

# Distilling Coarse-Grained Representations of Molecular Electronic Structure with Continuously Gated Message Passing

J. Charlie Maier,<sup>1</sup> Chun-I Wang,<sup>2</sup> and Nicholas E. Jackson<sup>2</sup>

<sup>1</sup>*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

<sup>2</sup>*Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

(\*Electronic mail: jacksonn@illinois.edu)

(Dated: 30 September 2023)

Bottom-up methods for coarse-grained (CG) molecular modeling are critically needed to establish rigorous links between atomistic reference data and reduced molecular representations. For a target molecule, the ideal reduced CG representation is a function of both the conformational ensemble of the system and the target physical observable(s) to be reproduced at the CG resolution. However, there is an absence of algorithms for selecting CG representations of molecules from which complex properties, including molecular electronic structure, can be accurately modeled. We introduce continuously gated message passing (CGMP), a graph neural network (GNN) method for atomically decomposing molecular electronic structure sampled over conformational ensembles. CGMP integrates 3D-invariant GNNs and a novel gated message passing system to continuously reduce the atomic degrees of freedom accessible for electronic predictions, resulting in a one-shot importance ranking of atoms contributing to a target molecular property. Moreover, CGMP provides the first approach by which to quantify the degeneracy of “good” CG representations conditioned on specific prediction targets, facilitating the development of more transferable CG representations. We further show how CGMP can be used to highlight multiatom correlations, illuminating a path to developing CG electronic Hamiltonians in terms of interpretable collective variables for arbitrarily complex molecules.

## I. INTRODUCTION

The identification of reduced representations of molecular systems is critical to extracting fundamental understanding across chemical space. Within this scope, coarse-grained (CG) modelling<sup>1–7</sup> forms an essential quantitative framework for deriving fundamental understanding of soft materials over mesoscopic spatiotemporal scales. In the process of averaging over or decimating superfluous molecular degrees of freedom, representations are simplified, interatomic interactions are smoothed out, and accessible timesteps for molecular dynamics (MD) simulations can be increased, greatly accelerating simulations. These advantageous features of CG modelling facilitate thermodynamic sampling that would otherwise require computationally intractable calculations, if performed at atomistic resolution.

While CG modelling is advantageous for long-time conformational sampling, the lower resolution of CG conformations introduces challenges for extracting the corresponding electronic structure of a system. In many applications, it is necessary to understand the electronic properties of disordered soft material morphologies,<sup>8</sup> which traditionally necessitates the reintroduction of all atomic degrees of freedom to enable quantum chemical (QC) methods. To accomplish this, the atomistic degrees of freedom lost in the coarsening process must be reconstructed and re-sampled in the original all-atom (AA) thermodynamic ensemble, conditioned on the specific CG conformation of interest. A variety of methods collectively termed “backmapping”<sup>9–18</sup> have been developed to address this problem, though the poor computational scaling and chemical transferability of this paradigm strongly limits its application to electronic property determination.

The incompatibility between reduced CG representations

and electronic structure analysis can be circumvented by identifying CG representations that preserve nuclear degrees of freedom critical to the targeted electronic prediction task. This approach reframes the challenge of selecting appropriate CG representations into one of identifying the atomic degrees of freedom that exhibit the strongest correlation with electronic structure variations over the full conformational ensemble. In this vein, quantitative approaches to CG representation identification have been explored using topology-based graph partitioning<sup>19,20</sup> and continuous representation optimization for topological<sup>21</sup> and thermodynamic properties.<sup>22</sup> While these results demonstrate the power of CG representation discovery, they lack extensibility to arbitrary molecular prediction targets (e.g. electronic structure) and cannot provide importance rankings for possible CG representations.

Machine learning (ML) has recently found purchase in computational chemistry.<sup>13,23–31</sup> Graph-based neural networks (GNNs)<sup>32–37</sup> in particular have enabled transferable model development across diverse chemistries and multiple prediction tasks by encoding known physical symmetries into the GNN architecture<sup>38–40</sup> and leveraging data enhancement techniques.<sup>41</sup> ML has also begun to enable CG representation identification for structural reproduction,<sup>22</sup> calculation of the mapping entropy,<sup>42–44</sup> and the matching of human-generated CG maps.<sup>45</sup> Such approaches aimed at identifying a minimal basis of molecular degrees of freedom are closely related to the emerging field of explainable artificial intelligence<sup>46–49</sup> that aims to explain the predictions of GNNs in terms of node, edge, and subgraph specific features. While the field is in its early phases, it is clear that such ML approaches hold tremendous promise for the future of CG modeling.

In this article, we develop a novel continuously gated message passing (CGMP) scheme for use with 3D-invariant

GNNs (Figure 1) to generate a hierarchy of CG representations of molecular graphs, optimized to specific molecular prediction targets. Within CGMP, the choice of whether or not to observe an atom in the GNN is defined by a differentiable model parameter referred to as the coarsening "score", rendering the GNN able to discriminate between important and unimportant sets of atoms for any combination of prediction tasks and resolution constraint. After detailing the CGMP method, we apply CGMP to extract the importance hierarchy of atomic degrees of freedom for a variety of electronic prediction tasks in a complex organic semiconductor molecule. First, through continuous lowering of the target resolution in the loss function, we measure the optimal information limit of arbitrary molecular observables for all possible CG resolutions. Second, over the course of training runs, we sample the inherent solution degeneracy of the coarsening problem and quantify the resolution dependence of this degeneracy. Finally, by calculating the covariance between atomic fluctuations in the final layer of the GNN, we investigate high-covariance links as a potential path to the development of CG electronic models expressed in terms of physically interpretable collective variables.

## II. METHODS

CGMP (Figure 1) uses a series of gates applied to a 3D-invariant GNN message-passing algorithm to exclude "unimportant" nodes (atoms) from contributing information to the GNN prediction. This message passing architecture is combined with a harmonic constraint on the total number of nodes (atoms) that may be used by the GNN, which is changed dynamically throughout training. The result is a one-shot training of a GNN on data sampled from a conformational ensemble that produces a hierarchy of atomic importances (and consequently CG representations) for all resolutions of interest. Importantly, CGMP is an atomic decimation style approach to coarse-graining as opposed to the commonly used averaging maps, in which CG beads are composed of linear combinations of atomic degrees of freedom.<sup>50</sup> In this work, we explore the use of CGMP by training on density functional theory (DFT)-derived variations in electronic structure from an MD-derived conformational ensemble, resulting in the production of a hierarchy of atomic importance rankings for a variety of electronic prediction tasks. The complete architecture of the GNN incorporating CGMP is illustrated in Figure 2.

CGMP employs a custom modification of the ComENet<sup>35</sup> 3D-invariant GNN architecture. ComENet defines a reference coordinate system for each graph node, based upon the directions to its nearest neighbors, then represents each graph edge using the interatomic distance, as well as the local angles and dihedrals with respect to the reference coordinates. This results in a representation which is still complete while maintaining invariance to global rotations and translations. Further mathematical details of ComENet and modifications appear in the Appendix.

### A. Node (Atom) Scores

The fundamental element of CGMP is its node-scoring method. For context, the node scoring method used in CGMP is distinct from common GNN pooling technique in two ways. First, atom scores act as a gate on the message-passing algorithm and directly exclude nodes from information aggregation at every layer of the GNN, simultaneously. This is in contrast to standard pooling approaches that generate a low-dimensional representation via projections of convolutions of higher-dimensional data. Second, pooling approaches operate using a discrete parameter to determine the pooled layer resolution. CGMP replaces this discrete parameter with a continuous resolution constraint using sparsity- and resolution-focused terms in the model loss function.

