Electronic Supplementary Material (ESI) for CrystEngComm.

Supplementary Information: Construction of Hypothetical MOFs using a Graph Theoretical Approach

Peter G. Boyd and Tom K. Woo*

Centre for Catalysis Research and Innovation, Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie, Ottawa, ON, K1N 6N5, Canada

Constructing the embedding for the dia and qtz nets

In the case of the labelled quotient graph for the **dia** net, we construct the B^* matrix using the 3 cycles $(-e_1+e_2)$, $(-e_1+e_3)$, $(-e_1+e_4)$ as a basis for the cycle space and one co-cycle, summing the outgoing edges from the vertex A, $(e_1+e_2+e_3+e_4)$. The image of these vectors, represented in the matrix α_{dia}, has its first 3 rows as the sum of the edge labels associated with the 3 cycle vectors, and the final row as the zero vector, which as stated in the main text, will yield the barycentric placement of the vertices.

$$B^*_{dia} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \alpha_{dia} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{(SE1)}$$

Then invoking Equation 1 from the main text, we obtain a representation of the positive orientations of the arcs in the lattice space.

$$\Omega^+_{dia} = \begin{pmatrix} -1/4 & -1/4 & -1/4 & 1/4 \\ 3/4 & -1/4 & -1/4 & 1/4 \\ -1/4 & 3/4 & -1/4 & 1/4 \\ -1/4 & -1/4 & 3/4 & 1/4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -1/4 & -1/4 & -1/4 \\ 3/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & 3/4 \end{pmatrix} \quad \text{(SE2)}$$

Each row in the matrix Ω^+ defines a line in the orthonormal lattice space of the net. With all of the lines defined, a lattice representation of the points and lines can be realized by placing the point corresponding to vertex A in any lattice position, say (5/8, 5/8, 5/8). The vertex B can be found by adding the first line in Ω^+ to our initial placement of A. Thus B will have the lattice coordinates (3/8, 3/8, 3/8).

The labelled quotient graph **qtz** contains 3 vertices and 6 arcs. The matrix B^* can be constructing using the cycles $(+e_1-e_2+e_3)$, $(-e_1+e_6)$, $(-e_2+e_5)$, and $(-e_4+e_3)$. We will take the outward oriented arcs from vertices A and B as the basis for the co-cycle space; $(+e_1-e_3-e_4+e_6)$ and $(-e_1-e_2-e_5-e_6)$ which are represented in the last two rows of the matrix B^*_{qtz}. By summing the labels associated with these edges, as was done in above for **dia**, we obtain the matrix α_{qtz}

$$B^{*}_{qtz} = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ -1 & -1 & 0 & 0 & -1 & -1 \end{pmatrix}, \alpha_{qtz} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ -1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{(SE3)}$$

Again, using equation 1 from the main text, we obtain lattice representations of the arcs for **qtz**

$$\Omega^{+}_{qtz} = \begin{pmatrix} 1/3 & -1/3 & -1/6 & -1/6 & 1/6 & -1/6 \\ -1/3 & -1/6 & -1/3 & 1/6 & -1/6 & -1/3 \\ 1/3 & 1/6 & -1/6 & 1/3 & -1/3 & -1/6 \\ 1/3 & 1/6 & -1/6 & -2/3 & -1/3 & -1/6 \\ -1/3 & -1/6 & 2/3 & 1/6 & -1/6 & -1/3 \\ 1/3 & 2/3 & -1/6 & -1/6 & 1/6 & -1/6 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ -1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & 0 \\ -1/2 & 0 \\ -1/2 & -1/ \\ 0 & -1/ \end{pmatrix} \quad \text{(SE4)}$$

VASP calculations

The Vienna Ab initio Simulation Package (VASP)[1–4] was used to compute the single point energies and stresses of each hypothetical MOF. The energy cutoff for the planewave basis was 500 eV, sampling the brillouin zone at the gamma point. The projector-augmented wave (PAW)[5,6] pseudopotentials were used with the PBE functional[7,8]. The electronic wave function was considered converged when the free energy difference between each electronic step was below 1×10^{-5} eV.

