THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

Improved Scaling of Molecular Network Calculations: The Emergence of Molecular Domains

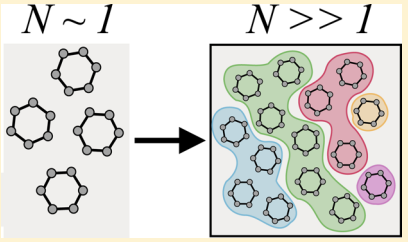
Adam G. Gagorik,[†] Brett Savoie,[†] Nick Jackson,[†] Ankit Agrawal,[‡] Alok Choudhary,[‡] Mark A. Ratner,[†] George C. Schatz,[†] and Kevin L. Kohlstedt*[,†]

[†]Department of Chemistry, Northwestern University, Evanston, Illinois 60208, United States

[‡]Department of Electrical Engineering and Computer Science, Northwestern University, Evanston Illinois 60208, United States

Supporting Information

ABSTRACT: The design of materials needed for the storage, delivery, and conversion of (re)useable energy is still hindered by the lack of new, hierarchical molecular screening methodologies that encode information on more than one length scale. Using a molecular network theory as a foundation, we show that to describe charge transport in disordered materials the network methodology must be scaled-up. We detail the scale-up through the use of adjacency lists and depth first search algorithms for during operations on the adjacency matrix. We consider two types of electronic acceptors, perylenediimide (PDI) and the fullerene derivative phenyl-C61-butyric acid methyl ester (PCBM), and we demonstrate that the method is scalable to length scales relevant to grain boundary and trap formations. Such boundaries lead to a decrease in the percolation ratio of PDI with system size, while the ratio for PCBM remains constant, further quantifying the stable, diverse transport pathways of PCBM and its success as a charge-accepting material.



The prospect of organic semiconductors replacing their inorganic counterparts, in a variety of applications, depends on a detailed understanding of charge transport on the mesoscale, while the chemical design of the organic moieties requires a molecular-level understanding. Rationalizing this dichotomy requires one to build a bridge from quantum-chemistry descriptors to the macroscopic observables of a functional semiconducting device. Recently, a molecular network theory has been developed that has successfully described charge-transport networks in a variety of organic materials.[1] Utilizing a graph theoretical approach incorporating intermolecular electronic couplings, a methodology for quantifying multimolecule charge transport networks in soft, disordered materials has been established using molecular couplings from a tight-binding Hamiltonian.[1] Conventional organic semiconducting molecules such as polyacenes and pyrenes showed promise as high electron mobility materials because of their capacity to strongly couple to adjacent molecules using π-orbital overlap, yet transport beyond the neighbor shell proved problematic. Next-generation derivative molecules such as rubrene,[2,3] TIPS pentacene,[4] and benzothiophene heterocyclic oligomers[5,6] rely less on just cofacial π-stacking due to their higher order molecular packing. However, multicrystalline interfaces such as domain boundaries that contain >1000 molecules still provide transport bottlenecks, such as recombination,[7] and theoretical methods must be held to account for them.[8]

Theoretical understanding of these devices is often bifurcated into distinct length scales. At one end (1−10 nm), single-molecule in vacuo studies probe HOMO/LUMO gaps and reorganization energy.[9,10] At the other end (10−1000 nm), bulk (device) studies probe packing, percolation, hopping transport, and device characteristics.[11,12] The size limitation of single-molecule studies and the lack of molecular details in bulk studies hinders our understanding and impedes rational design. Motivated by these challenges, we have previously developed a charge-transport model based on graph theory[1] that couples bulk sampling (molecular dynamics) with n-wise molecular couplings (tight binding Hamiltonian) to bridge the length-scale divide and to accurately measure the charge transport on length scales competitive with recombination.

Using the network methodology, we advance the model for intermolecular couplings toward the length scale of an entire active layer of an organic photovoltaic cell, thin-film transistor, or light-emitting diode, that is, ∼50 nm. This requires improving the method scaling, a process on which we have made significant progress. We show the effect of network fragmentation due to the emergence of (dis)order on larger length scales. This points to the importance of modeling larger system sizes to capture longer length scale phenomena, such as multicrystalline domains and the interfacial effects of ordered domains[13] and their effects on charge transport. We report the percolation ratio of two high-quality electron acceptors for system sizes where domain effects like domain boundaries become evident and discuss their disruptive behavior to charge transport. We connect the need for larger scale modeling to the desire to utilize the next-generation rational design methods of machine learning. Machine learning and data mining offer a

Received: December 13, 2016
Accepted: December 30, 2016
Published: December 30, 2016