The purpose of a node score in CGMP is to decide whether or not to include an atom in the CG representation. If all scores are set to 1, then the GNN is able to compute over the full AA graph. When a node score is set to 0, it is removed from consideration, and unable to contribute to the prediction task (e.g. regression DFT-derived electronic structure). Then, when only  $M$  beads are allowed to have non-zero scores, the GNN is acting directly on the CG representation of the molecule. Importantly, our approach is constructed in such a way that the transition from AA to CG is entirely differentiable, enabling a continuous optimization process, and a one-shot rendering of a CG map for each CG resolution.

In CGMP there are two separate node-level features: the standard embedding vector and the score. The embedding vector  $\mathbf{x}_{i,1}$  is a learnable vector in the hidden space. This embedding vector is the initial representation for each atom, and shown at the top of the flow chart in Figure 2. The score parameter  $x_{i,2}$  is a single-valued parameter, and the score value is the sigmoid of this parameter

$$g_i = \sigma(x_{i,2}) \quad (1)$$

For any given training batch, we sample the node score values using injected noise to aid training. The noise is only used for calculating the output (regression) prediction; the score value passed to the regularization loss is not stochastic.

$$x_{i,2} = x_{i,2} + \lambda \varepsilon \quad (2)$$

$$\varepsilon \sim N(\mathbf{0}, \mathbf{1}) \quad (3)$$

Throughout the description of training,  $\varepsilon$  always refers to a normally distributed variable. Numbering within any equation is only to indicate i.i.d. (independent and identically distributed) samples. The dimension of the random variable is always that which matches the variable it is being added to. Since  $\lambda$  will be a repeated term throughout the methods, a list of all hyperparameter choices is provided in Table S1. The purpose of the injected noise is to create higher batch variance, and thus lower regressive performance, when scores are undecided, or near 0.5. When scores are near 0 or 1, the sigmoid curve flattens, and this noise has trivial impact. The noise also has the effect of lowering the barrier in the loss function, which encourages scores to select 0 or 1.

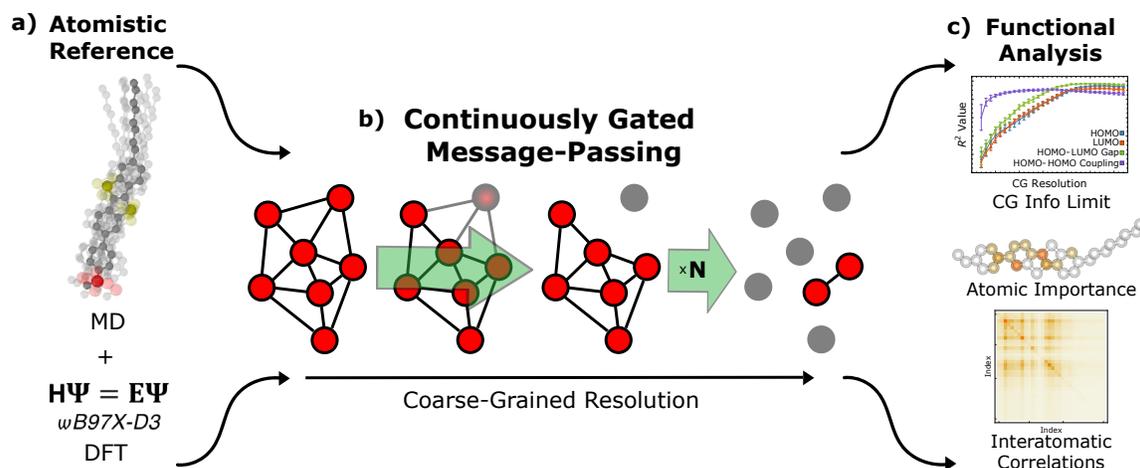


FIG. 1. Description of the computational procedure employing CGMP for the identification of optimal reduced molecular representations. a) A dataset is generated through a combination of atomistic MD simulations and DFT calculations on the sampled conformations. b) The GNN with CGMP is trained with a continuously decreasing target resolution to determine the optimal representation for all resolutions in one-shot. c) Interpretation and subsequent analysis of the atomic contributions to molecular electronic structure to inform CG representation identification and model Hamiltonian elucidation.

## B. Node Embeddings

The CGMP procedure is performed using a message-passing neural network (MPNN). In this network, the input to the MPNN,  $\mathbf{z}_i$ , is a combination of the embedding and score, with the tanh function bounding its range to disallow possible competition between itself and the scoring function.

$$\mathbf{z}_i^{(0)} = \tanh(\mathbf{x}_{i,1})g_i + \lambda_1 \epsilon_1(1 - g_i) + \lambda_2 \epsilon_2 \quad (4)$$

The first noise term in Eq. 4 means that the node score is used to interpolate the node vector from pure signal to pure noise as the node is coarsened out. This is one way in which the GNN can smoothly transition from inferring on the full molecule to subsets of atomic positions. The second noise term in Eq. 4 is an additional noise to mitigate overfitting.

## C. Graph Edges

The edges of the graph used for the GNN are derived from a distance-based graph using the atomic positions and a cut-off radius  $r_C$ . For our experiments, we choose  $r_C$  such that its length multiplied by the number of message-passing layers spans the molecular diameter. It is important to note that the graph edge feature vectors depend on a larger neighborhood than just two nodes. The reference coordinates for the calculation are based on neighbors of each node, which may or may not be included in the CG representation. If these nodes are not removed from the graph, then AA-level information can leak into the CG GNN. For example, an edge feature vector contains torsional information, which should not be able to be expressed with a CG representation that only contains two nodes. Therefore, the distance calculation used to determine whether an edge is formed is modified to include score values.

$$d(i, j) = \sqrt{\|\mathbf{r}_i - \mathbf{r}_j\|^2 + 2r_C((1 - g_i)^2 + (1 - g_j)^2)} \quad (5)$$

This metric virtually shifts the "off" nodes to be far apart from the "on" nodes, and thus keeps them from being included in the GNN once score values become sufficiently low. This metric further isolates each coarsened node and forces it to degree 0, unable to communicate, regardless of how informative any message could be.

## D. Message Passing

The GNN employs an edge-weighted message-passing scheme. The message is the hidden vector on the source node. The edge weight is a combination of two weights. The initial weight comes from the previously discussed graph featurization.

$$\mathbf{e}_{ij,k} = NN(v_k(i, j)) \quad (6)$$

The feedforward neural network,  $NN$ , is a learnable model component which maps the geometric information of the edge feature vector to hidden space. The second edge weight is a scalar gating function based on the score values for the nodes involved in the edge.

$$g_{ij} = (g_i g_j)^2 \quad (7)$$

The purpose of the nonlinearity in Eq. 7 is to separate the score domains where coarsening a node begins to have an effect on GNN communication and where edges begin to be removed from the graph. The MPNN update is then expressed as

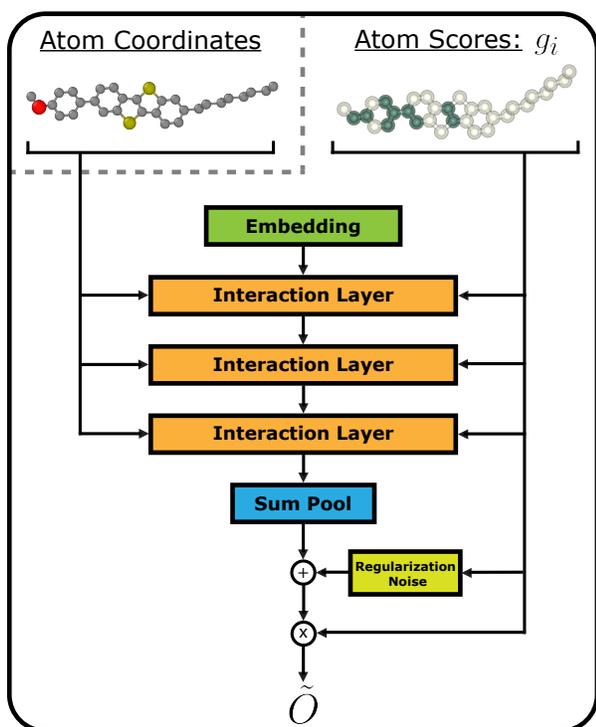


FIG. 2. The GNN architecture with score-based gated message passing used in CGMP. In contrast to pooling-based methods for representation projection, these per-node parameters control information flow at all layers of the network to ensure the GNN infers over a strict subset of atomic positions.

