Cutting materials in half: a graph theory approach for generating crystal surfaces and its prediction of 2D zeolites

Matthew Witman,[†] Sanliang Ling,[‡] Peter Boyd,[¶] Senja Barthel,[¶] Maciej Haranczyk,[§,‖] Ben Slater,[‡] and Berend Smit[*,†,¶]

[†]*Department of Chemical and Biomolecular Engineering, University of California, Berkeley 94720, USA*

[‡]*Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK*

[¶]*Laboratory of Molecular Simulation, Institut des Sciences et Ingénierie Chimiques, Valais, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Switzerland*

[§]*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

[‖]*IMDEA Materials Institute, C/Eric Kandel 2, 28906 - Getafe, Madrid, Spain*

E-mail: berend.smit@epfl.ch

Python implementation of the Balcioglu and Wood algorithm

A Python implementation of the Balcioglu and Wood algorithm[1] is provided in the "Nearmincut" github repository (`https://github.com/mwitman1/nearmincut.git`). This routine has been incorporated into the private version of the LammpsInterface[2] (`https://github.com/peteboyd/lammps_interface.git`) program which handles all other logistics of structure processing using both Pymatgen[3] and its own internal functionality. This new version may be publicly released in the near future, but for immediate inquiries regarding the use of this new functionality please contact the developers of LammpsInterface.

Formal definitions of min cuts are provided in the main manuscript, but a brief overview is provided here to understand the functionality of nearmincut. Nearmincut accepts a Networkx graph and a source and target node to compute the min s-t edge cut. Denoted $w(C_0)$, this is the sum of the weights of all edges in the cut, where a cut is a set of edges that, if removed, interrupts all paths form s to t. If C_0 is the min cut, then $w(C_0) \leq w(C \in \mathbf{C})$, where $\mathbf{C}$ is all possible s-t cuts. Then the algorithm of Balcioglu and Wood recursively searches for all cuts in the set $\mathbf{C}_\epsilon$ whose weights have $w(C) \leq (1+\epsilon)w(C_0)$, where $\epsilon > 0$ is a user specified parameter that determines how "near" a cut is to the min cut. A trivial modification was introduced that also allows the user to output all cuts with weight $w(C) \leq k + w(C_0)$, where again $k > 0$ is a user specified parameter. A simple test script in the repository shows how to use the basic functionality of nearmincut.

The recursive algorithm of Balcioglu and Wood is briefly summarized here and further details can be found in Ref. 1. Consider the directed graph in Figure S1. Note that this graph is slightly different than the example presented in the manuscript since, instead of anti-parallel, directed edges between nodes, only a single directed edge exists. This example represents the identification of all cuts with $w(C) \leq k + w(C_0)$ where $k = 2$. Each graph in Figure S2a represents the min cut found at each step of the recursive enumeration of the

cutset, and Figure S2b visualizes the corresponding recursion tree. At each iteration edges have forced exclusion (by setting the weight to infinity) or forced inclusion (by adding an edge with infinite weight from both the source and target to both nodes in the edge). For example, after the first min cut is identified (Iter. 1) with $w(C_0) = 6$, e_{10} is excluded from the next possible solution by setting its weight to infinity. This produces a min cut with value of 8 in Iter. 2, which satisfies $w(C) \leq k + 6$. Hence we now go one level deeper into the recursion tree. Excluding e_4 does not produce a min cut with $w(C) \leq k + 6$ in Iter. 3. In Iter. 4 we must include e_4 and exclude e_3. In Iter. 4 one can see that the inclusion of e_4 is forced by setting $w(e_1) = w(e_8) = \text{inf}$. This again leads to a min cut with a value greater than 8, leading to termination at this node in the recursion tree. The examples are continued for Iter. 5 and Iter. 6. Upon completion of the recursion tree, one would find in addition that $w(e_{10}e_1) = 8$, yielding four possible solutions for all (near) min cuts.

