

# A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis

Matthew J. McDermott<sup>1,2</sup>, Shyam S. Dwaraknath<sup>1</sup>, and Kristin A. Persson<sup>1,2,\*</sup>

<sup>1</sup>Energy Technologies Area, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

<sup>2</sup>Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA

\*Corresponding author: Kristin Persson (email: [kapersson@lbl.gov](mailto:kapersson@lbl.gov))

June 24, 2020

# Abstract

Accelerated synthesis of inorganic materials remains a significant challenge in the search for novel, functional materials. Many of the chemical principles which enable “synthesis by design” in synthetic organic chemistry do not exist in solid-state chemistry, despite extensive computed/experimental thermochemistry data. We present a chemical reaction network model constructed from thermochemistry databases that captures features of the thermodynamic phase space which synthesis reactions traverse. Directed edges in the network are assigned weights via a transformation that maps reaction parameters to costs. We devise a computationally tractable approach for suggesting likely reaction pathways via application of pathfinding algorithms and linear combination of lowest-cost paths in the network. We demonstrate initial success of the reaction network in predicting a complex metathesis reaction pathway toward yttrium manganese oxide ( $\text{YMnO}_3$ ). The reaction network presents new opportunities for enabling reaction pathway prediction, rapid iteration between experimental/theoretical results, and ultimately, control of synthesis of solid-state materials.

## Introduction

Dating back to 18th century mineralogy,<sup>1</sup> solid-state inorganic chemistry is a cornerstone in the design of novel, functional materials and continues to be driven by pressing technological demands. Consequently, the development of new techniques that accelerate materials synthesis/processing is vital for achieving multifunctional materials with complex properties that satisfy today’s technological needs. Solid materials with target functionality are often thermodynamically metastable, which can limit their accessibility via conventional solid-state synthesis routes such as the classic “shake and bake” ceramic methods that typically require high temperatures to overcome diffusion barriers and often proceed to global thermodynamic equilibrium.<sup>2</sup> Indeed, solid-state chemistry itself has been dubbed a “black box” which is most effectively probed via systematic and extensive iteration, requiring significant experimental expertise akin to apprenticed artistry.<sup>3</sup> The optimization of synthesis procedures for new materials is hence both highly time- and resource-consuming, demanding human-guided iteration over many combinations of precursors, processing steps, and environmental conditions.

A more efficient approach to synthesizing novel inorganic materials is “synthesis by design”, in which a set of guiding principles and relationships is used

to quickly devise a synthesis method towards a target material, much like the paradigm central to synthetic organic chemistry.<sup>4,5</sup> Recent work, fueled by advances in solid-state in situ characterization techniques,<sup>6,7</sup> has taken steps in this direction by exploring reaction pathways in select case systems and identifying mechanistic relationships that explain how synthesis conditions (e.g. precursor selection and reaction environment) alter the reaction pathway, leading to selective formation of different target products. For example, Neilson and coworkers demonstrated the use of unconventional solid-state metathesis reactions to kinetically control the reaction pathway towards metastable polymorphs of  $\text{CuSe}_2$ <sup>8</sup> and  $\text{YMnO}_3$ .<sup>9,10</sup> Jiang et al. explored the use of iron silicide reactants to bypass kinetic limitations and achieve low-temperature synthesis of  $\text{Fe}_2\text{SiS}_4$ .<sup>11</sup> Miura et al. demonstrated the synthesis of  $\text{MgCr}_2\text{S}_4$  thiospinel via a metathesis route using novel precursors, which was shown to be thermodynamically favorable through computational phase diagram construction.<sup>12</sup> Bianchini et al. showed that the first phase formed in the synthesis of P2 type  $\text{Na}_{0.67}\text{MO}_2$  ( $\text{M}=\text{Co},\text{Mn}$ ) can be predicted by minimizing compositionally unconstrained reaction energies, and that formation of the initial phase can drastically alter the kinetics of the subsequent reaction and final phase selectivity.<sup>13</sup> Each of these studies elucidates an important concept: chemical reaction pathways follow a complex thermodynamic free energy surface which can be carefully manipulated and navigated via thoughtful selection of precursors, processing, and environmental conditions.

Explicit modelling of the free energy surface at an atomistic level (i.e. the potential energy surface of atomic interactions) has been successful in predicting chemical reaction pathways/dynamics in molecules.<sup>14</sup> However, in solid-state chemistry reactions, monitoring the time dependence of each atom’s spatial coordinates and interactions over the much larger scale ( $\sim 10^{23}$  atoms per mole) becomes intractable. Despite these limitations, modelling of bounded solid-state reaction mechanisms at the atomistic level has been achieved in particular with molecular dynamics (MD)<sup>15</sup> and kinetic Monte Carlo (KMC)-based<sup>16</sup> approaches. Reactive force fields, such as ReaxFF<sup>17</sup> further permit the breaking of chemical bonds and can be used to study specific chemical reaction mechanisms and kinetic parameters.<sup>18</sup> KMC-based methods also explore parts of the potential energy surface, given reaction rate constants that can be approximated with quantum mechanical calculations. However, such methods are ultimately confined to an a priori selection of the relevant domains of the high

106 dimensional solid-state potential energy surface.

107 To aid in the development of materials synthesis  
108 by design, we propose to leverage recent advances in  
109 data-driven methods which have resulted in computa-  
110 tional/experimental thermochemical databases<sup>19–22</sup>  
111 covering hundreds of thousands of materials and mil-  
112 lions of associated reaction energies.<sup>23</sup> In the remain-  
113 der of this work, we describe a framework for pre-  
114 dicting and suggesting solid-state inorganic reaction  
115 pathways, which combined with experimental efforts,  
116 will ultimately realize inorganic synthesis by design.  
117 We propose a chemical reaction network which blends  
118 typical thermodynamic phase diagrams with the con-  
119 nectivity and kinetic heuristics derived from transi-  
120 tion state theory. The network model serves as a  
121 convenient data structure for exploring the underly-  
122 ing free energy surface of thermodynamic phase space  
123 in solid-state chemistry via the power and efficiency of  
124 existing computational infrastructure for large graph  
125 networks. We outline the methodology used to create  
126 the chemical reaction network from thermochemical  
127 databases and demonstrate its predictive power as a  
128 generator of probable reaction pathways, using the re-  
129 cent metathesis synthesis of the multiferroic YMnO<sub>3</sub>  
130 as a demonstrative case study.

## 131 Results

132 The following subsections describe the i) construction  
133 of the chemical reaction network and its relationship  
134 to previous models of thermodynamic phase space,  
135 ii) prediction of reaction pathways using pathfinding  
136 methods, and iii) demonstration in predicting reac-  
137 tion pathways in the experimental metathesis syn-  
138 thesis of YMnO<sub>3</sub>.

