ADVANCED FUNCTIONAL MATERIALS

# Supporting Information

for *Adv. Funct. Mater.*, DOI 10.1002/adfm.202301594

In Search of Covalent Organic Framework Photocatalysts: A DFT-Based Screening Approach

*Beatriz Mourino, Kevin Maik Jablonka, Andres Ortega-Guerrero* and Berend Smit**

# Supporting Information: In search of Covalent Organic Framework photocatalysts: A DFT-based screening approach

Beatriz Mourino, Kevin Maik Jablonka, Andres Ortega-Guerrero,* and Berend Smit*

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

E-mail: oandresg15@gmail.com; berend.smit@epfl.ch

## Contents

# 1 Methods

## 1.1 DFT calculations

DFT calculations were performed with the QuickStep code of CP2K 9.1.[S1] The Perdew-Burke-Ernzehof (PBE) exchange-correlation functional was employed with DFT-D3(BJ) van der Waals correction. Plane-wave and relative energy cutoffs of 600 Ry and 60 Ry, respectively, were chosen with a 4-level multigrid mapping. The orbital transformation (OT) method was used. Calculations were performed at Γ-point over the irreducible Brillouin zone due to the large cell parameters of the COFs. When necessary, supercells were built to ensure all cell parameters have a distance above 10 Å, and guarantee proper sampling.

Goedecker-Teter-Hutter (GTH) pseudopotentials and mixed Gaussian and plane-waves basis sets were used, with double-$\zeta$ and triple-$\zeta$ polarization MOLOPT basis sets for describing non-metals and metals, respectively. For PBE0 calculations on the 20-COFs dataset, MOLOPT-ADMM auxiliary basis functions with cFIT11 for closed-shell metals (Ni, Zn, Si) and pFIT3 for non-metals were employed along with the auxiliary density matrix method (ADMM) to reduce computational cost.

At all stages of optimization, COFs with band gaps below 0.5 eV were excluded to

compensate for errors from PBE, and this must be taken into account for structures with metallic behavior.

For band structure calculations, the choice of k-points mesh was mostly at Γ-point, but when needed a Monkhorst-Pack scheme was chosen to guarantee proper sampling of structures with lattice parameters smaller than 10Å. The reciprocal space path within the first Brillouin zone (BZ) was computed by connecting the high-symmetry points with the Seek-path package.[S2] As default, we chose to sample 20 k-points along the line between two high-symmetry points.

# 2 Datasets

Table S1: Dataset of 20 CURATED COFs chosen for PBE and PBE0 calculations of IP, EA, band gap values, and Λ between frontier orbitals by UKS electron and hole injection

| **Name** | **DOI** | **2D or 3D** | **Heteroatom** |
|---|---|---|---|
| COF-5 | 10.1126/science.1120411 | 2D | B |
| COF-103 | 10.1126/science.1139915 | 3D | B, Si |
| TFPT-COF | 10.1039/C4SC00016A | 2D | - |
| HBC-COF | 10.1038/ncomms8786 | 2D | - |
| TAT-COF-1 AA | 10.1039/C5TC02256H | 2D | - |
| Ni-DBA-3D-COF-bor | 10.1021/jacs.6b10316 | 3D | B, Ni |
| TPE-Ph COF | 10.1021/jacs.6b02700 | 2D | B |
| 3D-Por-COF | 10.1021/jacs.7b04141 | 3D | - |
| LZU-111 | 10.1126/science.aat7679 | 3D | Si |
| TTCOF-Zn AA | 10.1002/anie.201906890 | 2D | Zn |
| GS-COF-1 | 10.1021/jacs.0c05970 | 2D | - |
| JUC-564 | 10.1021/jacs.0c06485 | 3D | - |
| COF-QA-2 | 10.1002/adma.202001284 | 3D | Cl |
| CCOF-15 | 10.1002/anie.202013926 | 3D | Cl |
| FCOF-5 | 10.1021/jacs.0c12505 | 3D | - |
| TAPP-COF | 10.1021/jacs.1c02932 | 2D | - |
| NKCOF-108 | 10.1021/acscatal.0c04820 | 2D | F, S |
| WBDT | 10.1039/D0SC03909H | 2D | S |
| JUC-569 2-fold acs | 10.1021/jacs.0c12499 | 3D | - |
| BP-COF-1 | 10.1126/science.abd6406 | 3D | B, P |

Table S2: Photocatalytic descriptors for the $N_x$-COF series (x=0-3): IP, EA, PBE0-adjusted band gap, m* and charge recombination descriptor ($\Lambda$). Experimental optical band gap values reported in the literature for this series range between 2.6–2.7 eV, and the same work states calculated EA for this series in the order of -2 eV[S3]

| **Name** | **IP [eV]** | **EA [eV]** | **BG [eV]** | **$m^*_e$ [$m/m_e$]** | **$m^*_h$ [$m/m_e$]** | $\Lambda$ |
|---|---|---|---|---|---|---|
| $N_0$-COF | -5.34 | -2.27 | 3.05 | 0.09 | 0.28 | 0.78 |
| $N_1$-COF | -5.18 | -2.57 | 2.53 | 0.11 | 0.05 | 0.37 |
| $N_2$-COF | -5.37 | -2.71 | 2.57 | 0.12 | 0.28 | 0.25 |
| $N_3$-COF | -5.74 | -2.74 | 2.97 | 0.40 | 0.22 | 0.64 |

# 3 Preliminary tests

## 3.1 PBE vs PBE0

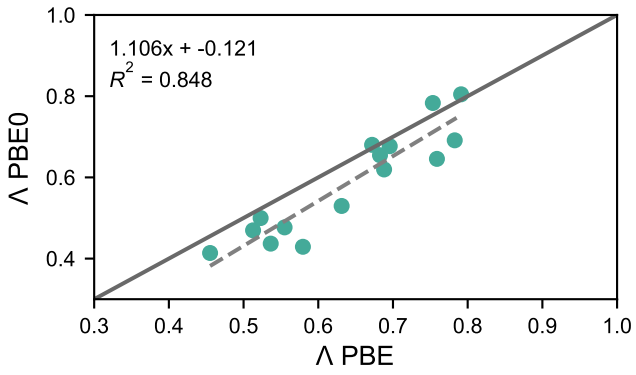


Figure S1: Evaluation of correlation between the averaged spatial overlap ($\Lambda$) computed with PBE0 UKS and PBE UKS calculations.

