plot between three variables — the average open circuit voltage ($\bar{V}$), the thickness of the negative current collector ($\delta_{-,cc}$), and the area specific impedance (ASI) — and five energy density metrics from 0 to 1,250 Wh/kg. A color bar describing the color and hatching in Figure S1a is shown in Figure S1c. The colors represent the magnitude of the slope calculated in the previous section, with dark colors corresponding to high slopes (*i.e.*, high correlations) and light colors corresponding to minimal slopes (*i.e.*, minimal correlations). A solid square in Figure S1 corresponds to a positive slope, which indicates a positive correlation between the variable and the specific energy target. A positive correlation implies that increasing the variable increases the probability of achieving the specific energy target. A hatched square corresponds to a negative slope, which indicates a negative correlation. A negative correlation implies that increasing the variable decreases the probability of achieving the target. This color and fill explanation applies to all correlation plots in this work. The same color bar is also used for all correlation plots, with the darkest color representing the highest correlation between a parameter and a metric.

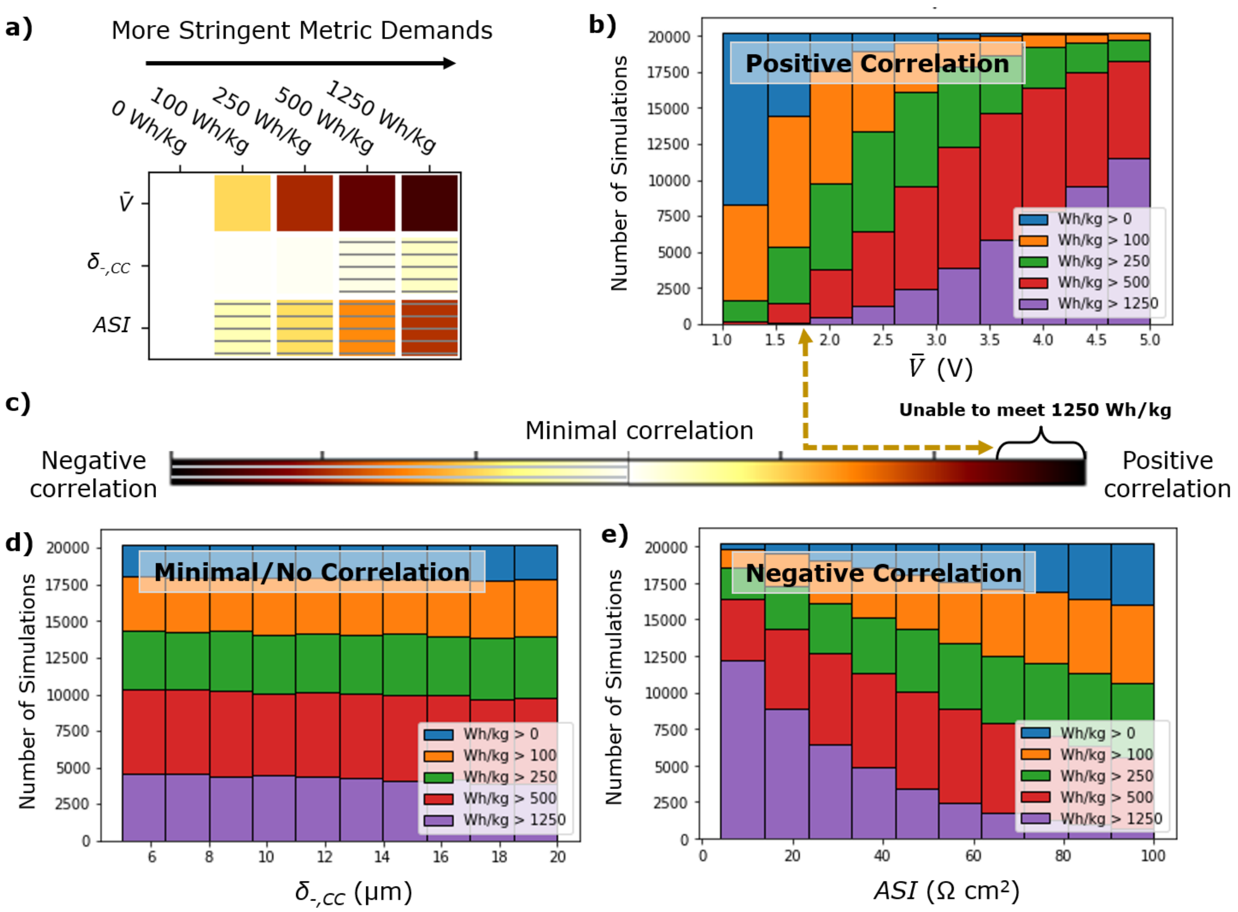


Figure S1. (a) Example correlation plot and histograms of probability distribution functions (f_{xy}) for (b) the open circuit voltage, (d) the negative current collector thickness, and (e) the reference area specific impedance. The plots show results for different specific energy metrics ($Y > y$) ranging from 0 to 1,250 Wh/kg. A color bar is provided in (c) to describe the meaning of the colors in (a), where dark and light colors correspond to a high and minimal correlation, respectively, and solid and hatched squares correspond to positive and negative correlation, respectively. Parts (b), (d), and (e) correspond to variables with positive correlation, minimal correlation, and negative correlation, respectively.

Figures S1b, S1d, and S1e provide stacked histograms of the probability distribution functions (f_{xy}) for $\bar{V}$, $\delta_{-,CC}$, and ASI, respectively. These histograms are representations of the raw data used to calculate the correlations in Figure S1a. $\bar{V}$ in Figure S1b represents a variable that has a positive correlation to specific energy targets. For all specific energies except 0 Wh/kg, the PDF increases with increasing values of $\bar{V}$, indicating the probability of achieving a specific energy target increases with increasing voltage. The PDF also decreases with