### 139 A weighted directed graph of chemical 140 reactions

141 Following the thermodynamic equilibrium approach  
142 developed by Gibbs,<sup>24</sup> we consider solid-state chem-  
143 ical reactions as traversing a thermodynamic phase  
144 space governed by a generalized thermodynamic po-  
145 tential or free energy,  $\Phi$ , where the global minimum  
146 represents thermodynamic equilibrium for the sys-  
147 tem. Figure 1 depicts several models of chemical reac-  
148 tions in thermodynamic phase space, ordered by  
149 increasing level of abstraction. The free energy con-  
150 vex hull construction of Figure 1(a) is a purely ther-  
151 modynamic model of a chemical reaction between a  
152 pair of two reactant phases,  $R_1$  and  $R_2$ . The convex  
153 hull yields the chemical reactions which result in the  
154 largest decrease in free energy for a given mole ratio of

155 the two reactants. Figure 1(b) abstracts the thermo-  
156 dynamic model further by incorporating the concept  
157 of activation energy,  $E_a$ , as defined in transition state  
158 theory.<sup>25</sup> This enables inclusion of simple kinetic be-  
159 havior of reactions, where the height of the activation  
160 energy barrier correlates with the rate of reaction.  
161 Abstracting further, we can consider these reaction  
162 coordinate diagrams as weighted directed graphs, as  
163 in the upper portion of Figure 1(b). In these graphs,  
164 the cost (or weight) of a chemical reaction edge rep-  
165 resents an a priori unknown function of synthesis pa-  
166 rameters such as the thermodynamic driving force,  
167 activation energy, etc. Figure 1(c) shows the inter-  
168 linking of many such graph representations within a  
169 set of phases, where the nodes represent a combina-  
170 tion of phases (e.g.  $R_1 + R_2$ ) and the edges repre-  
171 sent chemical reactions with a designated cost. This  
172 weighted directed graph, or chemical reaction net-  
173 work, is a densely connected model of thermodynamic  
174 phase space where thermodynamic/kinetic features  
175 can be combined and transformed into a unique cost  
176 representation for each reaction.

177 Figure 2 illustrates the generalized graph struc-  
178 ture of a reaction network for any chemical system.  
179 Here, “chemical system” refers to the set of all  $N$   
180 phases  $p_i$  ( $i = 1, 2, \dots, N$ ), which can be produced  
181 from a designated set of chemical elements. Each re-  
182 actant/product node on the graph is created by con-  
183 sidering combinations of distinct phases up to a maxi-  
184 mum size,  $n$ . This corresponds to the set of all nodes,  
185  $P$ , given by:

$$\begin{aligned} P = & \{p_i | i \leq N\} \\ & \cup \{p_i + p_j | i, j \leq N; i \neq j\} \cup \dots \\ & \cup \{p_i + p_j + \dots + p_n | i, j, \dots, n \leq N; \\ & i \neq j \neq \dots \neq n\} \end{aligned} \quad (1)$$

186 Each of these phase combinations is added twice: once  
187 as reactants node and again as a products node. While  
188 higher values of  $n$  enable more complex reactions,  
189 in general it suffices to choose  $n = 2$ , since truly  
190 simultaneous reactions among three or more reac-  
191 tants are less likely due to kinetic and steric con-  
192 straints in a solid composite.

193 To create the dense set of directed edges at the cen-  
194 ter of the network, we algorithmically iterate through  
195 every possible chemical reaction between all pairs of  
196 reactants and product nodes. Using a reaction bal-  
197 ancing algorithm, we then solve for the stoichiometric  
198 coefficients and add a weighted, directed edge from  
199 the reactant node to corresponding product node for  
200 every chemical reaction which is successfully bal-  
201 anced. Note that many of the generated trial reac-  
202 tions cannot be stoichiometrically balanced and hence

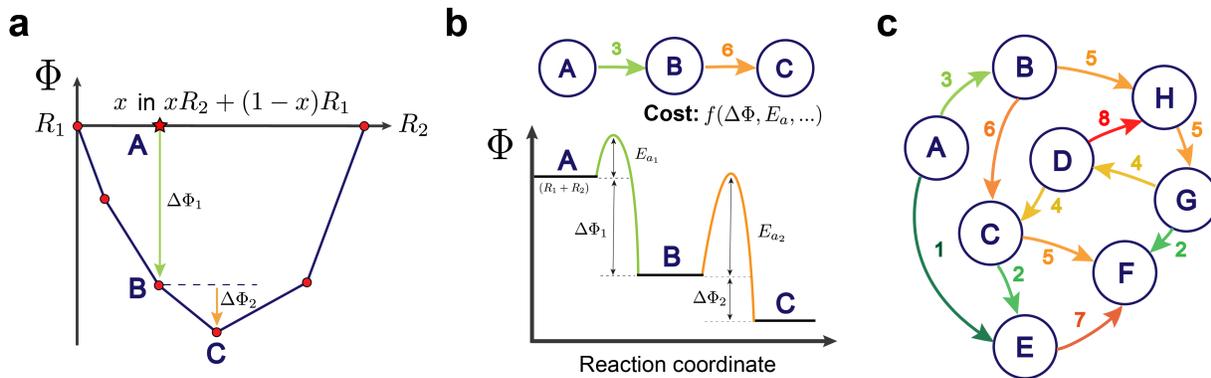


Figure 1: **Models of chemical reactions in thermodynamic phase space, ordered by increasing degree of abstraction.** The label, *A*, indicates a combination of arbitrary amounts of phases  $R_1$  and  $R_2$ . Similarly, the other labels (*B-F*) are combinations of arbitrary amounts of other phases which result from the chemical reactions. (a) The convex hull construction for consecutive reactions between two reactants,  $R_1$  and  $R_2$ . The points drawn indicate chemical reactions at different stoichiometric mixtures of the reactants, and the lines trace the convex hull indicating the thermodynamic equilibrium (minimum free energy) for all ratios of mixing,  $x$ . (b) A traditional reaction coordinate diagram used to represent both the free energies of reaction,  $\Delta\Phi$ , and activation energies,  $E_a$ . This is generalized by a weighted, directed graph connecting the three states (top). The cost/weight of the directed edges, shown as the colored number adjacent to each edge, is some function of the free energy change, activation energy, and other reaction features. (c) A chemical reaction network linking together many such possible reaction pathways that may occur within a set of phases.

203 are excluded from the graph, e.g. there are no  $x, y, z$   
 204 that satisfy  $xY_2O_3 + yMnO_2 \longrightarrow zYMnO_3$ , but the  
 205 reaction is balanceable if  $O_2$  is included as an addi-  
 206 tional product. We also exclude trivial “identity-like”  
 207 reactions between identical reactants and products,  
 208 e.g.  $Y_2O_3 \longrightarrow Y_2O_3$ . The weight of the reaction  
 209 edge is determined by a “cost function” that maps  
 210 features of the chemical reaction (e.g.  $\Delta\Phi_{rxn}$ ) to a  
 211 single cost value. To facilitate product phases being  
 212 capable of reacting again (e.g. autocatalytic reac-  
 213 tions), zero-weight edges are added which connect  
 214 each product node to all reactant nodes that contain,  
 215 as a subset, at least one of the product phases and/or  
 216 starting reactant phases (regardless of consideration  
 217 of stoichiometric coefficients). This creates a large  
 218 degree of cycles in the network, enabling the network  
 219 to capture multiple step reaction pathways.

