Machine learning band gaps from the electron density

Javier Robledo Moreno [1,2,*] Johannes Flick [1,†] and Antoine Georges [1,3,4,5,‡]

[1]*Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA*
[2]*Center for Quantum Phenomena, Department of Physics, New York University, 726 Broadway, New York, New York 10003, USA*
[3]*Collège de France, 11 place Marcelin Berthelot, 75005 Paris, France*
[4]*CPHT, CNRS, École Polytechnique, IP Paris, F-91128 Palaiseau, France*
[5]*DQMP, Université de Genève, 24 quai Ernest Ansermet, CH-1211 Genève, Switzerland*

A remarkable consequence of the Hohenberg-Kohn theorem of density functional theory is the existence of an injective map between the electronic density and any observable of the many-electron problem in an external potential. In this work, we study the problem of predicting a particular observable, the band gap of semiconductors and band insulators, from the knowledge of the local electronic density. Using state-of-the-art machine learning techniques, we predict the experimental band gaps from computationally inexpensive density functional theory calculations. We propose a modified Behler-Parrinello (BP) architecture that greatly improves the model capacity while maintaining the symmetry properties of the BP architecture. Using this scheme, we obtain band gaps at a level of accuracy comparable to those obtained with state-of-the-art and computationally intensive hybrid functionals, thus significantly reducing the computational cost of the task.

I. INTRODUCTION

The Hohenberg-Kohn (HK) theorem of density functional theory (DFT) establishes a one-to-one correspondence between the ground-state local charge density $\rho(\vec{r})$ and the external potential of a many-electron system $v(\vec{r})$ [2]. Thus, the Hamiltonian of this system is uniquely and unequivocally determined by $\rho(\vec{r})$. This remarkable statement implies that any observable of the system must be a functional of the ground-state density. However, the functionals are unknown in most cases [3,4], and in practice physically motivated approximations are used, such as the local density approximation (LDA) [4] or the Perdew-Burke-Ernzerhof (PBE) [5] exchange correlation functionals. Despite the success of these approximations in certain regimes, they can fail to describe relevant observables, such as band gaps of semiconducting materials [6].

For modern electronic, optoelectronic, and photovoltaic applications [7,8], the reliable estimation of band gaps of solids is of great relevance. The *ab initio* computation of accurate band gaps tends to be a resource-intensive task, as the cheaper exchange correlation functionals like the LDA or generalized gradient approximations (GGAs) underestimate the value of the band gap due to the vanishing of the so called derivative discontinuity [5,6,9,10]. Meta and hybrid GGA functionals, like the Heyd-Scuseria-Ernzerhof (HSE), have a nonzero derivative discontinuity, providing more accurate band gap predictions at a much higher computational

cost [1,6]. Data-driven approaches have been used to improve the band gap prediction at a lower computational cost. Some works use the PBE band gap predictions as a starting point to compute corrections on the level of HSE or *GW* accuracy [11,12]. Another popular approach is to predict the band gap from the positions of the atoms in the unit cell and the crystalline structure [13–18] using synthetic data sets for training. Another common approach is the prediction of band gaps from a set of relevant physical quantities such as boiling and melting points, atomic radii, and bond lengths, among others [17,19–23]. Previous studies have successfully demonstrated that data-driven approaches can be used to approximate the HK maps between the electron density and observables of interest [24–28]. More recent studies have considered the prediction of accurate band gaps from an approximate description of the electronic density [29–31]. Despite outstanding progress, most works rely on synthetic training sets containing a relatively small number of different materials in different configurations.

In this article, we study the problem of inferring experimental band gaps from DFT computed valence electron density distributions in a wide range of bulk semiconductors and band insulators, using deep neural networks (NNs). We propose a Deep Sets [32] aided modification of the Behler-Parrinello architecture, improving the model capacity while maintaining the permutation invariance in the output. We provide empirical evidence suggesting that learning from the density is advantageous as opposed to learning directly from atomic positions, due to the chemical bonding information explicitly encoded in the electronic density. We find that learning band gaps from PBE-computed densities yields errors comparable to the direct calculation of the band gap using the computationally intensive and accurate HSE06

*jrm874@nyu.edu
†jflick@flatironinstitute.org
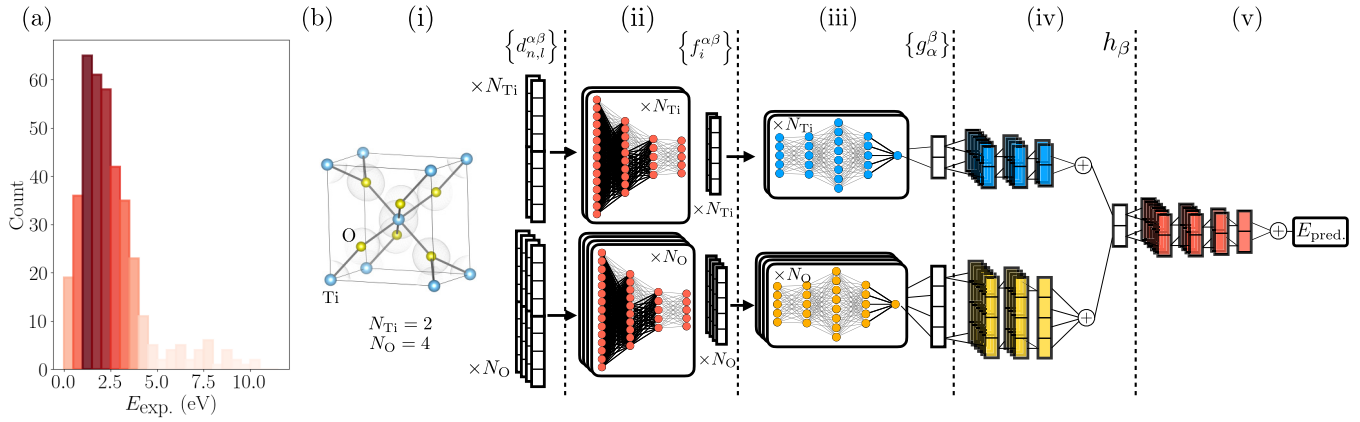‡ageorges@flatironinstitute.org

