Density Functional Theory Analysis of hBN and MoS2 with SIESTA

NING Siyuan A0275835R

2023-11-24

1. Purpose

This report is dedicated to the computational study of two-dimensional (2D) materials, specifically hexagonal boron nitride (hBN) and molybdenum disulfide (MoS2), using the Density Functional Theory (DFT) methodologies implemented in the SIESTA software.

The report discusses practical DFT calculations in SIESTA[1], aligned with input file (i.fdf). It starts with pseudopotentials and then dives into the selection and implications of different exchange-correlation functionals, particularly emphasizing the generalized gradient approximation (GGA) PBE functional. The report also explores the basis set configuration in SIESTA, discussing the size, range, and shape of atomic orbitals and their impact on calculation precision and accuracy.

In terms of practical application, the report details the process of lattice optimization and band structure calculations for hBN and MoS2. This includes structural relaxation configurations, optimization methods (such as the Conjugate Gradient method), adjustments of lattice constants, and analysis of band structures as defined in the SIESTA simulations.

The findings from these computational analyses are then compared with experimental data and results from the Computational 2D Materials Database (C2DB)[2,3]. This comparison points out limitations of the DFT methodologies, particularly in the estimation of electronic bandgaps. Specifically, this report notes the discrepancies between DFT results and experimental observations, acknowledging the underestimation of bandgaps by Standard DFT and the varying accuracy when applying alternative methods like the HSE06 functional and the GW method.

Overall, the report aims to provide a thorough understanding of the computational strategies and outcomes in the study of 2D materials, offering valuable insights into their electronic properties and the potential of DFT methods in accurately modeling these systems.

2. Computational Method

In this section, we delve into the detailed Density Functional Theory (DFT) methods applied within the SIESTA framework. This discussion is specifically tailored to align with the parameters set in the i.fdf file, ensuring a coherent and targeted exploration of the computational techniques. Notably, for a better understanding and verification, the complete i.fdf file and a screenshot of the submission have been made available in the Supplementary Information section of the report.

2.1 Pseudopotential

SIESTA employs norm-conserving pseudopotentials for representing electron-ion interactions, a common approach in most plane-wave codes, contrasting with "all-electron" programs. These pseudopotentials can be generated using the Atom program, as detailed in its manual. Specifically designed for SIESTA, these pseudopotentials are typically saved in the .psf format. SIESTA reads the pseudopotentials from separate files for each species, which are named according to the convention: Chemical_label.psf (ASCII), where "Chemical_label" is the identifier specified in the ChemicalSpeciesLabel block as presented below.

```
%block ChemicalSpeciesLabel
  1  42 Mo
  2  16 S
%endblock ChemicalSpeciesLabel
```

2.2 Exchange-Correlation Functional and Solver

The computationally most efficient functionals are the local density approximation (LDA), the generalized gradient approximations (GGA) and the meta-GGA functionals. The GGA and metaGGA functionals are of the semilocal type and a plethora of them have been proposed. The GGA PBE functional, implemented in SIESTA, is the most commonly used in solid-state physics. However, note that the LDA and GGA functionals do not provide a correct description of the band gap. Whether the Kohn-Sham (KS) orbitals $\psi_{n\mathbf{k}}(\mathbf{r})$ reflect the right physics crucially depends on the type of the used exchange-correlation (xc) functional. In this respect meta-GGA and hybrid functionals are more appropriate for band gap calculation.

Hybrid functionals make up a special category of xc functionals that mix some fraction of Fock exchange into a semilocal functional. This approach is inspired by the well-known Hartree-Fock (HF) method, which is the exact solution for uncorrelated systems. Regarding extended systems, it aims to cure the infamous underestimation of band gaps by the LDA and GGA. The nonlocal Fock potential reads as

$$V_{\mathrm{X}}^{\mathrm{HF}}(\mathbf{r},\mathbf{r}') = -\frac{e^2}{2}\sum_{n\mathbf{k}} f_{n\mathbf{k}} \frac{\psi_{n\mathbf{k}}^{*}(\mathbf{r}')\psi_{n\mathbf{k}}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|},$$

where $f_{n\mathbf{k}}$ is the occupation of band n and at a specific $\mathbf{k}$ point, $\mathbf{r}$ is the position and e the charge of an electron. Above equation depends on the KS orbitals instead of the density and thus, hybrid functionals are not density functionals.