220 Finally, two more nodes are added: one for the  
 221 synthesis precursors and one for the selected target.  
 222 These two external nodes act as single-source and  
 223 destination nodes, defining a net (overall) synthesis  
 224 reaction. The precursors node connects into the net-  
 225 work via zero-weight edges directed towards all reac-  
 226 tants nodes that contain, as a subset, at least one of  
 227 the precursor phases. To allow for an open system,  
 228 or excess with respect to precursors or specific reac-  
 229 tants, zero-weight edges are optionally added from

each from each product node to all reactant nodes  
 230 which include the excess precursor phases. This extra  
 231 layer of connectivity enables the precursors to react at  
 232 different steps along the reaction pathway. Lastly, the  
 233 network links into the target node via a set of zero-  
 234 weight edges directed from all product nodes which  
 235 contain the target phase.  
 236

The cost function used to determine the weight-  
 237 ing of edges is critical to its performance in gener-  
 238 ating probable reaction pathways. In this work, we  
 239 employ Dijkstra’s algorithm,<sup>26</sup> which uses a priority  
 240 queue structure to determine the shortest path from a  
 241 single-source node to destination node. The shortest  
 242 path is defined as the path which has the smallest sum  
 243 of all its edge weights. The simplest, and possibly  
 244 most intuitive, cost function is a direct mapping onto  
 245 the thermodynamic landscape, such as the measured  
 246 or calculated Gibbs free energy of reaction,  $\Delta G_{rxn}$ .  
 247 However, using reaction energies alone poses several  
 248 problems: 1) negative reaction energies result in infi-  
 249 nite cycles during pathfinding which preclude the use  
 250 of Dijkstra’s algorithm and many other pathfinding  
 251 methods, 2) kinetic effects and other known heuristics  
 252 about the reaction are excluded, and 3) reaction costs  
 253 are affected by stoichiometric scaling. Instead, here  
 254 we choose a single, positive cost function that maps  
 255 the Gibbs free energy of reaction, per reactant atom,  
 256

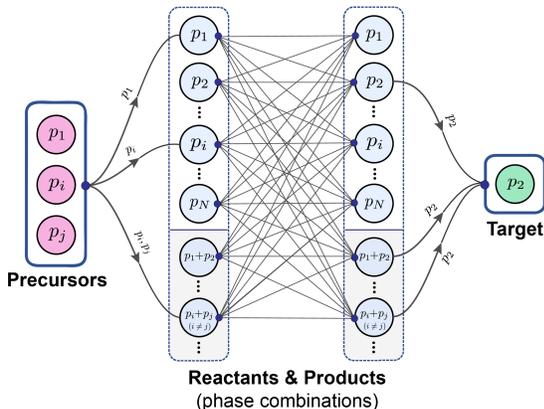


Figure 2: **The generalized graph architecture of a solid-state chemical reaction network.** The reaction network is constructed for a hypothetical chemical system containing  $N$  phases  $p_i$  ( $i = 1, 2, \dots, N$ ), with nodes made of distinct phase combinations up to maximum size  $n$ . The precursors and target nodes link into and out of the densely connected network of reactions, respectively. The edge directions between the reactant and product nodes have been omitted for clarity. These central edges include chemical reactions weighted by their cost value, as well as zero-weight edges which create loops from product nodes back to reactant nodes, such that multiple step reaction pathways can be captured.

257 to a positive value for each reaction. The choice of a  
 258 functional mapping also provides the opportunity to  
 259 create different cost functions for chemical reactions  
 260 where additional information is known, such as exper-  
 261 imental data, kinetic factors and/or other heuristics.  
 262 One example of a cost function that captures reaction  
 263 thermodynamics is the *softplus* function, which was  
 264 originally developed for use as an activation function  
 265 in neural networks.<sup>27</sup> This function maps the Gibbs  
 266 free energy of a chemical reaction,  $\Delta G_{rxn}$ , to a posi-  
 267 tive cost value,  $C$ , via:

$$C = \ln\left(1 + \frac{273 \text{ K}}{T} e^{\Delta G_{rxn}}\right) \quad (2)$$

268 where  $T$  is absolute temperature in Kelvin and  
 269  $\Delta G_{rxn}$  is the Gibbs free energy of reaction in units  
 270 of eV per reactant atom, divided by unity to be di-  
 271 mensionless. Since molar reaction energies scale with  
 272 the stoichiometric balancing of the reaction,  $\Delta G_{rxn}$   
 273 must be normalized on a per-atom basis independent  
 274 of the stoichiometric coefficients. Figure 3 demon-  
 275 strates the effect of the cost transformation on the  
 276 distribution of reaction free energies in a chemical sys-  
 277 tem. The softplus cost function transforms highly ex-

ergonic ( $\Delta G_{rxn} \ll 0$ ) reactions into low (near zero) 278  
 cost events, whereas endergonic ( $\Delta G_{rxn} > 0$ ) reac- 279  
 tions exhibit a finite cost that smoothly approaches 280  
 a linear scaling as  $\Delta G_{rxn} \rightarrow \infty$ . Note that different 281  
 environmental boundary conditions, such as open el- 282  
 ements, can be modelled by replacing  $\Delta G_{rxn}$  with 283  
 $\Delta \Phi_{rxn}$ , where  $\Phi$  represents a customized thermo- 284  
 dynamic potential. Other monotonically increasing 285  
 functions were tested (Supplementary Figure 1) how- 286  
 ever the smooth and differentiable softplus function 287  
 was preferred. We hypothesize that there indeed ex- 288  
 ists a *true* cost function for each chemical reaction, 289  
 which depends on local and global environmental con- 290  
 ditions and transforms the thermodynamic/kinetic 291  
 factors to an appropriate cost that allows the net- 292  
 work to accurately capture the real behavior of the 293  
 system. 294

## Predicting reaction pathways via graph pathfinding

Ideally, predicting a reaction pathway using the net- 297  
 work would be equivalent to leveraging existing, pow- 298  
 erful algorithms for solving the single-source shortest 299  
 path problem from graph theory. However, chemical 300  
 reactions rarely trace a set of linear steps, even in 301  
 simple syntheses; instead the precursor phases often 302  
 undergo reactions concurrently in parallel, or react 303  
 again via autocatalytic reactions. Within the net- 304  
 work, parallel reactions can be modeled as simultane- 305  
 ous travel along multiple reaction edges. In the phys- 306  
 ical world these reactions must obey mass conserva- 307  
 tion, and phases produced in one reaction may react 308  
 with phases in another. These so-called ‘‘crossover’’ 309  
 reactions, along with the possibility of parallel reac- 310  
 tions, prohibit the *direct* application of shortest path 311  
 algorithms. 312

To solve this issue of parallel paths and crossover 313  
 reactions, we identify not only the single shortest 314  
 path from precursors to target phase, but all  $k$ - 315  
 shortest paths in the network, and then attempt to 316  
 find linear combinations of and interactions between 317  
 paths which satisfy the stoichiometric constraints of 318  
 a net/overall reaction which is known a priori. For 319  
 syntheses that involve multiple targets or byproducts 320  
 (e.g.  $\text{CO}_2$ ), pathfinding is performed towards each 321  
 target phase separately, and then all paths are accu- 322  
 mulated together; this ensures that generated paths 323  
 access all targets. With the set of likely paths, we 324  
 identify intermediate phases and supplement the set 325  
 with possible crossover reactions which result in fur- 326  
 ther production of the target phase(s). To computa- 327  
 tionally generate all  $k$ -shortest paths, we utilize Yen’s 328  
 algorithm<sup>28</sup> which iteratively produces the next  $k - 1$  329