Figure S2 shows the well-known opening of the band gap when using a hybrid functional (PBE0) in comparison to a GGA functional (PBE), and the nature of the orbitals contributing to the band gap in COFs is mostly unchanged between PBE to PBE0.

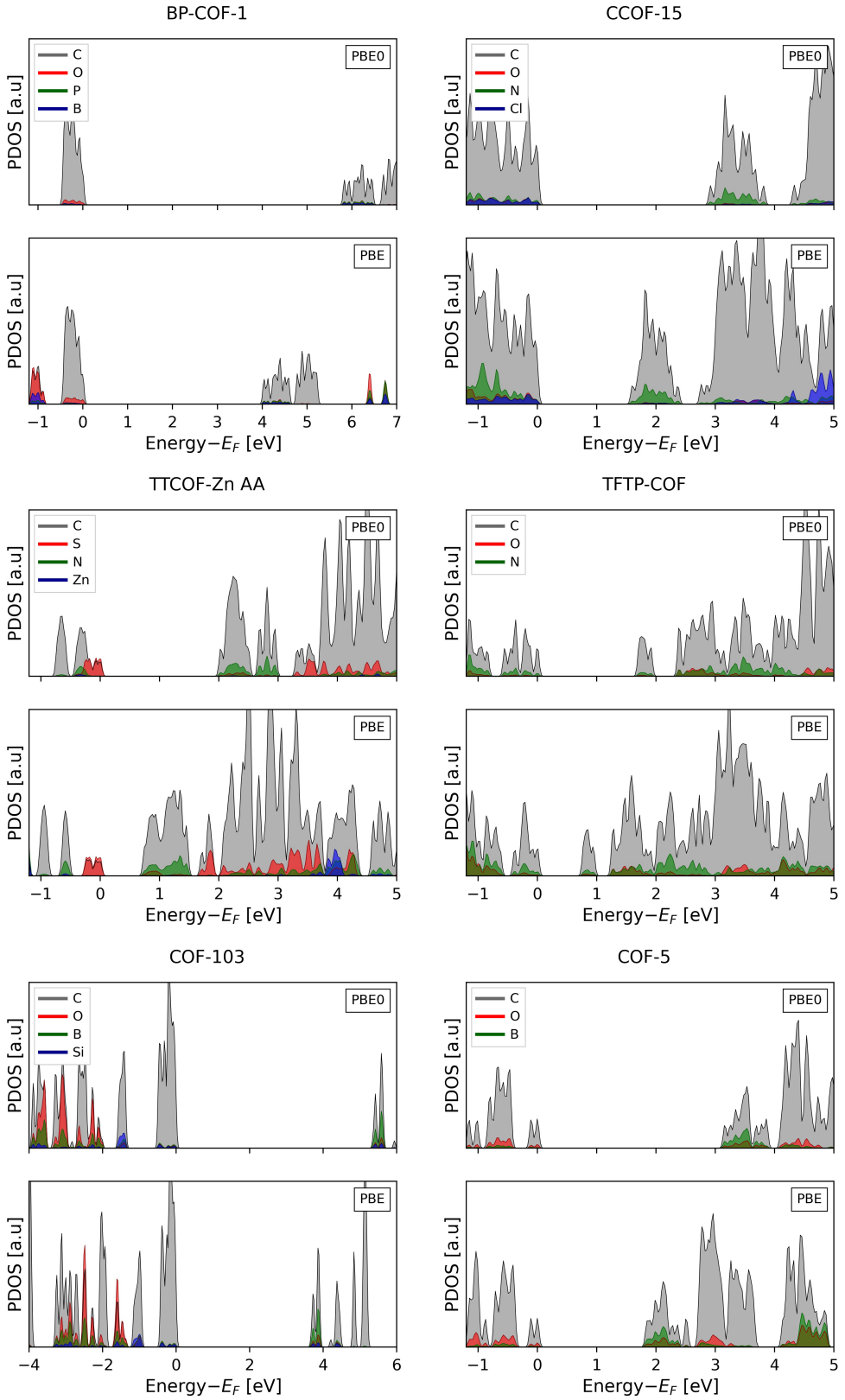


Figure S2: PBE and PBE0 partial density of states for 6 COFs in Table S1.

## 3.2 Extra structure optimization

Figure S3 shows a preliminary test used to validate the ground-state geometry of the structures from the CURATED database. We used as a starting point the already optimized structures from the CURATED COFs database, and performed PBE and PBE0 single-point calculations on those structures, as well as on structures that were further optimized with tighter parameters using PBE. The deviation of band gap values with respect to the structures as reported in the database (displayed by the height of the histograms) suggested that performing an extra structure optimization step could be crucial, and total energy values are lower (more negative) when the extra optimization is performed (see Figure S4). Therefore, we opted for a tighter optimization to guarantee a better ground-state geometry. Figure S5 shows that the correlation between PBE and PBE0 values for the structures before this extra optimization step is slightly lower than what we obtained for the extra optimized structures (Figure 1), thus further confirming our choice for extra structure optimization.