FIG. 1. (a) Histogram of the distribution of experimental band gaps used for training and validation, from Ref. [1]. Histogram box width is 0.5 eV. Colors represent the height of each bin, matching the color labels in Fig. 3. (b) Diagram of the neural network architecture used to approximate the density to band gap map in the specific example of TiO_2. Red modules are shared among all atomic species. Blue and yellow modules are specifically applied to Ti and O densities, respectively. From left to right: (i) Rotationally invariant atom-centered density descriptors [see Eq. (3)]; (ii) feature extraction for dimensional reduction; (iii) Behler-Parrinello block and concatenated output $\{g_\alpha^\beta\}$; (iv) species-specific Deep Sets neural network and concatenation of output of each species h_β; (v) species-combiner Deep Sets network whose output is the band gap.

hybrid functional [33]. Finally, we test the neural network performance in a collection of monolayer materials, including examples of the molybdenum family of the transition metal dichalcogenides (TMDCs) and hexagonal boron nitride.

II. TRAINING SET AND DENSITY DESCRIPTORS

The data set we use to train and validate our neural network is the set of bulk materials compiled in Ref. [1]. It contains 472 materials for which a reliable and accurate experimental band gap measurement is available in the literature. The reported band gaps are mostly from optical absorption experiments; hence we only consider the determination of direct gaps in this article and not of the fundamental (photoemission) or transport gaps. We note furthermore that the reduction of the optical gap by excitonic effects, while relevant in principle when comparing to DFT or HSE determinations of the gaps, is typically only a few tens of milli-electron-volts in bulk systems, much smaller than the typical error on the band gaps from these methods. Most materials contain less than 24 atoms per unit cell, while none exceeds 32 atoms. No magnetic materials are included, as the band gap can be highly affected by the magnetic configuration in antiferromagnets. The experimental band gap distribution is shown in Fig. 1(a). This data set also provides the Materials Project [34] identification number. This number allows for the extraction of the relevant unit cell parameters to compute the density distribution.

The density distribution corresponding to the valence electrons for each material is computed from a lattice relaxation *ab initio* DFT calculation using the projector-augmented wave (PAW) method implemented in the Vienna *Ab initio* Simulation Package (VASP) [35–37]. Different levels of approximation are used to compute the density distributions. Specifically, we use LDA and PBE exchange correlation functionals. In the latter, we also include van der Waals corrections using the DFT-D2 method [38].

The obtained density values are specified on a uniform grid in the unit cell. Different materials have different unit cell geometries and therefore different grids. Hence, the density value in each grid cell cannot be used as the input of a traditional neural network. Instead, we project the density onto an atomic centered set of orbitals:

$$\rho^{\alpha\beta}(\vec{r} - \vec{R}^{\alpha\beta}) \approx \sum_{\substack{0 \leqslant n \leqslant n_{\max} \\ 0 \leqslant l \leqslant l_{\max} \\ -l \leqslant m_l \leqslant l}} C_{n,l,m}^{\alpha\beta} P_n^{\mathrm{orth}}(r) Y_{l,m}(\theta, \phi), \quad (1)$$

where $Y_{l,m}(\theta, \phi)$ are real spherical harmonics, $P_n^{\mathrm{orth}}(r)$ are orthogonal polynomials in the radial coordinate, and $C_{n,l,m}^{\alpha\beta}$ are the expansion coefficients for atom α of atomic species β in position $\vec{R}^{\alpha\beta}$, and therefore a set of density descriptors. Recent works have demonstrated this approach to yield accurate density descriptors for machine learning (ML) applications [39,40]. The set $\{P_n^{\mathrm{orth}}(r)\}$ is obtained by the orthogonalization of the polynomial basis:

$$P_n(r) = \frac{2n}{r_0^{2n+1}} \left(r + \frac{r_0}{2n} \right) (r - r_0)^{2n}, \quad (2)$$

with r_0 a cutoff, chosen to be the radius where the value of the density is 5% of the value of the density at $r = 0$. We follow the standard orthogonalization procedure proposed in Ref. [39]. In order to impose rotational symmetry, i.e., ensure that the band gap prediction does not not change if the unit cell is rotated, we choose the set of rotationally invariant features [40–43]

$$d_{n,l}^{\alpha\beta} = \sum_{m=-l}^{l} \left[c_{n,l,m}^{\alpha\beta} \right]^2. \quad (3)$$

While the above rotationally invariant descriptors are not the only choice to construct invariant features from spherical harmonics [42], they are widely used in ML applications and have been demonstrated to yield accurate results in the context

TABLE I. Detailed description of the modified Behler-Parrinello architecture and the intermediate data structures generated. See Fig. 1(b) for a schematic representation of the architecture and data structures. Bold symbols label trainable neural networks. In the data structures field, upper indices label the objects on a given set whereas lower indices label elements of an array.