$$\mathbf{z}_{i,k}^{(l+1)} = \mathbf{W}_1 \mathbf{z}_i^{(l)} + \mathbf{W}_2 \sum_{j \in N(i)} \mathbf{z}_j^{(l)} \mathbf{e}_{ij,k} g_{ij} \quad (8)$$

$$\Delta \mathbf{z}_i^{(l+1)} = NN(\text{cat}(NN_1(\mathbf{z}_{i,1}^{(l+1)}), NN_2(\mathbf{z}_{i,2}^{(l+1)}))) \quad (9)$$

$$\mathbf{z}_i^{(l+1)} = \mathbf{z}_i^{(l)} + g_i \Delta \mathbf{z}_i^{(l+1)} + \lambda \varepsilon (1 - g_i) \quad (10)$$

Where  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are weight matrices,  $NN_1$  and  $NN_2$  are feedforward neural networks, with all involving trainable parameters of the full network. This update step, combining all aggregated messages, is referred to as the interaction layer in Figure 2. The final predicted energy is then generated by gated sum-pooling over the graph.

$$\tilde{O} = \sum_i (NN(\mathbf{z}_i) g_i + \lambda \varepsilon g_i (1 - g_i)) \quad (11)$$

There are multiple distinct noisy components being added to the message at every step. Each of these components is always designed to encourage decisive scores, and to have zero magnitude at the appropriate limits. In the discrete limit of this continuous model, all gates would take on binary values, and noise terms would go to 0.

## E. Loss Function

The CGMP loss function is comprised of a weighted sum of three functions described by

$$\mathcal{L} = \mathcal{L}_{MSE} + \lambda_\mu \mathcal{L}_\mu + \lambda_n \mathcal{L}_n \quad (12)$$

The first term ( $\mathcal{L}_{MSE}$ ) is the regression error, measured by the mean-squared error (MSE) associated with the property prediction task (here, DFT-derived electronic structure values). The second component ( $\mathcal{L}_\mu$ ) is a penalty on the batch mean of each node-level energy to prevent node contributions artificially drifting in energy space; training data with nonzero mean is accounted for with a learnable biasing parameter added to the output.

$$\mathcal{L}_\mu = \sum_i \langle E_i \rangle_{batch}^2 \quad (13)$$

The third component ( $\mathcal{L}_n$ ) is the node score term.  $\mathcal{L}_n$  is composed of local entropic loss terms on each node score that encourage binary score values, as well as a global harmonic constraint on the number of CG beads.

$$\mathcal{L}_n = \sum_i |(g_i - \Delta_L)(1 - \Delta_R - g_i)| + \lambda \left( \sum_i g_i - M \right)^2 \quad (14)$$

The activation function for the score function is sigmoid, which cannot reach 0 or 1 in a compact domain. This results in irrelevant nodes, with little coupling to other model parameters, quickly reaching extremely large values in early training before having to retread the same path when they are inevitably coarsened out of the graph. By shifting the right-hand minimum of the entropic loss by  $\Delta_R$  (set to 0.01 during our experiments), we keep the solution space for all resolutions within a bounded domain and allow for faster transitions between resolutions. Shifting the score for an irrelevant node only introduces minimal additional output noise in the final GNN layer, while shifting an important node produces larger effects in the message-passing layer. Therefore, some score separation between relevant and irrelevant nodes is introduced before any coarsening pressure is applied. We set  $\Delta_L$  to 0, as during our experiments we only reduce the target resolution.

Optimization of network parameters is performed with the Adam optimizer, which was designed in part to be robust to non-stationary objective functions.<sup>51</sup> Empirically, we observe that the ability of the network to rapidly adjust atomic scores to a changing target resolution is greatly diminished when using other optimizers such as SGD with momentum. Momentum-based optimization methods may carry over diminishing information from higher-resolution gradients into the network parameter training trajectories even after the resolution is lowered. However, this information cannot carry over into network predictions, which are constrained by the GNN's computational graph. The restraint-based dynamic loss function used in CGMP is a novel approach compared to top-k

techniques<sup>52–54</sup> which might form lower dimensional representations using internal network constraints. We found stability in our optimization method over a large range of hyperparameter choices and prediction tasks. Hydrogens are excluded from all molecular representations to simplify the coarsening procedure.

### III. RESULTS

#### A. Resolution Dependence of Electronic Prediction Accuracy

To demonstrate the performance of CGMP, we apply it to identify optimal CG representations of the liquid crystalline molecular semiconductor BTBT<sup>55</sup> for four distinct electronic prediction targets (encompassing 3 single-molecule and 1 pair-molecule electronic properties): the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energy levels, the HOMO-LUMO gap, and the HOMO-HOMO intermolecular electronic coupling between a dimer of BTBT molecules. BTBT was selected as a model system due to both its complex dependencies of electronic structure on molecular configuration and the availability of a high quality dataset for model training.<sup>56</sup> Model training, hyperparameters, and convergence behavior are discussed in Figure S1 of the Supplementary Materials (SM).

We first show that the CG representations learned by a one-shot CGMP trajectory exhibit high electronic prediction accuracy at all sampled resolution. We performed a 1000-epoch training run in which the number of CG beads,  $M$ , was discretely interpolated from 32 heavy atoms to 2 over the course of training by systematically changing the harmonic constraint in the loss function. Note that for the HOMO-HOMO coupling, the same CG map is being applied to each molecule within the dimer, making it a prediction over  $2M$  atomic positions. At the last epoch before reducing resolution, the model was evaluated on the test set and used to calculate the coefficient of determination ( $R^2$ ). To ensure statistical robustness in the development of CGMP, training was repeated 100 times with different CGMP network initializations.

Figure 3 demonstrates that CGMP achieves high predictive accuracy across a broad set of electronic prediction tasks sampled from a broad conformational space. As the results in Figure 3 are averaged over ten repeated trial runs, the low standard deviation of the resulting predictive accuracy ensures robustness across CGMP training.  $R^2$  decreases with the coarsening of the CG resolution, as is expected by the decimation of atomic information the GNN can use to accomplish the regression task, with the raw values of  $R^2$  being consistent with previous work.<sup>56</sup> Moreover, it is important to note that for a specific CG resolution,  $R^2$  can be systematically improved by expliciting learning the noise associated with the CG prediction task using deep kernel learning-based Electronic Coarse-Graining methods.<sup>57</sup> However, as this technique is significantly more data intensive, here we only report the  $R^2$  on the mean prediction, resulting in an expected decrease in  $R^2$  with the coarsening of the CG representation.

