Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.
This journal is © The Royal Society of Chemistry 2023

FURTHER DETAIL ON TRAINING DATA GENERATION AND CLUSTER EXPANSION FITTING

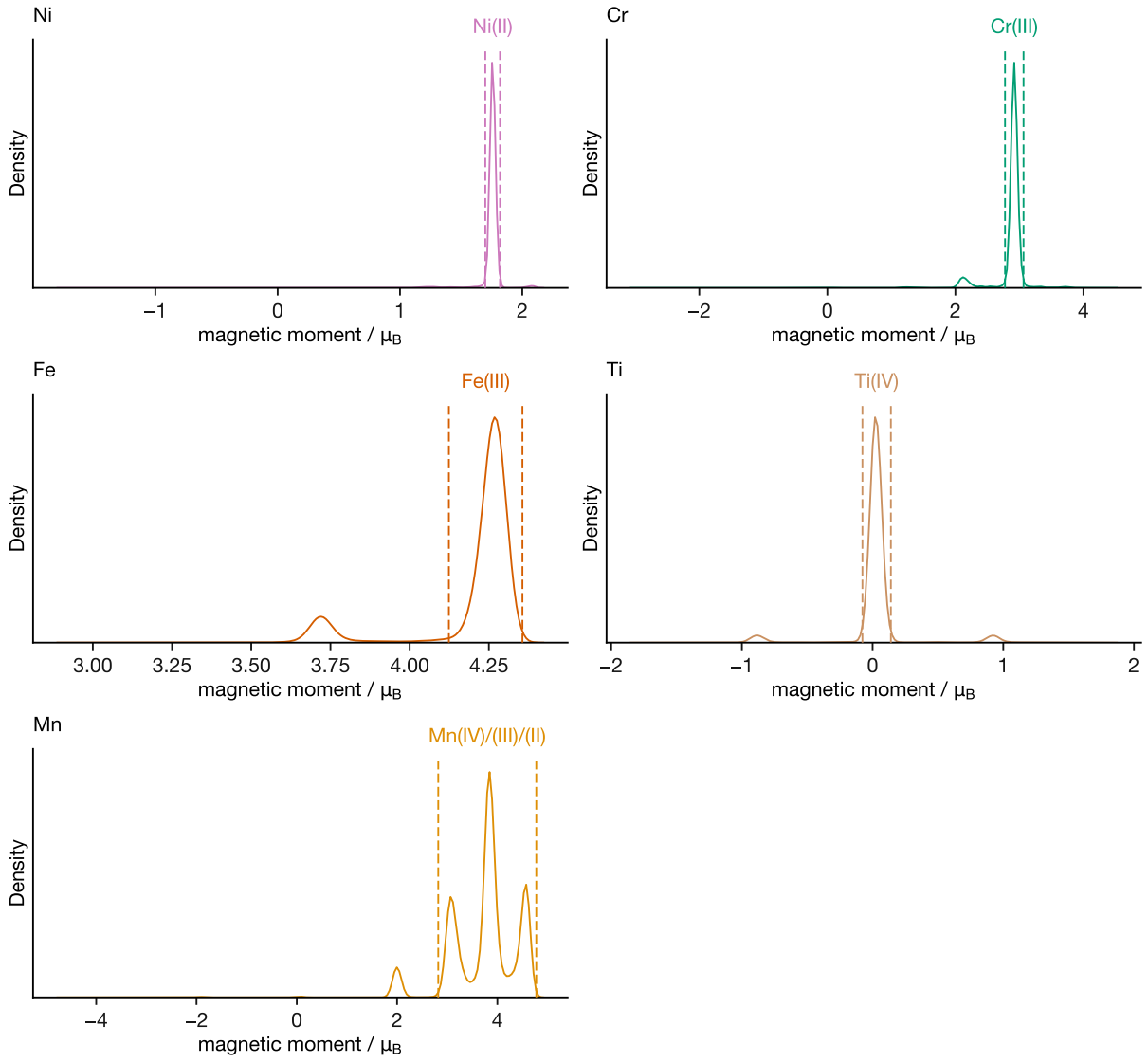

Figure 1. Probability density distributions of the atom-projected magnetic moments obtained from the training data which were used to assign oxidation states. The majority magnetic moments are show with their upper and lower bounds as dashed lines and labelled with their corresponding oxidation states.

Oxidation states were assigned from the atom-projected magnetic moments of the DFT-relaxed structures, with the corresponding probability densities shown in fig. 1. For all transition metals, barring manganese, a predominant oxidation state was identifiable and easily de-convoluted from the minority oxidation states based on the magnetic moments alone. Structures containing atoms deviating from these oxidation states were omitted from the training data. While a truly comprehensive model would include all possible oxidation states for each transition metal, the added complexity and increased data required to account for minority species were deemed excessive, and we do not believe that this potential model shortcoming impacts our ability to capture general trends in short range order suppression in many component disordered rock salt cathode materials. For the more complex case of manganese, we employed a Bayesian black-box optimization process to assign the variable oxidation state, using the number of charge balanced structures (with the oxidation states of all other species preassigned) as the loss

function [1, 2]. The magnetic moment cut offs determined via this approach are $\mu_B^{Mn(IV)} = 2.28\,\mu_B$ to $3.53\,\mu_B$, $\mu_B^{Mn(III)} = 2.53\,\mu_B$ to $4.38\,\mu_B$, and $\mu_B^{Mn(II)} = 2.38\,\mu_B$ to $4.78\,\mu_B$.