Name	Description
Operations:	
$\boldsymbol{PP}(x)$	Preprocessing fully connected NN for feature extraction.
$\boldsymbol{BP}^\beta(x)$	BP fully connected NN of atomic species β.
$\boldsymbol{DS}^\beta(x)$	Deep Sets NN of atomic species β.
$\boldsymbol{DSsc}(x)$	Deep Sets NN combining the output of each species block.
$\oplus^\gamma$	Concatenation across index γ.
Data structures:	
$\{d_{n,l}^{\alpha\beta}\}$	Rotationally invariant density descriptors; see Eq. (3).
$\{f_i^{\alpha\beta}\} = \{\boldsymbol{PP}(d_{n,l}^{\alpha\beta})\}$	Preprocessed density features. $i = 1, \ldots, 5$.
$\{g_\alpha^\beta\} = \{\oplus^\alpha \boldsymbol{BP}^\beta(f_i^{\alpha\beta})\}$	Output of the BP block concatenated for atoms in the same species.
$h_\beta = \oplus^\beta \boldsymbol{DS}^\beta(g_\alpha^\beta)$	Concatenation of the output of the species-specific Deep Sets across the different species.
$E_{\mathrm{pred}} = \boldsymbol{DSsc}(h_\beta)$	Output of the Deep Sets species combiner. Predicted band gap.

of ML for DFT [40]. Even though the directionality of the principal axes of the density distribution is not present in $d_{n,l}^{\alpha\beta}$ as noted in Ref. [41], the density descriptors still preserve the features of the bonding.

III. NEURAL NETWORK ARCHITECTURE

The goal is to infer the value of the experimental band gap given the set of vectors of rotationally invariant density descriptors $\{d_{n,l}^{\alpha\beta}\}$. As a preprocessing step, the set of descriptors $\{d_{n,l}^{\alpha\beta}\}$ are compressed into a set of dimensionally reduced features $\{f_i^{\alpha\beta} : i = 1, \ldots, 5\}$ by a fully connected NN with trainable parameters (the same for all atoms), as depicted in Fig. 1 and Table I.

As noted before, different materials have different numbers of atoms in the unit cell, belonging to different atomic species. In order to have the ability to handle the variable size of input density descriptors, yet respecting the permutation invariance of the input set, we use a modified Behler-Parrinello [44] architecture. The NN architecture is depicted in Fig. 1(b) and detailed in Table I. Bold symbols label trainable neural networks. In the data structures field, upper indices label the objects on a given set whereas lower indices label elements of an array.

The use of Deep Sets [32] after the BP block ensures the permutation invariance of the predicted gap with respect to permutations of the density descriptors of different atoms, while improving the model capacity by combining in a nonlinear manner the outputs of the BP block. Deep Sets layers consist of several channels of the composition of a nonlinear activation function and a restricted affine transformation. The restricted affine map imposes permutation equivariance in the output features. It maps an M-dimensional input feature x to an M-dimensional output feature y in the following form:

$$ y = [\lambda \mathbf{I} + \gamma (\mathbf{1} \cdot \mathbf{1}^\mathrm{T})]x + b\mathbf{1}, \tag{4} $$

with λ, γ, and b optimizable parameters, $\mathbf{I}$ the $M \times M$ identity matrix, and $\mathbf{1}$ the M-dimensional column vector of ones.

A reduction of overparametrization effects, owing to the small size of the training set, is achieved via the so called early stopping [45] and the use of minibatch gradient descent [46,47], with 40 materials per batch.

A collection of statistical measurements of the prediction error is chosen to test the performance of the band gap prediction. Defining the error as the difference between experimental and target band gaps, $E_{\mathrm{exp}} - E_{\mathrm{pred}}$, we consider the mean absolute percentage error (MAPE), mean percentage error (MPE), mean absolute error (MAE), mean error (ME), the standard deviation of the error distribution (σ), the interquartile range (IQR), and the parameters (a and b) of fitting E_{pred} vs E_{exp} to $y = ax + b$. We employ stratified 10-fold validation in order to mitigate the possible bias in the error statistics induced by the splitting of the data set in training and validation partitions. Within each partition, we make sure that the distribution of validation band gaps is similar to the band gap distribution of the training set [see Fig. 1(a)] in order to mitigate error bias. The error statistics correspond to the averages of each statistical measurement over the different partitions.

