

Supplementary Information:

Multi-fidelity machine learning predictors for
gas permeability through polymers: an
example of experimental and simulation data
fusion

Brandon K. Phan,[†] Kuan-Hsuan Shen,[†] Rishi Gurnani,[†] Huan Tran,[†] Ryan Lively,[‡]
and Rampi Ramprasad^{*,†}

[†]*School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta,
Georgia 30332, United States*

[‡]*School of Chemical and Biomolecular Engineering, Georgia Institute of Technology,
Atlanta, Georgia 30332, United States*

E-mail: rampi.ramprasad@mse.gatech.edu

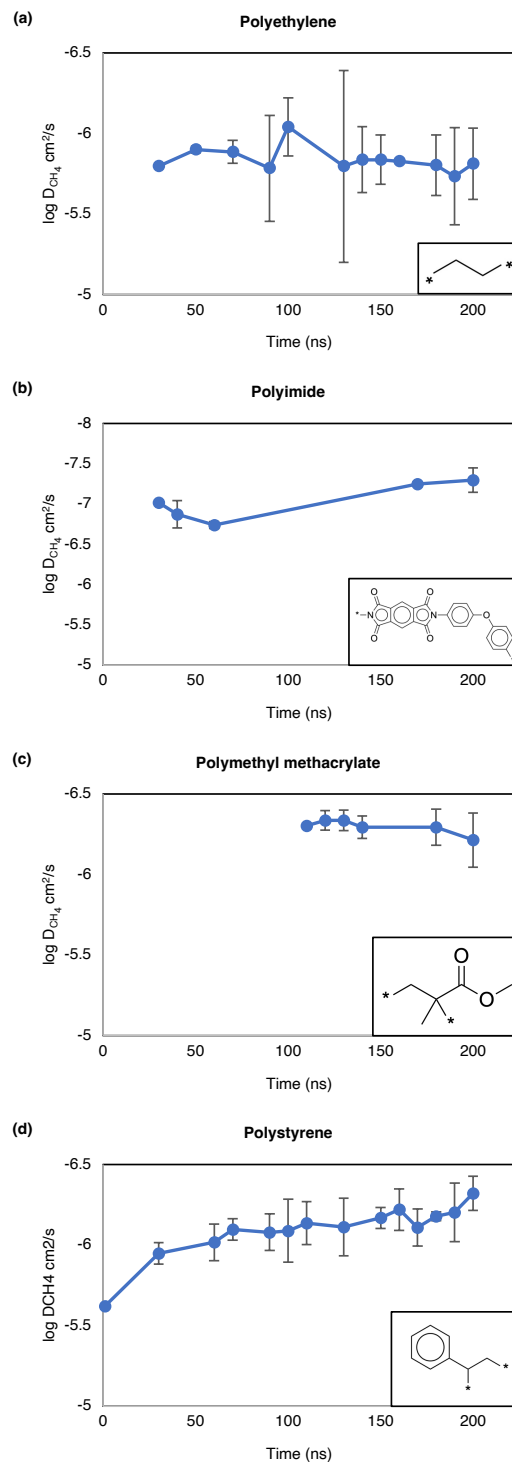


Figure S1: **Simulation time vs methane diffusivity for (a) Polyethylene, (b) Polyimide, (c) Polystyrene, (d) Polymethyl methacrylate.** The log-log slopes of all the systems' MSD curves are in the range of 0.95-1.05. Our investigations revealed that running simulations in the range of 100–200ns tends to yield stable diffusivity results. Polymer structures are also illustrated.

Table S1: **polyGNN training parameters.** Parameters used to train benchmark and production polyGNN models.

Model parameters	Values
RANDOM_SEED	100
HP_EPOCHS	200
SUBMODEL_EPOCHS	1000
N_FOLDS	5
HP_NCALLS	25
MAX_BATCH_SIZE	300
capacity_ls	2 to 14
weight_decay	0
N_PASSES	10