PBE[4] is one of the hybrid functionals available in SIESTA and has the following composition:

$$E_{\mathrm{xc}}^{\mathrm{PBE0}} = \frac{1}{4}E_{\mathrm{x}}^{\mathrm{HF}} + \frac{3}{4}E_{\mathrm{x}}^{\mathrm{PBE}} + E_{\mathrm{c}}^{\mathrm{PBE}}.$$

Here, $E_{\mathrm{X}}^{\mathrm{HF}}$ is the Fock exchange energy, while $E_{\mathrm{X}}^{\mathrm{PBE}}$ and $E_{\mathrm{c}}^{\mathrm{PBE}}$ are the PBE exchange and correlation energies, respectively. Compared to many other functionals, PBE contains no parameter that were fitted to experimental data.

In the SIESTA software, there are many solvers available for calculating the density kernel, including different flavors of Diagonalization, the Orbital Minimization Method (OMM), PEXSI, and CheSS. The choice of the most appropriate method depends on several factors such as the size of the system, the sparsity of the matrices, the HOMO-LUMO gap, and the basis set used, among others. Among these methods, the most straightforward approach to calculate the density kernel is to first solve the generalized eigenvalue problem $\boldsymbol{H}c_i = \epsilon_i \mathbf{S}\mathbf{c}_i$, and then compute the density kernel as K= $\sum_i \mathbf{c}_i\, \mathbf{c}_i^T$. This method boasts the advantage of being an exact calculation without any approximations and is universally applicable, with highly optimized libraries available. However, it also has its shortcomings, such as the inability to exploit the potential sparsity of the matrices, cubic scaling with the size of the system, and challenges in parallelization.

In the `i.fdf` file of this project, the classic GGA PBE functional and the basic Diagonalization solver have been selected.

```
XC.functional   GGA              #exchange-correlation functional
XC.authors      PBE

SolutionMethod   diagon          #default diagonalization solution of the KS Hamiltoni
an
```

2.3 Basis Set

In the following basis sets that SIESTA generates. The starting point is the solution of Kohn-Sham's Hamiltonian for the isolated pseudo-atoms, solved in a radial grid, with the same approximations as for the solid or molecule (the same exchange-correlation functional and pseudopotential), plus some way of confinement (as below). Three main features of a basis set of atomic orbitals will be discusse: size, range, and radial shape.

Size (number of orbitals per atom):

SIESTA offers a range of basis sets[5], from **single-ζ (SZ)** to more complex ones, allowing for varying degrees of calculation precision. The SZ basis set includes one radial function per angular momentum channel for each atom, focusing on the valence electrons. It's suitable for quick, qualitative analyses.

The **double-ζ (DZ)** basis set enhances precision by adding a second function per channel. It's constructed by splitting one atomic orbital's Gaussian expansion into two: the more contracted Gaussians define the first orbital, and the more extended ones form the second.

In SIESTA, achieving convergence in the basis set often requires higher levels of multiple-ζ, compared to other methods. To improve this, SIESTA introduces a numerical orbital to the first ζ, which replicates the tail of the original **pseudo atomic orbital (PAO)** outside a matching radius r_m. This numerical orbital transitions smoothly to the form $r^l(a - br^2)$ at the origin, with parameters a and b ensuring continuity and differentiability at r_m. The second orbital in the basis set is then defined as the difference between this smooth orbital and the original PAO, restricted within r_m.

Users can manually introduce r_m for each orbital via `PAO.Basis`, or define all r_m values simultaneously by setting a `split norm`. This parameter, typically optimized to around 15%, dictates the tail value of the original PAO beyond r_m. This concept extends naturally to multiple-ζ sets by adding an additional matching radius for each new ζ.

SIESTA also allows for angular flexibility by adding higher angular momentum shells, or polarization orbitals[6]. These can be generated in two ways: using confined atomic PAO's of higher angular momentum, or by solving the pseudoatom in an electric field and deriving the l + 1 orbitals from the l orbitals' perturbation by the field. The first method generally yields better results. Additionally, for open systems like surfaces, diffuse orbitals can be included by adding extra "n" shells.

SIESTA supports the inclusion of off-site (ghost) orbitals, not centered on any atom, which are useful in certain calculations like the counterpoise correction for basis-set superposition errors. The software also allows the addition of Bessel functions for any radius and excitation level to the basis set.

Range (Cutoff Radii of Orbitals):
In SIESTA, strictly localized orbitals, terminating at a specific cutoff radius, enable sparse Hamiltonian and overlap matrices for efficient linear scaling. Each species requires a distinct cutoff radius for each angular momentum channel. A unified approach to setting these radii involves the `energy shift` parameter, indicating the energy increase when an orbital is confined. Typically, an `energy shift` of approximately 100 meV aligns with the precision of General Gradient Approximation (GGA) functionals.