IV. MODEL CAPACITY

We compare the model capacity of the bare BP architecture against the Deep Sets improved BP architecture in Fig. 2(a). The bare BP architecture does not possess enough expressivity to even interpolate the data. This deficiency in the expressive power stems from lack of physical justification to take the band gap to be the sum of individual contributions of the different atoms in the unit cell, as done in a bare BP architecture. The improved Deep Sets architecture achieves significantly lower training error by combining in a nonlinear and parametrized manner the outputs of the BP architecture. These results show that the addition of the Deep Sets module provides a solution to the model rigidity of the bare BP architecture. This modification, not limited to the problem of band gap prediction, can be implemented in the many applications of the BP architecture.
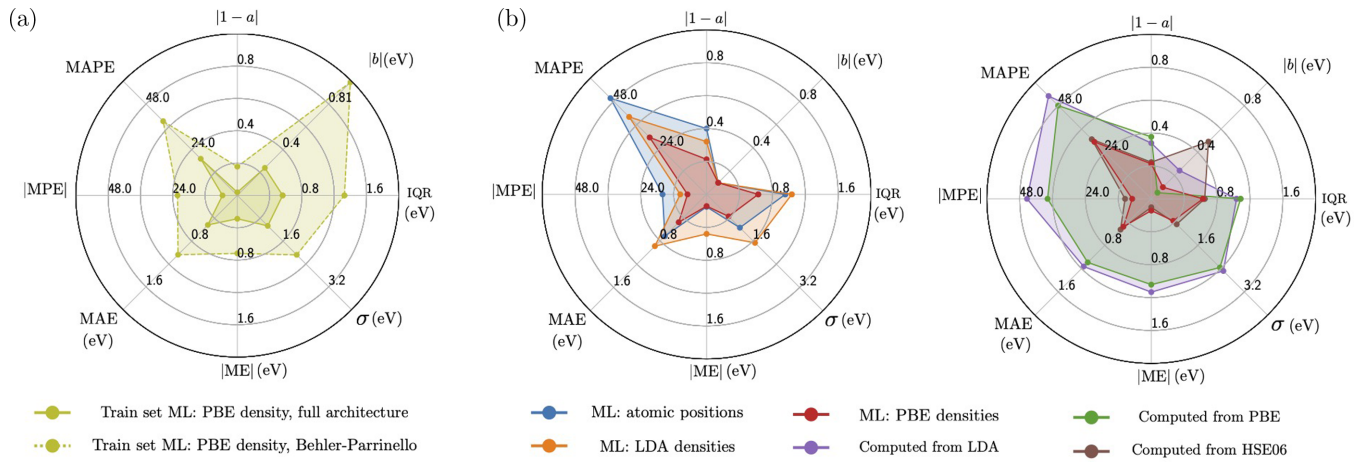
FIG. 2. Radar plots of the test statistics chosen to measure the band gap prediction accuracy. (a) Train set performance comparison between a Behler-Parrinello architecture—with preprocessing feature extraction—and the full architecture including the Deep Sets modules. Input densities are computed from DFT-PBE. (b) Validation set performance in the full architecture including the Deep Sets modules. Left: Band gaps are learned from atomic positions in the unit cell, from densities computed using DFT-LDA, and from densities computed using DFT-PBE. Right: Comparison of the best performing ML scheme with DFT-calculated band gaps, using LDA, PBE, and HSE06 [6] functionals.

V. LEARNING PERFORMANCE

The map between the density and physical properties of interest stems from the one-to-one correspondence between the electronic density and the external potential, which fixes the specific form of the Hamiltonian and therefore uniquely determines the whole collection of observables. Obviously, there is also a map from the atomic positions and their nuclear charges to the band gap. We compare the prediction of band gaps obtained directly from the atomic positions against the prediction from electronic densities, computed to different levels of approximation (LDA and PBE) in the left panel of Fig. 2(b). The test error is larger if atomic positions are used as inputs to the ML model. This difference is particularly noteworthy in the the MAPE, which is reduced from ∼48% in the case of atomic positions to ∼28% in the case of PBE densities. The low MPE and ME values indicate that the ML approach does not suffer from a systematic overestimation—or underestimation—of

the band gap. A natural explanation of the prediction improvement arising from the use of the density is that the chemical bonding between the atoms in the unit cell is better taken into account in this approach. The electronic structure of crystals is strongly dependent on the nature of the chemical bonding between their atoms, which is explicitly encoded in the electronic density distribution. Our results thus provide empirical evidence that an explicit encoding of the chemical bonding is beneficial for the task of band gap prediction. This claim is supported by the prediction improvement achieved by using more accurate input densities, as the neural network trained from PBE densities has better error statistics compared to the model trained from LDA densities. While in some specific cases the LDA provides more accurate ground-state densities, the PBE functional provides more accurate densities on average.

The right panel of Fig. 2(b) shows the error statistics comparison between various DFT approaches to directly com-
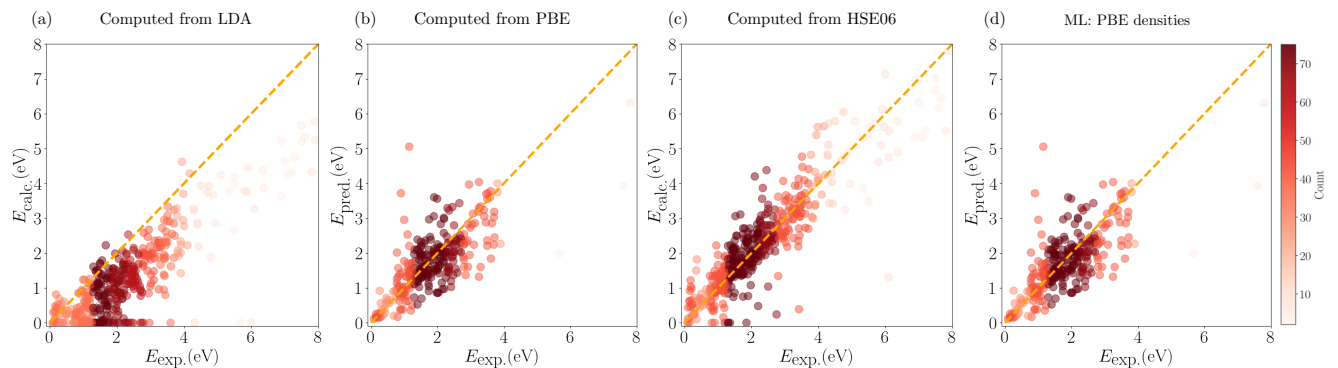