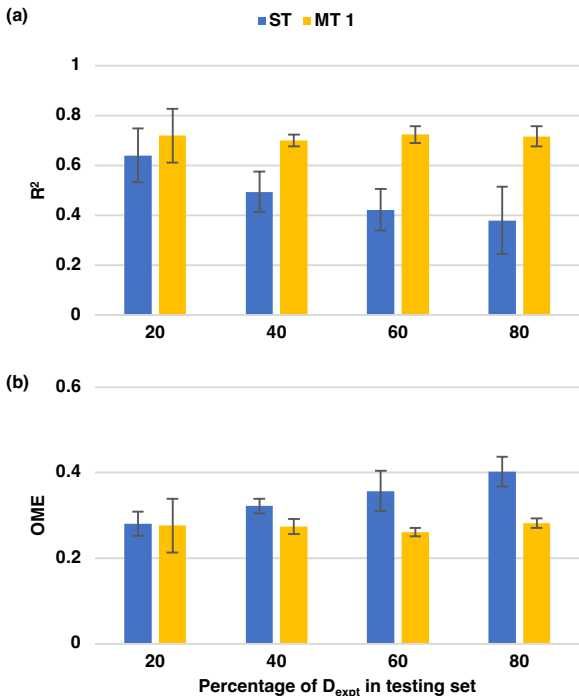


Figure S2: **Predictions of Diffusivity at various train test splits (a) Coefficient of determination (R^2) and (b) order of magnitude error (OME)**, comparisons between the ST and MT models as a function of percentage (%) of the unseen test set. The different unseen test set sizes demonstrate the effect of having less and less data in the training set. At 80%, the model only sees 20% of the dataset and is tested on the remaining 80%. This 80% replicates a data-scarce region, where there is less chemical coverage. A comprehensive improvement is seen in MT over ST models, especially at higher unseen test sets %.

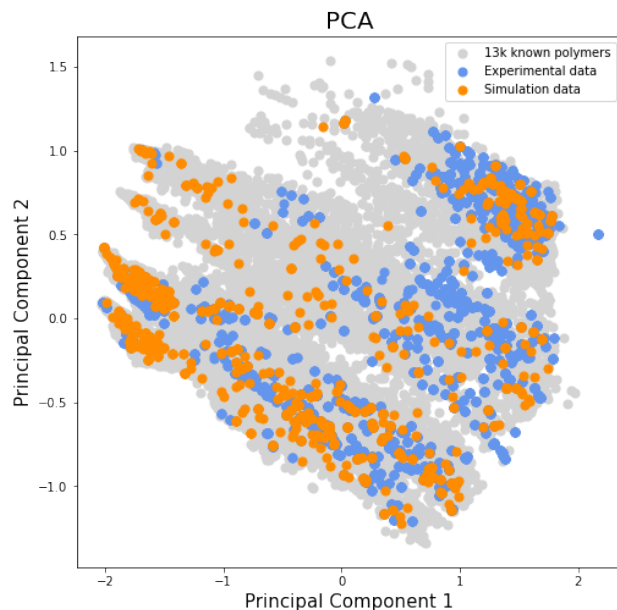


Figure S3: **Principal Component Analysis (PCA) plot.** The PCA plot demonstrates the chemical coverage of our simulation data in comparison to experimental gas transport data and 13,000 known polymers in our database. The blue and orange dots correspond to experimental and simulation data respectively. The grey dots represent the 13,000 known polymers in our database.

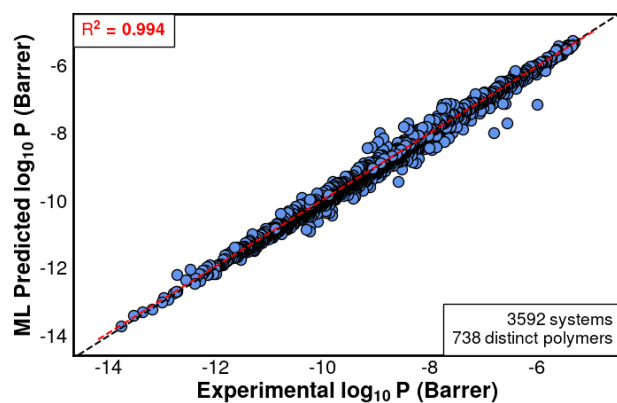


Figure S4: **Train Parity Plot.** Comparison of predicted and experimental permeability values in train set using the MT production model.

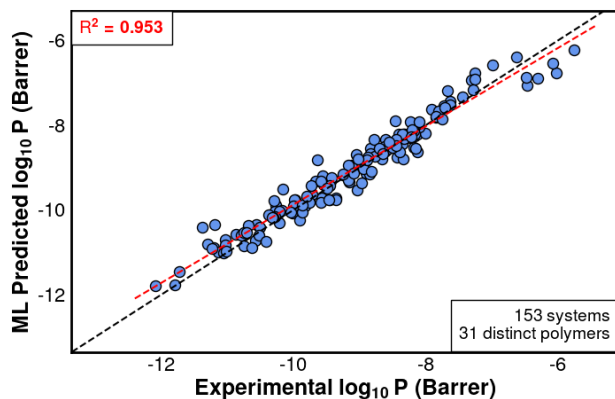


Figure S5: **Test Parity Plot.** Comparison of predicted and experimental permeability values in the test set using the MT production model.