Shape (Orbital Consistency and Localization):
Consistency between pseudopotentials and the shape of pseudoatomic orbitals, especially in the core region, is crucial within the pseudopotential framework. The orbital shape at larger radii is influenced by the cutoff radius and the method of localization enforcement.

The standard approach in SIESTA involves using an infinite square-well potential, which, despite its effectiveness, often results in orbitals with a discontinuous derivative at the **cutoff radius** r_c. This issue is more pronounced for smaller r_c values and less so for larger ones, but it remains noticeable for realistically sized orbitals. Surprisingly, such discontinuities have minimal impact on the total energy of condensed systems.

Addressing the limitations of this scheme, especially for free atoms with extremely extended orbitals, SIESTA offers alternatives. One method involves radial scaling of orbitals, which adheres to the virial theorem in finite bases for Coulombic potentials. However, with pseudopotentials, this method permits only minor deviations from unity in scale factors, except for hydrogen atoms.

A more effective solution is to add a soft confinement potential to the atomic Hamiltonian for basis orbital generation. This potential, which starts at an internal radius r_i and diverges at r_c, ensures strict localization and smoothens the discontinuity. The potential is defined as:

$$V(r) = V_{\mathrm{0}} \frac{e^{-\frac{r_c - r_i}{r - r_i}}}{r_c - r}$$

with r_i and V_o being adjustable parameters. SIESTA's default soft confinement settings, typically adequate, are $ri = 0.9r_c$ and $V_o = 40$ Ry.

For orbitals that are typically empty in the atom, like polarization orbitals, SIESTA uses charge confinement to ensure more accurate shapes. This approach adds a Z/r potential to the soft confinement, with the potential given by:

$$V_{\mathrm{Q}}(r) = \frac{Ze^{-\lambda r}}{\sqrt{r^2 + \delta^2}}$$

Here, δ prevents singularity, and λ allows for tail length adjustment.

Lastly, the ionic character of an atom affects orbital shape. Cation orbitals tend to shrink, while anion orbitals expand. Introducing a δQ in free-atom calculations for basis generation can produce orbitals more suited to these ionic situations in condensed systems.

The following part details the meanings of the parameters due to specific setting in i.fdf.

```
#---Options for Basis Set------------------------------------------------------
PAO.BasisType   split    # Default split, split-valence scheme for multiple-zeta, based
```

```
 on different radii
PAO.BasisSize   DZP      # Default DZP double zeta with polarization orbitals basis set
PAO.SplitNorm   0.15     # Default 0.15
PAO.EnergyShift 0.02 Ry  # Default 0.02 Ry  a standard for orbital-confining cutoff rad
ii
```

`PAO.BasisType split` split defines the kind of basis to be generated is chosen. All are based on finite-range pseudo-atomic orbitals. The original PAO's were described only for minimal bases. SIESTA generates extended bases (multiple- ζ, polarization, and diffuse orbitals) applying different schemes of choice:

- Generalization of the PAO's: uses the excited orbitals of the finite-range pseudo-atomic problem, both for multiple- ζ and for polarization. Adequate for short-range orbitals.
- Multiple- ζ in the spirit of split valence, decomposing the original PAO in several pieces of different range, either defining more (and smaller) confining radii, or introducing Gaussians from known bases.

The string `split` stands for split-valence scheme for multiple-zeta. The split is based on different radii.

`PAO.BasisSize DZP` specifies usual basis sizes. `DZP` is for standard Like DZ with polarization orbitals. Polarization orbitals are constructed from perturbation theory, and they are defined so they have the minimum angular momentum l such that there are not occupied orbitals with the same l in the valence shell of the groundstate atomic configuration. They polarize the corresponding $l-1$ shell. There are also other choices:

- `SZ` | minimal Use single- ζ basis.
- `DZ` Double zeta basis, in the scheme defined by PAO.BasisType.
- `SZP` Single-zeta basis plus polarization orbitals.

`PAO.EnergyShift 0.02 Ry`indicates the orbital-confining cutoff radii. It is the excitation energy of the PAO's due to the confinement to a finite-range. It offers a general procedure for defining the confining radii of the original (first-ζ) PAO's for all the species guaranteeing the compensation of the basis. It only has an effect when the block PAO. Basis is not present or when the radii specified in that block are zero for the first zeta.