The DFT+U approach used in this study to correct for the inability of semi-local DFT to capture the electronic structure of transition metal ions is subject to a few limitations. In principle, the U parameter is sensitive to chemical environment, meaning that a universal U parameter across all species in the system is dubious. We acknowledge that hybrid functional calculations are generally more reliable than the DFT+U for describing mixed oxidation state transition metal systems, but the computational cost of running many thousands of hybrid-DFT studies was deemed prohibitive. We note that this same approach has been applied in multiple studies of disordered rocksalt cathode materials, with those results then successfully correlated to experiment [3–6].

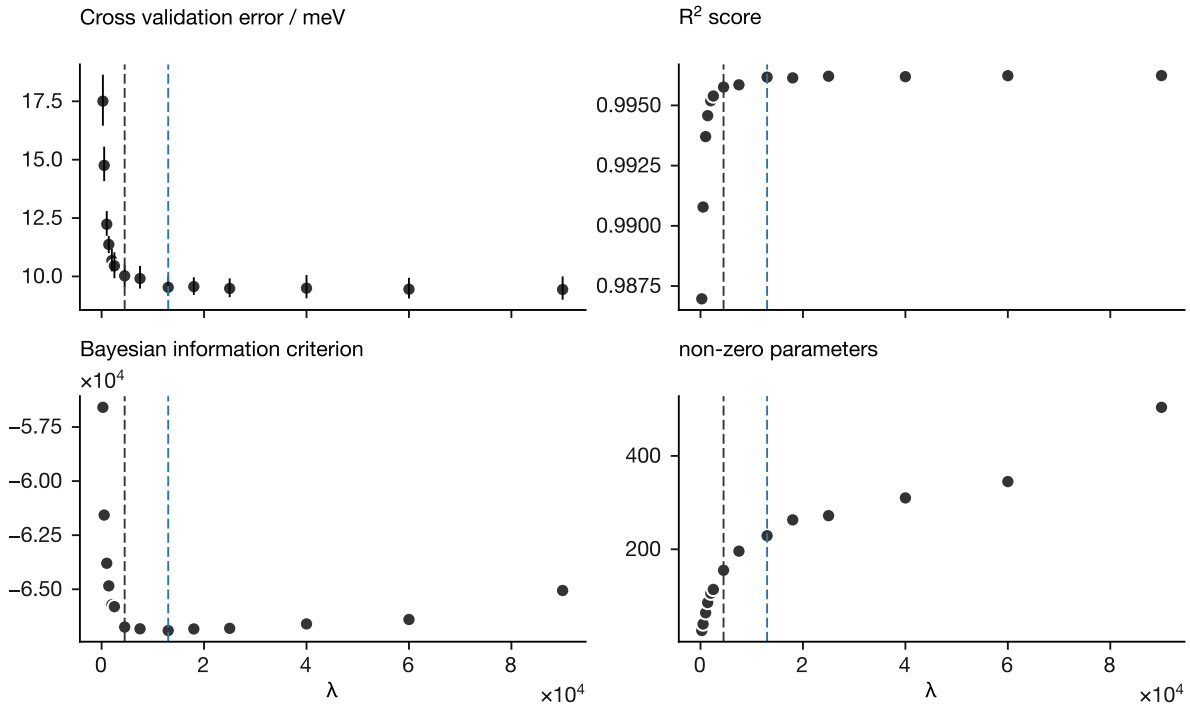

Figure 2. Cross validation errors, R^2 score, Bayesian information criterion and number of non-zero parameters as a function of the automatic relevance detection regression parameter λ. The dashed blue line highlights the fit with the lowest Bayesian information criterion, and the black dashed line shows the values associated with the production fit used in this work.

We use automatic relevance detection regression (ARDR) to fit the expansion based on the determined clusters. The cross validation errors, R^2 score, Bayesian information criterion (BIC) and number of non-zero parameters in the fit are shown as a function of the regularization hyper-parameter (λ) in Figure 2. As the cluster expansion is very high-component (8 species on the cation site, and 2 on the anion site), to prioritise model sparsity, we select the value of λ used in the production fit according to the "one standard error" rule, taking the most sparse fit within one standard deviation of the lowest Bayesian information criterion calculated from the λ scan indicated as blue dashed line on fig. 2; this fit has a cross validation error is 9.60 meV. This final production fit is shown in λ scan indicated as black dashed line on fig. 2. The overall model performance for the lowest BIC and production models are summarised in fig. 3.

[1] J. H. Yang, T. Chen, L. Barroso-Luque, Z. Jadidi, and G. Ceder, Approaches for handling high-dimensional cluster expansions of ionic systems, npj Computational Materials **8**, 1 (2022).
[2] L. Barroso-Luque, P. Zhong, J. H. Yang, F. Xie, T. Chen, B. Ouyang, and G. Ceder, Cluster expansions of multicomponent ionic materials: Formalism and methodology, Phys. Rev. B Condens. Matter **106**, 144202 (2022).