FIG. 3. Predicted versus experimental band gaps in the following cases: (a)–(c) Band gaps are computed from DFT calculations using LDA, PBE, and HSE06 [6] functionals. (d) ML-predicted experimental band gaps from PBE-computed densities. Only validation set examples are shown. Orange dashed lines are for visual reference, showing what the exact band gap prediction corresponds to. The color scale for each point corresponds to the number of examples in the training set for that experimental band gap, matching the color scheme of the histogram in Fig. 1(a). The error statistics of the predicted band gaps shown in this figure are those shown in the right panel of Fig. 2(b).
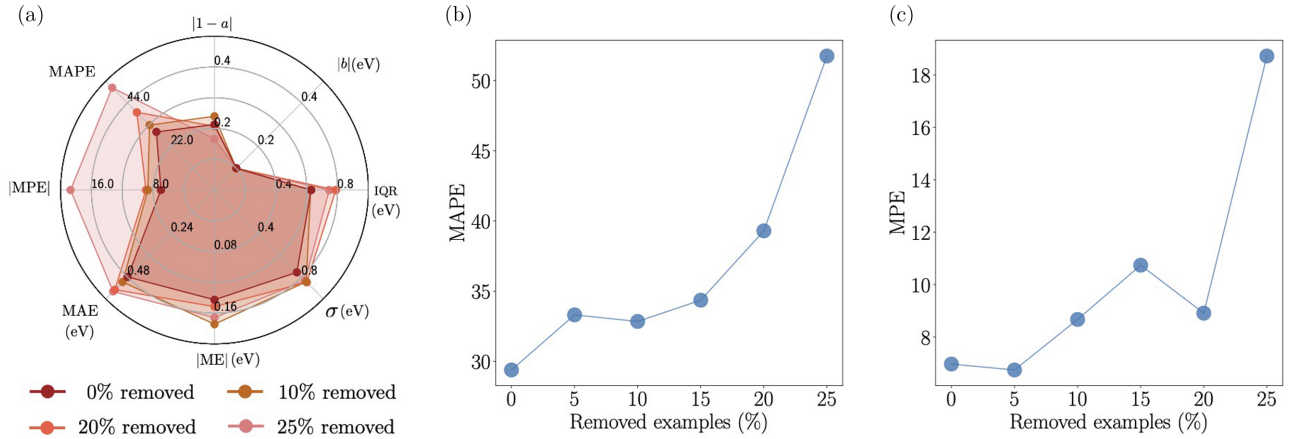
FIG. 4. (a) Radar plots of the test statistics chosen to measure the band gap prediction accuracy. Validation set performance comparison with the data set with removed examples, as indicated by the color labels. (b) Mean absolute percentage error (MAPE) in the validation set as a function of the number of removed examples expressed as a percentage of the total number of examples of the original data set. (c) Mean percentage error (MPE) in the validation set as a function of the number of removed examples expressed as a percentage of the total number of examples of the original data set.

pute the band gap and our best performing ML scheme. The DFT-calculated band gaps come from three levels of approximation: LDA, PBE, and HSE06 energy functionals. The LDA and PBE band gaps were obtained from the same DFT calculations used to compute the input densities. The errors are consistent with those in Ref. [6]. The HSE06 band gaps are found in Ref. [6]. The ML-predicted band gaps greatly outperform standard DFT-calculated band gaps in the LDA and PBE approximations. The error in the ML-predicted band gaps is comparable to that of the accurate [6] and computationally intensive HSE06 functional.

These results demonstrate that, using the proposed statistical learning setting, it is possible to infer the band gap at HSE06 level of accuracy, while benefiting from the much cheaper PBE estimation of the electronic density.

Panels (a)–(c) of Fig. 3 show the values of the DFT-calculated band gaps vs the experimental values for LDA, PBE, and HSE06 functionals. The LDA and PBE functionals systematically underestimate the correct experimental value, providing a biased estimate of the band gap. This observation is also clear in the corresponding values of the MPE and ME shown in Fig. 2(b). The HSE06 provides band gap values with significantly smaller error and bias. The ML-predicted band gaps in Fig. 3(d) are centered around the correct experimental values, thus providing a low bias estimate. In this case the band gaps show more spread compared to the HSE06 values. The higher spread is related to the small number of available training examples as discussed in the Appendix. The spread is larger for larger band gaps due to the smaller number of training examples with band gaps larger than 4 eV [see Fig. 1(a) and the Appendix].