`PAO.SplitNorm 0.15` defines sensible default radii for the split-valence type of basis. It gives the amount of norm that the second- ζ split-off piece has to carry. The split radius is defined accordingly. If multiple- ζ is used, the corresponding radii are obtained by imposing smaller fractions of the SplitNorm $(1/2,1/4 \ldots)$ value as norm carried by the higher zetas. For hBN calculations, a default SplitNorm of 0.15 is suitable, while for MoS2, it should be adjusted to 0.15392.

```
PAO.SplitNorm   0.15392     # Default 0.15
```

2.4 K-sampling

Normally, there are two sampling methods for K point mesh.

In ***Γ*-centered mesh**, k-points in the Brillouin zone are sampled starting from the Γ point. The k-points are generated using the equation $\mathbf{k} = \sum_{i=1}^{3} \frac{n_i + s_i}{N_i} \mathbf{b}_i$, where n_i ranges from 0 to $N_i - 1$, s_i is a shift parameter, and $\mathbf{b}_i$ are the reciprocal lattice vectors. This method is typically used when the Γ point's inclusion is essential for the study.

The **Monkhorst-Pack** method creates a uniform k points grid with introducing an offset to Γ point. The k-points are determined by the equation $\mathbf{k} = \sum_{i=1}^{3} \frac{n_i + s_i + \frac{1-N_i}{2}}{N_i} \mathbf{b}_i$, where n_i are integers within the range $[0, N_i)$, and s_i accounts for any shifts. This method is widely adopted for its comprehensive coverage of the Brillouin zone.

For the k sampling part in SIESTA `.fdf` file, the `%block kgrid_Monkhorst_Pack` section defines a Monkhorst-Pack mesh with a specific configuration. Here, a 4x4x1 grid is set with half-grid shifts in the first and second directions (indicated by `0.5`) and no shift in the third direction (`0.0`). This setup ensures a uniformly distributed set of k-points within the defined parameters. K-point sampling often requires testing to balance efficiency and accuracy. In this project, a 4x4x1 grid has been found to effectively balance these aspects, yielding good computational results.

```
#---Options for k-Points Sampling---------------------------------------------
WriteKpoints                F     #default false, not writing the coordinates of k vec
tors for ksampling
%block kgrid_Monkhorst_Pack
    4    0   0   0.5               # K-sampling kpoints 4x4x1
    0    4   0   0.5
    0    0   1   0.0
%endblock kgrid_Monkhorst_Pack
```

2.5 SCF Loop

A typical Self-Consistent Field (SCF) cycle within Density Functional Theory (DFT) codes, such as SIESTA, involves a series of systematic steps. Initially, the Hamiltonian matrix $H_{\alpha\beta} = \langle \phi_\alpha | \mathcal{H} | \phi_\beta \rangle$ and the overlap matrix $S_{\alpha\beta} = \langle \phi_\alpha \mid \phi_\beta \rangle$ are calculated. These matrices are crucial in determining the electronic properties of the system. Following this, the density kernel **K** is computed from the Hamiltonian **H** and overlap **S** matrices. The subsequent step involves calculating the energy of the system as $E = \mathrm{Tr}(\mathbf{KH})$ and the electronic density $\rho(\mathbf{r}) = \sum_{\alpha,\beta} \phi_\alpha^* (\mathbf{r}) K_{\alpha\beta} \phi_\beta(\mathbf{r})$. Once these calculations are complete, the Hamiltonian operator H is updated based on the new electronic density ρ, and the cycle starts over.

The SCF loop in SIESTA depicted in the flow diagram below:

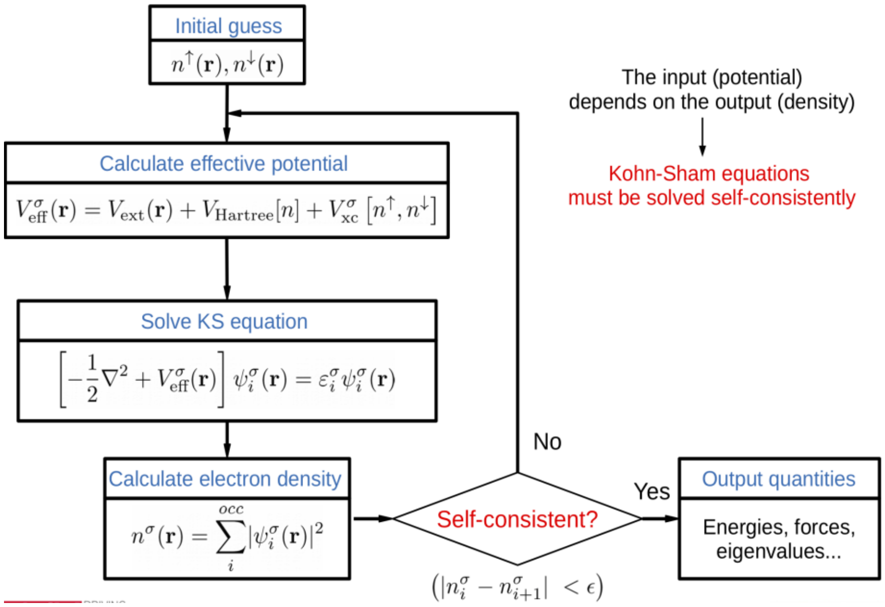

Initial guess
$n^{\uparrow}(\mathbf{r}), n^{\downarrow}(\mathbf{r})$
Calculate effective potential
$V^{\sigma}_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + V_{\text{Hartree}}[n] + V^{\sigma}_{\text{xc}}\left[n^{\uparrow}, n^{\downarrow}\right]$
Solve KS equation
$\left[-\frac{1}{2}\nabla^2 + V^{\sigma}_{\text{eff}}(\mathbf{r})\right]\psi^{\sigma}_i(\mathbf{r}) = \varepsilon^{\sigma}_i\psi^{\sigma}_i(\mathbf{r})$
Calculate electron density
$n^{\sigma}(\mathbf{r}) = \sum_{i}^{occ}|\psi^{\sigma}_i(\mathbf{r})|^2$
Self-consistent?
$(|n^{\sigma}_i - n^{\sigma}_{i+1}| < \epsilon)$
No
Yes
Output quantities
Energies, forces, eigenvalues...
The input (potential) depends on the output (density)
Kohn-Sham equations must be solved self-consistently

Figure 1 SCF circle in SIESTA

In most calculations, particularly in larger systems, the focused part within SIESTA is the computation of the density matrix. In the SIESTA `i.fdf` file, DM refers to the Density Matrix, which is used as perfix for many SCF parameters . `ElectronicTemperature 300 K` sets the temperature for electronic calculations. `OccupationFunction FD` defines the Fermi-Dirac function for electronic state occupation. `MeshCutOff 200 Ry` establishes the plane wave energy cutoff. `DM.UseSaveDM .True.` indicates the use of a saved density matrix. `DM.MixingWeight 0.02` specifies the mixing weight for the density matrix. `DM.NumberPulay 5` sets the number of iterations for Pulay mixing. `DM.Tolerance 1.d-5` defines the convergence tolerance for the density matrix. `DM.RequireEnergyConvergence .True.` requires energy convergence in the calculations. `DM.EnergyTolerance 1.d-4 eV` sets the energy convergence threshold. `MaxSCFIterations 500` limits the number of self-consistent field iterations.

```
#---Options for Initialisation and SCF Loop------------------------------------
ElectronicTemperature        300 K
OccupationFunction           FD          # function determining the occupation of the e
lectronic states
MeshCutOff                   200 Ry      # default 100 Ry, plane wave energy cutoff

DM.UseSaveDM                  .True.               # whether or not to read previous de
nsity matrix
DM.MixingWeight               0.02                 # Default 0.25
DM.NumberPulay                5                    # Default 0
DM.Tolerance                  1.d-5                # Default convergence criteria for D
M
DM.RequireEnergyConvergence   .True.
DM.EnergyTolerance            1.d-4 eV             # Energy convergency criteria
MaxSCFIterations              500                  # Maximum number of SCF iterations
```

3. Results and Discussion

In the thsi section, lattice optimization and structural analysis of 2D materials, with a focus on hBN and MoS2, are discussed. Utilizing the Conjugate Gradient method for optimization in SIESTA simulations, parameters are carefully adjusted to stabilize structures. This involves automated scripts for lattice constant testing and energy calculations, leading to subtle yet crucial modifications in atomic coordinates and bond lengths. The band structures of hBN and MoS2 are calculated, each showing unique band gaps.

3.1 2d-BN

3.1.1 Lattice Optimization

In the `i.fdf` file for SIESTA simulations, structural relaxation configurations are defined by a series of parameters.

```
#---Options for Structural Relaxations-----------------------------------------
MD.TypeOfRun                 CG                    # structure optimization option
MD.NumCGSteps                500
MD.VariableCell              .False.               #default, no lattice relaxation
MD.ConstantVolume            .False.               # default, no change in lattice volume
MD.MaxForceTol               0.02 eV/Ang           # Default 0.04 eV/Ang, force tolerance
MD.MaxCGDispl                0.03 Bohr             # Default 0.20 Bohr
MD.UseSaveCG                 T
MD.UseSaveXV                 T
#SaveRho                     T
DM.UseSaveDM                 T
```