Table S2: **Data used for benchmark test.** An overview of the polymer classes, the number of data points, and the performance metrics used in the comparison of the original and new model.

Polymer Class	Polymer Name	Datapoints	R ²
Conjugated Polymers	Poly(phenylene vinylene)	4	0.99
Polycarbonates	Hexafluoro Bis A polycarbonate	6	0.98
Polyarylates	TBHFBA-I PAR TMBPA-tBI polyarylate	12	0.97
Polynorbornenes	PB3FMPNDI	5	0.96
Polyimides and Polypyrrolones	6FDA-6FpDA PI TADATO-TMBZ PI 6FDA-4BDAF BTDA-6FpDA PMDA-3,3'-ODA PI DSDA-DDBT PI BPDA-BAPP PI 6FDA-DMDMA PI 6FDA/PPTI-4 PI HQDPA-TMDMA1 PEI BPI-BTDA PI	53	0.99
Vinyl and Vinylidene Polymers	Poly(vinyl acetate) Poly(p-trimethylsilyl styrene)	9	0.96
Poly(aryl ethers) and Poly(aryl ether ketones)	6H-PEN poly(aryl ether) Poly(aryl ether ketone) TMPEK-C	10	0.98
Polyphosphazenes	Polyphosphazene 4	5	0.90
Polysulfones	DMPSF	5	0.98
Polyamides and poly(amide- imides)	APA1 aromatic polyamide	5	0.99
Parylenes	Poly(methyl-p-xylylene)	4	0.98
High Temperature Polymers	PIM-PI-7 Poly(1,2,4-triazole)	10	0.99
Polypropynes, substituted polyacetylenes, polypentynes	Polyacetylene Poly[1-(n-hexyllthio)-1-propyne] Fluorenyl-poly(diarylacetylene) Poly(diphenyl acetylene)Poly(2-nonyne)	26	0.87

