

PAPER

# Electronic structure calculations of twisted multi-layer graphene superlattices

To cite this article: Georgios A Tritsaris *et al* 2020 *2D Mater.* **7** 035028

View the [article online](#) for updates and enhancements.

## Recent citations

- [The 2021 quantum materials roadmap](#)  
Feliciano Giustino *et al*
- [Progress on band structure engineering of twisted bilayer and two-dimensional moiré heterostructures](#)  
Wei Yao *et al*
- [Effect of bilayer stacking on the atomic and electronic structure of twisted double bilayer graphene](#)  
Xia Liang *et al*

# Electronic structure calculations of twisted multi-layer graphene superlattices



RECEIVED  
31 January 2020

REVISED  
23 April 2020

ACCEPTED FOR PUBLICATION  
1 May 2020

PUBLISHED  
18 June 2020

Georgios A Tritsaris<sup>1</sup> , Stephen Carr<sup>2</sup> , Ziyang Zhu<sup>2</sup>, Yiqi Xie<sup>1</sup>, Steven B Torrisi<sup>2</sup> ,  
Jing Tang<sup>3</sup>, Marios Mattheakis<sup>1</sup> , Daniel T Larson<sup>2</sup> and Efthimios Kaxiras<sup>1,2</sup>

<sup>1</sup> John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States of America

<sup>2</sup> Department of Physics, Harvard University, Cambridge, Massachusetts 02138, United States of America

<sup>3</sup> Department of Physics, Nanjing University, Nanjing, 210093, People's Republic of China

E-mail: [gtritsaris@seas.harvard.edu](mailto:gtritsaris@seas.harvard.edu) and [kaxiras@physics.harvard.edu](mailto:kaxiras@physics.harvard.edu)

**Keywords:** tight binding, machine learning, high-throughput calculations, electronic band structure, twistrionics, materials informatics, quantum materials

## Abstract

Quantum confinement endows two-dimensional (2D) layered materials with exceptional physics and novel properties compared to their bulk counterparts. Although certain two- and few-layer configurations of graphene have been realized and studied, a systematic investigation of the properties of arbitrarily layered graphene assemblies is still lacking. We introduce theoretical concepts and methods for the processing of materials information, and as a case study, apply them to investigate the electronic structure of multi-layer graphene-based assemblies in a high-throughput fashion. We provide a critical discussion of patterns and trends in tight binding band structures and we identify specific layered assemblies using low-dispersion electronic bands as indicators of potentially interesting physics like strongly correlated behavior. A combination of data-driven models for visualization and prediction is used to intelligently explore the materials space. This work more generally aims to increase confidence in the combined use of physics-based and data-driven modeling for the systematic refinement of knowledge about 2D layered materials, with implications for the development of novel quantum devices.

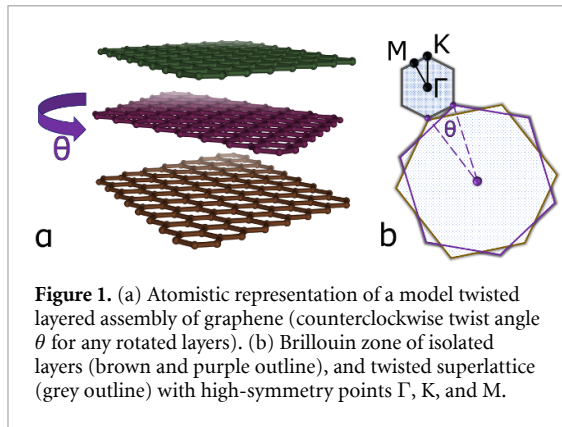
## 1. Introduction

The successful isolation of graphene has motivated sustained efforts to elucidate the properties and functionality of graphene and graphene-like nanostructures [1]. Quantum confinement endows these two-dimensional (2D) layered materials with exceptional physics and novel properties compared to their bulk counterparts. Over the years, the library of 2D materials has expanded significantly to encompass a broad range of electronic behavior from metals to insulators, with promise to induce transformational advances in applications from energy storage to quantum computing [2].

Atomically thin, single-layer forms of layered materials constitute building blocks for layered assemblies held together by weak but important interactions between the individual layers [1, 3–5]. The capability to fabricate 2D architectures with specific combinations of stacking or layer orientations presents almost unlimited possibilities for devices with novel functionality emerging from the coupling of layer-specific elementary excitations. Manipulating

the electronic properties of two-dimensional layered structures through their twist angle has emerged as a new paradigm in controlling the behavior of 2D layered materials and devices; this new field has been named "twistrionics" [6]. For example, 2D superlattices created by layers of graphene twisted relative to each other (figure 1) have provided a platform for the study of complex electronic phenomena—including correlated electrons and superconductivity in bilayer graphene when the two layers are twisted at the special, namely the "magic", angle of  $\sim 1.1^\circ$  [7, 8].

There remain severe barriers that hinder the development and deployment of quantum technology based on layered graphene. An important challenge is related to the high sensitivity of electronic transport properties and current-voltage characteristics of devices on the twist angle [9]. Although certain two- and few-layer configurations of graphene have been realized and studied [10, 11], the physics and properties of layered graphene assemblies remain largely unexplored: the configuration space quickly becomes enormous even with a restricted number of layers since the twist angle is a continuous parameter



**Figure 1.** (a) Atomistic representation of a model twisted layered assembly of graphene (counterclockwise twist angle  $\theta$  for any rotated layers). (b) Brillouin zone of isolated layers (brown and purple outline), and twisted superlattice (grey outline) with high-symmetry points  $\Gamma$ , K, and M.

that critically controls measured properties. In addition, the evaluation of a novel assembly design can be resource intensive, but it remains a more tractable task compared to the inverse problem of identifying specific layered assemblies with desirable electronic behavior. Fundamental understanding of the microscopic processes that govern the electronic properties of these quantum materials is often difficult to obtain solely by electronic transport measurements.

Materials modeling and computation have provided atomistic insights into 2D layered materials. In the prototypical case of twisted bilayer graphene, electronic band structure calculations have been used to identify electronic bands of low dispersion (or almost flat) at the Fermi level as an indicator of interesting electronic phenomena [7–9]. With regards to materials discovery, it is highly desirable to be able to efficiently search large materials spaces [12–16]. High-throughput (HT) techniques are particularly useful for assessing the scope and limitations of novel materials concepts and discovering new materials, and various approaches have been developed tailored for 2D layered materials. For instance, Mounet *et al* [5] relied on HT calculations to assess the ease of exfoliation of single-layer forms of 2D layered materials from their experimentally known bulk counterparts; Bassman *et al* [17] combined a Gaussian regression model with density functional theory (DFT) calculations to identify layered assemblies with target band gap; the work of Haastrup *et al* [18] used HT *ab initio* calculations based on DFT and many-body perturbation theory to establish a 2D materials database for the computational modeling and design of new 2D layered materials. This modality of materials discovery for graphene-based layered assemblies is now possible owing to the continuing development of quantum mechanical methods for electronic structure calculations of layered graphene [7, 8, 19, 20]. In contrast, a HT approach for the study and discovery of layered assemblies remains impractical in the lab due to the time required to fabricate and characterize a device, with typical rates of production being only few experiments on a new material in a month or year.