VI. EXTRAPOLATION: MONOLAYER MATERIALS

Lastly, we test the generalization capabilities of the neural network on a collection of monolayer materials, including the molybdenum family of transition metal dichalcogenides and the hexagonal boron nitride. We choose structures with

available freestanding experimental band gaps or on substrates with lattice mismatch <2%, which leads to almost no lattice distortion. Thus, the presence of the substrate is not expected to significantly alter the value of the measured band gap. The training set does not contain monolayer materials. It includes some examples of the family of TMDCs in the bulk: $MoSe_2$, MoS_2, and $MoTe_2$.

The mean absolute percentage errors of the ML-predicted band gaps are shown in Table II. Except for the case of hBN, we obtain accurate predictions of the band gap of these two dimensional structures. Given that $E_{exp} = 6.1$ eV, the reason behind the lack of accuracy in the hBN is believed to stem from the reduced number of training examples in that range of band gaps (see Appendix for a detailed discussion). As the ML model is only trained on bulk structures, the accurate predictions in monolayers shows the generalization capabilities of the proposed statistical learning approach. The results in the two-dimensional structures also show the improvement of accuracy as the input density is improved, supporting the claim that better predictions are related to the better explicit description of the chemical bonding in the density distribution.

TABLE II. Generalization error in the monolayer materials. Mean absolute percentage error of the ML band gaps is shown averaged across the 10 training partitions. Error bars correspond to the standard deviation of the MAPE across the 10 training partitions.

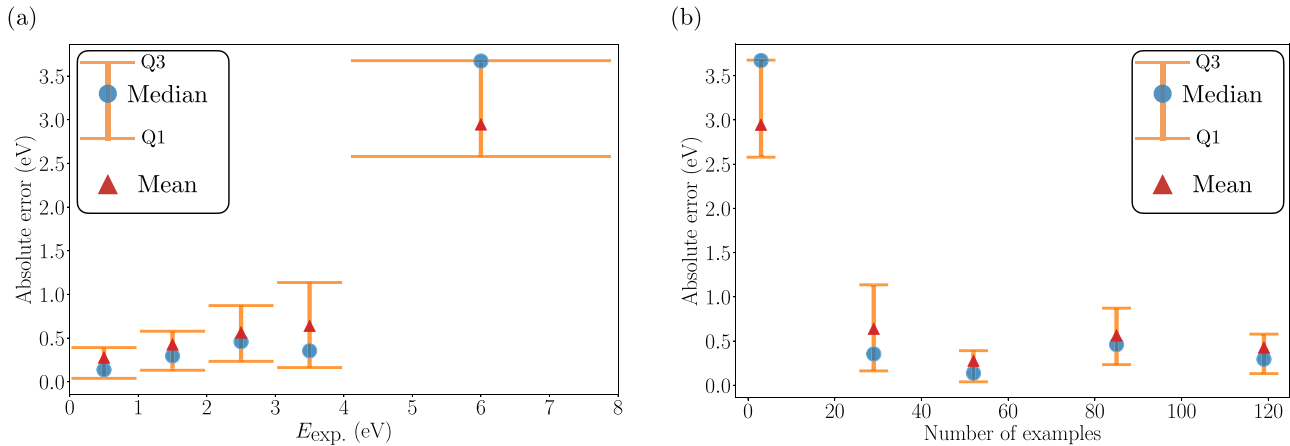|	ML from atomic positions	ML from LDA densities	ML from PBE densities
MoS_2 [48]	$(52 \pm 17)\%$	$(15 \pm 5)\%$	$(7 \pm 2)\%$
$MoSe_2$ [49]	$(44 \pm 15)\%$	$(40 \pm 12)\%$	$(15 \pm 5)\%$
hBN [50]	$(65 \pm 6)\%$	$(62 \pm 6)\%$	$(51 \pm 16)\%$

FIG. 5. (a) Absolute error statistics grouped by band gap value with group separations [0, 1, 2, 3, 4, 8] (eV) as indicated by the quartile Q1 and Q3 caps. (b) Same as panel (a) with the horizontal axis showing the number of examples in each band gap group. It shows that a small number of training examples in a certain band gap range leads to poor performance.

VII. CONCLUSION

In this paper we propose a supervised deep learning approach to predict the experimental band gaps of solids from the knowledge of an estimate of the electron density. This approach is justified by the Hohenberg-Kohn theorem of density functional theory. Owing to the fundamental lack of model capacity of a bare Behler-Parrinello architecture, we introduce a modification to the model architecture based on the utilization of permutation-invariant Deep Sets modules. The modified architecture shows a significant improvement in model capacity. This improved architecture can be implemented in other applications where the bare BP architecture is used [39,40]. We also show that the explicit encoding of chemical bonding information in the electron density provides an advantage over learning the experimental band gaps directly from atomic positions. This observation is supported by the prediction improvement when using more accurate estimates of the input densities. The ML-predicted band gaps achieve an accuracy comparable to state-of-the-art DFT-HSE06 estimation of the band gaps, at the much lower computational cost of PBE calculations. Finally, we test the neural network in a collection of monolayer materials, which are not present in the training set, finding good generalization power when the bad gap is in the range 0–4 eV. We believe that the largest source of error in our approach is the small size of experimental data sets available for training (see Appendix for a detailed discussion).

APPENDIX: EFFECT OF THE TRAINING SET SIZE

In this Appendix we explore the effect of the training set size on the performance of the ML approach proposed in the main text. Starting from the full data set, we remove 5%, 10%, 15%, 20%, and 25% of the examples (randomly selected) and study the validation performance using 10-fold validation as described in the main text. Examples are removed in a nonuniform fashion in order to maintain the same band gap distribution of the original data set [see Fig. 1(a)].