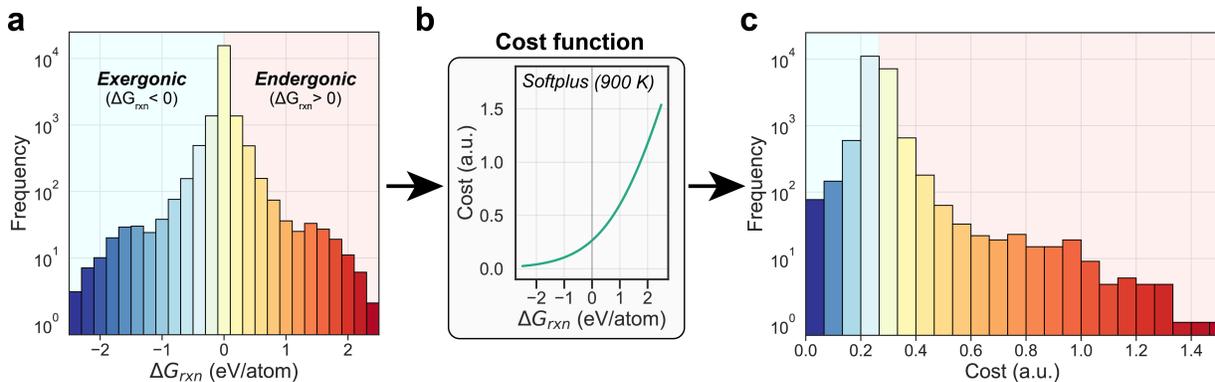
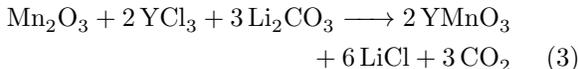


Figure 3: **The effect of the cost function transformation on reaction energies.** (a) Distribution of reaction Gibbs free energies (normalized per reactant atom) in the C-Cl-Li-Mn-O-Y chemical system. (b) Transformation of reaction energies via the softplus cost function described in equation (2). (c) Final distribution of reaction costs (in arbitrary units) after transformation. A reaction free energy of zero corresponds to a cost of  $\sim 0.265$ .

330 shortest paths via deviations from the first shortest  
 331 path, as calculated with Dijkstra’s algorithm. Given  
 332 that the cost function transformation provides only  
 333 an approximation to the thermodynamics and kinet-  
 334 ics of evolving reactions, the identification of *many*  
 335 low-cost paths is also practical in the creation of  
 336 a candidate set of possible paths. While the top  
 337 shortest paths may not be experimentally feasible,  
 338 the use of Yen’s algorithm to generate many such  
 339 low-cost paths narrows down the range of probable  
 340 reactions and facilitates tractable exploration of a  
 341 combinatorially-dense phase space.

### 342 **Demonstration of reaction network** 343 **model for synthesis of $\text{YMnO}_3$**

344 To demonstrate the capabilities of the reaction net-  
 345 work for predicting possible reaction pathways in a  
 346 synthesis procedure, we consider the synthesis of yt-  
 347 trium manganese oxide,  $\text{YMnO}_3$ , through the solid-  
 348 state assisted metathesis reaction reported by Todd  
 349 & Neilson.<sup>9</sup> The overall reaction,



350 was found to exhibit several steps with distinct inter-  
 351 mediate compounds, as determined through in situ  
 352 temperature-dependent x-ray diffraction performed  
 353 at a synchrotron beamline.<sup>10</sup> Figure 4(a) shows the  
 354 chemical reaction network generated for the C-Cl-  
 355 Li-Mn-O-Y chemical system. The C-Cl-Li-Mn-O-Y  
 356 phase diagram constructed from Materials Project  
 357 contains 768 entries; of these, 47 are predicted to

358 be stable at  $T = 0$  K. The machine-learned Gibbs  
 359 descriptor reported by Bartel et al.<sup>29</sup> and NIST-  
 360 JANAF tables<sup>21</sup> were further used to transform the  
 361 DFT-derived formation enthalpies into Gibbs free  
 362 energies of formation at 900 K – the approximate tem-  
 363 perature at which  $\text{YMnO}_3$  was observed to form. The  
 364 transformation to Gibbs free energy at 900 K reduces  
 365 the number of stable phases in the phase diagram to  
 366 39. We include all of these stable entries as well as  
 367 metastable entries (not including polymorphs) up to  
 368 a filter of +20 meV/atom above the hull, resulting  
 369 in a total of 56 phases considered (Supplementary  
 370 Table 1). The final reaction network contains 3,194  
 371 nodes and 33,428 edges, where 20,116 of these edges  
 372 represent chemical reactions with a maximum phase  
 373 combination size of  $n = 2$ . Costs for all reaction edges  
 374 were mapped using the softplus function described in  
 375 equation (2) with  $T = 900$  K and plotted in Figure  
 376 3b.

377 Reaction pathway prediction was performed given  
 378 the initial reactants and final products of the assisted  
 379 metathesis reaction in equation (3). The 60 short-  
 380 est paths (20 to each product of  $\text{YMnO}_3$ ,  $\text{LiCl}$ , and  
 381  $\text{CO}_2$ ) were identified via Yen’s algorithm, resulting  
 382 in a set of 47 unique reactions. All 60 shortest paths  
 383 can be seen in Supplementary Tables 2-4. Combined  
 384 pathways were generated via this set of reactions by  
 385 solving for linear combinations of reactions, up to a  
 386 maximum size of four reactions, that obey stoichi-  
 387 ometric constraints. Of the total 195,708 pathways  
 388 considered, only *two* combined reaction pathways fit  
 389 the stoichiometric constraints of the net reaction.