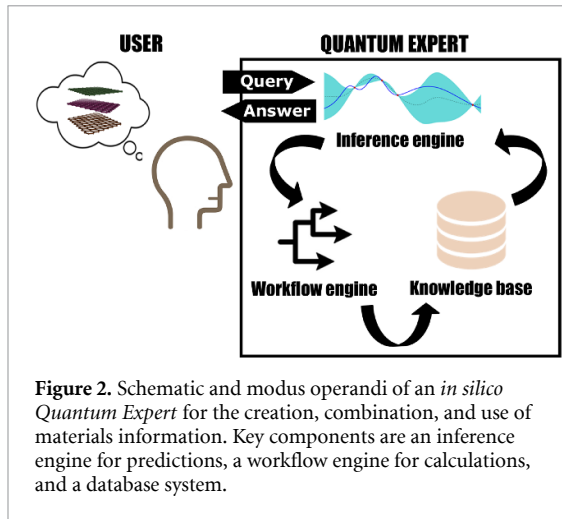
A HT computational approach necessitates dedicated tools to efficiently and seamlessly execute, manage, and visualize hundreds or thousands of calculations. A sufficient number of computational frameworks exist for HT materials calculation [21–23], although they tend to rely on extensions to take into consideration some important aspects of the materials knowledge creation cycle such as optimization and decision-making for materials selection and design. In the paradigm of materials informatics, data-driven approaches are used to extract broadly applicable physical insights from large materials data sets by identifying minimal sets of structural and functional descriptors. They are also used to expand the design space and produce informative visualizations of materials spaces for the accelerated discovery of not-yet-developed materials systems [17, 24–28].

Here we introduce theoretical concepts and methods for the creation, combination, and use of materials information and apply them to specifically investigate the electronic structure of graphene-based layered assemblies. To that end, we use workflows that combine physics-based (tight binding; TB) and data-driven (e.g. machine learning) models to obtain insights into the electronic structure of layered graphene and intelligently search the space of multi-layer assemblies for potentially interesting electronic behavior.

The manuscript is organized as follows: section 2 introduces concepts, models and methods for the creation, combination, and use of materials information. Algorithmic, computational, and implementation details are provided therein. Section 3 provides a critical discussion of trends in the TB band structures of multi-layer assemblies of graphene, obtained and analyzed using the methodology introduced in section 2. We also attempt to identify specific layered assemblies that might exhibit unusual electronic properties. Finally, section 4 summarizes findings and proposes future research directions.

## 2. Concepts, models and methods

The creation of materials knowledge begins with the introduction of a materials concept to be subsequently evaluated. The findings are then codified and combined with existing materials knowledge to generate novel concepts, completing a cycle. At the confluence of artificial intelligence and knowledge management, knowledge-based systems such as decision support systems and expert systems constitute a technology that seeks to facilitate the accessibility and dissemination of information within a particular knowledge domain [29, 30]. We propose the unifying conceptual framework of an *in silico Quantum Expert* (figure 2) for the processing of materials information, here implemented as a system specifically for the study and design of 2D superlattices of graphene.



The key components are:

- an inference engine for materials predictions,
- a workflow engine for quantum mechanical calculations,
- and a database system for storing materials data and information.

Figure 2 presents a schematic of the modus operandi:

- The user queries the system for information about one or more materials. For our particular application, the queries correspond to questions of the form: “*what is the likelihood of low-dispersion electronic bands near the Fermi level in the single-particle band structure of a specified layered assembly of graphene?*”
- The system responds with the requested information, which in our test case is a prediction for interesting electronic properties. These predictions require prior calculation and analysis of the low-energy band structures of layered assemblies for which an atomistic representation based on commensurate supercells exists (arbitrary rotations generally involve incommensurate, i.e. not periodic, 2D lattices even for layers of the same material) [31].
- If these calculations have not been already performed, the system proceeds to execute a set of predefined workflows for quantum mechanical calculations and add the results to the knowledge base. In this application, the calculations of layered assemblies consist of setting up and solving an *ab initio* TB model.

We provide application-specific implementation details for each component below.

**Inference engine.** The inference engine uses available materials information to produce new information. Multiple data-driven models are used to obtain a prediction for interesting electronic properties for

an input twisted layered assembly. Each model,  $m = (e, h)$ , is defined by an estimator,  $e$ , and a set of hyperparameters,  $h$ , and relies on a measure of band flatness in calculated TB band structures of (approximate) commensurate supercells to provide predictions for arbitrary twist angle  $\theta$  by means of quadratic interpolation. As we explain in more detail below, one estimator we employed relies on the Hough transform and another on the standard deviation of band energy eigenvalues. For each model  $m$ , a prediction  $p_m(\theta_m^*) = 1$  is assigned to all twisted layered assemblies that maximize band flatness, and the prediction

$$p_m(\theta) = \exp\left(-\frac{|\theta_m^* - \theta|}{0.1^\circ}\right)$$

is assigned to all twist angles near a magic angle  $\theta_m^*$ . The denominator  $0.1^\circ$  is fixed and chosen to reflect the sensitivity of calculations and measurements with respect to the twist angle. For each of the two estimators  $e$ , the average value,  $p_e(\theta)$ , over the corresponding  $h$  is calculated, and the prediction

$$p(\theta) = \max_e(p_e(\theta))$$

is assigned to the input layered assembly, in a blend that purposefully favors false positives. In the limiting case  $p_e(\theta) = \{0, 1\}$  such blending produces the same results as the logical OR operation—another reasonable choice would have been  $p(\theta) = 1 - \prod_e(1 - p_e(\theta))$ . The absolute difference between the two averages  $p_e(\theta)$  is used to quantify uncertainty across estimators in terms of precision.

**Workflow engine.** The function of this component is to augment the knowledge base using quantum mechanical calculations. In this work we elect to use FireWorks (version 1.9.2) as the workflow engine, a Python-based library for defining, managing, and executing workflows in a decentralized fashion on different types of high-performance computing resources [21].

Materials calculations that are based on a quantum mechanical description of electronic structure can be computationally demanding, but DFT favorably balances computational cost with accuracy [32]. Even for superlattices of the same material however a twist can render atomic-scale modeling and calculation practically intractable due to the size of the resulting atomistic models. For example, commensurate supercells of bilayer graphene with twist angles near the magic angle comprise more than 10,000 atoms. We use electronic structure calculations based on an TB model using parameters extracted from simple and accurate DFT calculations. The corresponding high-throughput workflow entails four main tasks (see also figure 3):

- Parse a string representing an input layered assembly into a list of rotations that is used to

specify an atomistic model and the corresponding TB Hamiltonian.

- (b) Generate an effective TB model for in-plane (8 nearest neighbors) and inter-plane  $p_z$  orbital interactions, without any adjustable parameters, using a basis of maximally localized Wannier functions. The electronic levels are then calculated by efficient diagonalization of the system's Hamiltonian. The memory required scales as  $\mathcal{O}(N^2)$ , where  $N$  is the number of atoms in the supercell. A detailed presentation of the method is provided in the works of Fang *et al* [20, 33]. Here we use 60 points to sample the Brillouin zone along the direction connecting the high-symmetry k-points  $\Gamma$ , M, and K.
- (c) Analyze the calculated energy levels in a window of 0.30 eV centered at the Fermi level using two different approaches to the detection of (almost-)flat bands in the TB band structure, generally with a bandwidth of a few meV. The first approach relies on an image-based Hough line transform, as implemented in OpenCV (version 4.1.1), an open source computer vision and machine learning software library [34]. The Hough line transform converts the problem of detecting collinear points into the problem of finding concurrent curves. The sensitivity of the algorithm depends on a so-called accumulator threshold, here between 50 and 200 (the smaller this number is, the more linear segments are detected, maybe incorrectly). The second approach relies on a calculation of the contribution of each k-point to the standard deviation of energy eigenvalues for each band. A band is classified as flat if the standard deviation is lower than a threshold value, here between 5 meV and 15 meV, after removing between 20% and 50% of k-points with the largest contribution to the standard deviation. The outputs of the two methods are complementary since the first circumvents the issue of artificial discontinuities in the electronic bands but the latter considers information about overlapping bands that is necessarily lost in a graphical representation although important near degeneracies.
- (d) Insert the electronic band structure and post-processing meta-data into the database for later access.

