cm CHEMISTRY OF MATERIALS

Comment on “Environmental Stability of Crystals: A Greedy Screening”

Cite This: *Chem. Mater.* 2023, 35, 801−803

Read Online

In a recent paper in this journal (*Chem. Mater.* **2022**, *34*, 2545−2552),[1] Twyman et al. studied the environmental stability of crystals using a greedy heuristic algorithm to identify oxidation products that approximately minimize reaction enthalpy. Reformulating their problem as a linear optimization allows it to be solved exactly, with less code and a comparable computational time. Use of the greedy heuristic results in quantitative and qualitative errors in the oxidation product prediction for 56.8% of the starting materials they consider. Consideration of copper oxidation illustrates the limitations of their underlying premise that thermodynamics determines environmental stability.

1. Oxidation Product Prediction as a Thermodynamic Minimization Problem. Twyman et al. posit that the necessary conditions for self-passivation are the formation of a thermodynamically stable oxide phase that is mechanically matched with the parent material. They describe the mechanical condition in terms of the Pilling−Bedworth ratio and the thermodynamic condition in terms of the following optimization problem: Given an oxygen-free reactant compound $\mathcal{M}$, and a set of all possible oxidation products $\{\Omega_1, \Omega_2, ..., \Omega_n\}$, the most stable set of products will be those that minimize the reaction enthalpy (alternatively, free energy) of the oxidation reaction

$$\mathcal{M} + yO_2 = x_1\Omega_1 + x_2\Omega_2 + ... + x_n\Omega_n \tag{1}$$

where $(x_1, x_2, ..., x_n) = \vec{x} \geqslant 0$ and y is an arbitrary non-negative number as the system is open to oxygen.

Twyman et al. note that this problem has the form of a *continuous* (or fractional) knapsack problem. Indeed, *integer* knapsack problems (also known as 0/1 knapsack problems) are NP-complete, and smaller problems are typically solved through dynamic programming, whereas larger problems require heuristics.[2] (A problem is in the nondeterministic polynomial-time (NP) complexity class if a proposed solution can be verified in polynomial time and a brute-force enumeration of possible candidates will find the true solution. An NP problem is NP-complete if all NP problems can be converted to it in polynomial time. Practically, solving NP-complete problems requires a time growing exponentially with input size to enumerate the brute-force search space.[3]) However, *continuous* knapsack problems can be solved exactly as a linear optimization problem (also known as linear programming, LP) in polynomial time (complexity class P).[4] Essentially, one solves for a vector of products minimizing the total formation enthalpy, subject to a set of stoichiometric constraints, expressed in canonical form as

$$\begin{aligned} &\text{minimize} && \vec{h}^T \cdot \vec{x} - h_0 \\ &\text{subject to} && \boldsymbol{s} \cdot \vec{x} = \vec{s_0} \\ & && \vec{x} \geqslant 0 \end{aligned} \tag{2}$$

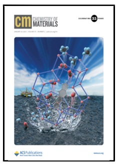
where $\vec{h}$ are the formation enthalpies of products in energy per atom (h_0 is the enthalpy of formation of the starting compound, $\mathcal{M}$) and $\boldsymbol{s} = (\vec{s_1}, ..., \vec{s_n})$ is a matrix whose column vectors are the stoichiometric coefficients for the elements to be conserved, divided by the total number of atoms (whether conserved or not), for each compound. (The superscript T indicates a transpose of the column vector to a row vector.) The corresponding vector of normalized stoichiometric coefficients for the starting compound, $\mathcal{M}$, is denoted $\vec{s_0}$. Only the stoichiometric constraints expressed by entries in $\boldsymbol{s}$ are imposed, and by omitting the row corresponding to an element (e.g., oxygen), no stoichiometric constraint is imposed.