390 The shorter of the two combined pathways (in both  
 391 total and average cost per reaction) involves the for-

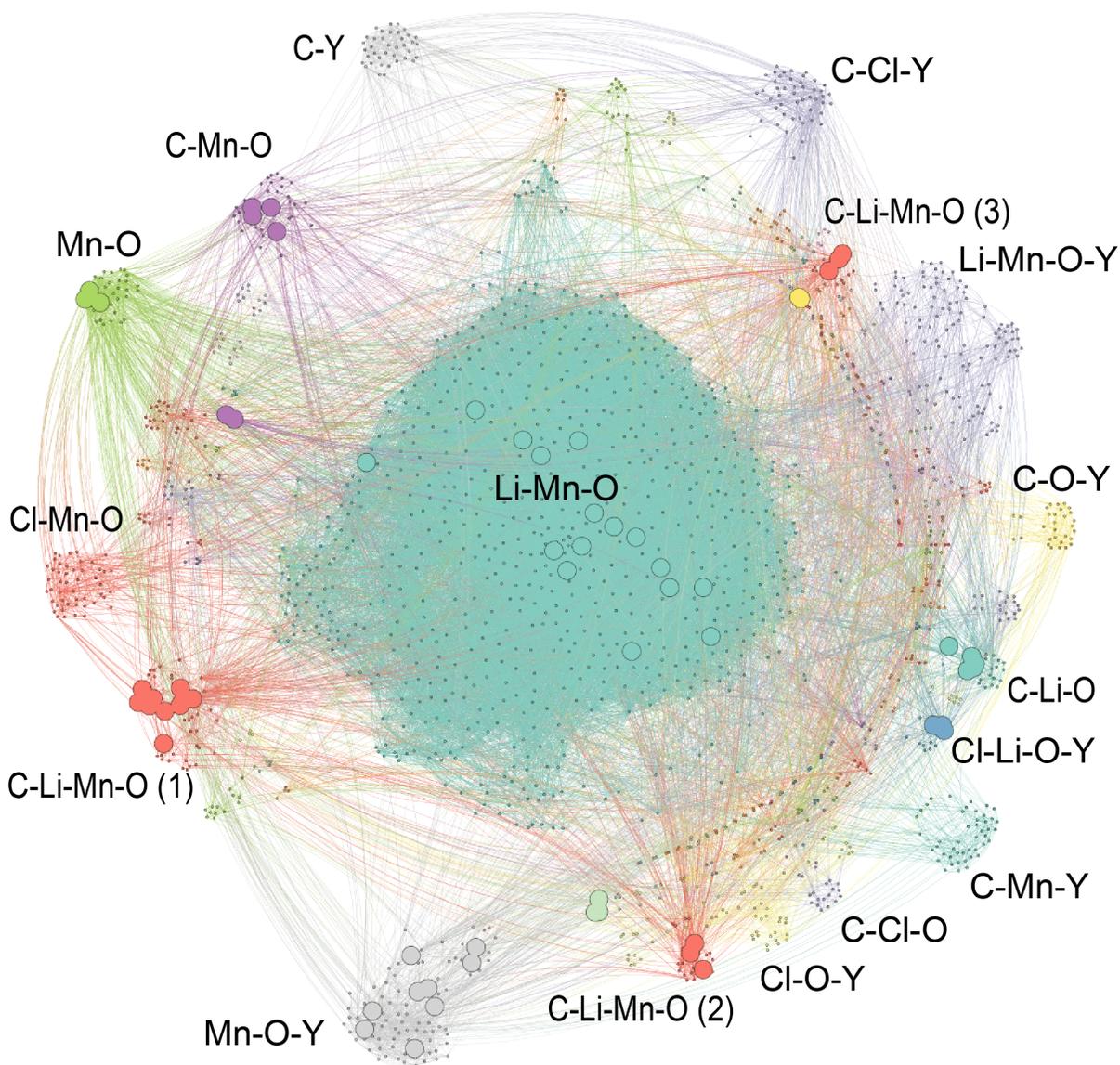
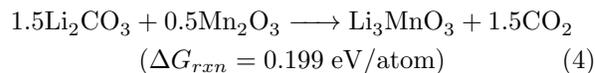
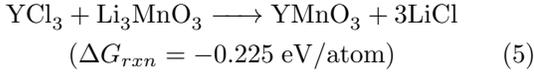


Figure 4: **The reaction network for the C-Cl-Li-Mn-O-Y chemical system.** The network contains 56 phases: 39 stable and 17 within +20 meV/atom above the hull. Chemical subsystems are labeled by color. The larger nodes indicate reactant nodes which are traversed on the 20 shortest pathways from precursors to targets in the  $\text{YMnO}_3$  assisted metathesis reaction given by equation (3).

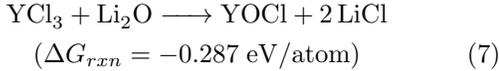
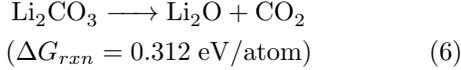
392 mation and reaction of  $\text{Li}_3\text{MnO}_3$  which, to our knowl-  
 393 edge, is an oxygen-deficit distorted rocksalt that is  
 394 yet to be experimentally verified; in fact, the Materi-  
 395 als Project contains tens of thousands of hypothetical  
 396 compounds, which may or may not be readily synthe-  
 397 sizable. According to available computed data, this  
 398 compound is metastable (+12 meV/atom above the  
 399 hull) at zero temperature, but is predicted to exhibit  
 400 thermodynamic stability at 900 K according to the

Gibbs free energy model. The proposed pathway is  
 reported below, along with the computed free energy  
 of reaction per reactant atom:

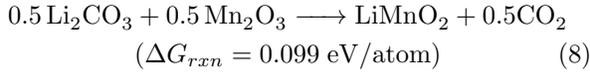




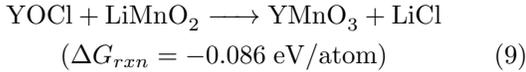
404 The second ranked pathway by cost, however,  
 405 closely matches the experimentally reported assisted  
 406 metathesis pathway. The  $\text{Li}_2\text{CO}_3$  first decomposes to  
 407  $\text{Li}_2\text{O}$  and then reacts to make  $\text{YOCl}$ :



408 while simultaneously,  $\text{LiMnO}_2$  is formed:



409 And these two intermediates finally react together in  
 410 a metathesis reaction to produce  $\text{YMnO}_3$ :



411 This pathway is nearly identical to the experimentally  
 412 reported pathway; with the added step that  $\text{Li}_2\text{CO}_3$   
 413 first decomposes to  $\text{Li}_2\text{O}$  and  $\text{CO}_2$  before reacting.

## 414 Discussion

415 The primary challenge in creating a reaction network  
 416 model is the high degree of complexity inherent to  
 417 thermodynamic phase space, which quickly leads to  
 418 a combinatorial explosion during both the creation of  
 419 the network and subsequent pathfinding steps. As an  
 420 example, consider a reaction network with  $N$  phases  
 421 and a maximum phase combination size,  $n$ . If during  
 422 the graph generation every possible chemical reaction  
 423 between any two nodes is considered, the number of  
 424 reactions,  $R$ , would be:

$$R = \left[ \sum_{i=1}^n \binom{N}{i} \right]^2$$

$$= \left[ \binom{N}{1} + \binom{N}{2} + \dots + \binom{N}{n} \right]^2 \quad (10)$$

425 For example, for the C-Cl-Li-Mn-O-Y reaction net-  
 426 work with  $N = 56$  distinct phases, the maximum  
 427 number of reactions described by equation (10), be-  
 428 fore stoichiometric balancing, is  $R \approx 3.14 \times 10^3$  ( $n =$   
 429  $1$ ),  $2.55 \times 10^6$  ( $n = 2$ ), and  $8.59 \times 10^8$  ( $n = 3$ ).