In summary, the above workflow transforms a string representing a layered assembly into measures of band flatness by means of efficient diagonalization of a TB Hamiltonian and featurization of the calculated band structures.

*Database system.* This component serves as the storehouse of materials data and information (the knowledge base). In this work the database relies on a document-based data model, as implemented

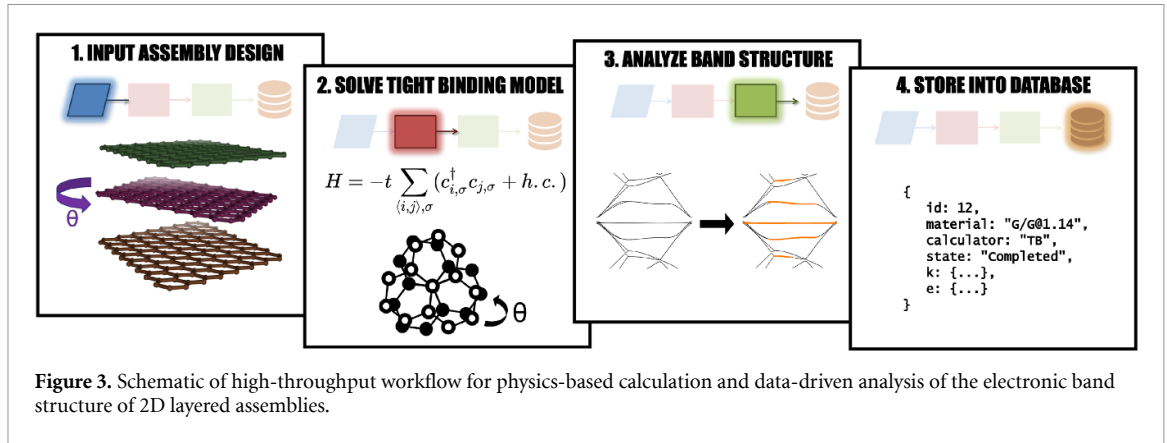
in the document-oriented (NoSQL) database MongoDB. Unlike relational databases, MongoDB is easy to scale while allowing the content and size of documents to differ from one to another. To reduce maintenance burden, the same database is used to support workflow management as required by FireWorks (e.g. workflow state and dependencies), and store the inputs and outputs of materials calculations, including structural details of the input layered assembly (e.g. rotations), its calculated TB band structure (i.e. a list of k-point dependent energy eigenvalues), and accompanying meta-data.

Overall, our computational strategy is based on scalable, parallelizable, decentralized materials calculations and it is distinctly higher up on the ladder of abstraction compared to other frameworks for computational materials science [21–23]. We note in passing that although the conceptual framework of Quantum Expert is useful for organizing materials calculations and directing the collection and analysis of data, it is not necessary for reproducing our specific findings. We also purposefully use only established approaches and open source libraries.

Coupling and linking the various components but also individual tasks within workflows pose challenges like devising systems and protocols for efficient codification and exchange of materials information. The first step in this direction is to be able to name a specific layered assembly in an unambiguous fashion. Despite the maturity of chemical language and notation such as SMILES and SMARTS for representing chemical compounds [35, 36] these cannot describe critical features of a layered assembly such as the twist angle between two neighboring layers. Previously, we introduced a flexible layered assemblies notation (LAN) [37], which we use to enumerate structures in the database, unambiguously describe trends across layered assemblies of arbitrary relative orientations of the constituting graphene layers, and construct structural descriptors for machine learning. The notation derives from a theoretical materials concept of layer-by-layer assembly of layered structures using a sequence of rotation, vertical stacking, and other operations on individual 2D layers. For example, the string 'G/G@1.08' describes a bilayer of graphene (often referred to in the literature as 'TBG') with twist angle  $\theta = 1.08^\circ$ , the string '(G/G)/(G/G)@1.23' (or the shorter '2G/2G@1.23') describes a twisted double bilayer of graphene ('TDBG' or 'TBBG') with  $\theta = 1.23^\circ$ , and so on. The symbols ' $'$ ' and '@' are used to describe the binary operations of the vertical stacking of a layer or layered (sub)structure on another, and counterclockwise rotation of a layer or layered (sub)structure by some angle about the stacking direction (in degrees;  $360^\circ = 2\pi$ ), respectively. The definition of the underlying grammar and basic operations is provided in Tritsarlis *et al* [37].

For the sake of simplicity, the queries to the system are restricted to strings representing twisted





**Figure 3.** Schematic of high-throughput workflow for physics-based calculation and data-driven analysis of the electronic band structure of 2D layered assemblies.

layered assemblies and each response is restricted to the corresponding predictions  $p(\theta)$ . In a Bayesian framework, the predictions  $p(\theta)$  constitute prior probabilities to be updated after more careful examination. For the purpose of materials discovery, after applying a cut-off value for  $p(\theta)$ , the task reduces to transforming a list of rotations ( $0$  or  $\theta$ ) into a yes/no response for further consideration.

### 3. Results and discussion

We proceed to apply our concepts, models and methods to investigate the electronic structure of twisted assemblies of graphene with two or more layers.

#### 3.1. The simplest assembly

We first concentrate attention on the prototypical system of twisted bilayer graphene,  $G/G@\theta$ . This configuration has been commonly used as a platform for the study of non-trivial emergent physical behavior that can be controlled with great precision with the relative rotation between the two layers. When the two layers are twisted at the magic angle of  $1.1^\circ$  it is possible to reduce electrical resistance significantly [7, 8]. Although elucidating the nature of superconductivity in graphene remains the goal of much ongoing research, features such as low-dispersion electronic bands in the single-particle band structure have been established as indicators of such unconventional physics [9].

We begin by requesting predictions  $p(\theta)$  for  $0.88^\circ \leq \theta \leq 21.79^\circ$ . For twist angles corresponding to commensurate 2D superlattices, the band structure is calculated using TB on the basis of which predictions are obtained for the entire range of twist angles. The twisted superlattices consist of alternating AA- and AB-stacked regions, while the Brillouin zone folds as shown in figure 1(b). Each atomistic model of commensurate superlattices consists of two graphene layers with in-plane lattice constant of  $2.47 \text{ \AA}$ , while the distance between the two graphene layers is fixed at  $3.35 \text{ \AA}$ . Twist angles that result into commensurate stackings are identified following the formalism of Uchida *et al* [38], which associates an angle  $\theta$  with a

pair of integers  $(M, N)$  that define the periodicity of the layers as:

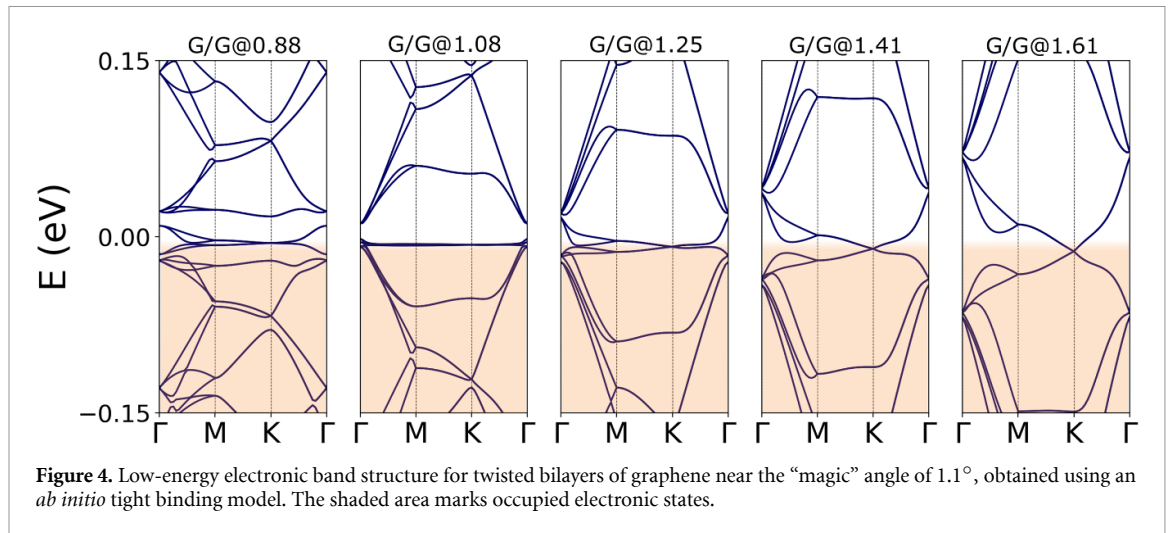
$$\cos(\theta) = \frac{N^2 + 4NM + M^2}{2(N^2 + NM + M^2)}.$$

For instance, in the regime of small twist angles, the three pairs  $(M, N)$  of  $(32, 31)$ ,  $(27, 26)$ , and  $(23, 22)$ , correspond to  $\theta$  of  $1.05^\circ$ ,  $1.25^\circ$ , and  $1.47^\circ$ , respectively. In total, 33 unique commensurate bilayers for  $0.88^\circ \leq \theta \leq 21.79^\circ$  are calculated (21 for  $1.05^\circ \leq \theta \leq 2.88^\circ$ ) and analyzed. For large  $\theta$  the low-energy band structure resembles that of an isolated graphene layer. As  $\theta$  becomes smaller, the number of bands increases and the electronic bands near the Fermi level become flatter. Figure 4 shows TB band structures of bilayers with  $\theta$  near the magic angle. In this range, strong interlayer hybridization leads to the formation of low-dispersion bands. For even smaller rotations ( $\theta < 1.0^\circ$ ) structural relaxation effects have been shown to modify the band structure significantly [31, 39, 40]. These effects however are not accounted for in the current high-throughput workflow to reduce computational burden and for that reason small twist angles are not discussed here.

An important observation is that the electronic bands evolve smoothly in the regime of small rotations (figure 4), which allows for predictions for incommensurate superlattices using interpolation as described in section 2. In accordance with previous theoretical and experimental results,  $p(\theta)$  is maximized for  $\theta^* = 1.1^\circ$  (see  $G/G@1.08$  in figure 4) [7–9].

#### 3.2. Three or four layers of graphene

We proceed to use our computational strategy to investigate twisted layered assemblies of graphene with three or four AB-stacked layers. We limit our investigation to configurations with a single twist angle for any rotated layers resulting in two and five configurations for an input angle for the tree- and four-layer assemblies, respectively. The three-layer configurations correspond to the assemblies of a twisted monolayer on a bilayer of graphene,  $2G/G@\theta$ , and a twisted graphene monolayer encapsulated by



**Figure 4.** Low-energy electronic band structure for twisted bilayers of graphene near the “magic” angle of  $1.1^\circ$ , obtained using an *ab initio* tight binding model. The shaded area marks occupied electronic states.

another two,  $G/G@θ/G$ . The four-layer configurations include assemblies such as a twisted monolayer on a trilayer,  $3G/G@θ$ , and a twisted bilayer on a bilayer,  $2G/2G@θ$ .

We obtain predictions  $p(θ)$  for  $0.88^\circ \leq θ \leq 21.79^\circ$ , which requires calculation and analysis of a total of 231 additional TB band structures of commensurate 2D superlattices. Using  $p(θ)$  as a guide, we inspect the band structures of commensurate supercells near  $θ^*$  and verify low-dispersion bands (figure 5). Some of the important features in the TB electronic band structures are discussed below:

$2G/G@θ$ . The electronic structure of a twisted monolayer on a AB-stacked bilayer of graphene resembles that of the twisted bilayer with  $θ^* = 1.1^\circ$ . The reduced symmetry causes the electronic bands near the Fermi level to be more dispersive than the twisted bilayer.

$G/G@θ/G$ . This sandwiched configuration achieves bands with very low dispersion near  $θ^* = 1.5^\circ$ . A distinguishing feature of this configuration’s band structure is a Dirac cone that resembles that of an isolated graphene layer. The Dirac cone feature is caused by an anti-symmetric combination of electronic orbitals (the top and bottom layers are in registry and couple identically to the middle layer), and is not sensitive to the twist angle [41]. These findings are in excellent agreement with the work of Khalaf *et al* [42], which used model Hamiltonians to argue that this trilayer configuration’s magic angle is obtained by multiplying the bilayer magic angle by  $\sqrt{2}$ .

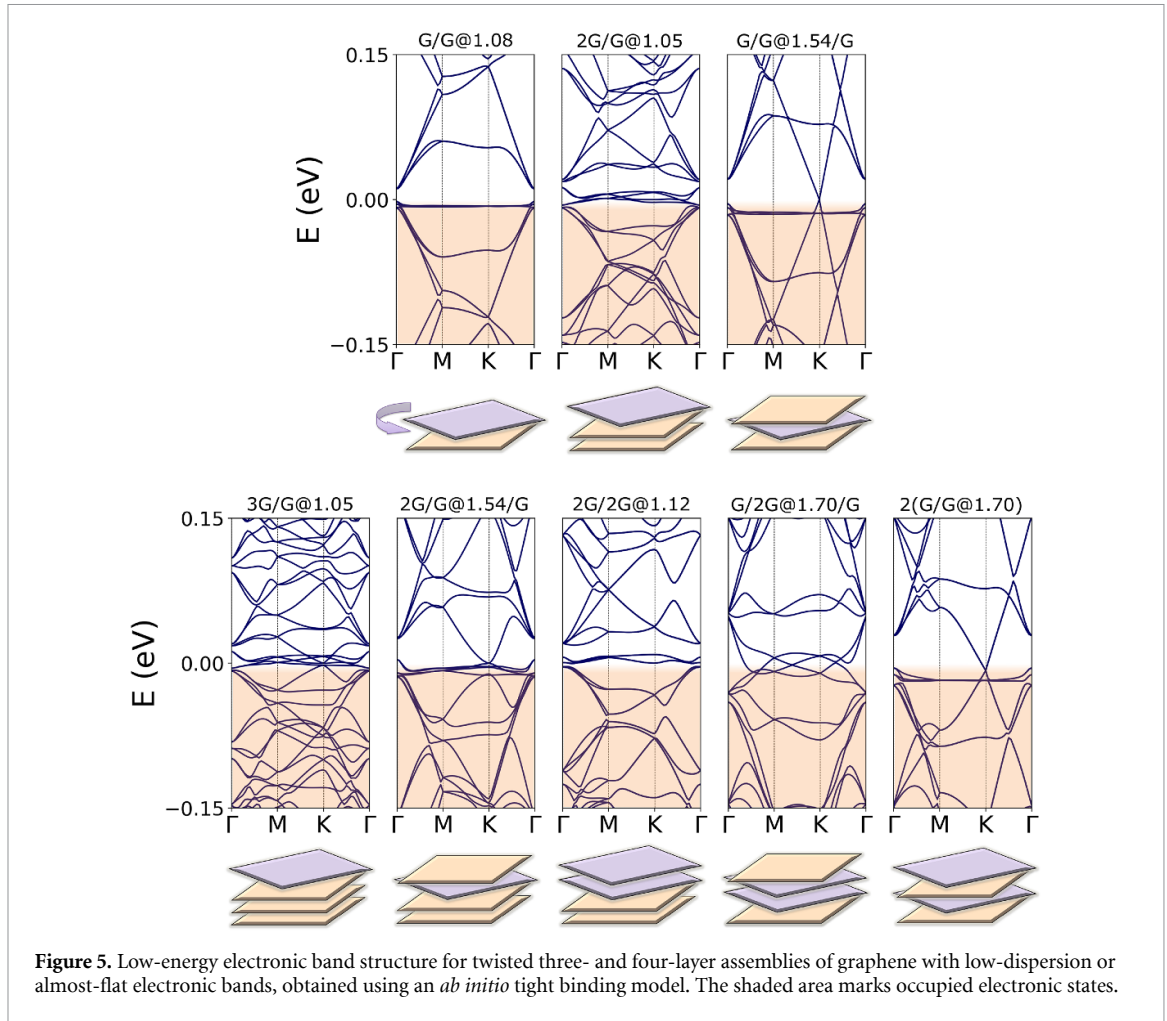
$3G/G@θ$ . The twisted monolayer on a trilayer of graphene has  $θ^* = 1.1^\circ$ . Unlike the twisted monolayer on bilayer, the flat bands are not gapped from the rest of the spectrum. A cone-like feature has begun to emerge at the K-point. As the number of layers in the unrotated part of the stack increases, the low-dispersion bands of the 2D twisted interface will coexist within a continuum of bands from bulk graphite [43].