DOI: 10.1021/acs.jpclett.6b02921
J. Phys. Chem. Lett. 2017, 8, 415−421

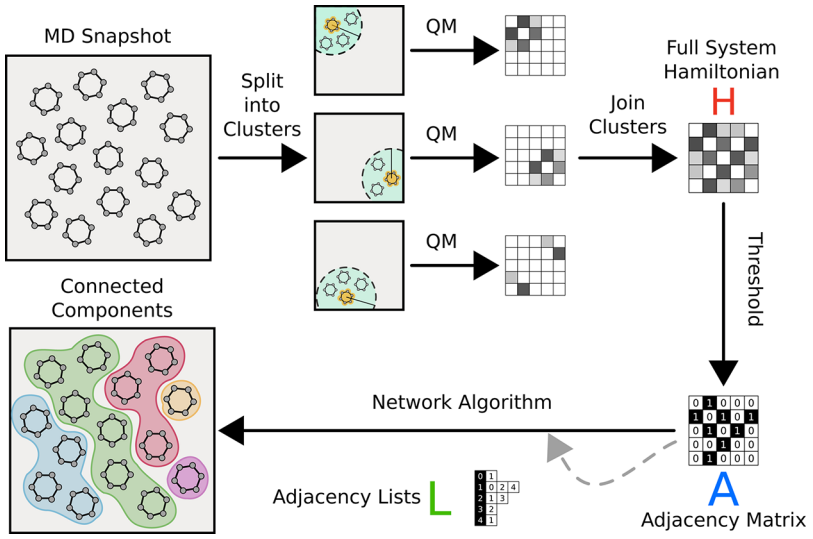


Figure 1. Overview of the molecular network theory method. Simulation snapshots of a molecular dynamics trajectory are partitioned into fragments based on a distance criterion. A quantum calculation is performed on each fragment. The fragments are joined together to produce the full-system Hamiltonian. A thresholding procedure is then used to transform the Hamiltonian into an adjacency matrix. This matrix is analyzed using graph theory. Optionally, adjacency lists can be used instead of the matrix for optimal calculation of graph properties.

promising route for achieving rational materials design.[14] In contrast with other data-driven methods such as rational design using cheminformatics,[15,16] where large training data sets must be manually constructed from literature, mesoscale chemical networks can rely on automated database generation as the macroscopic properties are predictive. Already, learning methods have been applied for inorganic[17] and crystalline[18] materials. We hope to use the network theory to enable similar progress in the area of (hybrid) organic electronics.

We consider two types of electron-accepting molecules, phenyl-C61-butyric acid methyl ester (PCBM) and a perylenediimide (PDI) with R = $-CH[(CH_2)_4CH_3]_2$. While both of these have been used with success in organic electronics, the charge networks of PDI and PCBM differ due to reduction in the possible coupling directions of PDI relative to PCBM. The core of PDI is perylene, a five-member fused benzene ring system that is planar. The addition of two imides at each end creates two more rings and allows the addition of solubilizing groups. Although solubilizing groups can stick out of the plane, charge transport is limited to the direction perpendicular to the plane, as π-stacking between adjacent PDIs leads to the most favorable coupling. This limits the number of coupling directions to two and leads to the terminology that PDI is a 1D material. In contrast, the PCBM core is a spherical fused ring system known as buckminsterfullerene, which can couple in more than two directions. Density functional and semiempirical methods show that PCBM can form one-, two-, or three-dimensional percolation networks, with higher energy virtual orbitals often participating in electron transfer.[19] For example, PCBM can couple to as many as six neighbors in the strong coupling regime.

We have made progress in scaling the network methodology from small systems of 64 PCBM or PDI molecules to as many as 256 molecules. This is accomplished through the use of sparse data structures for a Hamiltonian and the storage of hierarchical couplings in the adjacency matrix. The adjacency matrix is used to represent the molecular system in graph theoretical terms and, when expressed as a sparse adjacency list, allows for scaling of the methodology to millions of graph nodes. Likewise, to perform a calculation on a Hamiltonian containing couplings for thousands of basis functions, a sparse matrix data structure optimized for matrix-vector products must be used. Below, we describe the methodology and discuss the bottlenecks present in the Hamiltonian calculation. We find that our choice of sparse matrix structure significantly speeds up the calculation, leading us to discover the steady decay in the percolation ratio (defined below) with system size for PDI, in contrast with PCBM. Crystalline domain boundaries, on length scales that form beyond systems sizes of ~50 molecules, are discussed as a explanation behind the trend. Finally, the time improvements made by considering the use of an adjacency list to represent the molecular graph are presented.

The network methodology involves a series of calculations that are shown in Figure 1. For each snapshot of a molecular dynamics trajectory, one must (1) calculate a molecular Hamiltonian, (2) produce an adjacency matrix via thresholding, and (3) compute graph theoretical descriptors. One can average the graph descriptors over the trajectory. We have chosen to compute the connected components, which represent the pathways through which charge transfer can occur in the system. Graph theory is the bridge between length scales, whose application is potentially independent of the dynamics and quantum methods chosen. The application of graph theory has previously found success in many chemistry subdisciplines,[20,21] for example, in molecular connectivity[22] and morphology characterization.[23]

The molecular dynamics simulation is straightforward, using the optimized potentials for liquid simulations (OPLS) force field. A box of molecules is first relaxed in the microcanonical ensemble (NVE) to remove any high-energy initial configurations. The relaxed molecules are then heated from 10 to 550 K over a 50 ps time frame, using a 1 fs time step, in the isothermal−isobaric ensemble (NPT) at 1 atm. After annealing the molecules at this temperature for 2 ns, we rapidly cool the system over 100 ps to 298 K. Finally, we sample the low-energy states every 100 ps for 10 ns. This creates 100 snapshots to average network properties over.