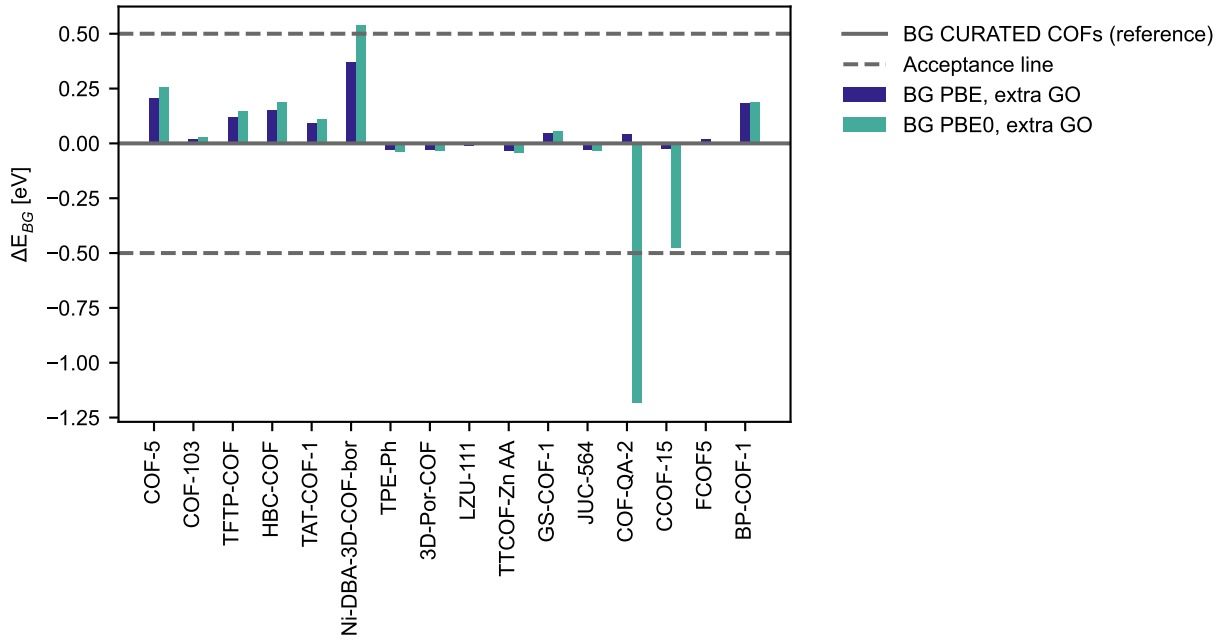


Figure S3: Band gap deviation from CURATED COFs when performing extra geometry optimization, as a preliminary test to decide on implementing an extra optimization step in the workflow described in Figure 2.

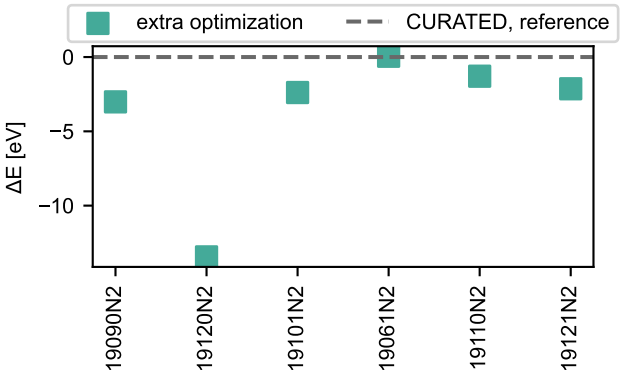


Figure S4: Total energy differences w.r.t. CURATED COFs when performing extra geometry optimization, as a preliminary test to decide on implementing an extra optimization step in the workflow described in Figure 2.

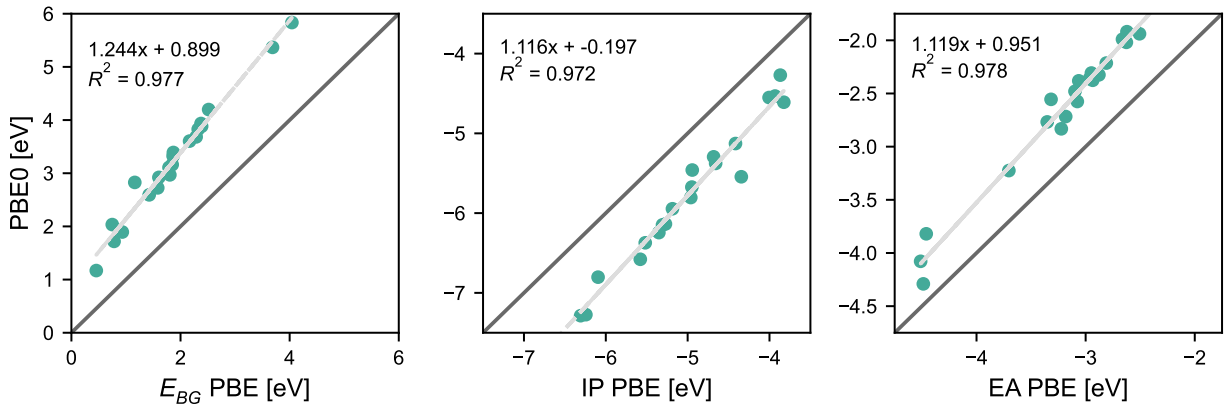


Figure S5: Correlation between PBE and PBE0 values for HOMO, LUMO, and band gap for CURATED COFs, without extra optimization step. $R^2$ values are improved when performing extra optimization step (Figure 1).