`MD.TypeOfRun CG` specifies the use of the Conjugate Gradient (CG) method, an optimization approach based on the principle of energy minimization, which iteratively adjusts atomic positions to reach the lowest energy state. The `MD.NumCGSteps`, set to a value like 100, determines the number of iterative steps in the CG algorithm, indicating the frequency of atomic position adjustments during optimization. The `MD.MaxForceTol` parameter, typically set to `0.04 eV/Ang`, defines the force tolerance limit. This value is the maximum force allowed on atoms during optimization, and convergence is assumed when the forces on atoms fall below this threshold. On the other hand, `MD.MaxCGDispl`, commonly set to `0.2 Bohr`, restricts the maximum displacement of atoms in each optimization step to prevent structural instability due to excessive movements. When `MD.VariableCell` is set to `true`, it indicates that the cell dimensions can vary during optimization, adapting the lattice parameters to minimize the total energy, essential for simulations where cell relaxation is required. However, in scenarios with fixed lattice parameters, like in simulations of layered materials such as hBN, this parameter is set to `false`, implying that only atomic positions are optimized while the cell dimensions remain unchanged.

In the context of SIESTA simulations, adjusting the `Lattice Constant` to find the most stable structure involves specific steps, especially when lattice parameters are fixed during structural optimization due to parameters like `MD.VariableCell` and `MD.ConstantVolume` being set to `false`. This setting necessitates manually adjusting the `Lattice Constant` in the `i.fdf` file and performing Self-Consistent Field (SCF) calculations for each varied lattice constant to compare the total energy after convergence, identifying the lowest energy structure. In practice, this involves modifying the `LatticeConstant` value and correspondingly adjusting the `LatticeVectors`. As the vectors are in fractional coordinates, only the Z-vector requires modification. Scripts such as `LatticeTest.sh` and `getdata.sh` are utilized to automate the submission of calculation jobs for different lattice constants and extract total energy from the results, determining the most stable lattice configuration.

To automate the process of finding the most stable lattice structure varying different lattice constants, custom scripts `LatticeTest.sh` and `getdata.sh` are used for bash submission and extract the total energy from the results.

In `LatticeTest.sh`, the script initializes the lattice constant range (`start=2.49`, `end=2.53`, `step=0.01`) and iterates over it using a `for` loop. The `z_coord` is calculated using `echo "scale=16; 12 / $lattice" | bc`, where `bc` is a command-line calculator, calculating the Z-vector length for the lattice vectors based on the current lattice constant. Directories for each lattice constant are created, and necessary template files are copied into them. The `sed` command, `sed -i "s/^LatticeConstant.*/LatticeConstant $lattice Ang/" $dir_name/i.fdf`, replaces the existing lattice constant and Z-component of the lattice vectorsin `i.fdf` with the current value. It also modifies the name of the PBS script correspondingly with sed. The script concludes each iteration by `qsub`. The script is displayed as follows:

```
for lattice in $(seq $start $step $end)
do
   z_coord=$(echo "scale=16; 12 / $lattice" | bc)
...
   sed -i "s/^LatticeConstant.*/LatticeConstant $lattice Ang/" $dir_name/i.fdf
   sed -i "/%block LatticeVectors/,/%endblock LatticeVectors/s/0\.0000000000000000  0\.0000000000000000  .*/  0.0000000000000000  0.0000000000000000  $z_coord/" $dir_name/i.fdf
   sed -i "s/#PBS -N hBN/#PBS -N hBN2_${lattice}/" $dir_name/hBN.pbs
...
done
```

The `getdata.sh` script extracts and compiles total energy data to .dat file. The script iterates over the same range of lattice constants, extracting the total energy from each simulation's output file. This is achieved with `grep "Total =" $lattice/siesta.out | awk '{print $4}'`, where `grep` searches for the line containing the total energy and `awk` extracts the specific energy value. The extracted energy is then appended to `energy.dat` using `echo -e "$lattice\t$energy" >> $output_file`. The script is displayed as follows:

```
...
for lattice in $(seq 2.49 0.01 2.53); do
    energy=$(grep "Total =" $lattice/siesta.out | awk '{print $4}')
    echo -e "$lattice\t$energy" >> $output_file
done
...
```

The obtained results indicate that the `Lattice Constant` at `2.51` yields the lowest `Total Energy` of `-386.657487`, suggesting it as the most stable structure. This optimized value is consistent with parameters from the C2DB database.