increasing specific energy targets, indicating it is more difficult to achieve higher targets. The slopes of the PDFs in Figure S1b *roughly* correspond to the correlation plots in Figure S1a, where higher average slopes correspond to darker colors. The term *roughly* is used because the colors in Figure S1a are calculated from the conditional probability in Eq. S7, not the PDF plotted in Figure S1b. The conditional probability is calculated by dividing each bin of a given color (*i.e.*, target) in Figure S1b by the total number of simulations with the same color. The PDF is shown in Figure S1b, and not the conditional probability, because it yielded similar conclusions (*i.e.*, magnitude of the slope) with a better visualization. The darker colors in Figure S1a occur at higher Wh/kg targets, indicating that maximizing $\bar{V}$ becomes more important at higher specific energy targets. Figure S1b also highlights a region where no simulations can achieve >1,250 Wh/kg at low voltages. This result is indicative of a highly correlated variable/metric pair, which was properly categorized by a dark solid square in Figure S1a.

Figure S1d represents a variable with minimal correlation. It corresponds to near-uniform distributions in the PDF, suggesting the value of $\delta_{-,CC}$ has a minimal impact on the probability of a battery cell achieving a specific energy target. The minimal correlation results in a light color in Figure S1a. The transition from a positive (solid square) to a negative (hatched square) in Figure S1a is caused by the cutoff criterion used for hatching and occurs because the correlation slope is close to zero. It is an artifact of the data analysis method and does not warrant further interpretation.

Figure S1e represents a variable that is negatively correlated to the specific energy. For all targets except 0 Wh/kg, increasing the value of *ASI* decreases the probability of achieving a specific energy target. The PDF also decreases with increasing specific energy targets, indicating it is more difficult to achieve higher targets. The slopes of the PDFs in Figure S1b *roughly* correspond to the correlation plots in Figure S1a, where slopes with higher magnitudes correspond to darker colors. The darker colors in Figure S1a occur at higher Wh/kg targets, indicating that minimizing *ASI* becomes more important at higher specific energy targets.

Variable Ranges for Chemistries in Figure 2

Table S4. Summary of materials parameters for battery chemistries in commercial use or in active development. Header definitions follow the table. See Table 2 in main text for symbol definitions.