Examination of the resolution dependence of the  $R^2$  val-

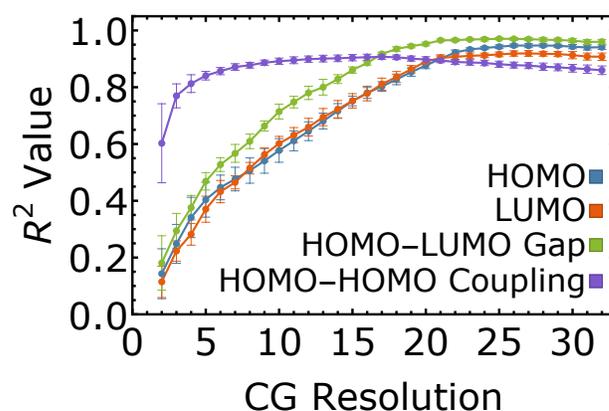


FIG. 3. Coefficient of determination for HOMO, LUMO, HOMO-LUMO Gap, and HOMO-HOMO coupling predictions using the CGMP GNN model following equilibration at each CG resolution. Error bars indicate standard deviation of ten repeated trials from independently trained CGMP GNN models.

ues in Figure 3 demonstrates complex dependencies of the optimized CG resolution on the targeted output. In the initial stages of coarsening (CG resolution between 32–24) all electronic prediction targets exhibit similar insensitivity to the coarsening procedure; subsequent analysis of the representations demonstrates that for all electronic prediction tasks the octyl sidechain is the first eight atoms to be coarsened out. This result is consistent with the general understanding in the literature that alkyl sidechains do not contribute directly to the electronic structure of the conjugated backbone.<sup>58</sup> After the side-chain has been decimated, the different electronic prediction tasks exhibit distinct sensitivities to the target CG resolution used within CGMP. HOMO and LUMO energy predictions as a function of CG molecular conformation exhibit the strongest dependence on CG resolution, which is likely related to the complex orbital structure that must be maintained by the CG representation.<sup>56</sup> Interestingly, the HOMO-LUMO gap appears significantly less sensitive to the coarsening process than either the HOMO or LUMO orbital - we hypothesize that this result is likely due to fortuitous error cancellation resulting from the simultaneous overlap of the HOMO and LUMO orbitals on similar atoms.

Surprisingly, the HOMO-HOMO coupling appears the most insensitive to CG resolution among all of the targeted electronic prediction tasks. While the electronic coupling result is unexpected provided the complexity of orbital overlap functions entering into the calculation of the electronic coupling, this result has a few potential explanations. First, the electronic coupling exhibits an exponential dependence on distance and orientation of the two molecules composing the dimer, whereas the regression loss here simply uses the raw value of the coupling. This will inevitably result in the CGMP GNN regression being biased towards the reproduction of large values of the coupling as opposed to the exponentially smaller values at the nodes of the wave function, leading to artificially induced insensitivity to the CG resolution. While this might appear strange, learning the raw cou-

pling (as opposed to e.g. the log) is a common practice in the literature<sup>59–61</sup> and motivates our inclusion of the raw coupling in Figure 3. Second, the insensitivity of the coupling to the CG resolution is likely due to the fact that the separation of the centers of mass and relative orientations of the two monomers are the largest contributors to the coupling, since BTBT is a relatively rigid molecule. Figure 3 shows that the HOMO-HOMO coupling performance drops precipitously at approximately 5–6 beads, where the preservation of the relative orientations of the rigid bodies is no longer possible.

We also note the apparent increase in  $R^2$  with decreasing resolution for resolutions above 15. Given that the GNN is likely focusing on broad intermonomer distance relationships and learning a much less complex function than the true, exponentially varying coupling function, this appears to be a convergence issue due to increased solution degeneracy at higher resolutions. Most of the atoms are able to capture some degree of the intermolecular distance, and are therefore given weight in the solution, but only a subset of these atoms are able to capture the relative molecular orientations. Since the intermolecular distance captures most of the variance of the electronic coupling function, this results in low-relevance atoms being given excessive weight in the solution space for high resolutions.

## B. Sampling over Mapping Degeneracy

With the intention of CGMP being to produce the “best” CG representation for a regression target over the full range of CG resolutions, we next examine the stochasticity of the training process in terms of the arrived at CG representations. There is no *a priori* guarantee that CGMP will arrive at the absolute best, or even the same CG representation, on every randomized training run. Moreover, within the framework of the generation of CG mappings, there is no theoretical precedent for there being a single best CG representation for any given target variable; indeed, strong degeneracy of CG representations is well-known in the literature.<sup>62</sup>

To explore CG representation degeneracy during CGMP model training, we examine the ensemble of CG representations produced by repeated runs of the CGMP method for BTBT using different initial weight and noise realizations. Figure 4 a) and b) depict the probability that a particular atom (labeling provided in Figure 4 c) is decimated at a given resolution, optimized for the HOMO and LUMO prediction tasks, respectively. These results are averaged over 100 repeated random network initializations and subsequent training runs. Zero probability in Figure 4 a) and b) indicates that an atom is either always included for large target resolutions or never included for low resolutions. The equivalent training trajectories for the HOMO-LUMO gap and HOMO-HOMO coupling are provided in Figure S7 of the SM. For every observable, the two most important atoms at the end of the coarsening procedure are always determined to be atom labels 8 and 16. As the GNN cannot make a meaningful prediction on a single atom, the coarsening procedure was stopped at the CG resolution of 2 atoms.

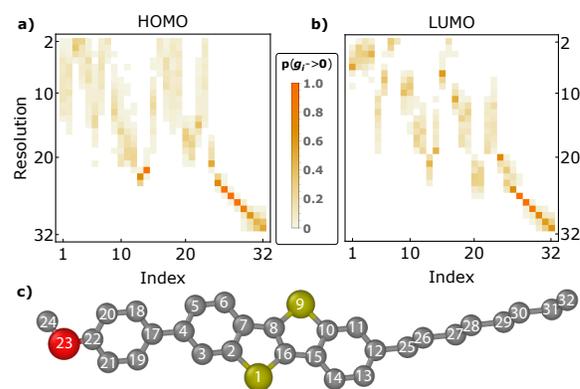


FIG. 4. Probability of a node (atom) being coarsened at a given CG resolution, sampled over 100 repeated random network initializations and training runs, shown here for (a) HOMO and (b) LUMO prediction tasks. The coarsening procedure is observed to be monotonic, such that at high resolutions any given atom is consistently included in the representation. As the target resolution decreases, it is then removed from the representation over a narrow distribution of resolutions, and is afterwards consistently not available to the GNN. (c) Indicates the correspondence between index numbering in (a) and (b) and the atomic indices.

Figure 4 shows that in the early coarsening process the carbons of the alkyl chain, which do not contribute significantly to any electronic prediction task, are decimated in order starting from the end of the chain. This ordering is likely due to a proximity-based hierarchy forming due to the distance-based structure of the GNN computational graph, but is consistent with the expectation that alkyl carbon atoms furthest from the conjugated core are less important to the electronic structure than those closest to the core. Moreover, the carbon atoms at the end of the alkyl chain, due to lower communication with the rest of the GNN, exhibit slightly higher degeneracy than those closer to the conjugated core. At lower CG resolutions, there exist similar clusters of atoms within the conjugated core that, when sampled over a large number of trials, represent degenerate coarsening choices which induce the same amount of information loss. These clusters are highly correlated within a connected block of resolutions and exhibit low overlap with other atomic clusters. However, it is important to note that atoms do not always exhibit close proximity in the chemical topology of the molecule. While this result might appear counterintuitive, it is consistent with a normal mode representation of molecular vibrations in which delocalized vibrational modes in the low-frequency domain can be coupled to the energetic changes of a single molecular orbital.

To develop more insight into the importance of particular atoms and functional groups to the prediction task, we examine samples of CG representations generated by independent CGMP runs for a target CG resolution (Figure 4). Figure 5 a) and b) shows five selected CG mappings produced by independent CGMP runs, for the HOMO and LUMO energy predictions, respectively. A CG resolution of 12 beads is selected to emphasize effects in regions of high CG mapping degeneracy, in which it can be understood if high degeneracy

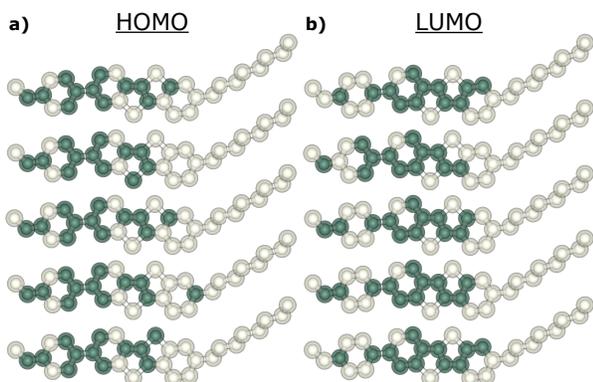


FIG. 5. Five randomly selected maps produced by independent CGMP runs fixed at a CG resolution of 12 beads, optimized for the (a) HOMO and (b) LUMO prediction tasks.

regions can be attributed to delocalized structural contributions or balances between localized core regions of the chemical structure. At the CG resolution of 12 beads, the preserved chemical motifs of the molecule are consistently represented for each prediction task: the C-C bond in the middle of the main aromatic group (atoms 8 and 16), the flexible phenyl dihedral (atoms 3, 4, 5, 18, 17, and 19), and the pendant oxygen (atom 23). The HOMO prediction task places greater emphasis on the C-O bond and the phenyl dihedral, whereas the LUMO prediction task intrinsically requires more information about the benzothienobenzothiophene core to make effective predictions. The difference between the families of maps generated for the HOMO and LUMO target variables, at a fixed CG resolution, highlight their commonalities, helping inform CG mapping design based on localized, physical variables as opposed to a random collection of atoms. For instance, in the HOMO prediction task, different atoms are kept on the right hand side of the benzothienobenzothiophene group, suggesting that the information contributing to the prediction isn't a particular atomic position or bond, but is instead the deformation of the plane on that side of the molecule.