To illustrate this approach, consider the example used by Twyman et al. in Figure 4 of their paper on the oxidation of a hypothetical compound $\mathcal{M} = AB$ into possible oxide products $\Omega = \{A_2BO, AB_2O, AO, BO\}$. The data for their example corresponds to

$$\begin{aligned} \vec{x} &= (x_1, x_2, x_3, x_4) \\ \vec{h} &= (-10, -9, -5, -2) \\ h_0 &= -1 \\ \vec{s_0} &= (1/2, 1/2) \\ \boldsymbol{s} &= \left(\begin{pmatrix}2\\1\end{pmatrix}/4, \begin{pmatrix}1\\2\end{pmatrix}/4, \begin{pmatrix}1\\0\end{pmatrix}/2, \begin{pmatrix}0\\1\end{pmatrix}/2\right) \end{aligned}$$

where formation enthalpies (h_0, $\vec{h}$) are in eV/atom. Notice how in $\boldsymbol{s}$, only the A and B stoichiometries are described, not the O stoichiometry. Solving the problem defined in eq 2 yields the optimal solution $\vec{x} = (0.667, 0.667, 0, 0)$, just as in their example.

Received: August 8, 2022
Revised: October 31, 2022
Published: January 6, 2023



https://doi.org/10.1021/acs.chemmater.2c02434

Whereas the greedy algorithm is approximate, LP finds the exact solution.[4] LP is simpler to implement because one can use existing, highly efficient convex optimization software packages. Excluding comments and formatting, Twyman's greedy heuristic requires 76 lines of Python code, whereas only 37 lines of Python code are needed to set up the problem and call an external LP package. LP solvers are highly efficient: Optimizing the stable oxidation products for 39,634 materials requires only 52 s on a personal computer. Thus, LP has the advantages of being exact, easier to implement, and fast.

The same mathematical structure can be used to minimize the free energy by transforming the variables in eq 2, as the reactants and products are either solids (with unit activity) or gaseous oxygen at constant pressure, so there is no need to include logarithmic terms in the free energy expression, and the expression remains linear. Grand-canonical free energy minimization problems, where the system is open to reservoirs of one or more components, are routinely solved using LP in *OQMD*[5−7] and *pymatgen*.[8,9] The method implemented in *pymatgen* constructs a convex hull defined by the thermodynamic inequalities and then examines the vertices, which is functionally equivalent to Dantzig's simplex algorithm for solving LP problems.

2. The Greedy Heuristic Gives Quantitative and Qualitative Errors in Oxidation Product Prediction. We repeated Twyman et al.'s calculations determining the most stable oxidation products for 39,634 starting materials from the Materials Project[10] using both their greedy algorithm ($\Delta H_{\text{ox}}^{\text{Greedy}}$) and a LP solution ($\Delta H_{\text{ox}}^{\text{LP}}$) using Gurobi.[11] The greedy algorithm is suboptimal for 22,504 (56.8%) starting materials; the remaining cases are the same. The distribution of nonzero $\Delta H_{ox}^{LP} - \Delta H_{\text{ox}}^{\text{Greedy}}$ is shown in Figure 1. The error (difference between ΔH_{ox}^{LP} and $\Delta H_{\text{ox}}^{\text{Greedy}}$) can be as large as −1.480 eV/atom, and the mean absolute error of the greedy approach is 0.119 eV/atom. In addition to *quantitatively* suboptimal reaction enthalpies, the greedy algorithm gives *qualitatively* incorrect predictions of the reaction products, i.e., it predicts products that are absent in the exact LP solution and vice versa. In fact, a nonzero $\Delta H^{\text{LP}} - \Delta H^{\text{Greedy}}$ value *always* corresponds to a set of qualitatively wrong products predicted for the greedy algorithm. As an example, Table 1 shows the predicted products of $Sr_{0.5}Se_{0.25}S_{0.25}$ (mp-1218707), the worst prediction by the greedy algorithm. The greedy algorithm predicts three binary oxides, whereas the exact calculation predicts two ternary oxides. More examples can be found in the notebook "whygreedy.ipynb" in the repository (see Data Availability section for access). Thus, the greedy algorithm unreliably predicts the oxidation products, undermining conclusions about environmental stability.