## 3.3 Band structure

COFs usually display a large unit cell, and 3D COFs often have all lattice vector lengths higher than 10 Å. In the reciprocal space, larger lengths in the real space are translated into smaller Brillouin Zones by a Fourier transform. Usually, smaller Brillouin Zones can be properly sampled at Γ-point. To test for that, we chose a 2D COF with one lattice vector (c) length smaller than 10 Å to perform band structure calculations at Γ-point and compare with different Monkhorst-Pack (MP) schemes accounting for the small c values, i.e., MP 112, 113.

Likewise, we also tested with MP 222 and 223. We tested how different integration grids can affect the effective masses, and we set as a reference line the scheme with the largest k-point mesh. Figure S6 shows that Γ-point calculations lead to m* values that deviate from MP 113 values by more than 0.1 $m_e$, but the difference is much less when using MP 112. Similar behavior can be observed in Figure S7, where we tested for larger k-point meshes. These findings suggest that the best approach for structures with lattice vectors smaller than 10 Å is to use an MP scheme to guarantee proper sampling.

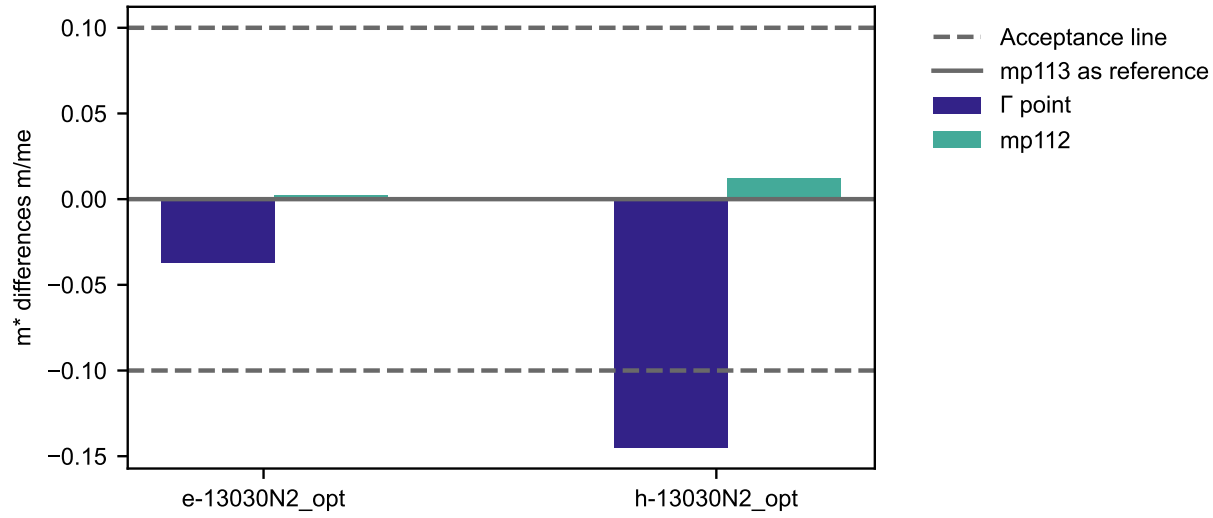


Figure S6: Band structure test for a COF with only the c lattice vector smaller than 10 Å. A Monkhorst-Pack (MP) 113 scheme is used as a reference to test for the difference in effective masses when performing calculations at Γ-point and with MP 112.

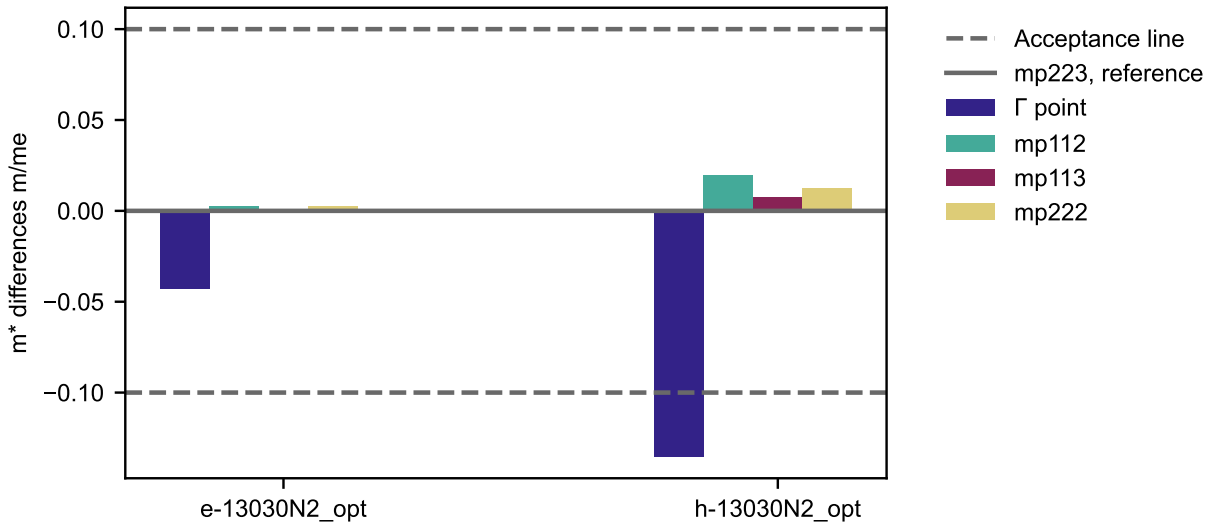


Figure S7: Band structure test for a COF with only the c lattice vector smaller than 10Å. A Monkhorst-Pack (MP) 223 scheme is used as a reference to test for the difference in effective masses when performing calculations at Γ-point and with MP 112, 113, and 222.