In this work, we present a computationally 430  
 tractable approach, which effectively introduces a se- 431  
 ries of filters which reduce the complexity and degrees 432  
 of freedom of the thermodynamic phase space. These 433  
 filters include: 1) restricting the number of phases 434  
 considered via thermodynamic stability arguments, 435  
 i.e. energy above the hull, 2) limiting the maximum 436  
 number of phases present on each side of the reaction 437  
 to a small number, e.g.  $n = 2, 3$ ) using a cost func- 438  
 tion to prioritize reactions which are likely to occur, 439  
 and 4) enforcing mass conservation via stoichiomet- 440  
 ric constraints. The first two filters work together 441  
 during graph generation to limit the combinatorial 442  
 size/complexity of the network. This number can be 443  
 reduced by decreasing either  $N$ ,  $n$ , or both. It is also 444  
 worth noting that the number of considered reactions 445  
 can be reduced by considering the connectivity of the 446  
 compositional phase diagram of the system; for exam- 447  
 ple, chemical reactions may be limited to only those 448  
 reactions which occur along facets of the phase dia- 449  
 gram. Since it is typically optimal to consider as 450  
 many phases as possible, it is more favorable to re- 451  
 duce  $n$ , rather than  $N$ . Therefore the choice of  $n = 2$  452  
 minimizes the combinatorics of the network without 453  
 inherently sacrificing the complexity of reactions that 454  
 can occur. Indeed, reaction pathways suggested by 455  
 the network adhering to the  $n = 2$  limit must consist 456  
 of pseudo-elementary steps which more closely follow 457  
 the free energy surface. This behavior is nicely il- 458  
 lustrated in the second reaction pathway suggested 459  
 by the C-Cl-Li-Mn-O-Y network (equations (6)-(9)), 460  
 which closely resembles the experimentally observed 461  
 pathway. In the experimental synthesis of  $\text{YMnO}_3$ , 462  
 however, the authors report a reaction step with three 463  
 products ( $\text{YCl}_3 + \text{Li}_2\text{CO}_3 \longrightarrow \text{YOCl} + 2\text{LiCl} + \text{CO}_2$ ). 464  
 Due to the selection of  $n = 2$ , this reaction was not 465  
 explicitly present in the network. However, the *net* 466  
 effect of the reaction is indeed incorporated in the 467  
 second suggested pathway, where this reaction is es- 468  
 sentially divided into two smaller steps featuring the 469  
 decomposition of  $\text{Li}_2\text{CO}_3$  into  $\text{Li}_2\text{O}/\text{CO}_2$  and the re- 470  
 action of  $\text{Li}_2\text{O}$  with  $\text{YCl}_3$  directly. It is reasonable 471  
 to postulate that this thermal decomposition step 472  
 might actually be occurring; the decomposition of 473  
 lithium carbonate has been well-studied and observed 474  
 to occur spontaneously at/above temperatures near 475  
 the maximum temperature of the assisted metathesis 476  
 synthesis route ( $T \sim 900 \text{ K}$ ).<sup>30</sup> 477

The cost function approach provides another sim- 478  
 plification of the complex thermodynamic phase 479  
 space. Here we have shown that a smooth, mono- 480  
 tonic transformation of the Gibbs free energy of re- 481  
 action is sufficient itself in capturing realistic behav- 482  
 ior in solid-state materials synthesis, as exemplified 483

484 by the network model of  $\text{YMnO}_3$  synthesis. The cost  
485 transformation to positive values allows for the uti-  
486 lization of existing shortest path algorithms such as  
487 Dijkstra’s algorithm, but also naturally incorporates  
488 several realistic/expected features of traversing the  
489 thermodynamic phase space. First, the cost function  
490 can in principle integrate multiple thermodynamic  
491 and kinetic heuristics. While we did not include any  
492 kinetic features in this work, we anticipate the fu-  
493 ture addition of kinetic features, such as the struc-  
494 tural (dis)similarity between phases,<sup>31,32</sup> the average  
495 number of bonds broken/created, change in the infor-  
496 mation entropy description of atomic configurations,  
497 change in atomic density, etc. The relative weights  
498 of each of these features within the cost function,  
499 of course, must be carefully examined and validated.  
500 For example, the fact that the reaction network en-  
501 codes modular pseudo-elementary steps is highly con-  
502 ductive to the inclusion of modeled or experimentally  
503 obtained kinetic barriers, where the cost associated  
504 with a particular step may be high enough to remove  
505 the entire pathway from consideration.

506 Second, shortest path algorithms are naturally  
507 cost-additive, which biases the pathway generation  
508 towards simpler reaction pathways with fewer steps.  
509 This also introduces a trade-off between the num-  
510 ber of steps in the pathway and the cost per step,  
511 i.e. paths with several low-cost steps may exhibit the  
512 same total cost as paths with only one medium/high  
513 cost step. The slope of the cost function determines  
514 the nature of this trade-off; however, the *softplus*  
515 function generally is found to favor shortest paths  
516 with fewer steps (Supplementary Tables 2-4). One  
517 unique result of the additive nature of the shortest  
518 path approach is that the shortest paths towards cer-  
519 tain products often involve unanticipated, endergonic  
520 ( $\Delta G_{rxn} > 0$ ) reaction steps. This is a major ad-  
521 vantage as compared to other thermodynamic mod-  
522 els where reaction pathways are often restricted to a  
523 cascade of monotonically decreasing free energy steps.  
524 Allowing endergonic steps introduces flexibility to the  
525 uncertainty (and lack) of thermochemistry data, lo-  
526 cal vs. global synthesis conditions, etc., allowing the  
527 full compositional space to be traversed when con-  
528 sidering chemical routes towards targets. This is ex-  
529 emplified in particular by the lithium carbonate de-  
530 composition reaction described previously, which is  
531 endergonic at low temperatures but is shown to oc-  
532 cur spontaneously at elevated temperatures.

533 Finally, the network can also be used to identify  
534 shortest paths to or from *any* nodes in the net-  
535 work. This mechanism may lead to alternative in-  
536 sights about the network, such as lists of the shortest  
537 pathways to any target. We anticipate this setup of

538 the network to be useful for speculating possible likely  
539 products, when no net reaction information is known  
540 a priori. Similarly, the network can be used in “re-  
541 verse” to identify promising precursors which yield  
542 efficient chemical routes towards desired targets.

## 543 Conclusions

544 A chemical reaction network model was designed, im-  
545 plemented with thermodynamic data from the Mate-  
546 rials Project, and demonstrated as a predictor of re-  
547 action pathways in solid-state chemistry. The frame-  
548 work effectively reduces the large, complex thermo-  
549 chemical landscape to a computationally tractable  
550 structure through i) creation of a weighted directed  
551 graph representation of the available thermodynamic  
552 phase space, ii) mapping of rigorous thermodynamic  
553 data and possible heuristics into a versatile cost func-  
554 tion and iii) application of existing pathfinding algo-  
555 rithms to identify probable reaction routes. While the  
556 framework explores reaction trajectories in the most  
557 general way possible, allowing for parallel combined  
558 pathways, the combinatorial complexity is reduced  
559 by chemically motivated filters such as: 1) restricting  
560 the number of phases considered via thermodynamic  
561 stability arguments, 2) limiting the maximum num-  
562 ber of simultaneously reacting phases, and 3) enforc-  
563 ing mass conservation via stoichiometric constraints.  
564 As a demonstration, the framework was shown to  
565 successfully identify a complex reaction pathway in  
566 a recently elucidated in-situ characterized solid-state  
567 synthesis of  $\text{YMnO}_3$ . We envision the methodology to  
568 be used to suggest possible synthesis routes and pre-  
569 cursors that allow for efficient thermodynamic condi-  
570 tions towards desirable target phases, as well as iden-  
571 tification of byproducts and possible thermodynamic  
572 sinks along synthesis routes. Future work will ben-  
573 efit tremendously by combining the framework ‘live’  
574 with automated data collection, *in situ* phase iden-  
575 tification, rapid analysis techniques, and automated  
576 feedback loops, moving towards active control of in-  
577 organic solid-state synthesis.