DOI: 10.1021/acs.jpclett.6b02921
J. Phys. Chem. Lett. 2017, 8, 415−421

To describe the graphical connectivities of topologically complex, disordered molecules, two data types have to be constructed: the molecular orbitals of the molecules and the adjacency matrix based on the couplings between molecular orbitals.[1] Methods for constructing electronic (molecular) orbitals for organic molecules are straightforward, yet to accurately simulate charge transport on the mesoscale, tens of thousands of calculated molecular orbitals must be stored dynamically throughout a simulation, presenting a practical data challenge. The scaling of the orbital data structure to the mesoscale is the first challenge we have started to address. In the past, we have used coarse-grained charge-hopping models and Monte Carlo lattice models. While these models are qualitatively descriptive, to move to quantitative descriptions there needs to be a scale-up in the capacity of molecular orbital sparse data structures.[13] Integration of sparse data operations into quantum-chemistry solvers is key to achieve the scale in orbital construction on the mesoscale. Complementary approaches in classical molecular simulations have been utilized for the description of sparsely populated (rare-event) properties of oligomeric soft materials, especially using high-throughput computational resources like parallel tempering molecular dynamics. As an example, simulated annealing in MD has often been utilized to simulate configurations that are sparsely distributed in highly disordered biomolecular systems, such as biopolymer (peptide or oligonucleotide) assembly,[24,25] and such methods could be adapted to construct the large sparsely occupied orbital data sets.

We use semiempirical extended Hückel theory to compute transfer integrals between all nearest-neighbor molecules in the system, creating a Hamiltonian of the aggregate. Details of extended Hückel can be found elsewhere.[26,27] As PDI and PCBM are used as electron acceptors, we consider the LUMO as the relevant charge-transport state. The basis set uses Slater Type Orbitals (STOs), which represent the valence electrons and approximate hydrogenic atomic orbitals centered on atoms in the system. Considering the valence orbitals on C, N, O, and H, there are 310 atomic orbitals on a single PCBM and 262 on PDI. Given a box of 32 to 256 molecules, the Hamiltonian will have 9920 to 79360 basis functions for PCBM and 8384 to 67072 basis functions for PDI. We utilize a semiempirical Hamiltonian due to the size of the system and its ability to scale to larger systems. Even so, in systems as small as 64 molecules, the orbital overlap calculation is still too expensive. Instead, a simulation snapshot is partitioned into clusters based on a distance criterion (<2 nm). A coupling calculation is then performed on each cluster. The justification is that molecules beyond a certain distance will have no orbital overlap and thus contribute nothing to the overlap matrix or Hamiltonian. The fragment calculations are then joined together to produce the full-system Hamiltonian.

The tight-binding Hamiltonian takes the form of eq 1, where ϕ_i is the ith basis function, ϵ_i is its eigenvalue, v_{ij} is the coupling between the ith and jth basis functions, and N is the total number of basis functions.

$$H = \sum_{i}^{N} \epsilon_i |\phi_i\rangle\langle\phi_i| + \sum_{ij}^{N} v_{ij} |\phi_i\rangle\langle\phi_j| \tag{1}$$

The form is similar to an adjacency matrix. An adjacency matrix is used to represent a graph or network. If two vertices v_i and v_j are connected, the corresponding entry in the adjacency matrix A_{ij} is set to one. If no such edge exists, the element A_{ij} is set to zero. The adjacency matrix of an undirected simple graph is symmetric; that is, the element $A_{ij} = A_{ji}$ and is redundant. Using an adjacency *list*, one can simplify the representation by only storing lists of vertices for every vertex v_i. While the redundancy is still present in such a representation, the explicit zeros in the adjacency matrix are not stored. More importantly, the question of which vertices share edges with a given vertex v_i changes from a $O(V)$ operation to a constant $O(1)$ operation. Such a question is crucial to labeling the connected components in the network algorithm.