Using TOPOS to identify the underlying net

Several steps are necessary to accomplish the goal of identifying the underlying net of the MOFs given only a description of their atomic positions and the lattice vectors. Structures were read in to the TOPOS 4.0 program[9] in the crystallographic information file (cif) format. Bonds in the MOFs were computed using the method of spherical sectors[10] within the AutoCN subprogram. Using the adjacency matrix computed from the previous step, Automatic Description of Structure (ADS) was then run with the cluster simplification method to reduce the MOF atoms to a series of enclosed cycles connected to each other by hydrogen bonds, which are used in the program to flag connectivity between distinct isolated molecules (the cycles). A standard simplification of the net is then computed in ADS, where molecules, flagged by hydrogen bonds, are reduced to single vertices within the structure. The resulting simplified structure is cleaned of any 0-, 1-, and 2-coordinate nodes to yield the final net. To classify this net into a particular topology, sequences are generated for all the vertices in the unit cell based on their first 10 coordination spheres. The sequences describe the number of vertices at each coordination sphere, and descriptions of the number of closed rings found within each angle of the vertex edges. These sequences unambiguously classify nets into a particular topology, so they can be compared with the nets within the TOPOS Topological Database (TTD) collection. The resulting topology for the example in Figure S1 was found to be **tbo**, which was the net used to build the structure in the structure generation program.

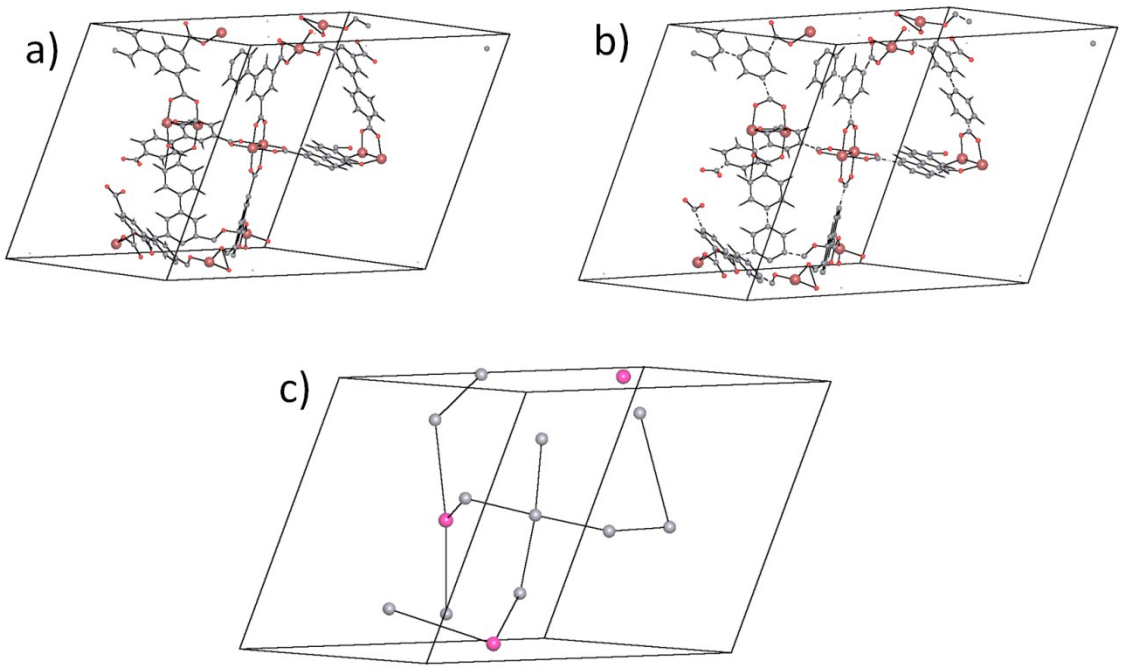

Figure S1: Graphical representation of progressing, in TOPOS, from a) the MOF structure, to b) the structure where molecules are isolated to clusters of enclosed rings, to finally c) the underlying net where the coordination sequence of each vertex is used to classify it as the **tbo** net.

Format of the labelled quotient graph

Labelled quotient graphs are read in as a single string. The string describes all of the arcs in the graph, where each arc is formatted 'v1 v2 L1 L2 L3' where v1 is the vertex origin of the arc, v2 is the terminal vertex of the arc, and L1 L2 L3 constitute three dimensional lattice coordinate label (e.g. 0 0 1).