Input Parameter	**G-NMCA**	**G-LFP**	**G-LMO**	**G-LCO**	**Si/G-NMC**	**Li-NMC**	**Li-S**
$Q_{+,act}$, mAh/g	160 – 230	150 – 170	110 – 130	130 – 150	160 – 230	160 – 230	1000 – 2000
$Q_{-,act}$, mAh/g	300 – 360	300 – 360	300 – 360	300 – 360	600 – 3000	1500 – 3000	1500 – 3000
$\bar{V}$, V	3.55 – 3.85	3.17 – 3.47	3.85 – 4.15	3.65 – 3.95	3.4 – 3.85	3.65 – 3.95	2.15 – 2.45
$\rho_{+,act}$, g/cm^3	4.5 – 4.8	3.3 – 3.6	4.1 - 4.4	4.5 – 4.8	4.5 – 4.8	4.5 – 4.8	1.9 – 2.2
$\rho_{-,act}$, g/cm^3	2.1 – 2.4	2.1 – 2.4	2.1 – 2.4	2.1 – 2.4	2.1 – 2.5	1 – 1.5	1 – 1.5
ε_{+}, %	0.2 – 0.4	0.2 – 0.4	0.2 – 0.4	0.2 – 0.4	0.2 – 0.4	0.2 – 0.4	0.2 – 0.5
ε-, %	0.2 – 0.4	0.2 – 0.4	0.2 – 0.4	0.2 – 0.4	0.3 – 0.5	0.01 – 0.4	0.01 – 0.4
$f_{+,act}$, %	90 – 100	90 – 100	90 – 100	90 – 100	90 – 100	90 – 100	80 – 90
$f_{-,act}$, %	90 – 100	90 – 100	90 – 100	90 – 100	70 – 100	90 – 100	90 – 100
δ_{sep}, µm	15 – 20	15 – 20	15 – 20	15 – 20	15 – 20	15 – 40	15 – 40
ρ_{sep}, g/cm^3	0.5 – 1.5	0.5 – 1.5	0.5 – 1.5	0.5 – 1.5	0.5 – 1.5	0.5 – 4	0.5 – 4
ρ_{elyte}, g/cm^3	1.0 – 1.5	1.2 – 1.5	1.2 – 1.5	1.2 – 1.5	1.2 – 1.5	0.5 – 4	0.5 – 4
ASI, Ω cm^2	10 – 30	10 – 30	10 – 30	10 – 30	15 – 40	20 – 50	20 – 50

Header Definitions

- **G:** Graphite negative active electrode material
- **Si/G:** Silicon/graphite composite negative active electrode material with up to 100% silicon
- **Li:** Lithium metal negative active electrode material
- **NMCA:** Nickel- and cobalt-containing layered oxide[*i.e.*, nickel-cobalt-aluminum oxide ($LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$, NCA) and nickel-manganese-cobalt oxide ($LiNi_{1-x-y}Mn_xCo_yO_2$, NMC)] positive active electrode materials
- **LFP:** Lithium iron phosphate ($LiFePO_4$) positive active electrode material
- **LMO:** Lithium manganese oxide ($LiMn_2O_4$) positive active electrode material
- **LCO:** Lithium cobalt oxide ($LiCoO_2$) positive active electrode material
- **S:** Sulfur positive active electrode material