To create the adjacency matrix (or list), a thresholding procedure is used on the Hamiltonian of eq 1. As shown in eq 2, elements $|H_{ij}| \geq v_t$ are set to 1 and elements $|H_{ij}| < v_t$ are set to 0. The value of v_t can be varied but typically is ∼5 meV. If elements of the Hamiltonian represent molecular states, this means significant coupling between molecules i and j and thus a charge- transfer pathway. A higher value of v_t corresponds to the strong coupling regime and leads to more edges, larger connected components, and percolative charge transfer.

$$\forall\, i, j \in [0, N),\ (|H_{ij}| < v_t \rightarrow A_{ij} = 0) \wedge (|H_{ij}| \geq v_t \rightarrow A_{ij} = 1) \tag{2}$$

Finally, the adjacency matrix is analyzed using graph theory. One key metric of interest to charge transfer theory is the presence of connected components in the graph. Connected components are subgraphs such that any vertex in the subgraph can be reached via traversal of edges within the subgraph *and* no vertex outside the subgraph can be reached. The idea is that charge transfer occurs between molecules with a baseline coupling, represented by the threshold value discussed above. This translates to the graph edges representing possible pathways for charge transfer, and a lack of connected components spanning the simulation box, and more so a semiconductor device active layer, leads to poor electron transport. The scalability of the method is crucial if rational design methods are to be deployed.

We now discuss the present limitations in the extended Hückel calculation, for which we have used the YAeHMOP package to perform. For a single molecular dynamics trajectory snapshot, tens of thousands of molecular STO orbitals must be stored dynamically throughout a simulation. Typically, one averages over 100 snapshots. This poses three challenges to the calculation. The first challenge is the size of the system. Even for a single snapshot, the calculation is impractical. Fortunately one can fragment the system into molecular components based on a distance criterion. The reasoning here is that molecular units outside a certain distance cutoff will have little to no orbital overlap. One can thus calculate the Hamiltonian and overlap matrix of the clusters and stitch them back together in a highly parallelized fashion.

The second challenge is that fragmenting the system becomes nontrivial as the system size increases. While the number of basis functions used to represent each fragment Hamiltonian remains small, the full-system Hamiltonian is a sparse matrix because distant molecules do not have significant interaction. The stitching of fragments together benefits from a List of Lists (LIL) sparse matrix representation. The LIL form is a row-based linked list of nonzero elements and is efficient for constructing sparse matrices incrementally. However, this form does not perform well when computing matrix multiplication or matrix-vector products. Stitching fragments using the LIL form, then switching to a compressed sparse row (CSR) form greatly improves performance. The CSR form

DOI: 10.1021/acs.jpclett.6b02921
J. Phys. Chem. Lett. 2017, 8, 415−421

allows fast matrix-vector products at the expense of changes in sparsity structure.[28] The benefit of switching matrix representation far outweighs the cost. Figure 2 shows the timing for a single snapshot calculation of systems of increasing size before and after this optimization. For example, for a single snapshot of a PDI trajectory with 256 molecules, the calculation time is reduced from 45.8 to 2.81 h.

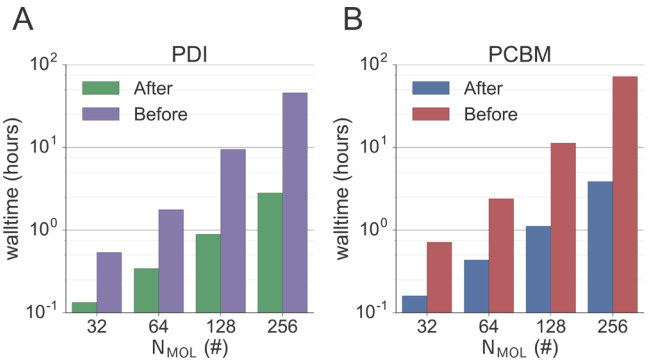


Figure 2. Walltime (hours) of the network code versus system size (number of molecules) before and after sparse matrix optimization for (A) PDI and (B) PCBM. The List of Lists (LIL, before) representation vastly slows down the calculation compared with the compressed structured row (CSR, after) representation. Although PCBM has fewer basis functions than PDI, calculations on PCBM take longer because there are more intramolecular interactions relative to PDI (i.e., the PCBM Hamiltonian has more nonzero elements).

The final challenge is the interface with the extended-Hückel calculation itself must be altered to scale up the fragmented $v_{ij}|\phi_i\rangle\langle\phi_j|$ molecular coupling calculations. Sparse matrix Hamiltonian and overlap integrals for each molecular fragment must be calculated as an intermediate calculation. For large molecules or large system sizes, it becomes intractable to save the data as an intermediate calculation. The number of fragments scales as N^2, where N is the number of molecules in the system. Additionally, as the number of orbitals increases, the size of even a single fragment can approach hundreds of GB. Fortunately, this is only an interface problem with the network calculation, and on-the-fly calculations of the overlap integrals will remove the bottleneck. Implementing the atomic overlap calculations is a work in progress, and it will eliminate the large amount of overhead in the coupling calculations.