For MP schemes that are not Γ-centered, we tested in an hexagonal COF whether centering it could originate a significant difference. Especially for hexagonal symmetries, a Γ-centered scheme is often employed to ensure that the symmetrized k-points have the same symmetry as the hexagonal lattice. Figure S8 takes a Γ-centered scheme as reference (*i.e.*, the same scheme as the color, but centering to include the Γ-point) shows that the differences are smaller than 0.04 eV. This suggests that such centering is not necessary, thus we opted for MP schemes for structures with lattice vector lengths smaller than 10Å, even when those schemes are not centered.

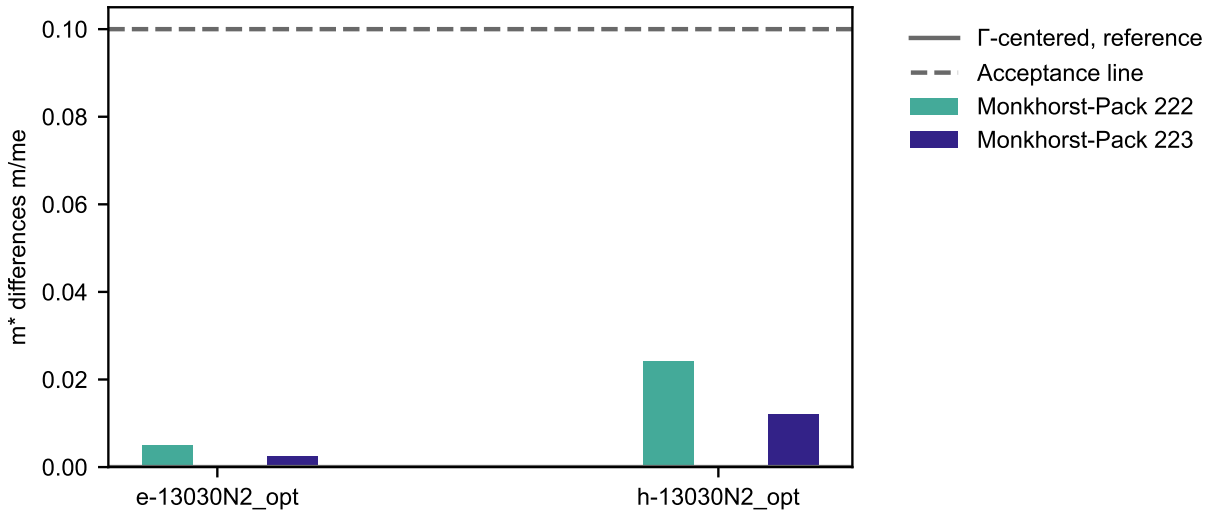


Figure S8: Band structure test for a COF with only the c lattice vector smaller than 10Å. A Γ-centered scheme is used as a reference to test for the difference in effective masses when performing calculations with MP 222 and 223 schemes, not Γ-centered.

Finally, we tested for the number of k-points along the integration line between high-symmetry points. Usually, a larger number of k-points ensures better sampling, but demands more computational resources, and finding a compromise is necessary. Moreover, to obtain the effective masses we need to compute the second derivative of the bands w.r.t. $\vec{k}$ around the band extrema. Therefore, to avoid abrupt changes in the fitting, it is important to have enough points. Figure S9 takes as reference the same k-mesh as each color but using 50 k-points. The findings suggest that a good compromise between cost and accuracy is to use 20 k-points (differences are smaller than 0.05 eV w.r.t 50 k-points).

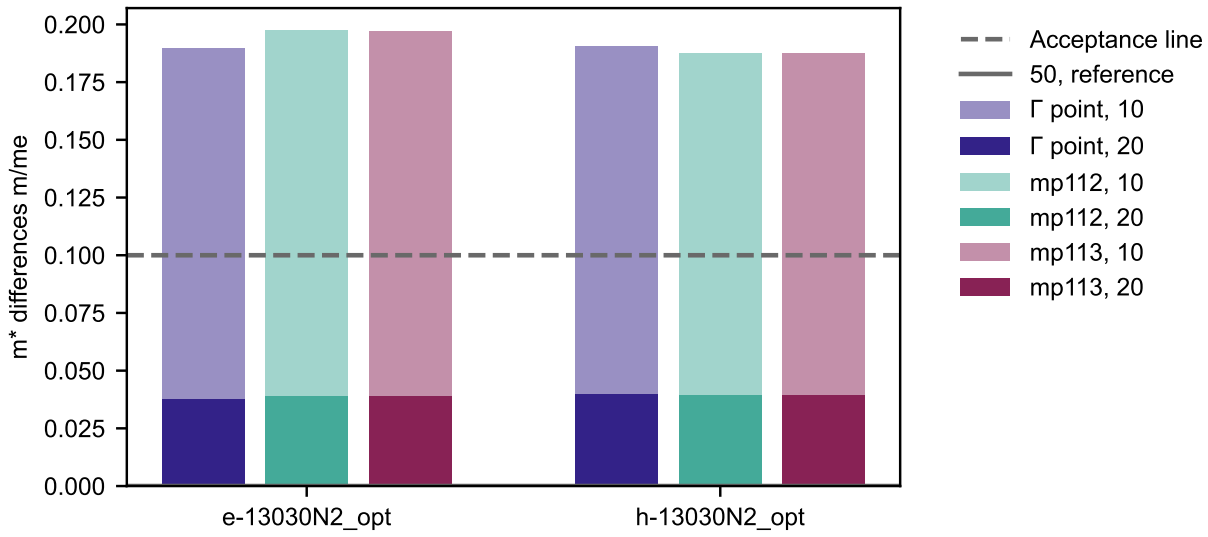


Figure S9: Band structure test for a COF with only the c lattice vector smaller than 10Å. Fixing calculations with 50 k-points along the line as a reference, we test for the difference in effective masses when performing calculations with 10 and 20 k-points with different schemes.