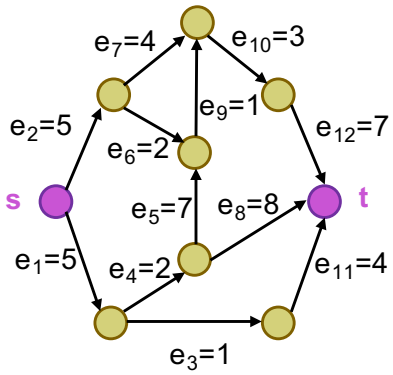

Figure S1: A sample directed graph for which we seek to find all $C \in \mathbf{C}$ with $w(C) <= k + w(C_0)$ where $k = 2$.

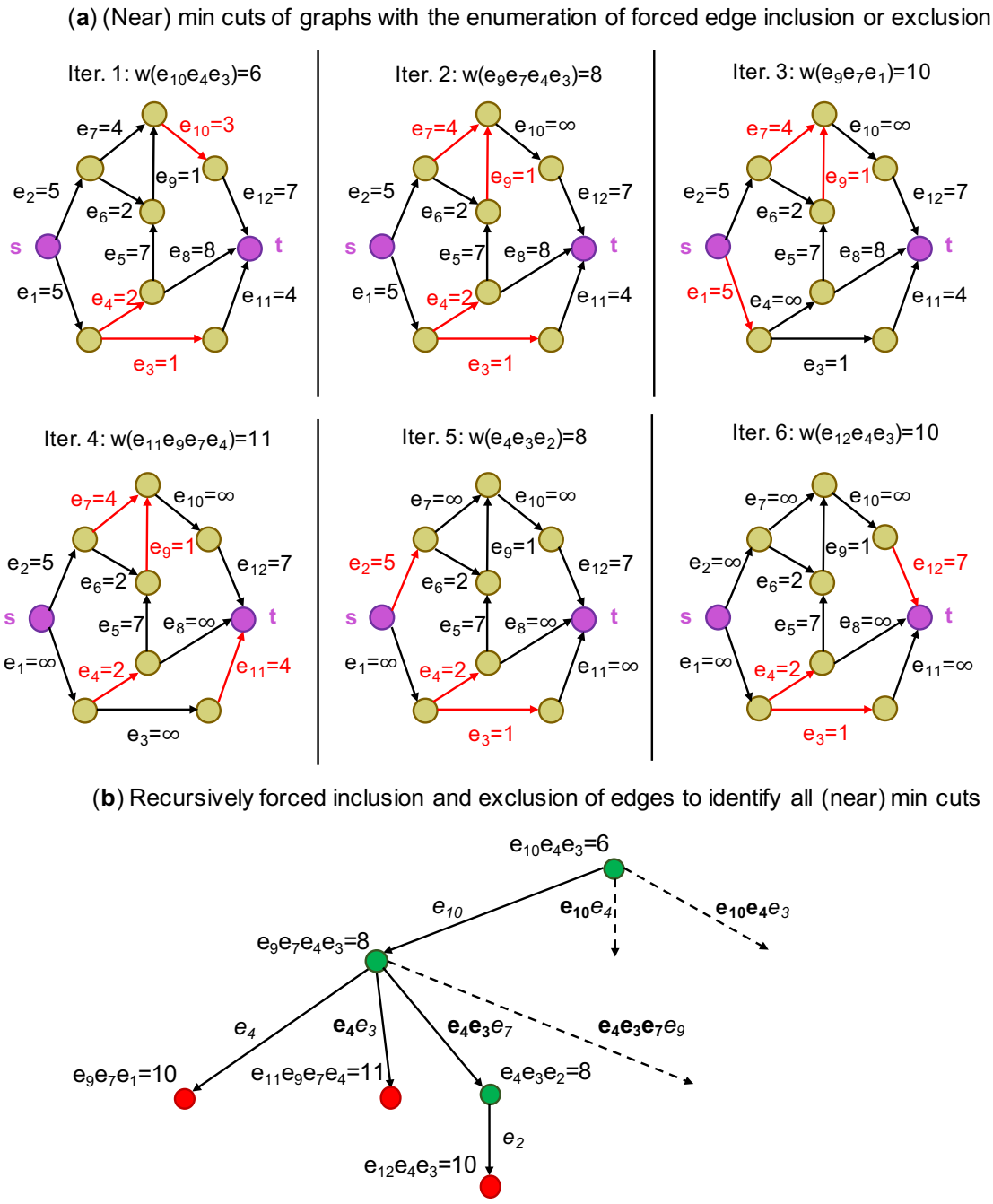

Figure S2: (a) Visualization of the solution to the min cut problem at the first six iterations in the recursion tree. Red edges represent the min cut in each particular iteration. (b) The partially completed recursion tree where a node represents a cut and the connection to its child shows which edges have forced exculsion (italicized) and forced inclusion (bold) in the next iteration. The base case returns occurs when $w(C) > k + w(C_0)$ (red nodes) for $k = 2$. Dashed lines indicate further exploration of the recursion tree is needed since the parent node satisfies $w(C) \leq k + w(C_0)$

Statistics of min cuts for IZA zeolites

Fig. S3 visualizes the ranking scheme demonstrated in Table S1 for all IZA zeolites in a single figure. To construct this figure, each IZA contributes one entry in each column of the histogram. For example, MWW contributes one purple unit in the (001) column since the (001) min cut density achieves the lowest cut density (rank R1) of all other Miller planes, which is also shown in Table S1. If two Miller surfaces are symmetrically equivalent, both are assigned the same rank for visualization purposes according to a dense ranking scheme (each surface's ranking number is 1 plus the number of items ranked above it that are distinct with respect to the ranking order). Anything above a rank of R16 is included in the R16 bin color scheme. The min cut density for each face of each zeolite, as well as the corresponding structure file, can be found in the supplementary data. According to Fig. S3, the majority of structures have R1 ranks occurring for (001), (010), and (100) Miller faces, while a