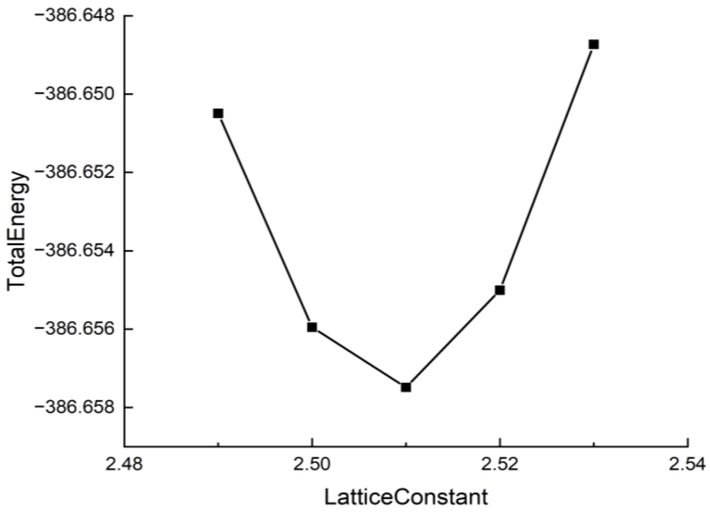


Figure 2 hBN total energy with different lattice constant

3.1.2 Optimized Geometry

The hBN.STRUCT_OUT file displays the optimized geometry coordinates as follows:

```
          2.512400000       0.000000000       0.000000000
         -1.256200000       2.175802224       0.000000000
          0.000000000       0.000000000      12.000000000
           2
  1     5       0.000217942      -0.000217942       0.166670170
  2     7       0.666458611       0.333541387       0.166670060
```

These vectors can also be extracted from the siesta.out file:

```
outcoor: Relaxed atomic coordinates (fractional):
    0.00021682   -0.00021683    0.16667010   1       1  B
    0.66645993    0.33354007    0.16667006   2       2  N
```

Compared to the input parameters:

```
%block AtomicCoordinatesAndAtomicSpecies
    0.00000   0.00000   0.16667     1        1    B
    0.66667   0.33333   0.16667     2        2    N
%endblock AtomicCoordinatesAndAtomicSpecies
```

There are only slight differences, which is expected since the input parameters are derived from an optimized CIF file.

SiESTA calculations also output bond information, available in the hBN.BONDS and hBN.BONDS_FINAL files.

Bond Lengths Before and After Structural Optimization

OUTPUT FILE	L1(bondlength)/Å	L2/Å	L3/Å

OUTPUT FILE	L1(bondlength)/Å	L2/Å	L3/Å
hBN.BONDS	1.4492	2.5100	2.8983
hBN.BONDS_FINAL	1.4482	2.5100	2.8992

Table 1 hBN bond lengths before and after optimization

L1 represents the bond length, and as seen, there is almost no difference before and after optimization.

Using Material Studio, the bond lengths L1, L2, and L3 can be visually demonstrated.

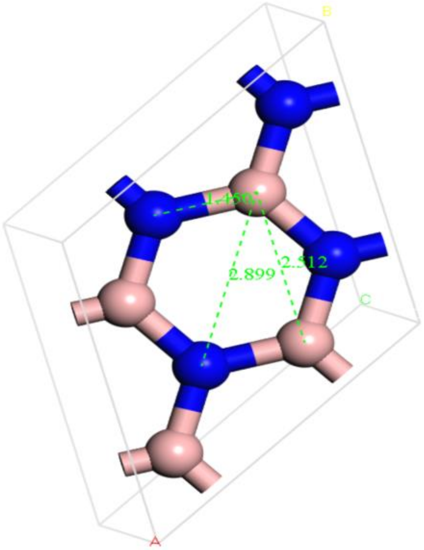


Figure 3 L1, L2, L3 in hBN Lattice

3.1.3 Bandstructure

The band structure in SIESTA simulations is defined in the `i.fdf` file under the `BandLines` block.

```
WriteBands          F                      # writing band information
BandLinesScale  ReciprocalLatticeVectors
%block BandLines
 1    0.000000   0.000000  0.000000   \Gamma
 30   0.333333   0.333333  0.000000   K
 30   0.500000   0.000000  0.000000    M
 45   0.000000   0.000000  0.000000   \Gamma
%endblock BandLines
```

The setting `BandLinesScale ReciprocalLatticeVectors` specifies that the coordinates for band structure calculations are in reciprocal lattice space. Within the `BandLines` block, the path through the Brillouin zone is outlined by key symmetry points. Starting from the `Gamma` point (`0.000000 0.000000 0.000000`), the path is segmented, moving to the `K` point (`0.333333 0.333333 0.000000`) over `30` segments. This segmentation allows for a detailed sampling along this high-symmetry direction. The path then proceeds to the `M` point (`0.500000 0.000000 0.000000`) with another `30` segments, and finally returns to the `Gamma` point over `45` segments. This detailed path through the reciprocal lattice space is essential to accurately map the electronic band structure, offering insights into the material's electronic properties.