$2G/G@θ/G$ . A twisted monolayer of graphene, when twisted on an bilayer of graphene, has a magic angle of  $θ^* = 1.5^\circ$ . Like the case of the sandwiched trilayer of graphene, the increase in the magic angle is caused by the enhanced interlayer coupling, but now parabolic bands meet near the Fermi level instead of a Dirac cone. This is comparable to the change in the electronic structure when going from the twisted bilayer to the twisted monolayer on bilayer, as the substitution of the bottom layer with a Bernal-stacked bilayer has reduced the symmetry and increased the dispersion of the low energy bands.

$2G/2G@θ$ . Twisted double bilayer graphene has been approached as an alternative to bilayer graphene in exploring correlated insulators and superconductivity [44]. Electrically tunable half-filled Mott-like insulating states for a wide range of twist angles, and superconductivity with critical temperature onset at 12 K, has been demonstrated in devices of this configuration [45, 46]. We identify a region of almost-flat bands at the Fermi level near  $θ^* = 1.1^\circ$ , although their dispersion is larger compared to bilayer graphene due to the absence of  $C_2$  symmetry.

$G/2G@θ/G$ . Electronic bands with low dispersion are identified around  $θ^* = 1.6^\circ$  in the case of a sandwiched bilayer of graphene. This generalization of the sandwiched monolayer lacks the symmetry-protected Dirac cone feature at the K-point, and the resulting electronic structure more closely resembles that of the twisted monolayer on bilayer configuration. The pairs of effective Dirac cones at both the M-point and K-point may warrant further study.

$2(G/G@θ)$ . This configuration belongs to the same class of layered assemblies of graphene with alternating relative twist angle as the sandwiched trilayer. It also exhibits magic angle flat bands, coexisting with a pair of Dirac bands at the moiré K-point. For this specific configuration, according to Khalaf *et al* [42] the magic angle is expected to be  $\sim 1.62$  times larger than this of bilayer graphene ( $θ^* = 1.8^\circ$ ).



**Figure 5.** Low-energy electronic band structure for twisted three- and four-layer assemblies of graphene with low-dispersion or almost-flat electronic bands, obtained using an *ab initio* tight binding model. The shaded area marks occupied electronic states.

Very flat electronic bands are indeed identified at  $\theta^* = 1.7^\circ$ .

As the number of calculated materials increases it becomes impractical to inspect calculations one at a time. To remedy the situation, we generate and inspect information-rich representations of the database in a state that contains all calculations of commensurate supercells (figure 6). To produce these visualizations we rely on data embedding using multidimensional scaling (MDS), as implemented in the scikit-learn package (version 0.20.3), a machine learning library for Python [47]. MDS as an approach to dimensionality reduction is particularly useful for visualizing the level of (dis)similarity of objects. Here, these objects are the TB band structures, predictions  $p(\theta)$ , and vectorized representations of the layered assemblies (264 in total). We use MDS to translate them into sets of 2D points in abstract Cartesian spaces by way of a 264-by-264 dissimilarity matrix  $K = [k_{ij}]$  with elements  $k_{ij}$  defined as:

$$k_{ij} = \|(ak_{ij}^{bs}, bk_{ij}^p, ck_{ij}^c)\|_2,$$

with  $a, b, c = \{0, 1\}$ . Its calculation relies on combinations  $(a, b, c)$  of three different metrics for evaluating pairwise distances between objects ( $k_{ii} = 0$ ):

- The calculation of  $K^{bs}$  relies on Euclidean distances between the TB band structures in an energy window of 0.30 eV centered at the Fermi level. The band structures are represented as binary images transformed by means of singular value decomposition to 200-dimensional feature vectors (covering 97% of variance).
- The calculation of  $K^p$  relies on absolute differences between the predictions  $p(\theta)$ .
- The calculation of  $K^c$  relies on cosine distances between unit-norm vectorized representations of the layered assemblies. In analogy to a “sum over bonds” vector for the featurization of chemical compound spaces [48], we construct a *sum over interfaces* (SOI) representation. These are vectors composed of sums, where each sum represents a counting of a particular interface type (G/G, G/G@ $\theta$ , G/G/G, G/G/G@ $\theta$ , etc). Two- and three-layer groupings (12 interfaces in total) suffice to uniquely describe the space of two- to five-layer assemblies, a reasonable choice given the weak but still important interactions between neighboring layers. By standard definition, cosine (dis)similarity metrics reduce the importance of the size of the assemblies.



Three visualizations are shown in figure 6, one for each of the above three matrices, which we use as building blocks for another four matrices corresponding to combinations  $(a, b, c)$  of  $(1,1,0)$ ,  $(1,0,1)$ ,  $(0,1,1)$ , and  $(1,1,1)$ . In figure 6, the points are color-coded with respect to  $p(\theta)$  and the symbol size is proportional to the associated uncertainty (larger for higher precision with respect to  $p_e(\theta)$ ). The black arrows in figure 6 indicate where the same layered assembly,  $G/G@1.05$ , can be found in each panel. As is often the case with visualizing materials information, the elected estimator MDS or approach to constructing the dissimilarity matrix  $K$  is neither unique nor necessarily optimal. Analogous representations are obtained by using the t-distributed stochastic neighbor embedding (t-SNE) method. The generated 2D plots together constitute *an atlas of interface complexity* that upon inspection reveals distinct groupings. The shorter the distance between any two structures, the more similar these are. With regards to materials discovery, such data-driven representation of the materials space simplifies the task of screening for interesting electronic behavior, drastically narrowing the list of potential candidates for more careful consideration.