With this implementation, we can observe the nature of networks in the larger systems. Figure 3A shows how the size of the largest network R_{NET}, measured by the radius of gyration R_G^{NET} of the largest network formed, scales as a function of the number of molecules N_{MOL}. We scale R_{NET} by the system size $R_{SYS} \equiv R_G^{TOT}$ to show the percolation ratio (R_{NET}/R_{SYS}) of the networks. The data are distributed over a set of coupling thresholds v_t for $1 \geq v_t \geq 50$ meV (blue for PCBM and green for PDI) for each system size N_{MOL}. The percolation ratio varies between 0 and 1 and is the fraction of the system of which is the largest charge network.[1] PDI shows a steady decrease in percolation ratio as the system size increases, in contrast with PCBM, which remains relatively constant over the range of system sizes up to $N_{MOL} = 256$. We attribute the decay in the percolation ratio for larger system sizes in Figure 3A to a competition between growing cofacial stacks of PDI, leading to stacking faults. At larger system sizes we had originally expected the that R_{NET} would either slightly increase or remain constant due to the minimization of the "surface effect", where network

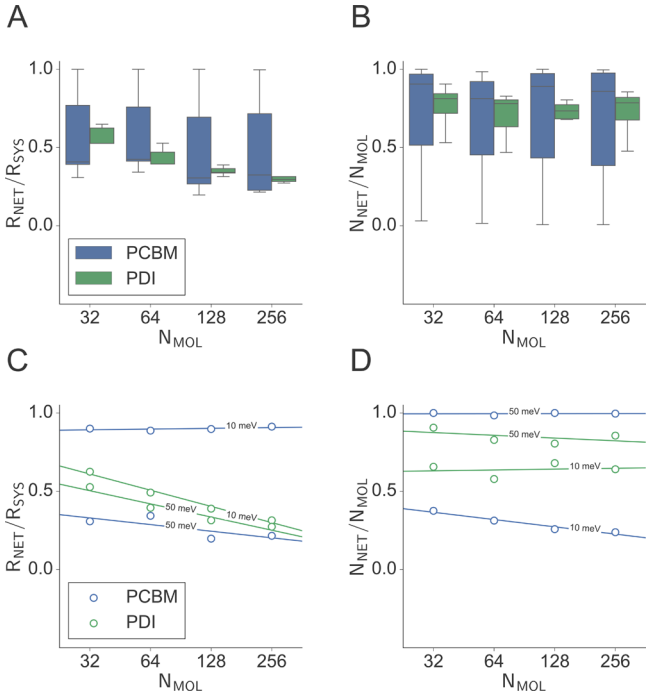


Figure 3. (A) Box plots of the percolation ratio R_{NET}/R_{SYS}, distributed over the coupling threshold values $1 \geq v_t \geq 50$ meV. Box plots are used as a guide for the eye on how the data points are distributed (shaded boxes show inner quartile and whiskers show outer quartiles). Distributions are shown for PCBM (blue) and for PDI (green) as well as for each system size N_{MOL}. (B) Box plots of the reduced number of networks N_{NET}/N_{MOL}. (C) Subset of R_{NET}/R_{SYS} and (D) subset of N_{NET}/N_{MOL} for low coupling ($v_t = 10$ meV) and high coupling ($v_t = 50$ meV).

fragmentation often starts at the surface of the molecular aggregate. The unexpected decrease in R_{NET} for PDI points to the importance of larger system sizes, where larger scale morphologies like domain boundaries can come into play, and their effect on charge transport is not negligible.[13]

To further show that this effect is not due to the fact that there are more molecules in the cluster and therefore the decay in the percolation ratio is just a scaling artifact due to the fact that there are additional (small) networks in the cluster, we plot in Figure 3B the number of total networks in the system N_{NET}. Again, the y axis in Figure 3B is scaled by the total system size, in this case the number of molecules N_{MOL}, to present the data as a ratio and to restrict it between 0 and 1. The plot is over the coupling threshold values $1 \geq v_t \geq 50$ meV. Note that a ratio of 1 would mean that the cluster is entirely fragmented and each molecule is its own network. It is found that the ratio of large networks does not significantly change for PDI (or PCBM) for larger system sizes. This is interpreted to mean that the decay of R_{NET}/R_{SYS} is due to the emergence of molecular order on another length scale, which we attribute to domain boundaries. We illustrate the role system size has in the coupled networks and the emergence of domain boundaries in PDI aggregates in the Supporting Information (see Figures S1 and S2). A detailed (zoomed) view at PDI domains is shown in Figure S3. The effect is highlighted for the modest threshold of $v_t = 10$ meV (Figure 3C,D), a threshold at which charge transfer occurs on time scales of $\tau \approx 5-7$ ps.[1] The decay of the percolation ratio for increasing N_{MOL} is evident even when the threshold for fragmentation is small. Spherically symmetric molecular

DOI: 10.1021/acs.jpclett.6b02921
J. Phys. Chem. Lett. 2017, 8, 415−421

topologies like PCBM may also have multicrystalline domains, but at these system sizes they are not evident. This indicates that PCBM forms robust percolation networks, while PDI does not. One should note that the number of networks, N_{NET}, and the percolation ratio consider only the membership of molecules in the respective networks; the number of connected neighbors per molecule is not directly reflected in the largest network size or total N_{NET}, yet fragmentation occurs when the average number of connections falls below 2, and that is the maximum number of connections for PDI, so the percolation ratio for PDI rigorously shows the fragmentation of the networks.

