SUPPORTING INFORMATION

Quantifying the Search for Solid Li-Ion Electrolyte Materials by Anion: A Data-Driven Perspective

Austin D. Sendek,[1,2] Gowoon Cheon,[2] Mauro Pasta,[3] Evan J. Reed[1*]

1. Department of Materials Science and Engineering, Stanford University, 496 Lomita Mall, Stanford, CA, USA
2. Department of Applied Physics, Stanford University, 348 Via Pueblo, Stanford, CA, USA
3. Department of Materials, Oxford University, 21 Banbury Road, Oxford OX2 6HT, UK

[*]Corresponding author: evanreed@stanford.edu; (650) 723-2971

S1. Data distributions in performance space

In order to quantify the likelihood of outlier discoveries in each material family, we employ two methods. In the first, we count the fraction of all materials in each family that possess predicted performance values beyond the minimum required metrics ("binning"). The second method entails assigning a Gaussian probability distribution for each family and computing the cumulative probability of the distribution that lies beyond the minimum required performance metrics ("integrating"). The integration approach assumes that the predictive metrics we employ here (data-driven ionic conductivity classifier and DFT-based electrochemical stability calculations) are sufficiently predictive as to elucidate the general expected performance distribution of each family on the whole, even if some error exists within individual predictions.

We find good agreement between both the counting method and the Gaussian distribution method, as provided in Table 1 of the main text, suggesting these calculations are fairly robust with respect to the two different methods.

The accuracy of the Gaussian distribution method relies on the data being approximately Gaussian distributed. In order to assess the Gaussian character of the data, we provide in Figs. S1-10 normalized histograms of all four performance metrics for each family: band gap, oxidation potential, reduction potential, and superionic likelihood. For each distribution we perform the Shapiro-Wilk test for normality[1] and provide the resulting *p*-values in the legend. We note that the Shapiro-Wilk test is very sensitive to outliers, which brings the *p*-value down in several cases despite a Gaussian-looking distribution. A minimum *p*-value of 0.05 for having confidence in the Gaussian character of the data is commonly used.

The Gaussian distribution for each family in two-dimensional Gaussian performance space is defined by mean μ and covariance Σ, which we denote $\mathcal{N}(\mu, \Sigma)$. We calculate P_{outlier} by integrating the Gaussian over the upper-right quadrant of the graph. For the case of superionic behavior and a wide electrochemical stability window width V_{w} (Section *II,i*), we assume $V_{\text{w}} \approx E_{\text{gap}}$ and integrate the section of the distributions over the desired minimum window width $v_{\text{w,min}}$:

$$P_{\text{outlier}}(V_{\text{w}} \geq V_{\text{w,min}}) = \frac{\int_{0.5}^{1} dP_{\text{superionic}} \int_{V_{\text{w,min}}}^{\infty} dV\, \mathcal{N}(\mu, \Sigma)}{\int_{0}^{1} dP_{\text{superionic}} \int_{-\infty}^{\infty} dV\, \mathcal{N}(\mu, \Sigma)}$$

The numerator represents the Gaussian integrated over the upper-right quadrant of the graph, and the denominator represents the Gaussian integrated over the entire graph as a normalization

factor. The P_{outlier} values corresponding to a minimum stability window width $V_{\text{w,min}}$ = 4 V represent the likelihood that a new material from that family will have an electrochemical stability window of 4 V or wider. The likelihoods associated with $V_{\text{w,min}}$ = 4 V are provided in Table 1 along with the values computed from the binning method. We see good agreement between these methods.

In Section *II,ii*, we compute the likelihood of discovering a material with superionic conduction and an oxidation potential V_{ox} above some minimum value $V_{\text{ox,min}}$:

$$P_{\text{outlier}}(V_{\text{ox}} \geq V_{\text{ox,min}}) = \frac{\int_{0.5}^{1} dP_{\text{superionic}} \int_{V_{\text{ox,min}}}^{\infty} dV\, \mathcal{N}(\mu, \Sigma)}{\int_{0}^{1} dP_{\text{superionic}} \int_{-\infty}^{\infty} dV\, \mathcal{N}(\mu, \Sigma)}$$

The likelihoods corresponding to $V_{\text{ox,min}}$ = 4V are provided in Table 1.