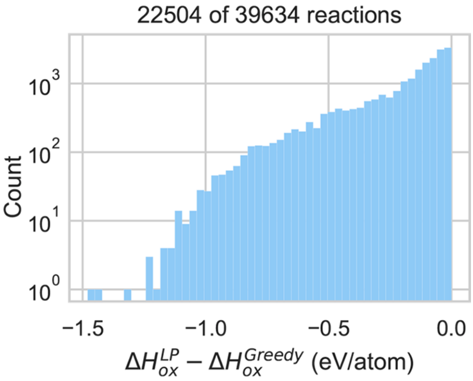


Figure 1. Distribution of nonzero (>10^{-5} eV/atom) values of $\Delta H_{\text{ox}}^{LP} - \Delta H_{\text{ox}}^{\text{Greedy}}$.

Table 1. Predicted Oxidation of $Sr_{0.5}Se_{0.25}S_{0.25}$ (mp-1218707) Predicted by the Greedy Approximation versus the Exact Solution

Method	Products	$\vec{h}^T\vec{x}$ (eV/atom)
Greedy	1 × $S_{0.25}O_{0.75}$ (mp-2414) + 0.875 × $Se_{0.286}O_{0.714}$ (mp-27358) + 1.5 × $Sr_{0.333}O_{0.667}$ (mp-2697)	−5.968
LP	0.25 × $Sr_{0.167}S_{0.167}O_{0.667}$ (mp-5285) + 0.25 × $Sr_{0.167}Se_{0.167}O_{0.667}$ (mp-4092)	−7.488

3. Copper Oxidation Illustrates the Limitations of the Thermodynamic Approach to Self-Passivation and Environmental Stability. Even with an exact calculation, using computed oxidation enthalpies can result in unreasonable predictions. Optimizing the oxidation reaction eq 1 using computed enthalpies of formation from the Materials Project, the oxidation product of pure copper is predicted to be Cu_2O_3 (mp-771359). Although computationally this compound is predicted to be stable (in the convex hull of Cu−O), Cu_2O_3 has never been observed experimentally. This illustrates limitations inherent in using computed data and/or the formation enthalpies at 0 K (as opposed to free energies at finite temperature). Furthermore, oxidation reactions of interest for environmental stability happen almost exclusively at the solid−air interface and are, thus, subject to diffusion mechanisms of all species involved. At room temperature, long-term air exposure of copper film yields an inner layer of cuprous oxide (Cu_2O) and an outer layer of cupric oxide (CuO), and the latter only develops when the former is thick enough such that Cu diffusion is partially blocked.[12−14] Even a thermodynamic optimization that limits the species to solely Cu, CuO, and Cu_2O would predict only CuO formation, not the intermediate Cu_2O layer. Fortuitously CuO and Cu_2O have similar Pilling−Bedworth ratios with respect to copper, so this would not change the environmental stability conclusion; however, this is not generally the case. One potential solution for such diffusion-limited oxidations is to solve eq 1 with a gradually increasing y instead of an unbounded y (Figure 2), which naïvely simulates the oxygen activity profile[15] from bulk copper to the outmost layer of copper oxide. This method assumes there is only one reductant or only one clear winner in the diffusion competition at the solid−air interface. For complex alloys, diffusion rates are required to calculate relative activities of reductants.

■ CONCLUSION: WHY BE GREEDY WHEN YOU CAN BE EXACT?

Greedy heuristics are a valuable algorithmic tool for finding approximate solutions to challenging optimization problems. However, for this problem, a greedy approximation is not only unnecessary—the exact solution is of comparable computational cost and easier to implement—but also leads to qualitative failures.

Our analysis was facilitated by open code availability. Full disclosure of program code allowed us to closely examine their algorithm, resolve narrative ambiguity about the implementation, and directly compare results and timings. This illustrates the value of code sharing for advancing scientific discourse, and

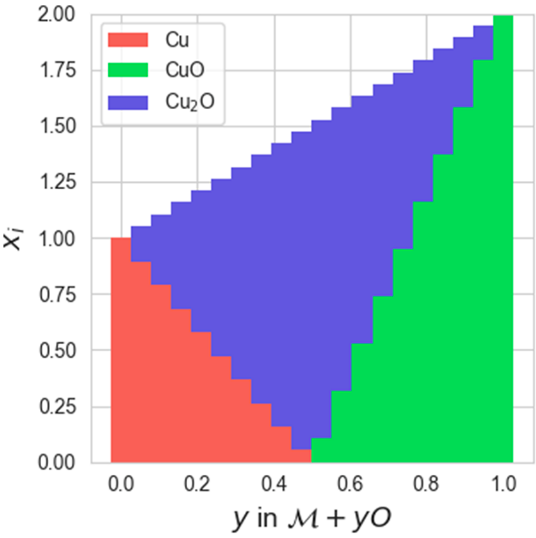


Figure 2. Predicted oxidation products for copper as a function of *y* in eq 1, limiting the product species to CuO and Cu_2O.