## 578 Methods

### 579 Thermochemical data

580 While the chemical reaction network can be created  
581 from any thermochemical data – computed, exper-  
582 imental, or a combination of both – in this work,  
583 we employ the Materials Project (MP), which con-  
584 tains well-benchmarked *ab initio* calculated forma-  
585 tion enthalpies for over one hundred thousand dif-

586 ferent materials as calculated with density functional  
587 theory (DFT).<sup>19,33</sup> To capture the temperature de-  
588 pendence of thermodynamic phase space, we employ  
589 the machine-learned Gibbs free energy descriptor re-  
590 ported by Bartel et al.,<sup>29</sup> which estimates the finite  
591 temperature contribution to the Gibbs free energy  
592 of formation of solids,  $\Delta G_f(T)$ . This contribution  
593 incorporates both temperature-dependent enthalpic  
594 and entropic effects, although the entropic contribu-  
595 tion ( $TS$ ) typically dominates. The elemental Gibbs  
596 free energies used for these formation energy calcula-  
597 tions are acquired from FactSage.<sup>20</sup> The Gibbs free  
598 energies of formation for non-elemental gases (e.g.  
599  $\text{CO}_2$ ) are acquired from NIST-JANAF experimen-  
600 tal thermochemical tables where possible.<sup>21</sup> Ther-  
601 modynamic stability (energy above the hull) is calcu-  
602 lated via phase diagram construction in the *pymatgen*  
603 package.<sup>34</sup>

## 604 Graph creation, traversal, and visual- 605 ization

606 All networks are implemented and analyzed using the  
607 *graph-tool* package.<sup>35</sup> Chemical reactions are bal-  
608 anced and combined in a high-throughput manner  
609 via the reaction balancing algorithm implemented in  
610 *pymatgen*. Graphs are visualized using Graphistry  
611 Hub.<sup>36</sup>

## 612 Combining chemical reactions via mass 613 conservation

614 When a net reaction is known a priori, reaction steps  
615 identified during the pathfinding can be linearly com-  
616 bined to satisfy the stoichiometric mass constraints  
617 of the overall reaction. These constraints correspond  
618 to numerically solving the linear system of equations  
619 given by:

$$Am = c \quad (11)$$

620 where  $\mathbf{m}$  is a vector containing the “multiplicity” of  
621 each reaction (i.e. the factor by which the entire  
622 reaction is multiplied),  $A$  is the matrix containing  
623 the stoichiometric coefficients of all phases present  
624 in all reactions where reactants/products have nega-  
625 tive/positive coefficients respectively, and  $\mathbf{c}$  is a vec-  
626 tor containing the stoichiometric coefficients of the  
627 net synthesis reaction. We solve this system of equa-  
628 tions for the multiplicity vector,  $\mathbf{m}$ , via application  
629 of the Moore-Penrose matrix pseudoinverse as imple-  
630 mented within the *SciPy* package.<sup>37</sup>

## References

- [1] James L Marshall and Virginia R Marshall. Re-  
discovery of the Elements: Cronstedt and Nickel.  
*The Hexagon of Alpha Chi Sigma*, 105(2):24–29,  
2014.
- [2] Francis J. Disalvo. Solid-state chemistry:  
A rediscovered chemical frontier. *Science*,  
247(4943):649–655, February 1990.
- [3] Holger Kohlmann. Looking into the Black Box of  
Solid-State Synthesis. *European Journal of In-  
organic Chemistry*, 2019(39-40):4174–4180, Oc-  
tober 2019.
- [4] L. Soderholm and J. F. Mitchell. Perspective:  
Toward “synthesis by design”: Exploring atomic  
correlations during inorganic materials synthe-  
sis. *APL Materials*, 4(5):053212, May 2016.
- [5] Andreas Stein, Steven W. Keller, and Thomas E.  
Mallouk. Turning down the heat: Design and  
mechanism in solid-state synthesis. *Science*,  
259(5101):1558–1564, 1993.
- [6] Daniel P. Shoemaker, Yung Jin Hu, Duck Young  
Chung, Gregory J. Halder, Peter J. Chupas,  
L. Soderholm, J. F. Mitchell, and Mercouri G.  
Kanatidis. In situ studies of a platform for  
metastable inorganic crystal growth and ma-  
terials discovery. *Proceedings of the National  
Academy of Sciences of the United States of  
America*, 111(30):10922–10927, 2014.
- [7] Daniel O’Nolan, Guanglong Huang, Gabrielle E.  
Kamm, Antonin Grenier, Chia-Hao Liu, Paul K.  
Todd, Allison Wustrow, Gia Thinh Tran, David  
Montiel, James R Neilson, Simon J. L. Billinge,  
Peter J. Chupas, Katsuyo S. Thornton, and  
Karena W. Chapman. A thermal-gradient ap-  
proach to variable-temperature measurements  
resolved in space. *Journal of Applied Crystal-  
lography*, 53(3), June 2020.
- [8] Andrew J Martinolich and James R Neilson.  
Toward Reaction-by-Design: Achieving Kinetic  
Control of Solid State Chemistry with Metathe-  
sis. *Chemistry of Materials*, 29(2):479–489, 2017.
- [9] Paul K. Todd and James R. Neilson. Selective  
formation of yttrium manganese oxides through  
kinetically competent assisted metathesis reac-  
tions. *Journal of the American Chemical Soci-  
ety*, 141(3):1191–1195, 2019.