In section *III,ii*, we compute the likelihood of discovering a material with superionic conduction and a reduction potential V_{red} below some maximum value $V_{\text{red,max}}$:

$$P_{\text{outlier}}(V_{\text{red}} \leq V_{\text{red,max}}) = \frac{\int_{0.5}^{1} dP_{\text{superionic}} \int_{-\infty}^{V_{\text{red}}} dV\, \mathcal{N}(\mu, \Sigma)}{\int_{0}^{1} dP_{\text{superionic}} \int_{-\infty}^{\infty} dV\, \mathcal{N}(\mu, \Sigma)}$$

The likelihoods corresponding to $V_{\text{red,max}}$ = 0V are provided in Table 1.

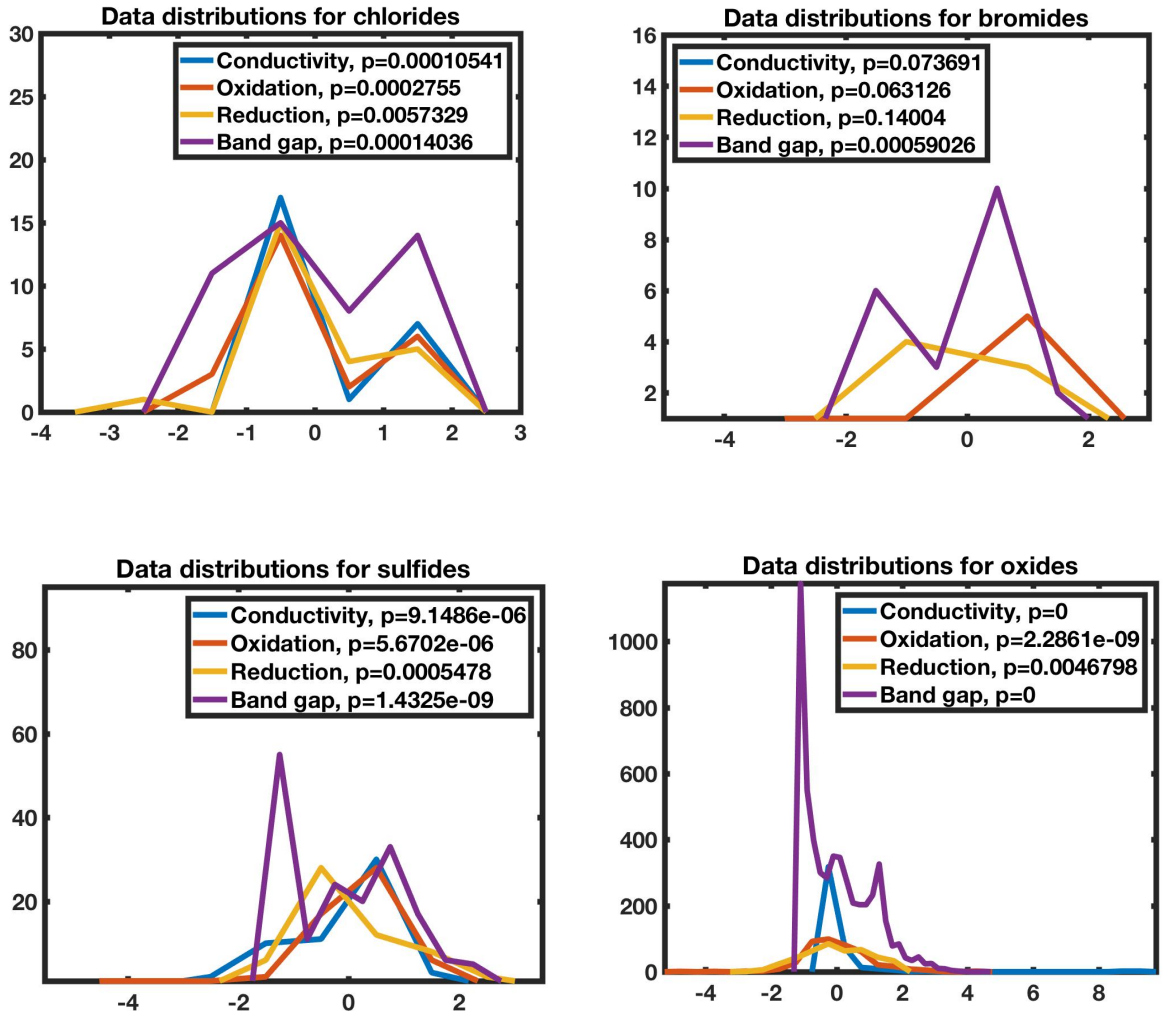
In section *II,iv*, we compute the likelihood of discovering an outlier material with superionic conduction that is stable against reduction at Li metal ($V_{\text{red,max}}$ = 0 V) and stable against oxidation to at least 4 V ($V_{\text{ox,min}}$ = 4 V). Zero of the candidate materials we study here satisfy this criteria, so we cannot employ the binning approach. Quantifying the likelihood of discovering such a "wonder material" requires extrapolation of the assumed