non-negligible number have R1 ranks occurring for (011), (101), and (110) faces. The general trend shows that, as might be expected, higher index Miller faces typically result in a higher min cut density. Yet clearly there are many outlying structures which have high index Miller faces with low ranks (R1, R2, or R3 for example). These subtleties highlight the necessity of an automated and robust approach for zeolite surface generation and screening: various IZA structures may require obtaining the min cut density of a relatively high index Miller plane to find the R1 surface. This task would be extremely arduous if not impossible by manual/visual inspection.

The power of the graph theory algorithm can also be appreciated by observing when it identifies min cuts that lead to faceted surfaces. When the min cut is formed by a highly faceted surface, which is especially true for high index Miller planes in the case of all silica zeolites, it becomes visually clear that only an advanced graph theory based algorithm can identify the solution to the min cut. Take for example several min cuts for various Miller planes in MFI, shown in Fig. S4. For high index Miller faces, the surface terminations become highly faceted to minimize the number of edges that are cut, but they are so complex that

Table S1: Ranking of the EMT and MWW Miller surfaces based on their min cut density, δ. Surfaces ranked higher than 10 are omitted for clarity.

	EMT		MWW	
Rank	Face	δ	Face	δ
R1	(001)	0.0234	(001)	0.0112
R2	(100)	0.0248	(102)	0.0314
R3	$(1\bar{1}0)$	0.0248	$(10\bar{2})$	0.0314
R4	(010)	0.0248	$(1\bar{1}2)$	0.0314
R5	(101)	0.0256	$(\bar{1}12)$	0.0314
R6	$(10\bar{1})$	0.0256	(012)	0.0314
R7	$(1\bar{1}1)$	0.0256	$(01\bar{2})$	0.0314
R8	$(\bar{1}11)$	0.0256	(100)	0.0331
R9	(011)	0.0256	$(1\bar{1}0)$	0.0331
R10	$(01\bar{1})$	0.0256	(010)	0.0331
...				