Another obstacle to scaling up the molecular network methodology is performing operations on the adjacency matrix. Under certain conditions, the adjacency matrix can be constructed directly from the molecular orbitals in a tight-binding (one-electron) Hamiltonian. By operating on the adjacency matrix, graphical properties such as the molecular connectivity,[29] percolative networks,[30,31] and network fragility can be found. Operations on adjacency matrices to find (sub)networks involve an upper limit on the order of hundreds of molecules. High-throughput techniques have been devised to operate on complex networks with thousands of nodes.[31,32] Utilizing these techniques will allow for the scale-up of the network descriptors in molecular systems. Additionally, alternative data structures, such as adjacency lists, need to be incorporated.

The original implementation of the network generation code, which determines the connected components in the graph scales as $O(V^3)$, where V is the number nodes in the graph.[1] The number of nodes in the graph is determined by the number of molecules in the system. If one chooses to represent the system Hamiltonian on a molecular basis, then V represents the number of molecules exactly, and the original scaling can be coped with, even for systems as large as 10^3 molecules. However, when attempting a simulation of an entire active layer of a semiconducting device, the number of molecules could easily approach 10^5 for nanoscale devices. Likewise, considering an atomic Hamiltonian, where states reside on atom centers instead of coarse-graining of states to the molecules, the number of states will increase significantly. Thus for a PCBM box with 128 molecules containing 39 680 STO basis functions, sparse storage and optimal graph techniques must be employed.

Just like a sparse matrix improves the Hamiltonian calculation performance, changing the graph data structure from an adjacency matrix to an adjacency list improves the running time of the algorithm. The adjacency list saves memory by not storing explicit zeros. Figure 4 compares the running time of connected component discovery versus network size for various implementations and the two graph data structures. It is found that adjacency lists with a recursive depth-first-search (DFS) implementation outperform the original implementation with the adjacency matrix.

DFS is a well known method for graph traversal. The complexity of DFS depends on the data structure used. While an adjacency matrix leads to $O(V^2)$, using an adjacency list leads to $O(V + E)$, with E being the number of edges traversed. Note that, in the worst possible case, $E = V(V - 1)$; that is, every single molecule has a significant interaction with every other molecule in the system, an unlikely scenario. Therefore, in this worst possible case, DFS would scale as $O(V + V^*(V - 1)) = O(V^2)$. In a more likely scenario, each

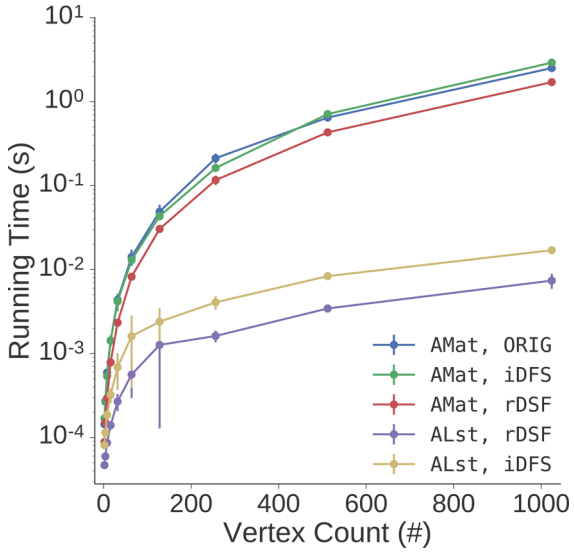


Figure 4. Running time (seconds) of the graph generation code versus system size (number of vertices). When using an adjacency matrix data structure, similar performance is observed for the original algorithm (AMat, ORIG, blue), the recursive depth first search algorithm (AMat, rDFS, red), and the iterative depth first search algorithm (AMat, iDFS, green). The performance increases drastically when an adjacency list data structure is used with iterative depth first search (ALst, iDFS, yellow). The best performance is found with the recursive form (ALst, rDFS, purple).

molecule would interact with a constant number of neighboring molecules. In this case, DFS would scale as $O(V + d^*V) = O(V)$, where d is the degree of each vertex, that is, the number of neighboring molecules with significant interaction. PCBM shows, on average, around six neighbors in the strong coupling regime, while PDI shows, on average, two neighbors, leading to different scaling prefactors. This optimization is expected to be very useful in large systems with highly coupled molecules similar to PCBM.