To further quantify the degeneracy of "good" CG representations at a target CG resolution, we compute an entropy measure associated with each resolution by summing the probability values within Figure 4 a) and b) over the atomic indices according to Eq. 15:

$$\Omega_j = -\sum_i p_{ij} \log p_{ij} \quad (15)$$

The computed resolution-dependent CG mapping entropies are shown in Figure 6 for the HOMO, LUMO, HOMO-LUMO gap, and HOMO-HOMO couplings. All observables begin with a common coarsening procedure eliminating the alkyl chain and follow a consistent entropy curve. As the CG resolution decreases further, the shape of the CG mapping entropy takes on its own character for each prediction target, with high entropy regions representing where there are many near-degenerate "good" CG representations. The observed high degeneracy (high entropy) of CG mappings for single-

property targets across broad regions of Figure 6 provides an opportunity for the quantitative development of multi-target-compatible CG representations. Practical CG representations should ideally be optimized to preserve as much information about downstream prediction tasks as possible. If, at a given resolution, all targets have unique, non-overlapping solutions to the optimal mapping problem, then the combined solution will necessitate a much higher CG resolution to achieve the same performance on all tasks simultaneously. However, as shown in Figure 6, all electronic targets considered in this work exhibit notable mapping degeneracy across resolutions. When combined with qualitative observations of similarities between different target solutions (Figure 5), we conclude that the solution space of CG representations for BTBT is broad and that transferable CG representations across multiple target properties should be accessible. Critically, CGMP provides the first numerical optimization approach with which to directly quantify the degeneracy of "good" CG representations optimized for specific prediction targets, facilitating the development of more transferable CG representations.

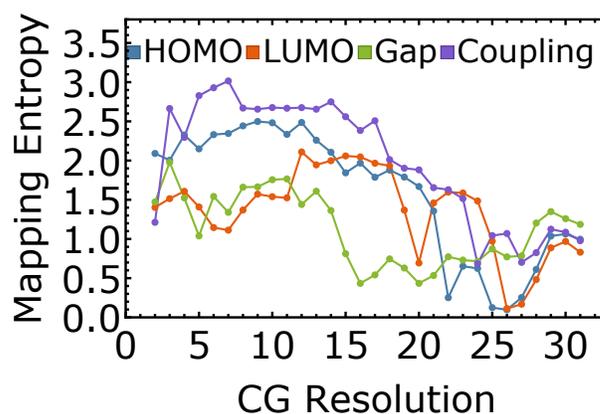


FIG. 6. Mapping entropy for the considered variables, calculated using Eq. 15. All observables initially exhibit low mapping degeneracy, as they all coarsen the alkyl chain. As the resolution decreases, degeneracy generally increases due to physical and informational symmetries of the problem. This high degeneracy region suggests higher likelihood of solution overlap between prediction tasks.

### C. Physical Significance of GNN Atomic Covariances

The output of the CGMP GNN is the sum of per-atom output functions learned according to information aggregation around each atom's neighborhood. While these functions are not intentionally optimized to distinguish one atom from another, they are constructed from each atom's local geometry, and therefore carry information specific to the atom. For the present interest of dissecting atomic importance and collective variables from the dense conformational sampling of specific chemistries, the GNN parameterization naturally converges to atom-centered functions whose individual fluctuations correspond to their relative contributions to the prediction task. We propose that the correlations between these atomic outputs

correspond to the underlying collective variables driving configurational changes in the molecular electronic structure.

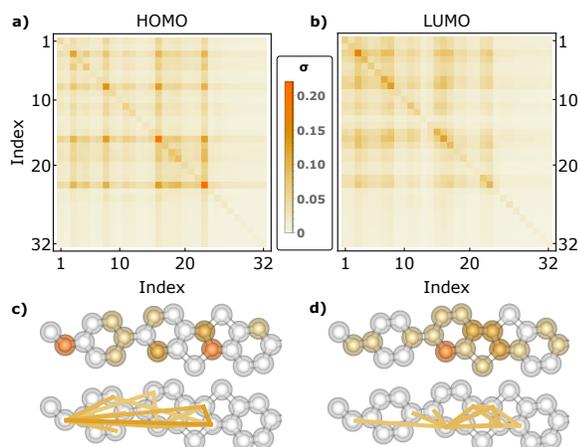


FIG. 7. Element-wise square root of the covariance matrix of the last layer of the GNN for (a) HOMO and (b) LUMO. Visualizations of the element-wise square root of the covariance matrix for the HOMO and LUMO projected onto the explicit molecular representation of BTBT are also shown in (c) and (d), respectively. The top molecular images in (c) and (d) show atomic standard deviations, which scale with their contribution to the prediction task. The bottom molecular images in (c) and (d) show the 12 highest covariance atomic relationships. Only the 12 largest edges are shown for visual clarity.

To explore the nature of these multi-atom correlations, we use the previously trained CGMP GNN at the all atom resolution (32), and extract the outputs of the last layer prior to sum pooling to generate a covariance matrix of outputs between each atom, sampled over the test set. Figure 7 shows the average of these covariance matrices for the HOMO and LUMO prediction tasks (Figure 7 a) and b), respectively) as well as an overlay of the 20 largest off-diagonal edges onto the molecular graph (Figure 7 c) and d)). The equivalent results for the HOMO-LUMO gap and HOMO-HOMO coupling are provided in Figure S8 of the SM. Large off-diagonal values in Figure 7 a) and b) indicate high communication between those atoms within the GNN; the set of these links forms a task-specific network much sparser than the full computational graph that CGMP is performed over. In particular, the strongest links correspond to physical bonds in the molecule between atoms which survive until very low resolution in the coarsening process. Atoms irrelevant to the target variable, for example the alkyl chain, never meaningfully contribute to the GNN output, even when there is no coarsening pressure applied. Further discussion of the quantitative relationship between the atomistic covariance matrix and the coarsening process can be found in sections III and IV of the SM.

Interpretation of the covariances presented in Figure 7 c) and d) provide further pathways to identify critical collective variables, and not just individual atoms, for CG representation development. Whereas this study has worked within the decimation style of CG representation generation, the results of Figure 7 c) and d) present a logical path forward to the development of CG maps that average over atomic positions.

Target-optimized average CG maps would likely maximize the covariance between between beads and minimize the covariance between atoms within a single bead. High covariance links also present a starting point for creating effective models in terms of human-interpretable collective degrees of freedom. The covariance matrix presents a much sparser graph, and therefore greatly reduced search space to look for high-importance collective variables of the molecular conformation, over which one could build effective Hamiltonians for the molecule's electronic properties.