These representations provide a clear picture of relationships across the entire materials database that is not readily observable by means of enumeration:

- $K^{bs}$  favors a clustering with respect to electronic bands,
- $K^p$  favors a one-dimensional ranking with respect to the predictions  $p(\theta)$ , and
- $K^c$  favors a clear clustering with respect to configurations (type of interfaces).

Combinations of the three matrices combine materials information to highlight multiple trends simultaneously: for example, a combination of  $(0,1,1)$  separates configurations while sorting structures according to  $p(\theta)$ , which can be interpreted as a score for selecting potentially interesting structures for closer inspection. Interestingly, a combination of  $(1,0,1)$  generates a sorting roughly corresponding to  $p(\theta)$  even if information from  $K^p$  is not used. Only very few predictions are associated with relatively high uncertainty. For example, one case is associated with multiple artificial band crossings at very low twist angles for  $3G/G@θ$ .

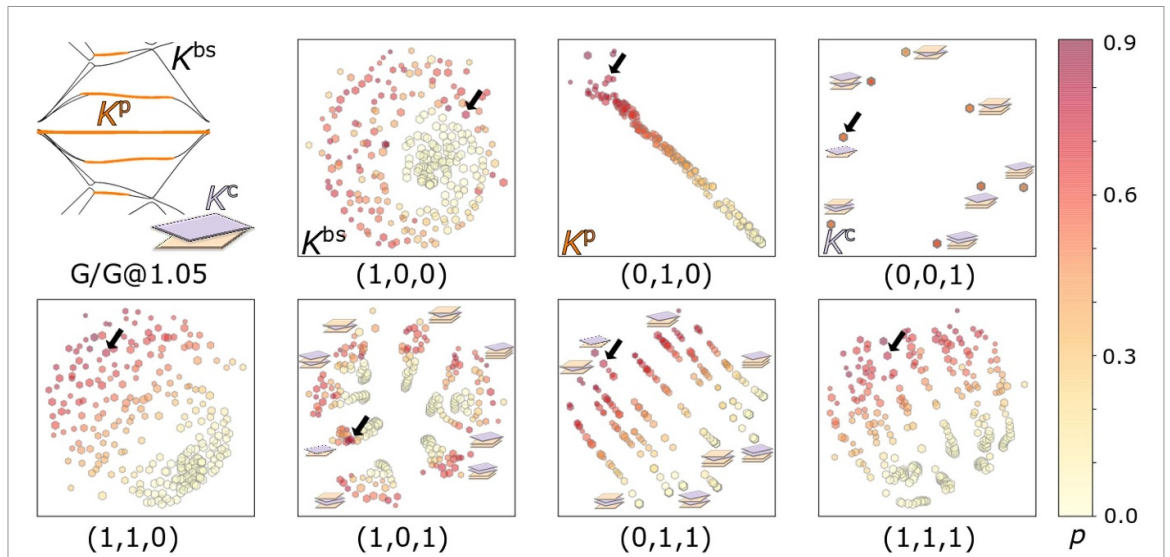
### 3.3. Other layered assemblies

We next extend our exploration to systems with more than four layers. Within the context of inverse materials design, our implementation of the Quantum Expert can also be used to answer questions of the form: "*what is the structure of a novel 2D layered assembly that could exhibit interesting electronic properties?*"

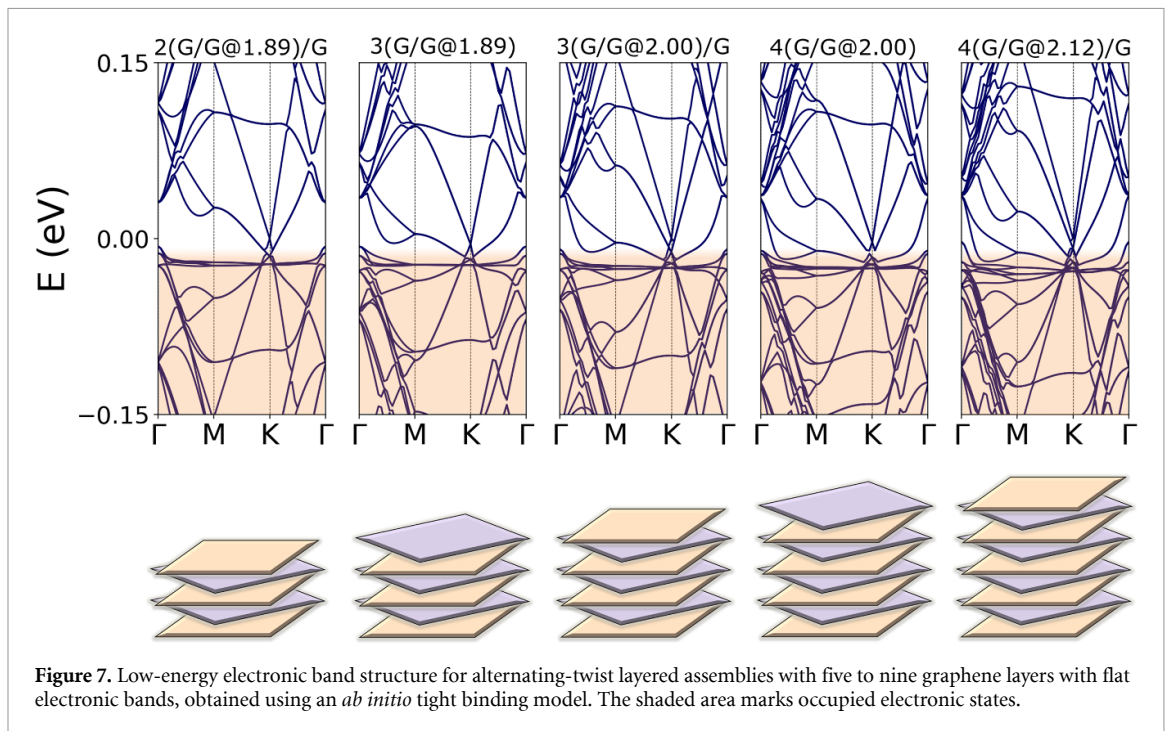
A pedestrian approach involves the use of the 2D embedding  $(0,1,1)$  shown in figure 6, i.e. using a combination of the dissimilarity matrices  $K^p$  and  $K^c$ . Informed by our findings for the three- and four-layer assemblies with alternating relative twist angle, we use it to identify  $\theta^*$  that maximize  $p(\theta)$  for the five-layer configuration  $2(G/G@θ)/G$ . First, we vectorize the structures according to SOI and then seek the nearest neighbors in the 2D embedding. With respect to cosine distances the two most similar configurations are  $G/G@θ/G$  and  $G/G@θ/G@θ$ , while  $3G/G@θ$  is the least. The nearest neighbors are associated with twist angles around  $1.1^\circ$  and  $1.6^\circ$ .

Another approach would rely on the use of a separate estimator (or optimizer) in the role of the researcher requesting and using the predictions  $p(\theta)$  for black-box optimization and adaptive search of the materials space [49, 50]. We demonstrate this modality of materials discovery by employing an extremely randomized trees (a.k.a. extra-trees) model for regression [51], as implemented in the scikit-learn machine learning library [47]. Picking interesting candidate structures for detailed investigation implies decision. Extra trees is an ensemble-based estimator that fits a number of randomized decision trees and predicts through averaging the predictions of all trees to control over-fitting. We use SOI vectors augmented with twist angles to train a forest of 100 trees of maximum depth of 8 for regression. These hyperparameters were obtained using 30-fold cross-validation (average coefficient of determination  $R^2$  of 0.7). By computing feature importances, we verify the twist angle to be by far the most important and therefore strongly correlated with  $p(\theta)$ , followed by the number of  $G/G@θ$  (or  $G@θ/G$ ) and  $G/G@θ/G$  interfaces. The trained meta-model codifies a structure-property relationship leading to results in good agreement with full TB calculations,  $\theta^* = 1.9^\circ$  (error of 15%; see also figure 7).