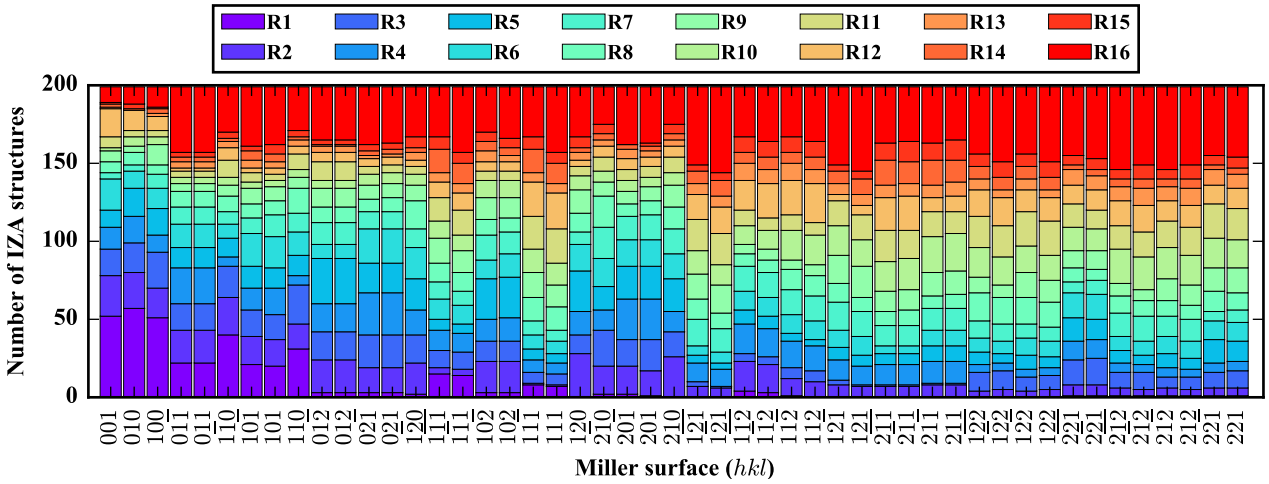

Figure S3: The tabulated ranks of all 2-maximum index Miller surfaces of each IZA zeolite. Here the minimum cut density corresponds to a rank of R1, and the maximum density corresponds to R16 (all surfaces with a higher density than R16 are included in the R16 color-coding).

only an advanced algorithmic approach would be able to determine them.