The highest-covariance edges in Figure 7 c) and d) span both local and long-range interactions. The marginal change in results from a reduced cutoff radius (see Figures S5 and S6 in the Supplementary Materials) suggests that the electronic properties of interest can be accurately calculated as a set of short-range interactions. Long-range covariances are then a product of correlated conformational changes across the molecule resulting in same-sign energy changes. Given the correspondence between the atomistic covariance matrix and lower-resolution performance, we hypothesize that the atom-centered functions comprising the covariance matrix are converging to a set of collective modes for the molecular function, as sampled over its thermodynamic distribution, which will constitute the subject of future work.

#### IV. DISCUSSION

Whereas this work focused on the decimation style of CG mapping generation, there is considerable future work to do to adapt the methodology to the generation of averaged style CG maps in a similar one-shot, target-optimized fashion. The symmetry in the indexing of CG beads presents challenges in optimizing an averaging coarse-graining projection, though initial work<sup>22</sup> has been done to machine learn this category of mappings. One limitation of applying such methods in the context of the current CGMP model arises due to the challenges of incorporating symmetries dynamically during a single training run - changing the resolution of an averaging map is subject to both index exchange and particle exchange between beads, while our CGMP method simply removes or adds another atom. Combining atoms into average CG beads requires designing policies acting on neighborhoods of atoms simultaneously, whereas CGMP designs a policy acting on individual atoms separately.

Emerging work<sup>63</sup> has challenged this notion, using an unsupervised approach to automatically form heterogeneous CG representations over a chemically diverse space. It remains an open question as to how to direct the training of this graph partitioning method in a more supervised manner; future work will seek to combine CGMP for learning electronically important interatomic relationships with other potential techniques for clustering-based representation learning. Obvious future directions utilizing the results CGMP towards these ends include employing the covariance matrix of node energies output by the GNN as an information-weighted graph over which graph partitioning could occur. The full resolution covariance matrix is predictive of coarsened model performance (SI), in-

dicating that this could form a potentially useful basis for various spectral partitioning schemes. In addition, this covariance data could be used in conjunction with normal mode analysis of molecular vibrations to develop further strategies for assessing map quality.<sup>56</sup>

Applications of CGMP to chemically diverse problems presents several new challenges. First, the CGMP gates will need to be expressed as functions of the atomic environment, rather than node-specific parameters. Gate parameters could also be replaced by an additional GNN layer, but such an architecture would need to be extensively explored to determine practicality. Second, we have focused on a dataset of diverse conformations for a single molecule, whereas increasing chemical diversity would introduce a new dimension to sample over, greatly increasing data costs. A more intense study of the transferability between the conformational dependencies of individual chemistries is needed to better understand the computational scaling of chemically diverse coarse-graining.

## V. CONCLUSION

We introduce a novel message passing scheme for 3D-invariant GNNs known as CGMP, from which atomic decompositions of complex prediction tasks can be extracted, facilitating the robust development of molecular CG representations. We illustrate how CGMP can be used to provide a full hierarchy of property-optimized CG representations with a one-shot training procedure. Further analysis of the generated representations provides an unprecedented look at the degeneracy of “good” CG mapping representations optimized to single-target prediction tasks. The observed high degeneracy of “good” CG mapping representations across four common electronic prediction targets provides considerable motivation for the pursuit of multi-target optimized CG mapping operators that can serve as a foundations for chemically transferable CG mapping generation. Moreover, we demonstrate how CGMP can be used to extract physical interpretations of important atomic degrees of freedom, and how this data can be used to seek average-based CG mapping representations. The ability of CGMP to delineate the landscape of multi-target CG representations in a numerically robust and high-performance fashion promises an era of generalizable CG models across chemical space.

## VI. SUPPLEMENTARY MATERIAL

Training details and hyperparameters, training convergence, interpreting coarsening through covariance, full-resolution score distributions, comparison with lower GNN cutoff radius, equivalent main text figures for other observables.

## ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Chemical Theory, Models, and Computation division under award CHE-2154916. We acknowledge support from the Dreyfus Program for Machine Learning in the Chemical Sciences and Engineering during this project. We thank the American Chemical Society Petroleum Research Fund Doctoral New Investigator award for additional support of this research.

## DATA AVAILABILITY STATEMENT

All data used in this work is freely available from the authors upon request. The CGMP implementation employed in this work can be found publicly at <https://github.com/TheJacksonLab>.

## Appendix A: ComENet Graph Featurization

The initial input to the GNN is the set of atomic indices and coordinates belonging to the molecule. A distance cutoff (model hyperparameter) is used to generate the edges of the graph representation of the molecule. Each node is assigned a learned vector feature according to atomic index, and edges are featurized using local geometric information. Specifically, each edge is represented via a tuple  $(d, \theta, \phi, \tau)$ , where  $d$  is the edge length,  $\theta$  and  $\phi$  are determined by the orientation of the edge with respect to the atom’s local neighborhood, and  $\tau$  is a representation of the relative orientation of the two neighborhoods involved in the edge.

### 1. Angle Calculation

Suppose a message is being passed from node  $j$  to node  $i$ . The algorithm finds node  $i$ ’s first- and second-nearest neighbors,  $f_i$  and  $f_j$ , respectively (if  $j$  is  $i$ ’s only neighbor, then  $j = f_i = s_i$ ). The edge from  $i$  to  $f_i$  and the  $(f_i, i, s_i)$  plane serve as a reference axis and plane to define the local neighborhood of  $i$ .  $\theta_{ij}$  is then the  $angle(f_i, i, j)$ , and  $\phi_{ij}$  is the angle between the planes  $(f_i, i, s_i)$  and  $(f_i, i, j)$ . To orient the neighborhood, the algorithm then finds nodes  $f_{ij}$  and  $f_{ji}$ , which are reference nodes that exclude the other member of the edge, and calculate  $\tau_{ij} = dih(f_{ij}, i, j, f_{ji})$ . Note the implication for low-degree nodes: if  $j$  is  $i$ ’s only neighbor, then  $\theta_{ij} = \phi_{ij} = \tau_{ij} = 0$ . If  $i$  is  $j$ ’s only neighbor, then  $\tau_{ij} = 0$ .

### 2. A Differentiable Reference Vector

The previous geometric calculations require that for each edge the closest neighboring nodes not involved in the edge are tracked, which is a multi-stage sorting process. As the goal of this work is to represent the coarsening process in a

continuous and differential fashion, we introduce an alternative atomic neighborhood defined by a first reference vector,  $\mathbf{f}_{ij}$ , based on a weighted sum of the full set of edge vectors. We first define a vector ( $\mathbf{u}_{ij}$ ) in the direction of each edge, scaled by a length ( $r_{ij}$ ) and an optional edge score ( $w_{ij}$ ). The second reference vector,  $\mathbf{s}_{ij}$ , is then generated in the same fashion, but projecting out any components parallel to the first reference vector. This method has the advantage that, instead of using an arbitrarily defined vector for each edge, one can employ a single set of reference vectors for every edge of a node.

$$\mathbf{u}_{ij} = \frac{w_{ij}}{r_{ij}^2} \mathbf{r}_{ij} \quad (\text{A1})$$

$$\mathbf{f}'_i = \sum_{j \in N(i)} \mathbf{u}_{ij} \quad (\text{A2})$$

$$\mathbf{s}'_i = \sum_{j \in N(i)} \left( \mathbf{u}_{ij} - \frac{\mathbf{u}_{ij} \cdot \mathbf{f}'_i}{10^{-9} + f_i^2} \mathbf{f}'_i \right) \quad (\text{A3})$$

We then modify these initial representations to account for decreasing node degrees through coarsening and remain consistent with the behavior of ComENet. We fix behavior such that a node with degree one has  $\mathbf{s}_i = \mathbf{f}_i$ , and a node with degree 0 has randomly assigned reference vectors. The values of a degree 0 node's reference vectors do not matter, as it does not communicate with the rest of the graph, and outputs will be gated to 0, but this assignment keeps gradient values stable during coarsening.

$$\mathbf{f}_i = \mathbf{f}'_i + 10^{-6} \boldsymbol{\varepsilon} \quad (\text{A4})$$

$$\mathbf{s}_i = \mathbf{s}'_i + 10^{-2} \mathbf{f}_i + 10^{-6} \boldsymbol{\varepsilon} \quad (\text{A5})$$

For an edge weight, we choose  $w_{ij} = 2\text{ReLU}(g_i g_j - 0.5)$ . This is a simple way of coupling node scores with the angle calculation, but is not strictly necessary; more sophisticated gating schemes could be developed, such as additional gates for individual edges in the graph to further prune the GNN computational graph. Edge gates could be a way to provide further interpretation of the computation happening within the GNN. As an example, partitioning the computational graph into two disconnected localized graphs over the molecule would be a clear indication that correlations between their outputs are purely due to delocalized, collective modes over the molecular conformations, rather than communication over long distances. Sparsifying the computational graph at a given resolution breaks the physical symmetries that may be present in the current representation and introduces a biased path for future coarsening. As the present work is interested in identifying the space of possible representations, we leave investigation into edge gating parameters to future work.