Figure 4 shows the data set size effect on the validation error statistics. Panel (a) shows a reduction in the error as the number of materials in the training set is increased. Panels (b) and (c) focus on the MAPE and MPE to quantify the error decrease as the size of the data set is decreased, confirming that the largest source or error of our approach is the small size of experimental data sets available for training.

Furthermore, we study the effect of the number of training examples in a certain band gap range on the performance of the proposed ML scheme. Figure 5(a) shows the absolute error statistics on different band gap ranges, with separation in [0, 1, 2, 3, 4, 8] (eV). The error is clearly smaller at smaller band gaps. This error reduction as the band gap is decreased is a direct consequence of having a larger number of training examples with smaller band gap values, as shown in Fig. 5(b).

[1] P. Borlido, T. Aull, A. Huran, F. Tran, M. Marques, and S. Botti, Large-scale benchmark of exchange-correlation functionals for the determination of electronic band gaps of solids, J. Chem. Theory Comput. **15**, 5069 (2019).

[2] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Phys. Rev. **136**, B864 (1964).

[3] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Challenges for density functional theory, Chem. Rev. **112**, 289 (2012).

[4] W. Kohn, Nobel Lecture: Electronic structure of matterwave functions and density functionals, Rev. Mod. Phys. **71**, 1253 (1999).

[5] J. P. Perdew, Density functional theory and the band gap problem, Int. J. Quantum Chem. **28**, 497 (1985).

[6] P. Borlido, J. Schmidt, A. W. Huran, F. Tran, M. A. L. Marques, and S. Botti, Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning, npj Comput. Mater. **6**, 96 (2020).

[7] S. Sze and K. N. Kwok, *Physics of Semiconductor Devices* (Wiley, Hoboken, New Jersey, 2006).

[8] Q. Jacquet, A. Iadecola, M. Saubanère, H. Li, E. J. Berg, G. Rousse, J. Cabana, M.-L. Doublet, and J.-M. Tarascon, Charge transfer band gap as an indicator of hysteresis in Li-disordered rock salt cathodes for Li-ion batteries, J. Am. Chem. Soc. **141**, 11452 (2019).

[9] J. P. Perdew, W. Yang, K. Burke, Z. Yang, E. K. U. Gross, M. Scheffler, G. E. Scuseria, T. M. Henderson, I. Y. Zhang, A. Ruzsinszky, H. Peng, J. Sun, E. Trushin, and A. Görling, Understanding band gaps of solids in generalized KohnSham theory, Proc. Natl. Acad. Sci. USA **114**, 2801 (2017).

[10] T. Dimitrov, H. Appel, J. I. Fuks, and A. Rubio, Exact maps in density functional theory for lattice models, New J. Phys. **18**, 083004 (2016).

[11] G. Pilania, J. Gubernatis, and T. Lookman, Multi-fidelity machine learning models for accurate bandgap predictions of solids, Comput. Mater. Sci. **129**, 156 (2017).

[12] J. Lee, A. Seko, K. Shitara, K. Nakayama, and I. Tanaka, Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques, Phys. Rev. B **93**, 115104 (2016).

[13] G. S. Na, S. Jang, Y.-L. Lee, and H. Chang, Tuplewise material representation based machine learning for accurate band gap prediction, J. Phys. Chem. A **124**, 10616 (2020).

[14] Y. Huang, C. Yu, W. Chen, Y. Liu, C. Li, C. Niu, F. Wang, and Y. Jia, Band gap and band alignment prediction of nitride-based semiconductors using machine learning, J. Mater. Chem. C **7**, 3238 (2019).

[15] B. Olsthoorn, R. M. Geilhufe, S. S. Borysov, and A. V. Balatsky, Band Gap prediction for large organic crystal structures with machine learning, Adv. Quantum Technol. **2**, 1900023 (2019).

[16] Y. Dong, C. Wu, C. Zhang, Y. Liu, J. Cheng, and J. Lin, Bandgap prediction by deep learning in configurationally hybridized graphene and boron nitride, npj Comput. Mater. **5**, 26 (2019).

[17] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, Accelerating materials property predictions using machine learning, Sci. Rep. **3**, 2810 (2013).

[18] G. A. Nemnes, T. L. Mitran, and A. Manolescu, Gap prediction in hybrid graphene-hexagonal boron nitride nanoflakes using artificial neural networks, J. Nanomater. **2019**, 6960787 (2019).

[19] S. K. Kauwe, T. Welker, and T. D. Sparks, Extracting knowledge from DFT: experimental band gap predictions through ensemble learning, Integrating Materials and Manufacturing Innovation **9**, 213 (2020).

[20] J. Liang and X. Zhu, Phillips-inspired machine learning for band gap and exciton binding energy prediction, J. Phys. Chem. Lett. **10**, 5640 (2019).

[21] X. Wang, Y. Xu, J. Yang, J. Ni, W. Zhang, and W. Zhu, ThermoEPred-EL: Robust bandgap predictions of chalcogenides with diamond-like structure via feature cross-based stacked ensemble learning, Comput. Mater. Sci. **169**, 109117 (2019).