# 4 Structural analysis

We chose the following building blocks and functional groups for structural analysis, based on chemical intuition and their occurence in COFs:[S4–S6] acrylonitrile, arylboronate ester, azo, azine, $\beta$-ketoenamine (with 2,4,6-trimethanimidoylbenzene-1,3,5-triol, TIBT), benzene, benzonitrile, benzoxazole, borazine, boroxine, glutarimide, halogens, hydrazone, imide, imidazole, phenazine, phthalimide, porphyrin, pyrene, pyrazolate, secondary amine, thiadiazole, thienothiophene, triphenylamine (ta), 2,4,6-triphenyl-1,3,5-triazine (triazine_3bz), and 1,3,5-(triphenyl)benzene (TBZ). Their SMARTS representation was used to look for their presence in the CURATED COFs.

# 5 Boostrapped effect sizes

We performed statistical analysis to determine whether the presence of a building block or functional group could have positive, negative, or no effects on our photocatalytic DFT

descriptors. We employed boostrapped effect sizes with the DABEST package[S7] for quantitative estimation of effect sizes. As an example, the following sections show the influence of $\beta$-ketoenamine linkages (with TIBT) on all descriptors, and thiadiazole and phthalimide on the charge separation descriptor, as discussed in the main text.

## 5.1 $\beta$-ketoenamine (TIBT)

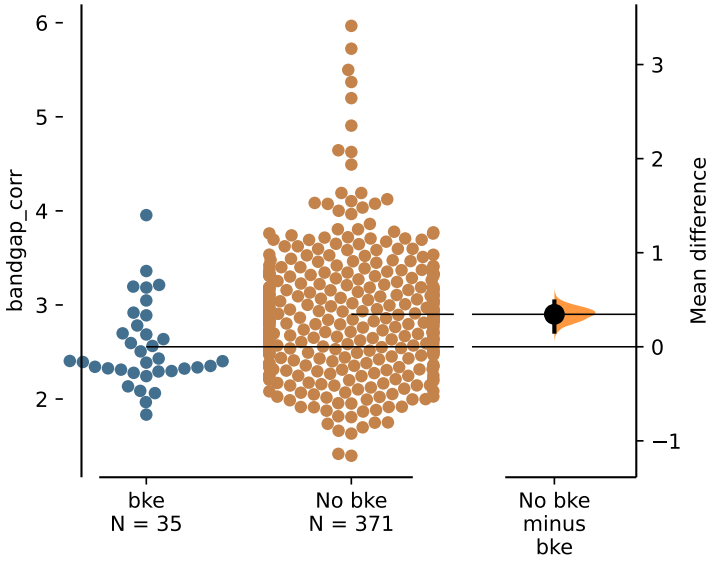


Figure S10: Bootstrapped effect sizes with DABEST[S7] comparing subsets of structures with and without the presence of $\beta$-ketoenamine (bke, with TIBT) linkage w.r.t. band gap values.

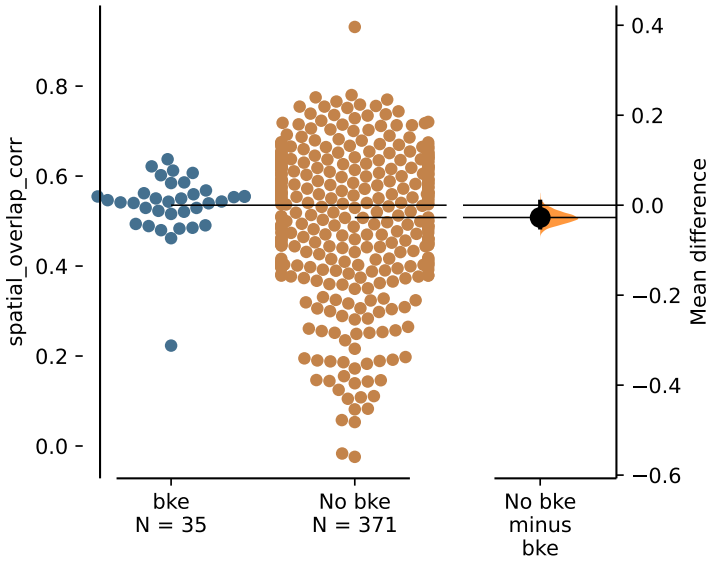


Figure S11: Bootstrapped effect sizes with DABEST[S7] comparing subsets of structures with and without the presence of $\beta$-ketoenamine (bke, with TIBT) linkage w.r.t. charge recombination descriptor values.

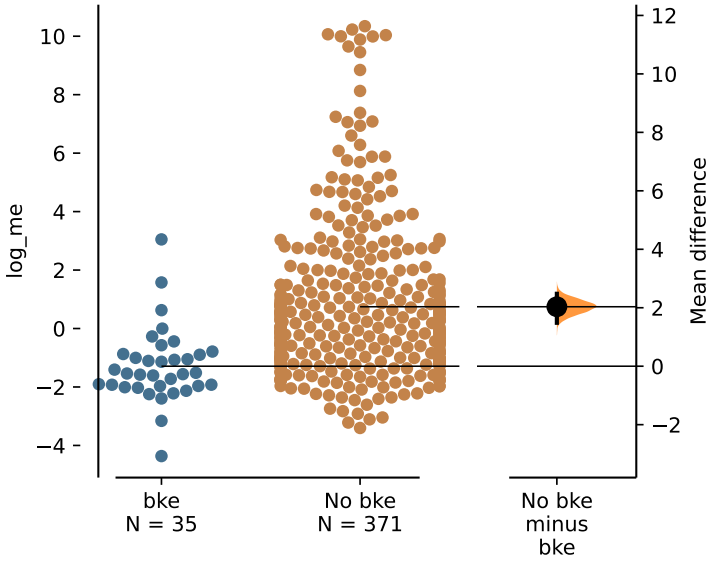


Figure S12: Bootstrapped effect sizes with DABEST[S7] comparing subsets of structures with and without the presence of $\beta$-ketoenamine (bke, with TIBT) linkage w.r.t. electron effective mass values.