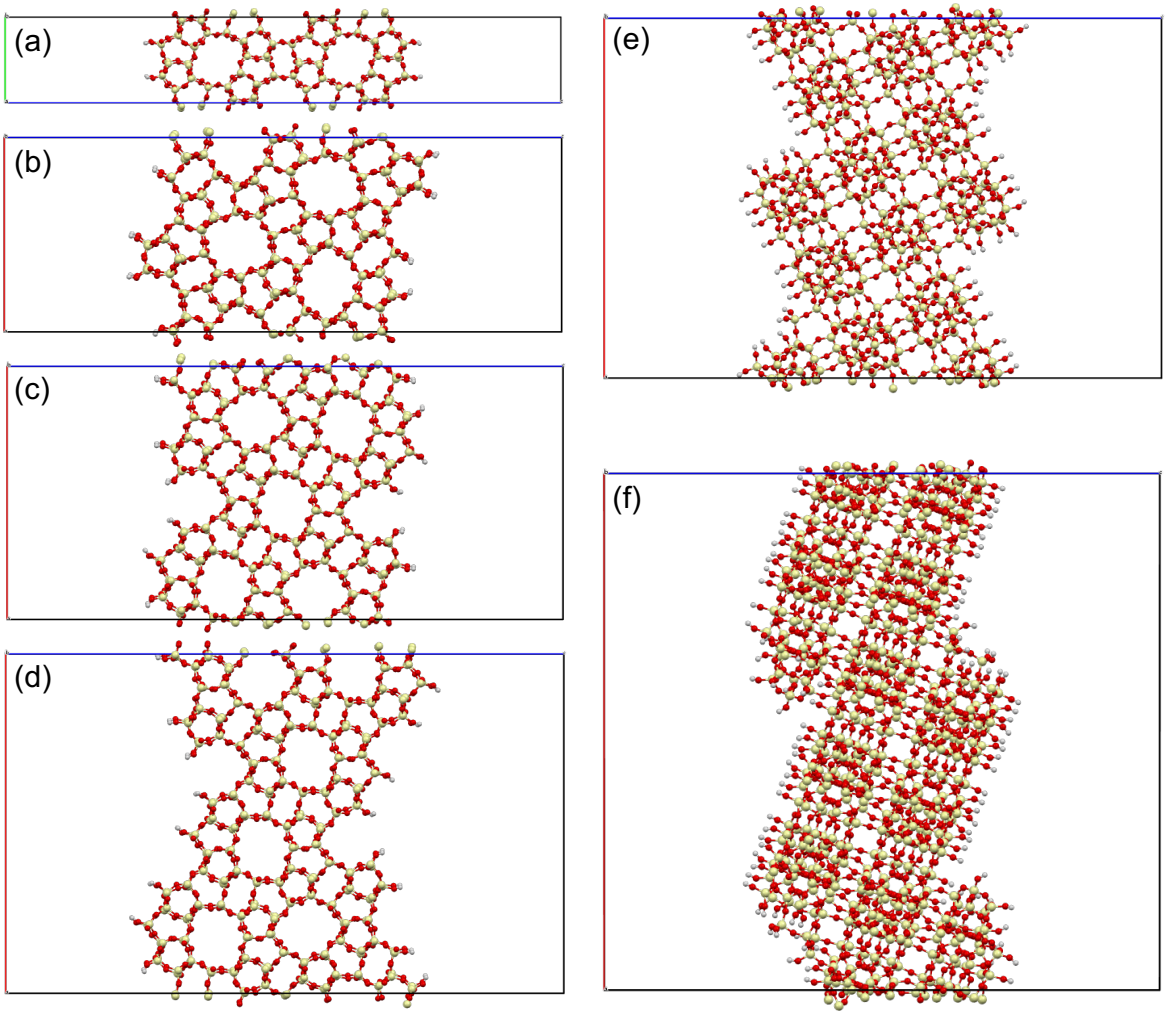

Figure S4: (a-f) show the min cut surface termination for the (100), (101), (201), (301), (120), and (511) Miller planes of MFI, respectively.

Visualization of zeolites with high potential for a layered 2D form

The candidates with the largest $\delta_{R2} - \delta_{R1}$ values are visualized in Fig. S5. The ranking of all IZA materials based on this descriptor is provided in supplementary data files as described in the following sections. Fig. S5 displays the R1 plane horizontally and orthogonal to the viewing page. These R1 surfaces are the predicted surfaces that would be expressed if the zeolite can be synthesized in a layered 2D form.