These findings motivate us to look at general trends in the electronic structure of alternating-twist layered assemblies. According to the work of Khalaf *et al* [42], for alternating-twist layered assemblies the magic angle is expected to reach the limiting value of  $2.2^\circ$  as the number of layers increases. To test this theoretical prediction, but also to demonstrate the capabilities of our models and methods, we obtain  $p(\theta)$  for layered assemblies with up to twenty layers of graphene with twist angle around  $2.0^\circ$ . We find that the assemblies with nine and up to twenty layers exhibit overlapping flat bands in the corresponding TB band structure at  $\theta^* = 2.1^\circ$  (figure 7). We ascribe this relatively small discrepancy (5%) to the higher accuracy of our *ab initio* TB approach compared to the model Hamiltonians used in the aforementioned study. As in the case of three- and four-layer assemblies, magic angle flat bands coexist with Dirac bands at the moiré K-point. The generally good agreement between theoretical predictions, full TB calculations,



**Figure 6.** Visualizations of the materials space using multidimensional scaling (MDS) with different combinations ( $a, b, c$ ) of three dissimilarity matrices  $K^{bs}$  (based on the calculated band energies),  $K^p$  (based on the probabilities  $p(\theta)$ ), and  $K^c$  (based on configurations). Points are color-coded with respect to predictions  $p(\theta)$  for low-dispersion electronic bands in the band structure (white for small; red for high) and the symbol size is proportional to the associated uncertainty (larger for higher precision). The black arrows indicate where the same layered assembly, G/G@1.05, can be found in each panel. The top left panel shows its band structure, with flat segments marked in orange (Hough transform).



**Figure 7.** Low-energy electronic band structure for alternating-twist layered assemblies with five to nine graphene layers with flat electronic bands, obtained using an *ab initio* tight binding model. The shaded area marks occupied electronic states.

and data-driven regression in retrospect validates SOI for representing structures.

In summary, the materials knowledge creation cycle can be outlined as follows:

- (a) Define the objective as the identification of twisted layered assemblies of graphene with potentially interesting electronic behavior.
- (b) Calculate and analyze the band structures of a set of layered assemblies and use them to predict such behavior.
- (c) Train an estimator for predictions of layered assemblies not contained in the database.
- (d) Identify promising candidate structures from step 3 and return to step 2 to assess them by full physics-based calculations.

In other words, the predictions  $p(\theta)$  generated in step 2 are used to support decision making in what can be thought as a basic active learning approach for autonomous (closed-loop) exploration of the materials space. Alternatively, the extra-trees regressor can be used as another predictive data-driven model

available directly to the inference engine for faster and broader predictions that become increasingly more accurate as the knowledge base expands. Conducting comprehensive model selection or an extensive exploration of the space of twisted layered assemblies with five or more graphene sheets nevertheless remains beyond the scope of this work.

#### 4. Conclusions

We introduced the unifying conceptual framework of an automated *in silico Quantum Expert* for the creation, combination, and use of materials information. We used it specifically to screen multi-layer graphene-based assemblies for interesting electronic behavior as an application. Materials calculations relied on an *ab initio* tight binding model. Their analysis was based on various featurizations, including an approach based on computer vision (Hough transform) to construct measures of band flatness, and *sum over interfaces* descriptors for the studied structures. We further produced informative visualizations of the materials space, or an *atlas of interface complexity*, by means of embeddings in two dimensions. The various transformations of materials information were enabled by a flexible *layered assemblies notation*, which we used to enumerate structures in the database and unambiguously describe trends across configurations.

Structural relaxation effects have been shown to modify the band structure significantly in the regime of small twist angles in twisted bilayer graphene: single-particle gaps near the Fermi level become wider, although the magic angle remains unchanged [41]. These effects are not accounted for in the current high-throughput workflow, but can be important even for configurations beyond the twisted bilayer. For instance, using a continuum model for interlayer interactions, Zhu *et al* [52] showed that the relaxation patterns of twisted trilayer of graphene are “moiré of moiré” as a result of the incommensurate coupling of two bilayer moiré patterns. An obvious extension of the current work therefore would include atomic relaxation. In addition, external conditions and perturbations such as pressure, temperature or a magnetic field could be considered as they are used to control the properties of a layered assembly. Here, we considered only high-symmetry directions in the Brillouin zone for calculation and analysis of electronic band structures, but more detailed physical insights could be obtained by densely sampling the entire Brillouin zone. The use of more refined methods for physics-based calculations, for example using more comprehensive model Hamiltonians and data-driven analysis with more extensive model selection, is also expected to improve the accuracy of predictions with regards to magic angles.

Since our approach relies on calculations within the single-particle picture, the TB band structures

alone cannot provide quantitative insights into correlated behaviors observed in experimental measurements. Combining materials data and information from virtual (computer) and physical experiments into common knowledge bases is expected to improve the usefulness of our approach: imagine a synthesis tool that enables rapid fabrication of layered assemblies on a controlled manner with integrated *in situ* characterization, acting either in the role of the researcher or augmenting the knowledge base for more reliable predictions. A common practical concern is related to the synthesizability of materials predicted using modeling and simulation. Our atlas (figure 6) provides a quick guide for picking promising candidates that are also structurally similar to known materials, making it easier to evaluate feasibility and needed resources for experimental verification. Moreover, in a Bayesian framework the knowledge obtained from unfruitful experiments can also be incorporated in the materials knowledge cycle by updating predictions, i.e. the probabilities  $p(\theta)$ , accordingly.

At the computational level, the presented framework is completely general and should be applicable to the accelerated investigation, optimization and screening of materials more broadly, describing in a uniform way such workflows [16, 17]. It is straightforward to extend our computational strategy to explore other scientifically and technologically interesting 2D layered materials, including layered assemblies of transition metal dichalcogenides. At the algorithmic level, constructing a Hamiltonian for other materials will necessarily require different sets of tight binding coefficients. Appropriate parametrizations could be tabulated or more generally obtained on demand by extending the capabilities of the workflow engine. At the implementation level, the modularity of our high-throughput workflow further facilitates the introduction of novel code functionality.

Our streamlined approach and specific findings provide insights into the electronic structure of twisted multi-layer graphene-based superlattices and offers reference for experimental measurements. Overall, this work aims to build confidence in the combined use of physics-based and data-driven modeling for the systematic refinement of knowledge about 2D layered materials, with implications for the development of novel quantum devices.

#### Acknowledgments

G.A.T. would like to acknowledge Pavlos Protopapas and Ricardo Hausmann for stimulating discussions. This work was supported in part by the U.S. Department of Energy (DOE) Office of Science (Basic Energy Science) under Award No. DE-SC0019300 and by the U.S. National Science Foundation's Science



and Technology Center for Integrated Quantum Materials under Grant No. DMR-1231319. S.B.T. received support from the DOE Computational Sciences Graduate Fellowship (DOE CSGF) under grant number DE-FG02-97ER25308.

Electronic structure calculations were performed on the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Pilot calculations were performed on the Cannon cluster, supported by the FAS Division of Science Research Computing Group at Harvard University. We relied on computational resources of the National Energy Research Scientific Computing Center, a DOE facility operated under Contract No. DE-AC02-05CH11231, for the provisioning of databases supporting the computational work.