### 3. Angles to Vector Features

After assigning each edge a tuple containing geometric information, these tuples are converted into feature vectors for input to a neural network. ComENet generates two separate

feature vectors for the tuples  $(d_{ij}, \theta_{ij}, \phi_{ij})$  and  $(d_{ij}, \tau_{ij})$ . This serves to have separate messages to communicate local neighborhood and neighborhood orientation information. The encoding of each tuple is

$$v_1(d, \theta, \phi) = j_l \left( \frac{\beta_{ln}}{c} d \right) Y_l^m(\theta, \phi) \quad (\text{A6})$$

$$v_2(d, \tau) = j_l \left( \frac{\beta_{ln}}{c} d \right) Y_l^0(\tau) \quad (\text{A7})$$

in which  $j_l$  and  $Y_l^m$  represent the Bessel and spherical harmonic functions of appropriate degree and order. These vectors are each projected via two-layer neural network to the hidden dimension of the GNN, and used as an edge weight for their own message-passing schemes (Equation 8). The resultant node update (Equation 9) is a neural network-projected concatenation of these two aggregated messages.

- <sup>1</sup>G. A. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems* (Taylor & Francis Ltd, Hoboken, NJ, 2008).
- <sup>2</sup>N. E. Jackson, "Coarse-Graining Organic Semiconductors: The Path to Multiscale Design," *J. Phys. Chem. B* **125**, 485–496 (2021).
- <sup>3</sup>W. G. Noid, "Perspective: Coarse-grained models for biomolecular systems," *J. Chem. Phys.* **139**, 090901 (2013).
- <sup>4</sup>S. Dhamankar and M. A. Webb, "Chemically specific coarse-graining of polymers: Methods and prospects," *J. Polym. Sci.* **59**, 2613–2643 (2021).
- <sup>5</sup>V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, "Versatile Object-Oriented Toolkit for Coarse-Graining Applications," *J. Chem. Theory Comput.* **5**, 3211–3223 (2009).
- <sup>6</sup>W. G. Noid, "Perspective: Advances, Challenges, and Insight for Predictive Coarse-Grained Models," *J. Phys. Chem. B* **127**, 4174–4207 (2023).
- <sup>7</sup>K. M. Lebold and W. G. Noid, "Dual-potential approach for coarse-grained implicit solvent models with accurate, internally consistent energetics and predictive transferability," *J. Chem. Phys.* **151**, 164113 (2019).
- <sup>8</sup>C.-I. Wang and N. E. Jackson, "Bringing Quantum Mechanics to Coarse-Grained Soft Materials Modeling," *Chem. Mater.* **35**, 1470–1486 (2023).
- <sup>9</sup>T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman, "Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models," *J. Chem. Theory Comput.* **10**, 676–690 (2014).
- <sup>10</sup>J. Peng, C. Yuan, R. Ma, and Z. Zhang, "Backmapping from Multiresolution Coarse-Grained Models to Atomic Structures of Large Biomolecules by Restrained Molecular Dynamics Simulations Using Bayesian Inference," *J. Chem. Theory Comput.* **15**, 3344–3353 (2019).
- <sup>11</sup>A. J. Rzepiela, L. V. Schäfer, N. Goga, H. J. Risselada, A. H. De Vries, and S. J. Marrink, "Reconstruction of atomistic details from coarse-grained structures," *J. Comput. Chem.* **31**, 1333–1343 (2010).
- <sup>12</sup>F. Müller-Plathe, "Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back," *Chem. Phys. Chem.* **3**, 754–769 (2002).
- <sup>13</sup>W. Wang and R. Gómez-Bombarelli, "Coarse-graining auto-encoders for molecular dynamics," *npj Comput. Mater.* **5**, 1–9 (2019).
- <sup>14</sup>M. Schöberl, N. Zabarab, and P.-S. Koutsourelakis, "Predictive coarse-graining," *J. Comput. Phys.* **333**, 49–77 (2017).
- <sup>15</sup>W. Li, C. Burkhart, P. Políńska, V. Harmandaris, and M. Doxastakis, "Backmapping coarse-grained macromolecules: An efficient and versatile machine learning approach," *J. Chem. Phys.* **153**, 041101 (2020).
- <sup>16</sup>Y. An and S. A. Deshmukh, "Machine learning approach for accurate backmapping of coarse-grained models to all-atom models," *Chem. Commun.* **56**, 9312–9315 (2020).
- <sup>17</sup>K. A. Louison, I. L. Dryden, and C. A. Laughton, "GLIMPS: A Machine Learning Approach to Resolution Transformation for Multiscale Modeling," *J. Chem. Theory Comput.* **17**, 7930–7937 (2021).
- <sup>18</sup>M. Stieffenhofer, T. Bereau, and M. Wand, "Adversarial reverse mapping of condensed-phase molecular structures: Chemical transferability," *APL Materials* **9**, 031107 (2021).