To summarize, we have previously applied the molecular network theory method to small systems, on the order of 64 molecules for the electron acceptors PCBM and PDI. By scaling up the methodology to reach length scales of 256 molecules, we observed the emergence of long length-scale domain boundary effects in the PDI system. This scalability was achieved by fragmenting the Hamiltonian calculation and taking care with the form of the sparse matrix representation. Additionally, the interface to the quantum results of any code must be streamlined to avoid large data files and redundant information. The network generation scaling was greatly improved from $O(V^3)$ to $O(V + E)$ by using an adjacency list and depth first search. As is the nature of disordered molecular systems, the significant coupling between molecules occurs only between neighbors, making $E = d * V$, with d being the average number of neighboring molecules. Molecules with deeply coupled charge networks will no longer suffer from poor scaling. We aim to target much larger systems in the future, on the order of a thousand molecules, to model the intermolecular couplings on the length scale of an entire active layer of a photovoltaic cell or organic light-emitting diode. This requires improving the method scaling, a process on which we have made significant progress via data structure optimization, as discussed in this paper. The use of sparse matrix data structures to optimize the quantum calculation and network generation

DOI: 10.1021/acs.jpclett.6b02921
J. Phys. Chem. Lett. 2017, 8, 415−421

has enabled us to study large systems, opening the door to true rational materials.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.6b02921.

Snapshots colored by network for PDI and PCBM systems at coupling threshold values of v_t = 1, 10, and 20 meV. A description of the network algorithm before using depth first search and a description of the network algorithm after using depth first search. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kkohlstedt@northwestern.edu.

ORCID

Kevin L. Kohlstedt: 0000-0001-8045-0930

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.G.G., K.L.K., and A.A. thank the Data Science Initiative (DSI) of Northwestern University for funding support and the Northwestern Institute of Complex Systems (NICO) for their support. The work and development of the electronic structure aspects of the work were supported by the Argonne-Northwestern Solar Energy Research (ANSER) Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0001059. The material informatics and scaling-up work was supported by the Data Science Initiative (DSI) of Northwestern University.

■ REFERENCES

(1) Savoie, B. M.; Kohlstedt, K. L.; Jackson, N. E.; Chen, L. X.; Olvera de la Cruz, M.; Schatz, G. C.; Marks, T. J.; Ratner, M. A. Mesoscale Molecular Network Formation in Amorphous Organic Materials. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10055−10060.

(2) Da Silva Filho, D. A.; Kim, E. G.; Brédas, J. L. Transport Properties in the Rubrene Crystal: Electronic Coupling and Vibrational Reorganization Energy. *Adv. Mater.* **2005**, *17*, 1072−1076.

(3) McGarry, K. A.; Xie, W.; Sutton, C.; Risko, C.; Wu, Y.; Young, V. G.; Bredas, J.-L.; Frisbie, C. D.; Douglas, C. J. Rubrene-Based Single-Crystal Organic Semiconductors: Synthesis, Electronic Structure, and Charge-Transport Properties. *Chem. Mater.* **2013**, *25*, 2254−2263.

(4) Eggeman, A.; Illig, S.; Troisi, A.; Sirringhaus, H.; Midgley, P. Measurement of Molecular Motion in Organic Semiconductors by Thermal Diffuse Electron Scattering. *Nat. Mater.* **2013**, *12*, 1045−1049.

(5) Takimiya, K.; Shinamura, S.; Osaka, I.; Miyazaki, E. Thienoacene-Based Organic Semiconductors. *Adv. Mater.* **2011**, *23*, 4347−4370.

(6) Illig, S.; Eggeman, A. S.; Troisi, A.; Jiang, L.; Warwick, C.; Nikolka, M.; Schweicher, G.; Yeates, S. G.; Henri Geerts, Y.; Anthony, J. E.; et al. Reducing Dynamic Disorder in Small-molecule Organic Semiconductors by Suppressing Large-amplitude Thermal Motions. *Nat. Commun.* **2016**, *7*, 10736.

(7) Jakowetz, A. C.; Böhm, M. L.; Zhang, J.; Sadhanala, A.; Huettner, S.; Bakulin, A. A.; Rao, A.; Friend, R. H. What Controls the Rate of Ultrafast Charge Transfer and Charge Separation Efficiency in Organic Photovoltaic Blends. *J. Am. Chem. Soc.* **2016**, *138*, 11672−11679.

(8) Pelzer, K. M.; Darling, S. B. Generation in Organic Photovoltaics: A Review of Theory, Charge and Computation. *Mol. Syst. Des. Eng.* **2016**, *1*, 10−24.

(9) Scharber, M. C.; Mühlbacher, D.; Koppe, M.; Denk, P.; Waldauf, C.; Heeger, A. J.; Brabec, C. J. Design Rules for Donors in Bulk-Heterojunction Solar Cells-Towards 10% Energy-Conversion Efficiency. *Adv. Mater.* **2006**, *18*, 789−794.

(10) Perez, M. D.; Borek, C.; Forrest, S. R.; Thompson, M. E. Molecular and Morphological Influences on the Open Circuit Voltages of Organic Photovoltaic Devices. *J. Am. Chem. Soc.* **2009**, *131*, 9281−9286.

(11) Savoie, B. M.; Movaghar, B.; Marks, T. J.; Ratner, M. A. Simple Analytic Description of Collection Efficiency in Organic Photovoltaics. *J. Phys. Chem. Lett.* **2013**, *4*, 704−709.