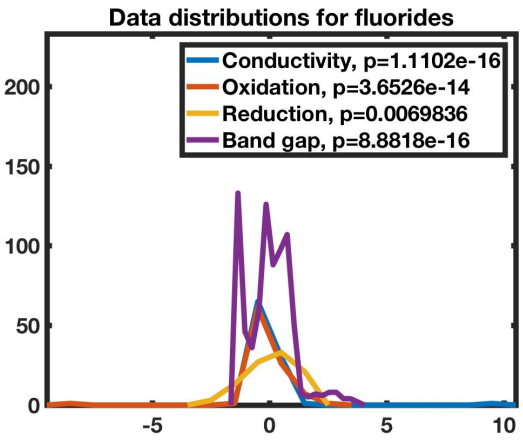
performance distributions, which the integration method allows. Treating the oxidation potential as independent of the reduction potential, the intersection of these two requirements requires a product of probabilities:

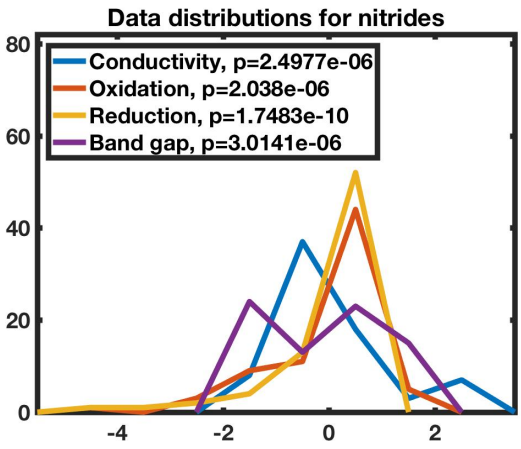
$$P_{\text{discovery}}\left(V_{\text{red}} \leq 0, V_{\text{ox}} \geq V_{\text{ox,min}}\right) = P_{\text{discovery}}(V_{\text{red}} \leq 0) \times P_{\text{discovery}}(V_{\text{ox}} \geq V_{\text{ox,min}})$$

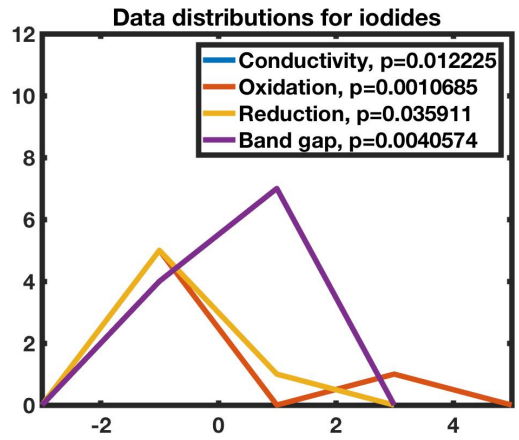
The results of this calculation are provided in Fig. 2(g) and Table 1.

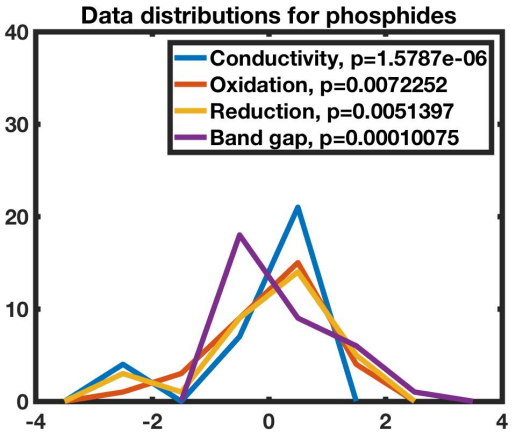
Figure S1. Distributions of performance data by family. For each of the ten families considered here, we visualize the (normalized) distributions of the band gap (electrochemical stability window width), oxidation potential, reduction potential, and superionic likelihood. For each distribution, we use the Shapiro-Wilks test to compute the likelihood of normality; the corresponding *p*-values are provided in the legends. The data looks approximately Gaussian in several cases, although the occasional presence of outliers can result in low *p*-values.