[22] A. C. Rajan, A. Mishra, S. Satsangi, R. Vaish, H. Mizuseki, K.-R. Lee, and A. K. Singh, Machine-learning-assisted accurate band gap predictions of functionalized MXene, Chem. Mater. **30**, 4031 (2018).

[23] Y. Zhuo, A. Mansouri Tehrani, and J. Brgoch, Predicting the band gaps of inorganic solids by machine learning, J. Phys. Chem. Lett. **9**, 1668 (2018).

[24] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Bypassing the Kohn-Sham equations with machine learning, Nat. Commun. **8**, 872 (2017).

[25] K. Ryczko, D. A. Strubbe, and I. Tamblyn, Deep learning and density-functional theory, Phys. Rev. A **100**, 022512 (2019).

[26] Q. Liu, J. Wang, P. Du, L. Hu, X. Zheng, and G. Chen, Improving the performance of long-range-corrected exchange-correlation functional with an embedded neural network, J. Phys. Chem. A **121**, 7273 (2017).

[27] J. Nelson, R. Tiwari, and S. Sanvito, Machine learning density functional theory for the Hubbard model, Phys. Rev. B **99**, 075132 (2019).

[28] J. R. Moreno, G. Carleo, and A. Georges, Deep Learning the Hohenberg-Kohn Maps of Density Functional Theory, Phys. Rev. Lett. **125**, 076402 (2020).

[29] L. C. Lentz and A. M. Kolpak, Predicting HSE band gaps from PBE charge densities via neural network functionals, J. Phys.: Condens. Matter **32**, 155901 (2020).

[30] F. Bruneval, I. Maliyov, C. Lapointe, and M.-C. Marinica, Extrapolating unconverged GW energies up to the complete basis set limit with linear regression, J. Chem. Theory Comput. **16**, 4399 (2020).

[31] B. Kolb, L. C. Lentz, and A. M. Kolpak, Discovering charge density functionals and structure-property relationships with PROPhet: A general framework for coupling machine learning and first-principles methods, Sci. Rep. **7**, 1192 (2017).

[32] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, Deep sets, arXiv:1703.06114 (2017).

[33] J. Heyd, G. E. Scuseria, and M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, J. Chem. Phys. **118**, 8207 (2003).

[34] M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. Krishna Ande, S. van der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson, and M. Asta, Charting the complete elastic properties of inorganic crystalline compounds, Sci. Data **2**, 150009 (2015).

[35] G. Kresse and J. Hafner, *Ab initio* molecular dynamics for liquid metals, Phys. Rev. B **47**, 558 (1993).

[36] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. **6**, 15 (1996).

[37] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[38] S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, J. Comput. Chem. **27**, 1787 (2006).

[39] S. Dick and M. Fernandez-Serra, Learning from the density to correct total energy and forces in first principle simulations, J. Chem. Phys. **151**, 144102 (2019).

[40] S. Dick and M. Fernandez-Serra, Machine learning accurate exchange and correlation functionals of the electronic density, Nat. Commun. **11**, 3509 (2020).

[41] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors, in *Symposium on Geometry Processing* (ACM, 2003).

[42] M. Zucchelli, S. Deslauriers-Gauthier, and R. Deriche, in *Computational Diffusion MRI*, edited by E. Bonet-Carne, F. Grussu, L. Ning, F. Sepehrband, and C. M. W. Tax (Springer International Publishing, Cham, 2019), pp. 77–89.

[43] D. S. Novikov, J. Veraart, I. O. Jelescu, and E. Fieremans, Rotationally-invariant mapping of scalar and orientational metrics of neuronal microstructure with diffusion MRI, NeuroImage **174**, 518 (2018).

[44] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[45] L. Prechelt, in *Neural Networks: Tricks of the Trade*, 2nd ed., edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer, Berlin, 2012), pp. 53–67.

[46] S. L. Smith, B. Dherin, D. G. T. Barrett and S. De, On the origin of implicit regularization in stochastic gradient descent, arXiv:2101.12176 (2021).

[47] D. Lei, Z. Sun, Y. Xiao, and W. Y. Wang, Implicit regularization of stochastic gradient descent in natural language processing: observations and implications, arXiv:1811.00659 (2018).

[48] A. R. Klots, A. K. M. Newaz, B. Wang, D. Prasai, H. Krzyzanowska, J. Lin, D. Caudel, N. J. Ghimire, J. Yan, B. L. Ivanov, K. A. Velizhanin, A. Burger, D. G. Mandrus, N. H. Tolk, S. T. Pantelides, and K. I. Bolotin, Probing excitonic states in suspended two-dimensional semiconductors by photocurrent spectroscopy, Sci. Rep. **4**, 6608 (2014).

[49] B. K. Choi, M. Kim, K.-H. Jung, J. Kim, K.-S. Yu, and Y. J. Chang, Temperature dependence of band gap in $MoSe_2$ grown by molecular beam epitaxy, Nanoscale Res. Lett. **12**, 492 (2017).

[50] C. Elias, P. Valvin, T. Pelini, A. Summerfield, C. J. Mellor, T. S. Cheng, L. Eaves, C. T. Foxon, P. H. Beton, S. V. Novikov, B. Gil, and G. Cassabois, Direct band-gap crossover in epitaxial monolayer boron nitride, Nat. Commun. **10**, 2639 (2019).