- 677 [10] Paul K Todd, Antoinette M M Smith, and  
678 James R Neilson. Yttrium Manganese Oxide  
679 Phase Stability and Selectivity Using Lithium  
680 Carbonate Assisted Metathesis Reactions. *In-*  
681 *organic Chemistry*, 58(22):15166–15174, Novem-  
682 ber 2019.
- 683 [11] Zhelong Jiang, Arun Ramanathan, and Daniel P  
684 Shoemaker. In situ identification of kinetic fac-  
685 tors that expedite inorganic crystal formation  
686 and discovery. *J. Mater. Chem. C*, 5:5709, 2017.
- 687 [12] Akira Miura, Hiroaki Ito, Christopher J. Bartel,  
688 Wenhao Sun, Nataly Carolina Rosero-Navarro,  
689 Kiyoharu Tadanaga, Hiroko Nakata, Kazuhiko  
690 Maeda, and Gerbrand Ceder. Selective metathe-  
691 sis synthesis of  $\text{MgCr}_2\text{S}_4$  by control of thermo-  
692 dynamic driving forces. *Mater. Horiz.*, 2020.
- 693 [13] Matteo Bianchini, Jingyang Wang, Raphaële J  
694 Clément, Bin Ouyang, Penghao Xiao, Daniil  
695 Kitchaev, Tan Shi, Yaqian Zhang, Yan Wang,  
696 Haegyeom Kim, Mingjian Zhang, Jianming Bai,  
697 Feng Wang, Wenhao Sun, and Gerbrand Ceder.  
698 The interplay between thermodynamics and kin-  
699 etics in the solid-state synthesis of layered ox-  
700 ides. *Nature Materials*, 2020.
- 701 [14] Jeffrey I Steinfeld, Joseph Salvatore. Francisco,  
702 and William L Hase. *Chemical kinetics and dy-*  
703 *namics*. Prentice Hall, Upper Saddle River, N.J.,  
704 1999.
- 705 [15] Michael P. Allen. *Introduction to Molecular Dy-*  
706 *namics Simulation*. Computational soft matter:  
707 from synthetic polymers to proteins (NIC Se-  
708 ries), Julich, 2004.
- 709 [16] Arthur F Voter. Introduction to the Kinetic  
710 Monte Carlo Method. In Kurt E Sickafus, Eu-  
711 gene A Kotomin, and Blas P Uberuaga, editors,  
712 *Radiation Effects in Solids*, pages 1–23, Dor-  
713 drecht, 2007. Springer Netherlands.
- 714 [17] Adri C T van Duin, Siddharth Dasgupta, Fran-  
715 cois Lorant, and William A Goddard. ReaxFF:  
716 A Reactive Force Field for Hydrocarbons. *The*  
717 *Journal of Physical Chemistry A*, 105(41):9396–  
718 9409, October 2001.
- 719 [18] Daniil V. Ilyin, William A. Goddard,  
720 Julius J. Oppenheim, and Tao Cheng. First-  
721 principles-based reaction kinetics from reactive  
722 molecular dynamics simulations: Application  
723 to hydrogen peroxide decomposition. *Pro-*  
724 *ceedings of the National Academy of Sciences*,  
725 116(37):18202–18208, 2019.
- [19] Anubhav Jain, Shyue Ping Ong, Geoffroy Hau- 726  
tier, Wei Chen, William Davidson Richards, 727  
Stephen Dacek, Shreyas Cholia, Dan Gunter, 728  
David Skinner, Gerbrand Ceder, and Kristin A. 729  
Persson. Commentary: The materials project: A 730  
materials genome approach to accelerating ma- 731  
terials innovation. 1(1), 2013. 732
- [20] C.W. Bale, E. Bélisle, P. Chartrand, S.A. 733  
Decterov, G. Eriksson, A.E. Gheribi, K. Hack, I- 734  
H. Jung, Y.-B. Kang, J. Melançon, A.D. Pelton, 735  
S. Petersen, C. Robelin, J. Sangster, P. Spencer, 736  
and M-A. Van Ende. Factsage thermochemi- 737  
cal software and databases, 2010–2016. *Calphad*, 738  
54:35–53, 2016. 739
- [21] Jr Malcolm W. Chase. *NIST-JANAF thermo-* 740  
*chemical tables*. Fourth edition. Washington, 741  
DC : American Chemical Society ; New York : 742  
American Institute of Physics for the National 743  
Institute of Standards and Technology, 1998., 744  
1998. 745
- [22] Christopher W. Bale and Gunnar Eriksson. Met- 746  
allurgical thermochemical databases—a review. 747  
*Canadian Metallurgical Quarterly*, 29(2):105– 748  
132, 1990. 749
- [23] Anubhav Jain, Yongwoo Shin, and Kristin A 750  
Persson. Computational predictions of energy 751  
materials using density functional theory. *Nat-* 752  
*ure Reviews Materials*, 1(1):15004, 2016. 753
- [24] J. W. Gibbs. On the equilibrium of heteroge- 754  
neous substances. *American Journal of Science*, 755  
16(96):441–458, December 1878. 756
- [25] Henry Eyring. The activated complex in chemi- 757  
cal reactions. *The Journal of Chemical Physics*, 758  
3(2):107–115, 1935. 759
- [26] E W Dijkstra. A note on two problems in con- 760  
nexion with graphs. *Numerische Mathematik*, 761  
1(1):269–271, 1959. 762
- [27] Charles Dugas, Yoshua Bengio, François Bélisle, 763  
Claude Nadeau, and René Garcia. Incorporat- 764  
ing second-order functional knowledge for better 765  
option pricing. In *Proceedings of the 13th Inter-* 766  
*national Conference on Neural Information Pro-* 767  
*cessing Systems*, NIPS’00, page 451–457, Cam- 768  
bridge, MA, USA, 2000. MIT Press. 769
- [28] Jin Y. Yen. Finding the k shortest loopless paths 770  
in a network. *Management Science*, 17(11):712– 771  
716, 1971. 772

- 773 [29] Christopher J. Bartel, Samantha L. Millican, 822  
774 Ann M. Deml, John R. Rumpitz, William Tu- 823  
775 mas, Alan W. Weimer, Stephan Lany, Vladan 824  
776 Stevanović, Charles B. Musgrave, and Aaron M. 825  
777 Holder. Physical descriptor for the Gibbs energy 826  
778 of inorganic crystalline solids and temperature- 827  
779 dependent materials chemistry. *Nature Commu-* 828  
780 *nications*, 9(1):4168, December 2018.
- 781 [30] A N Timoshevskii, M G Ktalkherman, V A 829  
782 Emel’kin, B A Pozdnyakov, and A P Zamyatin. 830  
783 High-temperature decomposition of lithium car- 831  
784 bonate at atmospheric pressure. *High Tempera-* 832  
785 *ture*, 46(3):414–421, 2008.
- 786 [31] Nils E. R. Zimmermann and Anubhav Jain. Lo- 833  
787 cal structure order parameters and site finger- 834  
788 prints for quantification of coordination environ-  
789 ment and crystal structure similarity. *RSC Adv.*,  
790 10:6063–6081, 2020.
- 791 [32] Sandip De, Albert P. Bartók, Gábor Csányi, 835  
792 and Michele Ceriotti. Comparing molecules and 836  
793 solids across structural and alchemical space. 837  
794 *Phys. Chem. Chem. Phys.*, 18:13754–13769, 838  
795 2016.
- 796 [33] Anubhav Jain, Geoffroy Hautier, Shyue Ping 839  
797 Ong, Charles J Moore, Christopher C Fischer, 840  
798 Kristin A Persson, and Gerbrand Ceder. Forma- 841  
799 tion enthalpies by mixing GGA and GGA + U 842  
800 calculations. *Physical Review B*, 84:45115, 2011.
- 801 [34] Shyue Ping Ong, William Davidson Richards, 843  
802 Anubhav Jain, Geoffroy Hautier, Michael 844  
803 Kocher, Shreyas Cholia, Dan Gunter, Vincent L. 845  
804 Chevrier, Kristin A. Persson, and Gerbrand 846  
805 Ceder. Python Materials Genomics (pymatgen): 847  
806 A robust, open-source python library for mate- 848  
807 rials analysis. *Computational Materials Science*, 849  
808 68:314–319, 2013.
- 809 [35] Tiago P. Peixoto. The graph-tool python library. 850  
810 May 2017.
- 811 [36] Graphistry. *Graphistry Hub*, 2020 (accessed June 851  
812 18, 2020).
- 813 [37] Pauli Virtanen, Ralf Gommers, Travis E. 852  
814 Oliphant, Matt Haberland, Tyler Reddy, David 853  
815 Cournapeau, Evgeni Burovski, Pearu Peterson, 854  
816 Warren Weckesser, Jonathan Bright, Stéfan J. 855  
817 van der Walt, Matthew Brett, Joshua Wilson,  
818 K. Jarrod Millman, Nikolay Mayorov, Andrew  
819 R. J. Nelson, Eric Jones, Robert Kern, Eric Lar-  
820 son, CJ Carey, İlhan Polat, Yu Feng, Eric W.  
821 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 Com-  
puting in Python. *Nature Methods*, 17:261–272,  
2020.

## Author Contributions

M.J.M. and S.S.D. conceived the idea of the presented work. M.J.M. designed and developed the *reaction-network* code with feedback from S.S.D. and K.A.P. M.J.M. wrote the manuscript with guidance of S.S.D. and K.A.P.

## Acknowledgements

This work was supported as part of GENESIS: A Next Generation Synthesis Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award Number DE-SC0019212. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

The authors would like to thank J. Neilson and P. Todd for their helpful discussion regarding the reaction network model, as well as E. Persson for math skills, and L. Meyerovich for assistance with graph visualization.

## Code availability

The *reaction-network* package was created in Python to implement the reaction network model described in this work. The code is free and available for use by the general community, via GitHub at <https://github.com/GENESIS-EFRC/reaction-network>.