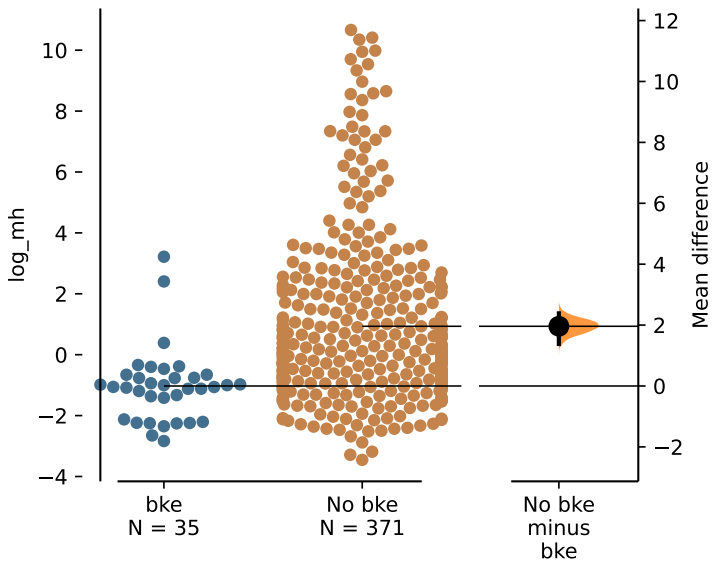


Figure S13: Bootstrapped effect sizes with DABEST[S7] comparing subsets of structures with and without the presence of $\beta$-ketoenamine (bke, with TIBT) linkage w.r.t. hole effective mass values.

## 5.2 Thiadiazole and phthalimide (charge recombination descriptor)

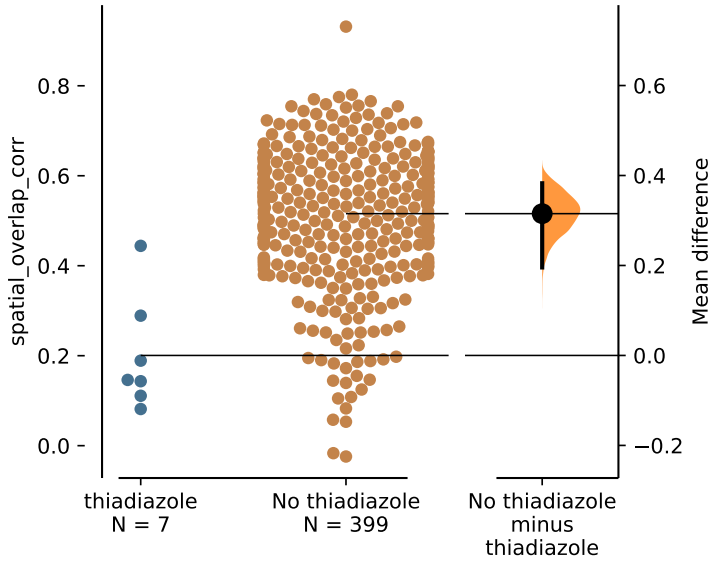


Figure S14: Bootstrapped effect sizes with DABEST[S7] comparing subsets of structures with and without the presence of thiadiazole group w.r.t. charge recombination descriptor values.

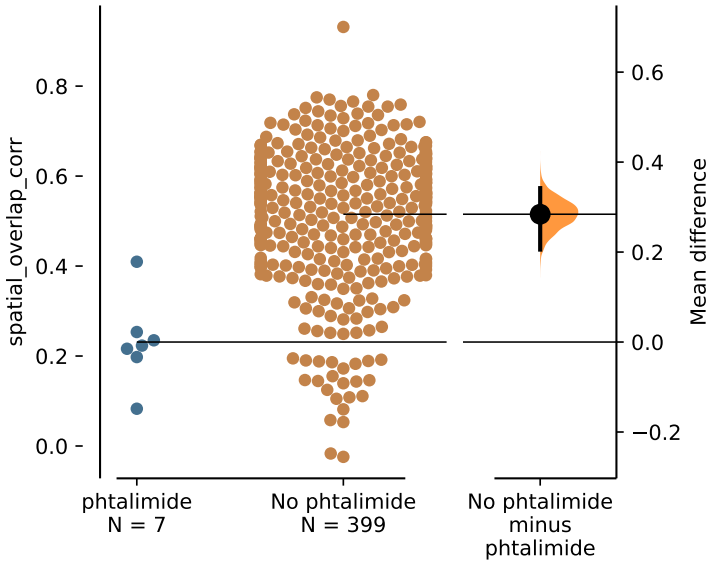


Figure S15: Bootstrapped effect sizes with DABEST[S7] comparing subsets of structures with and without the presence of phthalimide group w.r.t. charge recombination descriptor values.

# 6 Topology

Overall, the topologies identified in 300 COFs were hcb, hcb-a, sqc11236, fes, dia, hnb, dia-a, hca-a, bor-a, ctn-a, pts-f, cpq, sql, lon, dmd-a, kgd-a, fxt, jvh, tfk-a, kgm, and pbc. For some structures, more than one topology could be associated. For the remaining COFs, we encountered "unknown" or "unstable" nets.