we hope it will encourage authors, editors, and publishers to mandate complete dissemination of the code used for scientific papers. A release of the source code can also be found at https://doi.org/10.5281/zenodo.7504862.

Qianxiang Ai orcid.org/0000-0002-5487-2539
Joshua Schrier orcid.org/0000-0002-2071-1657

■ ASSOCIATED CONTENT

Data Availability Statement

Python 3.9 source code used for all calculations, details regarding the data set, and related documentation are available at https://github.com/qai222/whygreedy.

■ AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chemmater.2c02434

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge support from the National Science Foundation (DMR-1928882), the Henry Dreyfus Teacher-Scholar Award (TH-14-010), and the resources of the MERCURY consortium (http://mercuryconsortium.org/) under NSF Grants CNS-2018427. We thank Dr. Shyam Dwaraknath and Dr. Joseph Montoya for referring us to examples of grand canonical free energy calculations for solid-state problems.

■ REFERENCES

(1) Twyman, N. M.; Walsh, A.; Buonassisi, T. Environmental. *Chem. Mater.* **2022**, *34* (6), 2545.

(2) Skiena, S. S.*The Algorithm Design Manual*; Springer: London, 2008; DOI: 10.1007/978-1-84800-070-4.

(3) Hartmanis, J. Computers and Intractability: A Guide to the Theory of Np-Completeness (Michael r. Garey and David s. Johnson). *Siam Rev.* **1982**, *24* (1), 90.

(4) Boyd, S. P.; Vandenberghe, L.*Convex Optimization*; Cambridge University Press: Cambridge, U.K.; New York, 2004.

(5) Akbarzadeh, A. R.; Wolverton, C.; Ozolins, V. First-Principles Determination of Crystal Structures, Phase Stability, and Reaction Thermodynamics in the Li-Mg-Al-H Hydrogen Storage System. *Phys. Rev. B* **2009**, *79* (18), 184102.

(6) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies. *Npj Comput. Mater.* **2015**, *1* (1), 15010.

(7) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65* (11), 1501−1509.

(8) Ong, S. P.; Wang, L.; Kang, B.; Ceder, G. Li−Fe−P−O_2 Phase Diagram from First Principles Calculations. *Chem. Mater.* **2008**, *20* (5), 1798−1807.

(9) Ong, S. P.; Jain, A.; Hautier, G.; Kang, B.; Ceder, G. Thermal Stabilities of Delithiated Olivine MPO4 (M= Fe, Mn) Cathodes Investigated Using First Principles Calculations. *Electrochem. Commun.* **2010**, *12* (3), 427−430.

(10) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. a. The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), 011002.

(11) Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*; 2022.

(12) Lim, J.-W.; Iijima, J.; Zhu, Y.; Yoo, J. H.; Choi, G.-S.; Mimura, K.; Isshiki, M. Nanoscale Investigation of Long-Term Native Oxidation of Cu Films. *Thin Solid Films* **2008**, *516* (12), 4040−4046.

(13) León, J. J. D.; Fryauf, D. M.; Cormia, R. D.; Kobayashi, N. P.Study of the Formation of Native Oxide on Copper at Room Temperature. In *Low-Dimensional Materials and Devices 2016*; SPIE: 2016; Vol. *9924*, pp 63−69.

(14) Pilling, N. B.; Bedworth, R. E. The Oxidation of Metals at High Temperatures. *J. Inst. Met.* **1923**, *29*, 529−591.

(15) Chatterjee, U. K.; Bose, S. K.; Roy, S. K.*Environmental Degradation of Metals: Corrosion Technology Series/14*; CRC Press: 2001.