## Supplementary information

Calculated tight binding band structures of the studied twisted graphene superlattices of up to four layers are available on Materials Cloud Archive 2020.0050/v1 (2020), doi: [10.24435/materialscloud:2020.0050/v1](https://doi.org/10.24435/materialscloud:2020.0050/v1).

## ORCID iDs

Georgios A Tritsarlis  <https://orcid.org/0000-0002-5738-4493>

Stephen Carr  <https://orcid.org/0000-0002-2749-8625>

Steven B Torrisi  <https://orcid.org/0000-0002-4283-8077>

Marios Mattheakis  <https://orcid.org/0000-0001-6450-1128>

Daniel T Larson  <https://orcid.org/0000-0001-8528-0280>

Efthimios Kaxiras  <https://orcid.org/0000-0002-4682-0165>

## References

- [1] Geim A K and Grigorieva I V 2013 *Nature* **499** 419
- [2] Novoselov K S, Mishchenko A, Carvalho A and Neto A H C 2016 *Science* **353** aac9439
- [3] Zhou J et al 2018 *Nature* **556** 355
- [4] Cheon G, Duerloo K-A N, Sendek A D, Porter C, Chen Y and Reed E J 2017 *Nano Lett.* **17** 1915
- [5] Mounet N et al 2018 *Nat. Nanotechnol.* **13** 246
- [6] Carr S, Massatt D, Fang S, Cazeaux P, Luskin M and Kaxiras E 2017 *Phys. Rev. B* **95** 075420
- [7] Cao Y, Fatemi V, Fang S, Watanabe K, Taniguchi T, Kaxiras E and Jarillo-Herrero P 2018a *Nature* **556** 43
- [8] Cao Y et al 2018b *Nature* **556** 80
- [9] Bistrizter R and MacDonald A H 2011 *Proc. of the National Academy of Sciences* vol **108** p 12233–12237
- [10] Liu X, Hao Z, Khalaf E, Lee J Y, Watanabe K, Taniguchi T, Vishwanath A, and Kim P arXiv:1903.08130 [cond-mat] (2019) arXiv:1903.08130
- [11] Suárez Morell E, Pacheco M, Chico L and Brey L 2013 *Phys. Rev. B* **87** 125414
- [12] Tritsarlis G A and Rossmeisl J 2012 *J. Phys. Chem. C* **116** 11980
- [13] Chen W et al 2016 *J. Mater. Chem. C* **4** 4414
- [14] Jóhannesson G, Bligaard T, Ruban A, Skriver H, Jacobsen K and Nørskov J 2002 *Phys. Rev. Lett.* **88** 2555061
- [15] Torrisi S, Singh A, Montoya J and Persson K 2019 arXiv:1912.09545v1
- [16] Tran K and Ulissi Z W 2018 *Nat. Catalysis* **1** 696
- [17] Bassman L et al 2018 *npj Comput. Mater.* **4** 74
- [18] Haastrup S et al 2018 *2D Mater.* **5** 042002
- [19] Tritsarlis G A, Shirodkar S N, Kaxiras E, Cazeaux P, Luskin M, Plecháč P and Cancès E 2016 *J. Mater. Res.* **31** 959
- [20] Fang S and Kaxiras E 2016 *Phys. Rev. B* **93** 235153
- [21] Jain A et al 2015 *Concurrency Comput. Pract. Exp.* **27** 5037
- [22] Curtarolo S et al 2012 *Comput. Mater. Sci.* **58** 218
- [23] Pizzi G, Cepellotti A, Sabatini R, Marzari N and Kozinsky B 2016 *Comput. Mater. Sci.* **111** 218
- [24] Sun S et al 2019 *Joule* **3** 1437
- [25] Botu V and Ramprasad R 2015 *Int. J. Quantum Chem.* **115** 1074
- [26] Cubuk E, Schoenholz S, Rieser J, Malone B, Rottler J, Durian D, Kaxiras E and Liu A 2015 *Phys. Rev. Lett.* **114** 108001
- [27] Janet J P, Duan C, Yang T, Nandy A and Kulik H J 2019 A quantitative uncertainty metric controls error in neural network-driven chemical discovery *Chem. Sci.* **10** 7913–22
- [28] Sparks T D, Gaultois M W, Oliynyk A, Brgoch J and Meredig B 2016 Data mining our way to the next generation of thermoelectrics *Scripta Materialia* **111** 10–15
- [29] Liao S-H 2005 *Expert Syst. Appl.* **28** 93
- [30] Turban E, Aronson J E and Liang T-P 2004 edn *Decision Support Systems and Intelligent Systems* 7th ed (address Upper Saddle River, NJ: Prentice Hall)
- [31] Carr S, Massatt D, Torrisi S B, Cazeaux P, Luskin M and Kaxiras E 2018 *Phys. Rev. B* **98** 224102
- [32] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
- [33] Fang S, Kuate Defo R, Shirodkar S N, Lieu S, Tritsarlis G A and Kaxiras E 2015 *Phys. Rev. B* **92** 205108
- [34] Bradski G 2000 The openCV library *Dr. Dobbs' Journal of Software Tools*
- [35] Weininger D 1988 *J. Chem. Inf. Comput. Sci.* **28** 31
- [36] D. Systems SMARTS—a language for describing molecular patterns (<https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>)
- [37] Tritsarlis G A, Xie Y, Rush A M, Carr S, Mattheakis M, and Kaxiras E 2019 arXiv:1910.03413
- [38] Uchida K, Furuya S, Iwata J-I and Oshiyama A 2014 Atomic corrugation and electron localization due to Moiré patterns in twisted bilayer graphenes *Phys. Rev. B* **90** 155451
- [39] Nam N N T and Koshino M 2017 *Phys. Rev. B* **96** 075311
- [40] Carr S, Fang S, Zhu Z and Kaxiras E 2019 *Phys. Rev. Res.* **1** 013001
- [41] Carr S et al 2020 Ultraheavy and Ultrarelativistic Dirac Quasiparticles in Sandwiched Graphenes *Nano Lett.* **20** 3030–8
- [42] Khalaf E, Kruchkov A J, Tarnopolsky G and Vishwanath A 2019 *Physical Review B* **100** 085109
- [43] Cea T, Walet N R and Guinea F 2019 *Nano Lett.* **19** 8683
- [44] Lee J Y, Khalaf E, Liu S, Liu X, Hao Z, Kim P and Vishwanath A 2019 *Nat. Commun.* **10** 5333
- [45] Shen C, Chu Y, Wu Q et al 2020 Correlated states in twisted double bilayer graphene *Nat. Phys.* **16** 520–525
- [46] Cao Y, Rodan-Legrain D, Rubies-Bigorda O, Park J M, Watanabe K, Taniguchi T, and Jarillo-Herrero P arXiv:1903.08596
- [47] Pedregosa F et al 2011 *J. Machine Learn. Res.* **12** 2825
- [48] Elton D C, Boukouvalas Z, Butrico M S, Fuge M D and Chung P W 2018 *Sci. Rep.* **8** 1
- [49] Roch L M, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker L P E, Hein J E and Aspuru-Guzik A 2018 (10.26434/chemrxiv.5953606.v1)
- [50] Dunn A, Brenneck J and Jain A 2019 *J. Phys.: Mater.* **2** 034002
- [51] Geurts P, Ernst D and Wehenkel L 2006 *Mach. Learn.* **63** 3
- [52] Zhu Z, Cazeaux P, Luskin M and Kaxiras E 2019 arXiv:1911.05324 [cond-mat] arXiv: 1911.05324