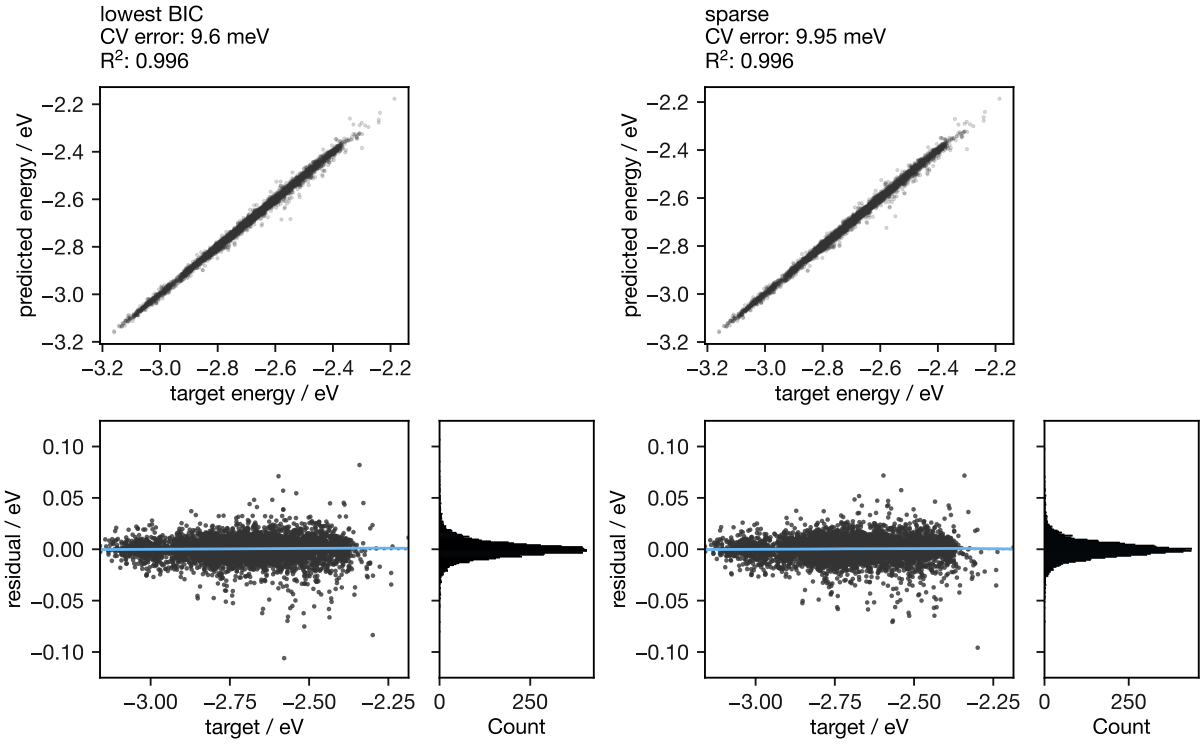

Figure 3. Visualisation of the model performance for the lowest Bayesian information criterion fit (lowest BIC, left) and the final model used in this work (sparse, right). The top plot in each case shows the target energies plotted against the predicted energies, with the plot underneath showing the model residuals with a Locally Weighted Scatterplot Smoothing (LOWESS) fit shown in blue, and the distribution of the residuals shown as a histogram.

[3] Z. Lun, B. Ouyang, D.-H. Kwon, Y. Ha, E. E. Foley, T.-Y. Huang, Z. Cai, H. Kim, M. Balasubramanian, Y. Sun, J. Huang, Y. Tian, H. Kim, B. D. McCloskey, W. Yang, R. J. Clément, H. Ji, and G. Ceder, Cation-disordered rocksalt-type high-entropy cathodes for Li-ion batteries, Nat. Mater. **20**, 214 (2021).

[4] K. McColl, R. A. House, G. J. Rees, A. G. Squires, S. W. Coles, P. G. Bruce, B. J. Morgan, and M. S. Islam, Transition metal migration and O_2 formation underpin voltage hysteresis in oxygen-redox disordered rocksalt cathodes, Nat. Commun. **13**, 5275 (2022).

[5] Z. Lun, B. Ouyang, Z. Cai, R. J. Clément, D.-H. Kwon, J. Huang, J. K. Papp, M. Balasubramanian, Y. Tian, B. D. McCloskey, H. Ji, H. Kim, D. A. Kitchaev, and G. Ceder, Design principles for High-Capacity Mn-Based Cation-Disordered rocksalt cathodes, Chem **6**, 153 (2020).

[6] H. Ji, A. Urban, D. A. Kitchaev, D.-H. Kwon, N. Artrith, C. Ophus, W. Huang, Z. Cai, T. Shi, J. C. Kim, H. Kim, and G. Ceder, Hidden structural and chemical order controls lithium transport in cation-disordered oxides for rechargeable batteries, Nat. Commun. **10**, 592 (2019).