Energy and Power Density Figures

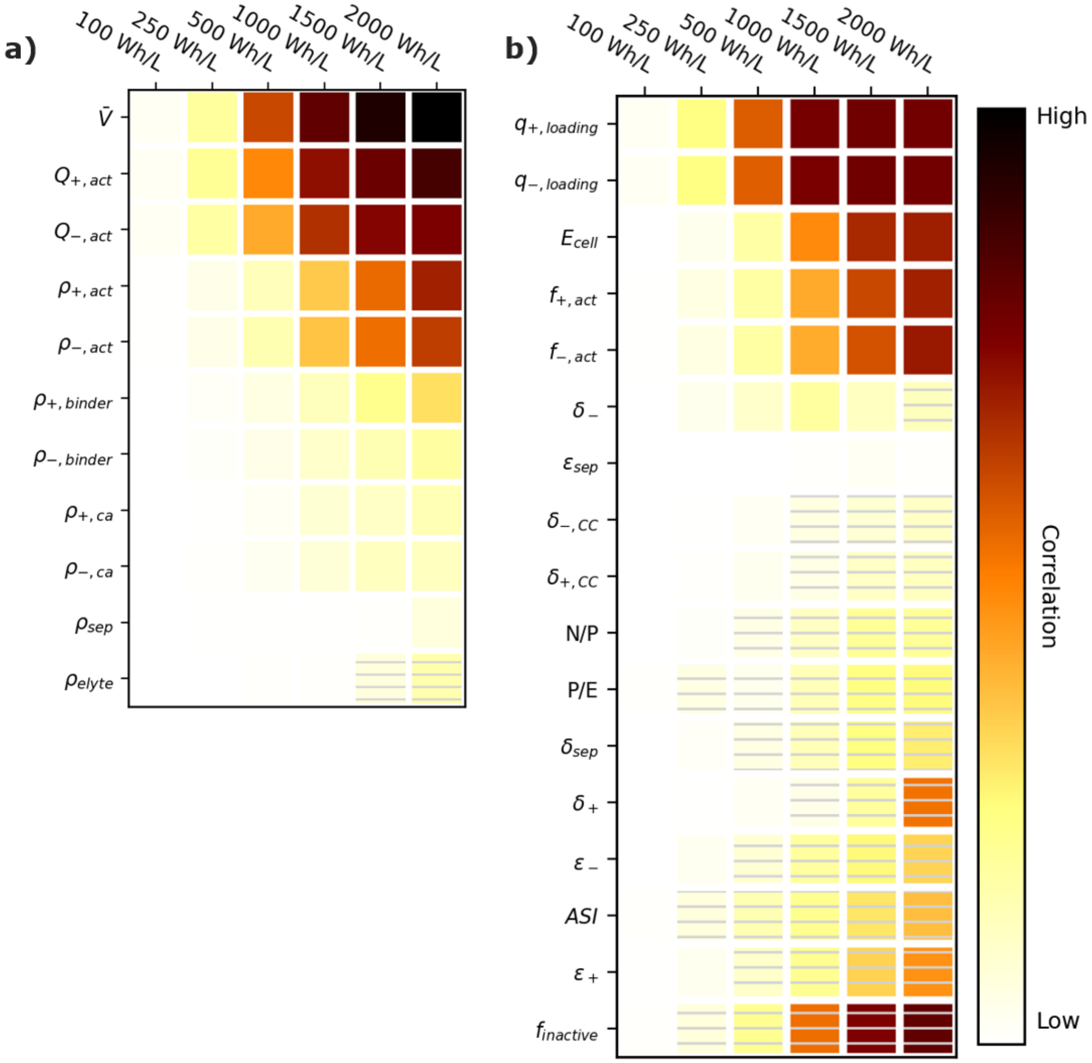


Figure S2. Correlation plots showing the relative importance of optimizing (a) material and (b) battery design parameters to achieving selected energy density (Wh/L) targets, for the input ranges given in Table 2. Solid darker colors indicate the variable has a high degree of positive correlation, where higher values are necessary to achieve the goal. Hatched darker colors indicate strong negative correlation, where lower values are needed to meet the target. Light colors indicate the variable has minimal or no correlation, and its value has a limited impact on achieving the target. The relative correlation of parameters to energy density remains similar with the correlation to specific energy (Wh/kg) observed in Figure 3 for most parameters. One difference is a significant decrease in the correlation for the electrolyte density ρ_{elyte} because mass is not a concern in energy density. Another difference is an increase in the correlation for the active material densities ($\rho_{\pm,act}$). This difference occurs because, for a fixed specific capacity ($Q_{\pm,act}$ in mAh/g), increases in $\rho_{\pm,act}$ provide decreases in the total volume, which are useful for achieving energy density targets.