(12) Koster, L. J. A.; Smits, E. C. P.; Mihailetchi, V. D.; Blom, P. W. M. Device Model for the Operation of Polymer/Fullerene Bulk Heterojunction Solar Cells. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *72*, 085205.

(13) Jackson, N. E.; Kohlstedt, K. L.; Chen, L. X.; Ratner, M. A. A *n*-Vector Model for Charge Transport in Molecular Semiconductors. *J. Chem. Phys.* **2016**, *145*, 204102−10.

(14) Hill, J.; Mulholland, G.; Persson, K.; Seshadri, R.; Wolverton, C.; Meredig, B. Materials Science with Large-Scale Data and Informatics: Unlocking New Opportunities. *MRS Bull.* **2016**, *41*, 399−409.

(15) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* **2011**, *2*, 2241−2251.

(16) Ryu, S.; Adachi, I.; Aihara, H.; Asner, D. M.; Aulchenko, V.; Aushev, T.; Bakich, A. M.; Bala, A.; Bhuyan, B.; Bobrov, A.; et al. Measurements of Branching Fractions of τ Lepton Decays with One or More K_S^0. *Phys. Rev. D* **2014**, *89*, 072009.

(17) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *npj Comput. Mater.* **2016**, *2*, 16028.

(18) Furmanchuk, A.; Agrawal, A.; Choudhary, A. Predictive Analytics for Crystalline Materials: Bulk Modulus. *RSC Adv.* **2016**, *6*, 95246−95251.

(19) Ide, J.; Fazzi, D.; Casalegno, M.; Meille, S. V.; Raos, G. Electron Transport in Crystalline PCBM-like Fullerene Derivatives: A Comparative Computational Study. *J. Mater. Chem. C* **2014**, *2*, 7313−7325.

(20) Garcia-Domenech, R.; Galvez, J.; de Julian-Ortiz, J. V.; Pogliani, L. Some New Trends in Chemical Graph Theory. *Chem. Rev.* **2008**, *108*, 1127−1169.

(21) Balaban, A. T. Applications of Graph Theory in Chemistry. *J. Chem. Inf. Model.* **1985**, *25*, 334−343.

(22) Pogliani, L. From Molecular Connectivity Indices to Semiempirical Connectivity Terms: Recent Trends in Graph Theoretical Descriptors. *Chem. Rev.* **2000**, *100*, 3827−3858.

(23) Wodo, O.; Tirthapura, S.; Chaudhary, S.; Ganapathysubramanian, B. A Graph-based Formulation for Computational Characterization of Bulk Heterojunction Morphology. *Org. Electron.* **2012**, *13*, 1105−1113.

(24) Kohlstedt, K. L.; Olvera de la Cruz, M.; Schatz, G. C. Controlling Orientational Order in 1-D Assemblies of Multivalent Triangular Prisms. *J. Phys. Chem. Lett.* **2013**, *4*, 203−208.

(25) Fawzi, N. L.; Kohlstedt, K. L.; Okabe, Y.; Head-Gordon, T. Protofibril Assemblies of the Arctic, Dutch, and Flemish Mutants of the Alzheimer's $A\beta_{1-40}$ Peptide. *Biophys. J.* **2008**, *94*, 2007−2016.

(26) Guseinov, I. I.; Özmen, A.; Atav, Ü.; Yüksel, H. Computation of Overlap Integrals Over Slater-Type Orbitals Using Auxiliary Functions. *Int. J. Quantum Chem.* **1998**, *67*, 199−204.

(27) Guseinov, I.; Mamedov, B. Evaluation of Overlap Integrals with Integer and Noninteger N Slater-Type Orbitals Using Auxiliary Functions. *J. Mol. Model.* **2002**, *8*, 272−276.

(28) Shahnaz, R.; Usman, A.; Chughtai, I. R. Review of Storage Techniques for Sparse Matrices. *2005 Pakistan Section Multitopic Conference* **2005**, 1.

(29) Estrada, E. Physicochemical Interpretation of Molecular Connectivity Indices. *J. Phys. Chem. A* **2002**, *106*, 9085−9091.
(30) Li, J.; Ray, B.; Alam, M. A.; Östling, M. Threshold of Hierarchical Percolating Systems. *Phys. Rev. E* **2012**, *85*, 021109.
(31) Pattabiraman, B.; Patwary, M. M. A.; Gebremedhin, A. H.; Liao, W.-k.; Choudhary, A. *Fast Algorithms for the Maximum Clique Problem on Massive Sparse Graphs*; Springer International Publishing: Cambridge, MA, 2013.
(32) Guimerà, R.; Amaral, L. A. N. Cartography of Complex Networks: Modules and Universal Roles. *J. Stat. Mech.: Theory Exp.* **2005**, *2005*, P02001.

DOI: 10.1021/acs.jpclett.6b02921
J. Phys. Chem. Lett. 2017, 8, 415−421