- <sup>19</sup>M. A. Webb, J.-Y. Delannoy, and J. J. de Pablo, "Graph-Based Approach to Systematic Molecular Coarse-Graining," *J. Chem. Theory Comput.* **15**, 1199–1208 (2019).
- <sup>20</sup>M. Chakraborty, C. Xu, and A. D. White, "Encoding and selecting coarse-grain mapping operators with hierarchical graphs," *J. Chem. Phys.* **149**, 134106 (2018).
- <sup>21</sup>W. Yang, C. Templeton, D. Rosenberger, A. Bittracher, F. Nüske, F. Noé, and C. Clementi, "Slicing and Dicing: Optimal Coarse-Grained Representation to Preserve Molecular Kinetics," *ACS Cent. Sci.* **9**, 186–196 (2023).
- <sup>22</sup>R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules," *ACS Cent. Sci.* **4**, 268–276 (2018).
- <sup>23</sup>F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, "Machine Learning for Molecular Simulation," *Ann. Rev. Phys. Chem.* **71**, 361–390 (2020).
- <sup>24</sup>V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, "Gaussian Process Regression for Materials and Molecules," *Chem. Rev.* **121**, 10073–10141 (2021).
- <sup>25</sup>S. T. John and G. Csányi, "Many-Body Coarse-Grained Interactions Using Gaussian Approximation Potentials," *J. Phys. Chem. B* **121**, 10934–10949 (2017).
- <sup>26</sup>L. Zhang, J. Han, H. Wang, R. Car, and W. E, "DeePCG: Constructing coarse-grained models via deep neural networks," *J. Chem. Phys.* **149**, 034101 (2018).
- <sup>27</sup>A. E. P. Durumeric and G. A. Voth, "Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining," *J. Chem. Phys.* **151**, 124110 (2019).
- <sup>28</sup>R. Alessandri and J. J. de Pablo, "Prediction of Electronic Properties of Radical-Containing Polymers at Coarse-Grained Resolutions," *Macromolecules* **56**, 3574–3584 (2023).
- <sup>29</sup>S. Zhang, M. Makoś, R. Jadrlich, E. Kraka, K. Barros, B. Nebgen, S. Tretiak, O. Isayev, N. Lubbers, R. Messlerly, and J. Smith, "Exploring the frontiers of chemistry with a general reactive machine learning potential," (2023).
- <sup>30</sup>S. Chennakesavalu, D. J. Toomer, and G. M. Rotskoff, "Ensuring thermodynamic consistency with invertible coarse-graining," *J. Chem. Phys.* **158**, 124126 (2023).
- <sup>31</sup>X. Fu, T. Xie, N. J. Rebello, B. Olsen, and T. S. Jaakkola, "Simulate Time-integrated Coarse-grained Molecular Dynamics with Multi-scale Graph Networks," *Transactions on Machine Learning Research* (2023).
- <sup>32</sup>B. E. Husic, N. E. Charron, D. Lemm, J. Wang, A. Pérez, M. Majewski, A. Krämer, Y. Chen, S. Olsson, G. de Fabritiis, F. Noé, and C. Clementi, "Coarse graining molecular dynamics with graph neural networks," *J. Chem. Phys.* **153**, 194101 (2020).
- <sup>33</sup>Z. Liu, L. Lin, Q. Jia, Z. Cheng, Y. Jiang, Y. Guo, and J. Ma, "Transferable Multilevel Attention Neural Network for Accurate Prediction of Quantum Chemistry Properties via Multitask Learning," *J. Chem. Inf. Model.* **61**, 1066–1082 (2021).
- <sup>34</sup>W. Zhu, Y. Zhang, D. Zhao, J. Xu, and L. Wang, "HiGNN: A Hierarchical Informative Graph Neural Network for Molecular Property Prediction Equipped with Feature-Wise Attention," *J. Chem. Inf. Model.* **63**, 43–55 (2023).
- <sup>35</sup>L. Wang, Y. Liu, Y. Lin, H. Liu, and S. Ji, "ComENet: Towards Complete and Efficient Message Passing for 3D Molecular Graphs," (2022).
- <sup>36</sup>K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet – A deep learning architecture for molecules and materials," *J. Chem. Phys.* **148**, 241722 (2018).
- <sup>37</sup>V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, "Benchmarking graph neural networks for materials chemistry," *npj Comput. Mater.* **7**, 1–8 (2021).
- <sup>38</sup>Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, and T. F. Miller, III, "OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features," *J. Chem. Phys.* **153**, 124111 (2020).
- <sup>39</sup>T. D. Loose, P. G. Sahrman, T. S. Qu, and G. A. Voth, "Coarse-Graining with Equivariant Neural Networks: A Path Towards Accurate and Data-Efficient Models," (2023).
- <sup>40</sup>S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," *Nat Commun* **13**, 2453 (2022).
- <sup>41</sup>A. Krämer, A. E. P. Durumeric, N. E. Charron, Y. Chen, C. Clementi, and F. Noé, "Statistically Optimal Force Aggregation for Coarse-Graining Molecular Dynamics," *J. Phys. Chem. Lett.* **14**, 3970–3979 (2023).
- <sup>42</sup>M. Giulini, R. Menichetti, M. S. Shell, and R. Potestio, "An Information-Theory-Based Approach for Optimal Model Reduction of Biomolecules," *J. Chem. Theory Comput.* **16**, 6795–6813 (2020).
- <sup>43</sup>F. Errica, M. Giulini, D. Bacciu, R. Menichetti, A. Micheli, and R. Potestio, "A Deep Graph Network-Enhanced Sampling Approach to Efficiently Explore the Space of Reduced Representations of Proteins," *Frontiers in Molecular Biosciences* **8** (2021).
- <sup>44</sup>R. Menichetti, M. Giulini, and R. Potestio, "A journey through mapping space: characterising the statistical and metric properties of reduced representations of macromolecules," *Eur. Phys. J. B* **94**, 204 (2021).
- <sup>45</sup>Z. Li, G. P. Wellawatte, M. Chakraborty, H. A. Gandhi, C. Xu, and A. D. White, "Graph neural network based coarse-grained mapping prediction," *Chem. Sci.* **11**, 9524–9531 (2020).
- <sup>46</sup>A. Mastropietro, G. Pasculli, C. Feldmann, R. Rodríguez-Pérez, and J. Bajorath, "EdgeSHAPer: Bond-centric Shapley value-based explanation method for graph neural networks," *iScience* **25**, 105043 (2022).
- <sup>47</sup>S. Zhang, F. Zhu, J. Yan, R. Zhao, and X. Yang, "DOTIN: Dropping Task-Irrelevant Nodes for GNNs," (2022).
- <sup>48</sup>R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "GNNExplainer: Generating Explanations for Graph Neural Networks," (2019).
- <sup>49</sup>H. Yuan, H. Yu, S. Gui, and S. Ji, "Explainability in Graph Neural Networks: A Taxonomic Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 5782–5799 (2023).
- <sup>50</sup>W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," *J Chem Phys* **128**, 244114 (2008).
- <sup>51</sup>D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," (2017).
- <sup>52</sup>Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical Graph Representation Learning with Differentiable Pooling," in *Advances in Neural Information Processing Systems*, Vol. 31 (Curran Associates, Inc., 2018).
- <sup>53</sup>J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 97, edited by K. Chaudhuri and R. Salakhutdinov (PMLR, 2019) pp. 3734–3743.
- <sup>54</sup>H. Gao, Y. Liu, and S. Ji, "Topology-Aware Graph Pooling Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43**, 4512–4518 (2021).
- <sup>55</sup>M. J. Han, D. Wei, Y. H. Kim, H. Ahn, T. J. Shin, N. A. Clark, D. M. Walba, and D. K. Yoon, "Highly Oriented Liquid Crystal Semiconductor for Organic Field-Effect Transistors," *ACS Cent. Sci.* **4**, 1495–1502 (2018).
- <sup>56</sup>C.-I. Wang, J. C. Maier, and N. E. Jackson, "Identifying Coarse-Grained Representations for Electronic Predictions," *J. Chem. Theory Comput.* **19**, 4982–4990 (2023).
- <sup>57</sup>J. C. Maier and N. E. Jackson, "Bypassing backmapping: Coarse-grained electronic property distributions using heteroscedastic Gaussian processes," *J. Chem. Phys.* **157**, 174102 (2022).
- <sup>58</sup>C. Risko, M. D. McGehee, and J.-L. Brédas, "A quantum-chemical perspective into low optical-gap polymers for highly-efficient organic solar cells," *Chem. Sci.* **2**, 1200–1218 (2011).
- <sup>59</sup>F. Musil, S. De, J. Yang, J. E. Campbell, G. M. Day, and M. Ceriotti, "Machine learning for the structure–energy–property landscapes of molecular crystals," *Chem. Sci.* **9**, 1289–1300 (2018).
- <sup>60</sup>C.-I. Wang, M. K. E. Braza, G. C. Claudio, R. B. Nellas, and C.-P. Hsu, "Machine Learning for Predicting Electron Transfer Coupling," *J. Phys. Chem. A* **123**, 7792–7802 (2019).
- <sup>61</sup>C.-I. Wang, I. Joanito, C.-F. Lan, and C.-P. Hsu, "Artificial neural networks for predicting charge transfer coupling," *J. Chem. Phys.* **153**, 214113 (2020).
- <sup>62</sup>T. T. Foley, K. M. Kidder, M. S. Shell, and W. G. Noid, "Exploring the landscape of model representations," *Proceedings of the National Academy of Sciences* **117**, 24061–24068 (2020).
- <sup>63</sup>J. Lederer, M. Gastegger, K. T. Schütt, M. Kampffmeyer, K.-R. Müller, and O. T. Unke, "Automatic Identification of Chemical Moieties," *Phys. Chem. Chem. Phys.* (2023), 10.1039/D3CP03845A.