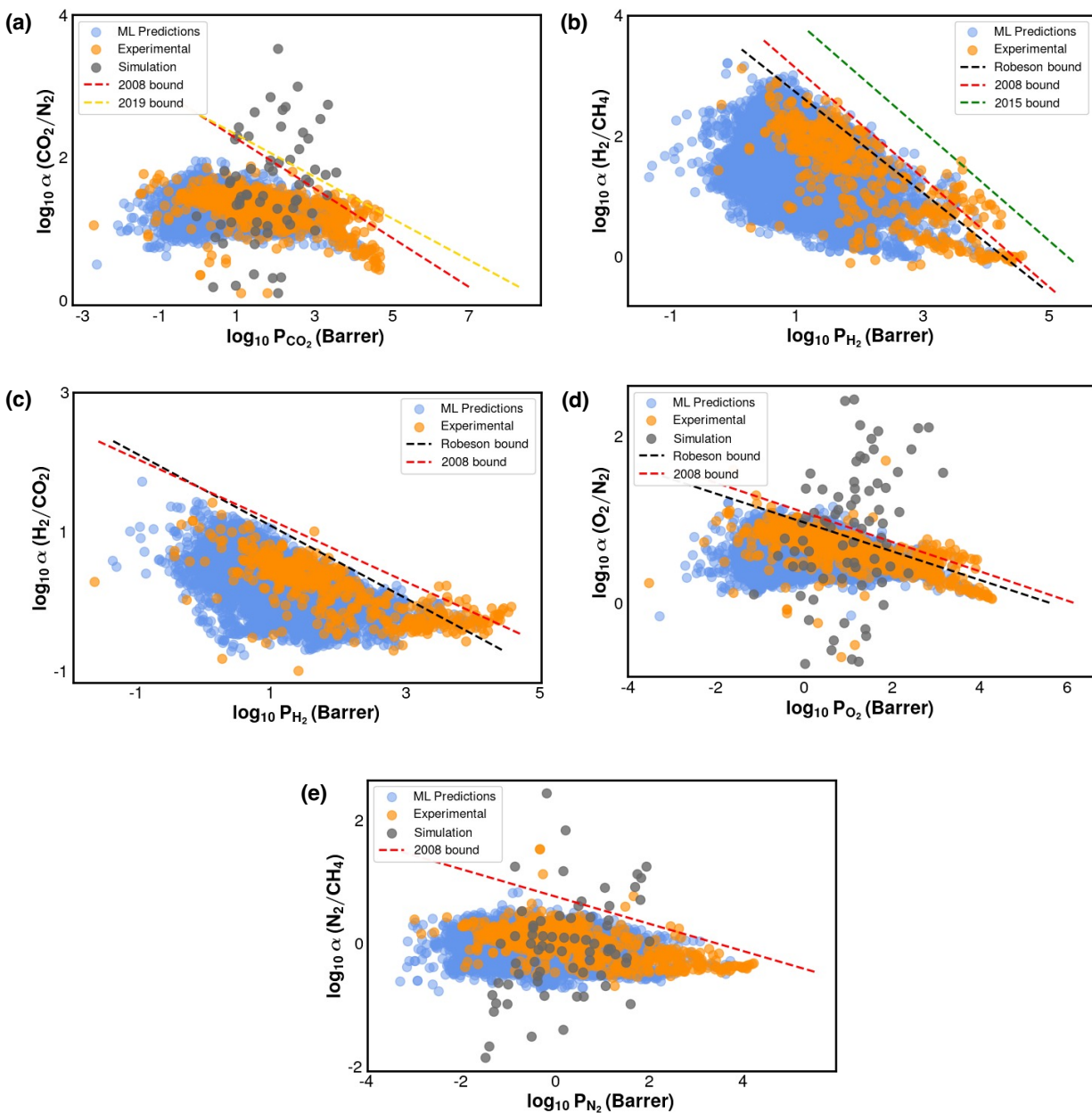


Figure S6: **Gas permeability trade-off plots for (a) CO_2/N_2 , (b) H_2/CH_4 , (c) H_2/CO_2 , (d) O_2/N_2 and (e) N_2/CH_4** Polymer upper bound with predicted and experimental gas permeability and selectivity data points. The orange and blue dots correspond to the experimental data and machine learning predictions, respectively, while the grey dots represent simulated values. The original Robeson upper bound (1991) is depicted as the dashed red line, with reevaluated bounds from 2008, 2015, and 2019 shown as red, green, and yellow dashed lines, respectively, for gas permeability.