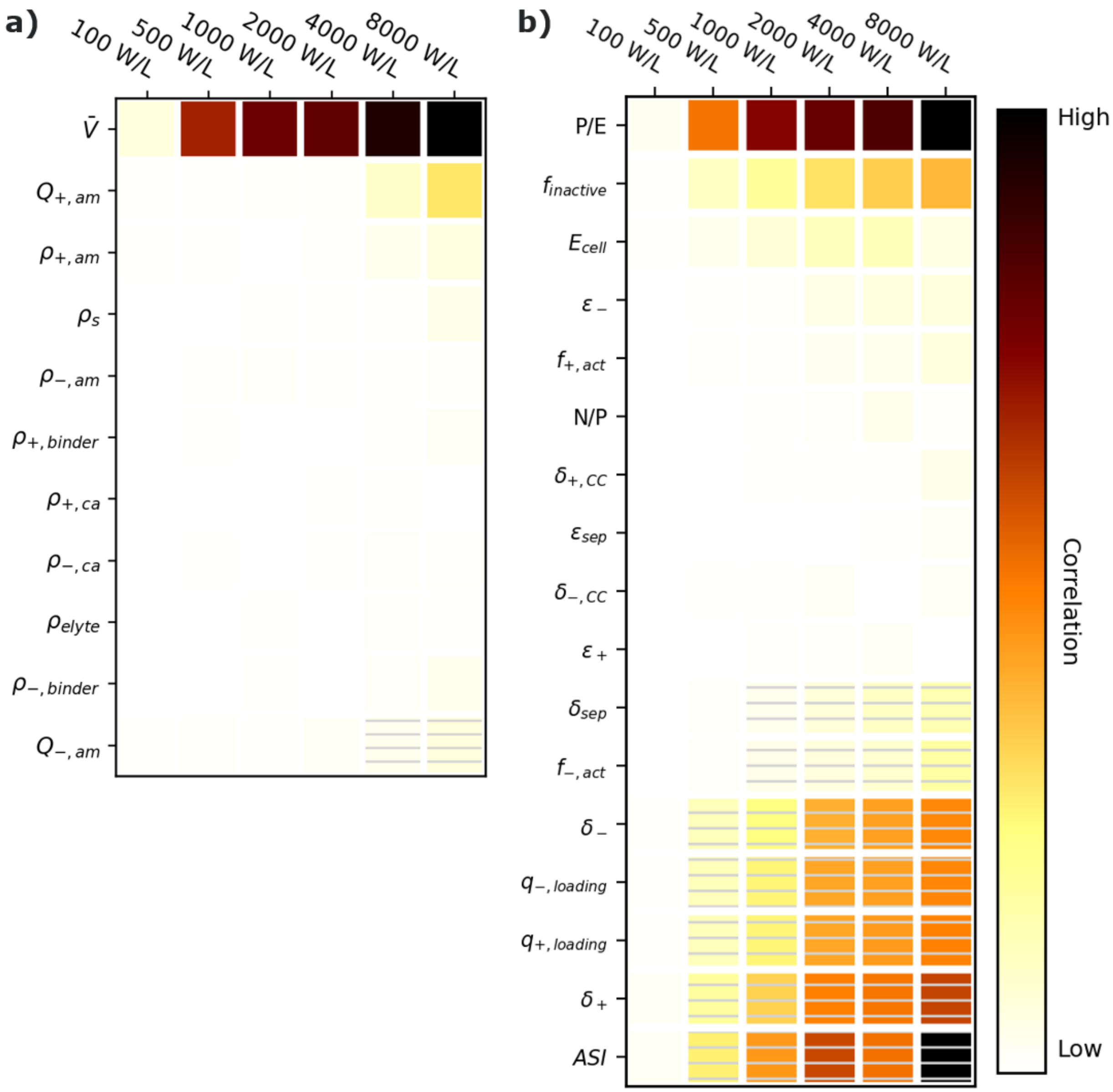


Figure S3. Correlation plots showing the relative importance of optimizing (a) material and (b) battery design parameters to achieving selected power density (W/L) targets, for the input ranges in Table 2. Solid darker colors indicate the variable has a high degree of positive correlation, where higher values are necessary to achieve the goal. Hatched darker colors indicate strong negative correlation, where lower values are needed to meet the target. Light colors indicate the variable has minimal or no correlation, and its value has a limited impact on achieving the target. The relative correlation of parameters to power density remains similar to the correlation with specific power (W/kg) observed in Figure 5 for most parameters. One difference is a decrease in the correlation for the electrolyte density ρ_{elyte}, because mass is not a concern in power density targets.

Supplementary References

1 J. C. F. de Winter, S. D. Gosling and J. Potter, *Psychol. Methods*, 2016, **21**, 273–290.