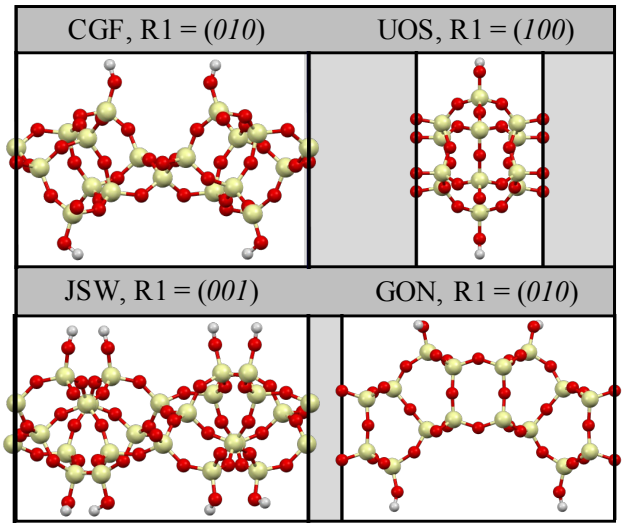

Figure S5: Visualization of the proposed layered precursor for the four IZA structures with the largest $\delta_{R2} - \delta_{R1}$ values that have not yet shown experimental verification of a 2D form. The structures are shown with the R1 Miller surface horizontal and orthogonal with respect to the viewing.

Supporting data files

The following sections explain the supplementary data files in *supplementary_data.zip* that include the structures/min-cut statistics/properties/data that was used to generate the figures in the manuscript.

Library of 2D zeolite nanosheets

A single min cut surface slab is provided for each symmetrically unique Miller face of each IZA zeolite in the all_slabs directory.

A single R1 surface slab is provided for each symmetrically unique Miller face of each IZA zeolite studied in IZA_all_slabs.zip. Note that, even if there are multiple unique terminations for a Miller face that have the same minimum cut density, only one slab is provided for that Miller face. Only data for symmetrically unique Miller faces is generated (where the symmetrical faces are determined by Pymatgen), so the symmetrically equivalent Miller faces are given in IZA_symm_equiv_M.txt. All hkl pairs appearing in the same line are equivalent.

Detailed data for IZA min cuts with normal weighting

For the normal edge weighting scheme where all edges between Si nodes have $w_e = 1$, the detailed data is given in IZA_min_cut_detailed.txt. For each IZA structure and symmetrically unique Miller face, M, the Miller indices and δ_M are listed. Column 1 = $w(C_0)$, Column 2 = area of the ab slab face, Column 3 = δ_M, Column 4 = a string denoting the IZA and M.

Summary data for IZAs min cuts with normal weighting

For the normal edge weighting scheme, the data is summarized in IZA_min_cut_summary.txt. Column 1 = IZA code, Column 2 = $\delta_{R2} - \delta_{R1}$, Column 3 = δ_{R1}, Column 4 = D_f (largest free sphere in crystallographic direction normal to R1).

Detailed data for IZAs with D4R special weighting

For the D4R edge weighting scheme, where any edge e with exactly one Si node in a D4R sub-unit has $w_e = 0$, the data is summarized in IZA_min_cut_summary_D4R.txt. Only IZAs that contain the D4R subunit are listed. Column 1 = number of edges in the min cut, Column 2 = area of the ab slab face, Column 3 = δ_M, Column 4 = a string denoting the

IZA and M.

Summary data for IZAs with D4R special weighting

For the D4R edge weighting scheme, the data is summarized in IZA_min_cut_summary_D4R.txt. Column 1 = IZA code, Column 2 = $\delta_{R2} - \delta_{R1}$, Column 3 = δ_{R1}, Column 4 = D_f (largest free sphere in crystallographic direction normal to R1).

References

(1) Balcioglu, A.; Kevin Wood, R. *Netw. Interdiction Stoch. Integer Program.*; Kluwer Academic Publishers: Boston, 2003; pp 21–49, DOI: 10.1007/0-306-48109-X_2.

(2) Boyd, P. G.; Moosavi, S. M.; Witman, M.; Smit, B. Force-Field Prediction of Materials Properties in Metal-Organic Frameworks. *J. Phys. Chem. Lett.* **2017**, *8*, 357–363, DOI: 10.1021/acs.jpclett.6b02532.

(3) Sun, W.; Ceder, G. Efficient creation and convergence of surface slabs. *Surf. Sci.* **2013**, *617*, 53–59, DOI: 10.1016/j.susc.2013.05.016.