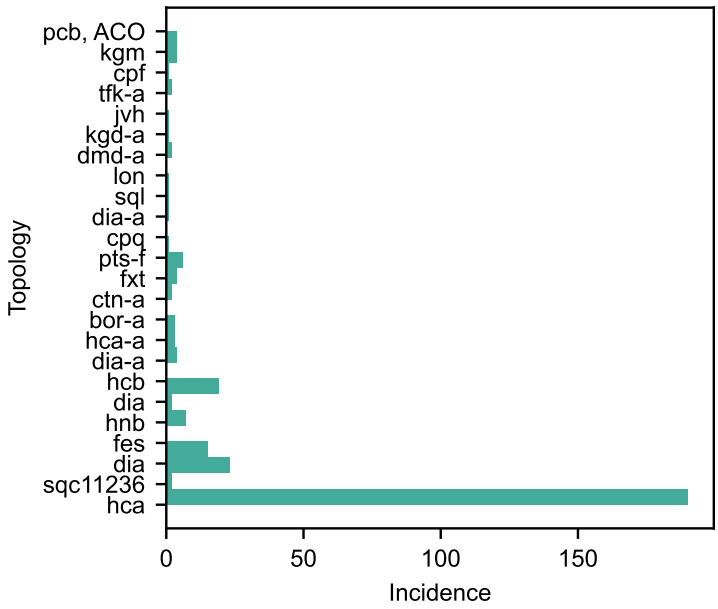


Figure S16: Distribution of topologies analyzed by CrystalNets.jl for 300 COFs.

# 7 IUPAC names

Table S3: Chemical compounds discussed throughout this work, abbreviations used and their IUPAC names (from PubChem website)

| **Name used** | **Abbrev.** | **IUPAC** |
|---|---|---|
| porphyrin | - | porphyrin |
| pyrene | - | pyrene |
| triphenylamine | - | N,N-diphenylaniline |
| 2,4,6-triphenyl-1,3,5-triazine | TPTA | 2,4,6-triphenyl-1,3,5-triazine |
| 2,3,6,7,10,11-hexahydroxytriphenylene | TP | triphenylene-2,3,6,7,10,11-hexol |
| benzene-1,4-dialdehyde | - | terephthalaldehyde |
| nitro-*p*-phenylenediamine | - | 2-nitrobenzene-1,4-diamine |
| *p*-phenylenediamine | - | benzene-1,4-diamine |
| 2,1,3-benzothiadiazole-4,7-diboronic acid | BTDADA | - |
| dicarboximide | - | - |
| triphenylbenzene | - | 1,3,5-triphenylbenzene |
| thiadiazole | - | 1,3,4-thiadiazole |
| phthalimide | - | isoindole-1,3-dione |
| 2,4,6-Tris(iminomethyl)benzene-1,3,5-triol | TIBT | - |
| glutarimide | - | piperidine-2,6-dione |
| 1,3,5-(triphenyl)benzene | - | 1,3,5-triphenylbenzene |
| pyrometallic diimide | - | pyrrolo[3,4-f]isoindole-1,3,5,7-tetrone |
| tetraphenylmethane | - | tritylbenzene |
| 1,3,5-tricyano-2,4,6-tris(vinyl)benzene | - | - |
| triethanolamine | - | 2-[bis(2-hydroxyethyl)amino]ethanol |
| arylboronate ester | - | - |
| imidazole | - | 1H-imidazole |

# 8 Packages

Packages used by this work include the cp2k-input- and cp2k-output-tools,[S8,S9] dabest for statistical analysis,[S7] pymatgen,[S10] spglib,[S11] NumPy,[S12] Pandas,[S13] SciPy,[S14] matplotlib,[S15] and bokeh.[S16]

# References


(S1) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöß, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics* **2020**, *152*, 194103.

(S2) Hinuma, Y.; Pizzi, G.; Kumagai, Y.; Oba, F.; Tanaka, I. Band structure diagram paths based on crystallography. *Computational Materials Science* **2017**, *128*, 140–184.

(S3) Vyas, V. S.; Haase, F.; Stegbauer, L.; Savasci, G.; Podjaski, F.; Ochsenfeld, C.; Lotsch, B. V. A tunable azine covalent organic framework platform for visible light-induced hydrogen generation. *Nat. Commun.* **2015**, *6*, 8508.

(S4) Li, Y.; Song, X.; Zhang, G.; Wang, L.; Liu, Y.; Chen, W.; Chen, L. 2D Covalent Organic Frameworks Toward Efficient Photocatalytic Hydrogen Evolution. *ChemSusChem* **2022**, *15*, e202200901.

(S5) Wan, Y.; Wang, L.; Xu, H.; Wu, X.; Yang, J. A Simple Molecular Design Strategy for Two-Dimensional Covalent Organic Framework Capable of Visible-Light-Driven Water Splitting. *Journal of the American Chemical Society* **2020**, *142*, 4508–4516, PMID: 32043354.

(S6) *Introduction to Reticular Chemistry*; John Wiley & Sons, Ltd, 2019; Chapter 8, pp 197–223.

(S7) Ho, J.; Tumkaya, T.; Aryal, S.; Choi, H.; Claridge-Chang, A. Moving beyond P values: data analysis with estimation graphics. *Nature Methods* **2019**, *16*, 565–566.

(S8) Cp2k input tools. `https://github.com/cp2k/cp2k-input-tools`, Accessed: 2022-12-07.

(S9) Cp2k output tools. `https://github.com/cp2k/cp2k-output-tools`, Accessed: 2022-12-07.

(S10) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **2013**, *68*, 314–319.

(S11) Togo, A.; Tanaka, I. `Spglib`: a software library for crystal symmetry search. 2018; `https://arxiv.org/abs/1808.01590`.

(S12) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.

(S13) Wes McKinney, Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference. 2010; pp 56 – 61.

(S14) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.;

Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.

(S15) Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **2007**, *9*, 90–95.

(S16) Bokeh Development Team, Bokeh: Python library for interactive visualization. 2018.