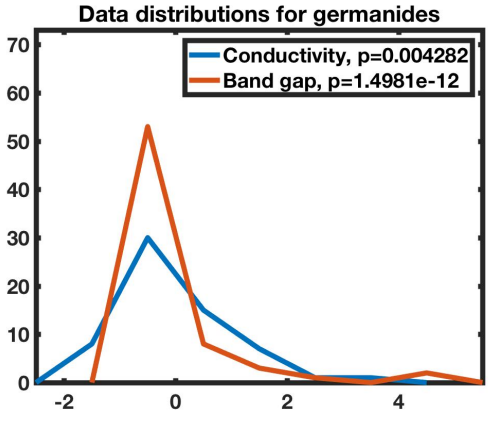



Data distributions for fluorides
Conductivity, p=1.1102e-16
Oxidation, p=3.6526e-14
Reduction, p=0.0069836
Band gap, p=8.8818e-16
200
150
100
50
0
-5
0
5
10


Data distributions for nitrides
Conductivity, p=2.4977e-06
Oxidation, p=2.038e-06
Reduction, p=1.7483e-10
Band gap, p=3.0141e-06
80
60
40
20
0
-4
-2
0
2


Data distributions for iodides
Conductivity, p=0.012225
Oxidation, p=0.0010685
Reduction, p=0.035911
Band gap, p=0.0040574
12
10
8
6
4
2
0
-2
0
2
4


Data distributions for phosphides
Conductivity, p=1.5787e-06
Oxidation, p=0.0072252
Reduction, p=0.0051397
Band gap, p=0.00010075
40
30
20
10
0
-4
-2
0
2
4


Data distributions for germanides
Conductivity, p=0.004282
Band gap, p=1.4981e-12
70
60
50
40
30
20
10
0
-2
0
2
4