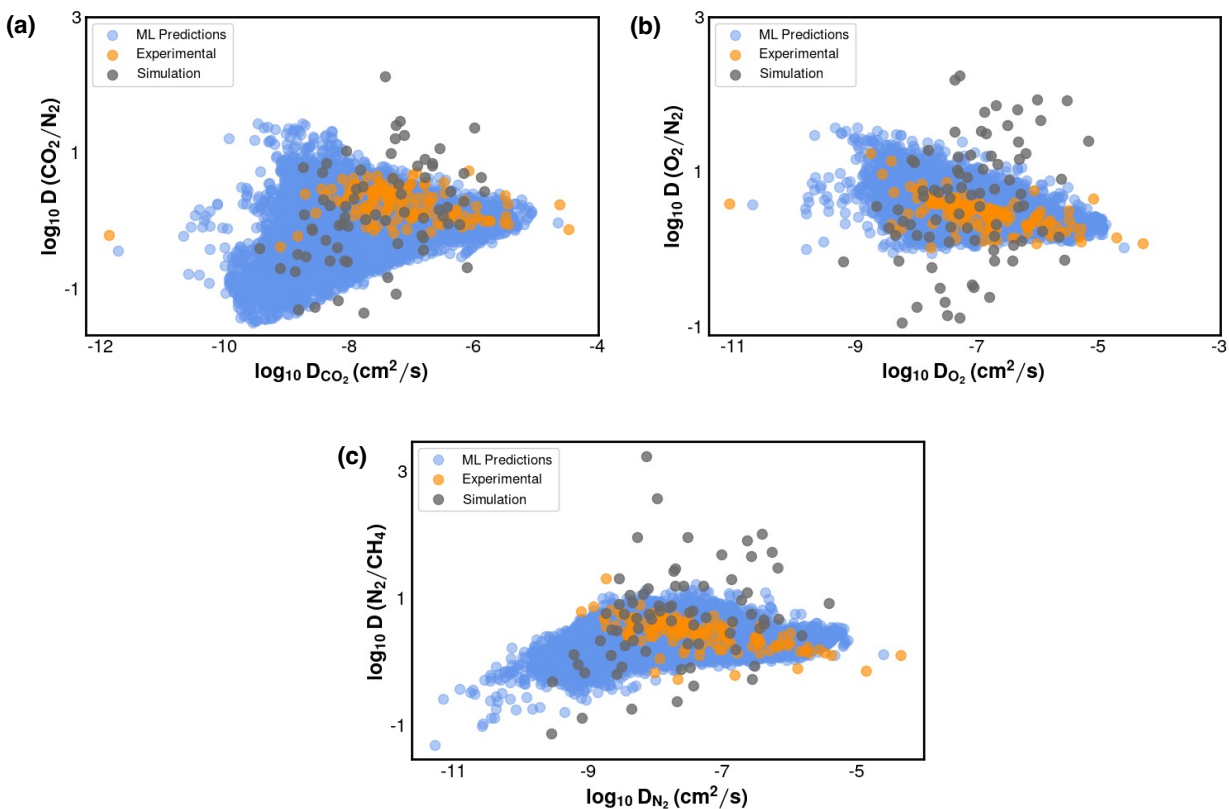


Figure S7: **Gas diffusivity trade-off plots for (a) CO₂/N₂, (b) O₂/N₂ and (c) N₂/CH₄.** Polymer trade-off plots with predicted and experimental gas diffusivity and selectivity data points. The orange and blue dots correspond to the experimental data and machine learning predictions, respectively, while the grey dots represent simulated values.

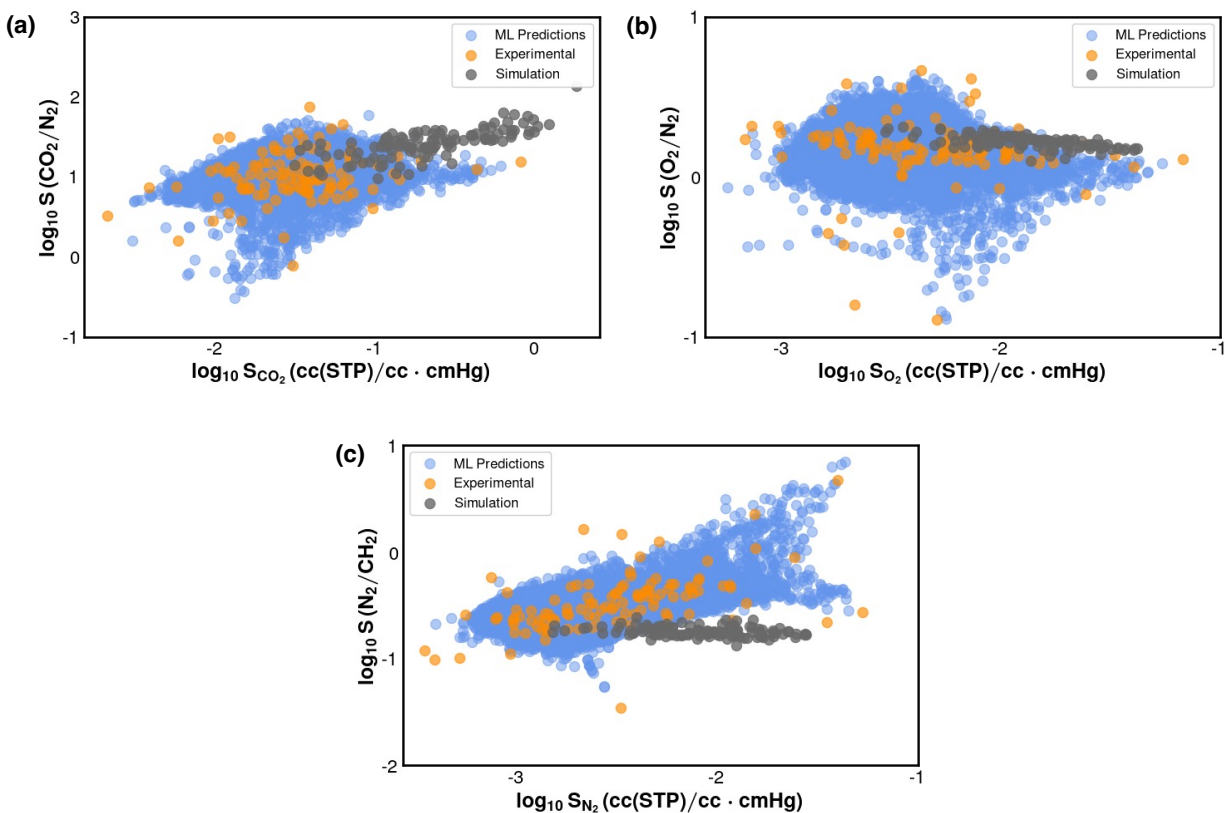


Figure S8: **Gas solubility trade-off plots for (a) CO₂/N₂, (b) O₂/N₂ and (c) N₂/CH₄.** Polymer trade-off plots with predicted and experimental gas solubility and selectivity data points. The orange and blue dots correspond to the experimental data and machine learning predictions, respectively, while the grey dots represent simulated values.