Re-creation of the experimental unit cell dimensions of MOF-210 and NU-110

Following generation of the hypothetical version of MOF-210 and NU-110, the raw structure produced from the algorithm possessed cell dimensions which did not agree with the experimental structure, and so to obtain the agreement in cell parameters shown in the first two columns of Table 2, the unit cell vectors were re-defined using a simple linear algebraic transformation.

$$Z_{exp} = Z_{hyp} \times b \qquad \text{(SE5)}$$

Where Z_{exp} is the 3×3 matrix of unit cell vectors of the experimental unit cell, and Z_{hyp} is the 'raw' 3×3 matrix of the as-built hypothetical unit cell, and b is a 3×3 transformation matrix, which we solve for by setting;

$$b = Z_{hyp}^{-1} \times Z_{exp} \quad \text{(SE6)}$$

To obtain an approximate supercell of the hypothetical structure which best matches that of the experimental structure, we round every element in *b* to it's nearest integer and then produce the new hypothetical unit cell dimensions via

$$Z_{hyp}^{'} = Z_{hyp} \times b \quad \text{(SE7)}$$

This way, we effectively generated a supercell from a linear combination of the original vectors to yield dimensions which more-or-less agreed with the experimental cell shape. As such the re-defined structures are in no way different from the output from the structure generation code, only their translational period was re-defined. The transformation matrices used to re-define the hypothetical versions of MOF-210 and NU-110 are shown in Table S1. The resulting larger hypothetical MOF was optimized with a molecular mechanics force field to yield the values reported in Table 2 of the main text as well as the agreement in the superposition of atoms of the hypothetical (blue) and experimental (red) structures shown in Figure 11b and 11d of the main text. The presented values for surface area and pore volume in Table 2 were identical both before and after the cell parameters were adjusted, and are in excellent agreement with the computed values for the experimental crystal structure.

Table S1. The cell vectors output from the generation program and the transformation matrices used to generate experimental-like cells.

	MOF-210	NU-110
original cell vectors (Z_{hyp})	$\begin{bmatrix} 51.4 & 0.0 & 0.0 \\ -25.7 & 44.5 & 0.0 \\ -0.1 & 29.9 & 63.8 \end{bmatrix}$	$\begin{bmatrix} 48.9 & 0.0 & 0.0 \\ -21.6 & 46.0 & 0.0 \\ -3.5 & -26.6 & 43.4 \end{bmatrix}$
transformation matrix (*b*)	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -2 & 3 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$
transformed vectors ($Z_{hyp}^{'}$)	$\begin{bmatrix} 51.4 & 0.0 & 0.0 \\ -25.7 & 44.5 & 0.0 \\ -0.3 & 0.7 & 191.4 \end{bmatrix}$	$\begin{bmatrix} 48.9 & 0.0 & 0.0 \\ 27.3 & 46.0 & 0.0 \\ 23.8 & 19.4 & 43.4 \end{bmatrix}$
experimental crystal vectors (Z_{exp})	$\begin{bmatrix} 50.7 & 0.0 & 0.0 \\ -25.4 & 43.9 & 0.0 \\ 0.0 & 0.0 & 194.2 \end{bmatrix}$	$\begin{bmatrix} 48.6 & 0.0 & 0.0 \\ 24.3 & 42.1 & 0.0 \\ 24.3 & 14.0 & 39.7 \end{bmatrix}$

References

1 G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169–11186.

2 G. Kresse and J. Hafner, *Phys. Rev. B*, 1994, 49, 14251–14269.

3 G. Kresse and J. Hafner, *Phys. Rev. B*, 1993, **47**, 558–561.

4 G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.

5 G. Kresse and D. Joubert, *Phys. Rev. B*, 1999, **59**, 1758–1775.

6 P. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953–17979.

7 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1997, **78**, 1396–1396.

8 J. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

9 V. A. Blatov, A. P. Shevchenko and D. M. Proserpio, *Cryst. Growth Des.*, 2014, **14**, 3576–3586.

10 E. V. Peresypkina and V. A. Blatov, *Acta Crystallogr. Sect. B Struct. Sci.*, 2000, **56**, 1035–1045.