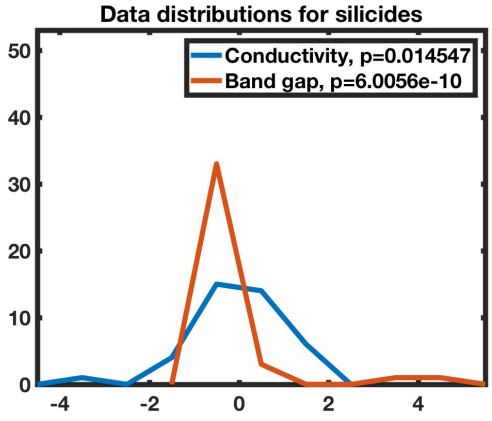
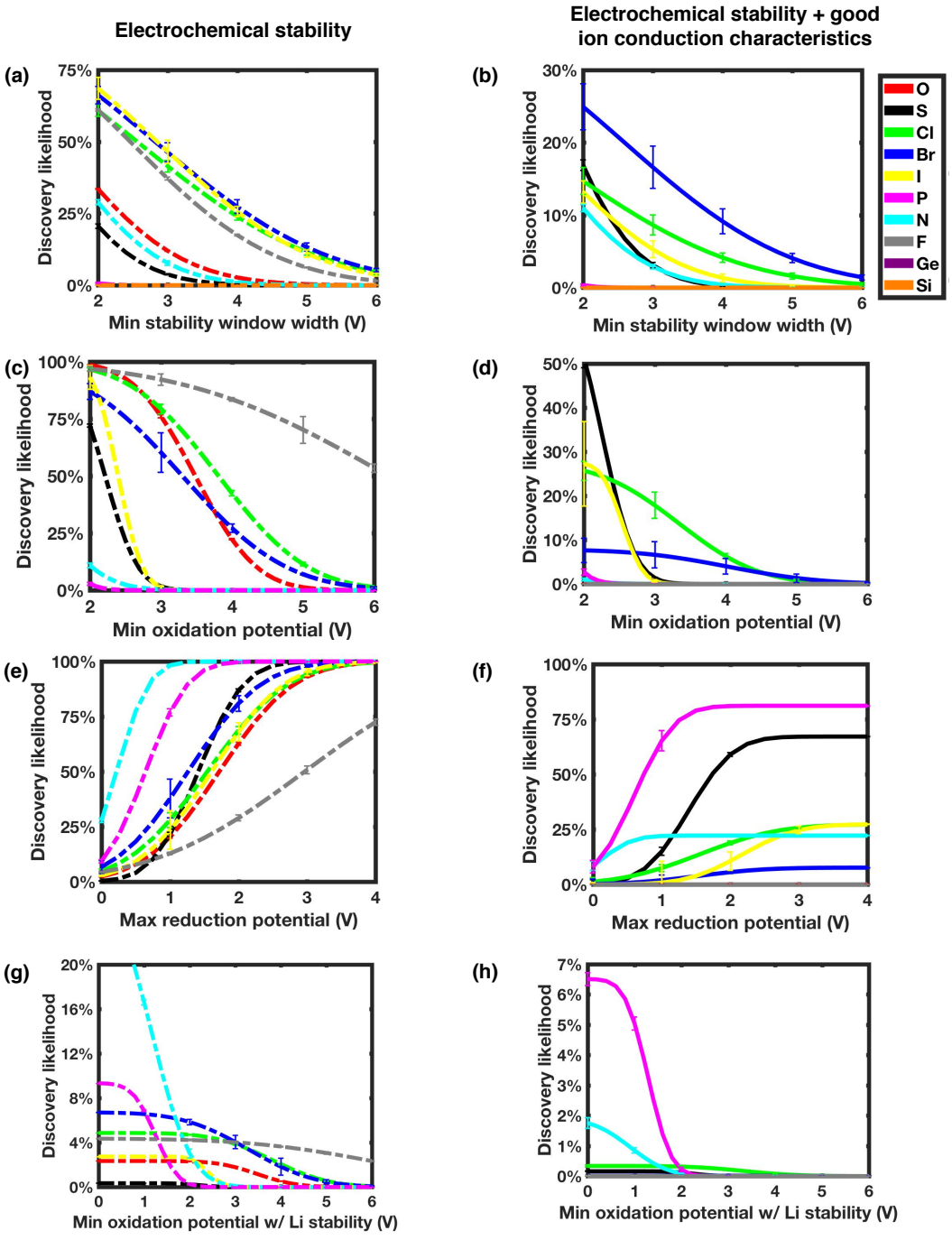

Data distributions for silicides
Conductivity, p=0.014547
Band gap, p=6.0056e-10
50
40
30
20
10
0
-4
-2
0
2
4

Figure S2: ***Outlier discovery likelihood by anion family.*** We use the Gaussian method to quantify outlier discovery likelihood for each of the four demands in Section II. In **(a)** we plot the likelihood of discovering an outlier material from each family that has an electrochemical stability window width exceeding the value represented on the *x*-axis. It can be seen that the fast conducting materials with window widths beyond 6V are rare, according to the structures and data available in the Materials Project database. In **(b)** we plot the likelihood of discovering an outlier material with fast ionic conductivity and a window width larger than the *x*-axis value. In **(c)** and **(d)** we perform the same calculation but replace the stability window width distribution with the oxidation potential distribution. In **(e)** we plot the likelihood of discovering a material in each family with a reduction potential below the value on the *x*-axis; x=0 corresponds to stability against Li metal. In **(f)** we add the ionic conductivity constraint. In **(g)** we plot the likelihood of discovering a material with stability against Li metal and an oxidation potential above the value on the x-axis; in **(h)** we add in the ionic conductivity constraint. In all plots, the error bars correspond to the standard deviations in likelihood values calculated upon removing 10% of the data at random ten times and recalculating the likelihoods.


Electrochemical stability
Electrochemical stability + good ion conduction characteristics
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Discovery likelihood
Min stability window width (V)
Min oxidation potential (V)
Max reduction potential (V)
Min oxidation potential w/ Li stability (V)
O
S
Cl
Br
I
P
N
F
Ge
Si

References

1. Shapiro, S. S., Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **52,** 591–611 (1965).