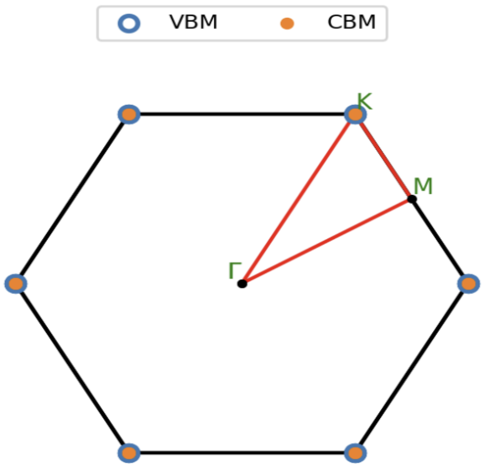


Figure 4 High symmetry points in reciprocal space

The calculated band structure is shown here, with the Fermi level adjusted to 0 eV. The band gap is 4.54 eV.

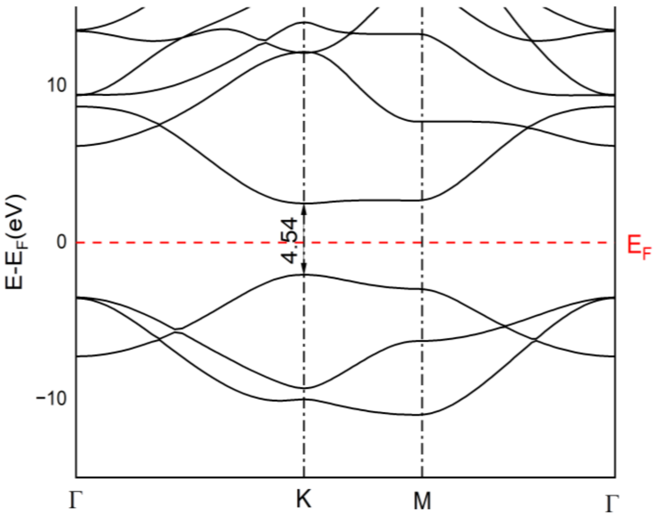


Figure 5 hBN calculated band structure

3.2 2d-MoS2

3.2.1 Lattice Optimization

As the methodologies and file configurations have been extensively discussed in the hBN section, this part will only briefly mention the results with adjustments specifically for MoS2. Despite the given CIF file suggesting a lattice constant of 3.20, computational results indicate that 3.21 is a more suitable structural parameter for MoS2.

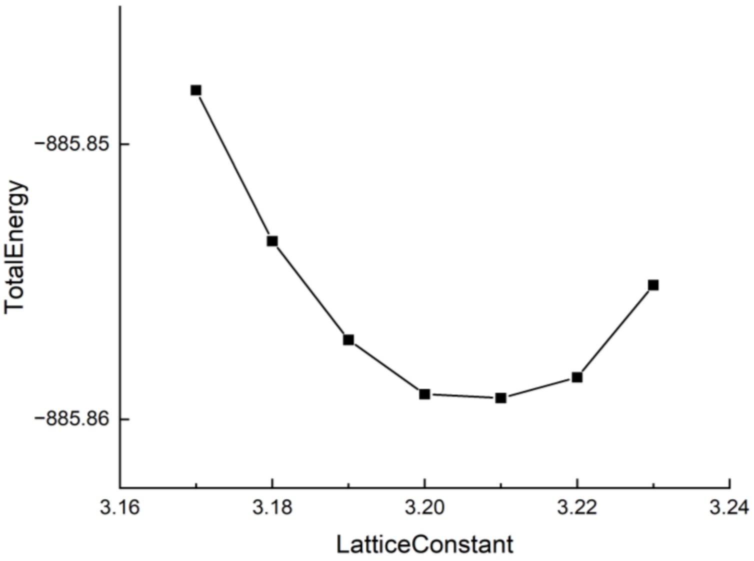


Figure 6 MoS2 total energy with different lattice constant

3.2.2 Optimized Geometry

The output MoS2.STRUCT_OUT file content:

```
          3.210000000       0.000000000       0.000000000
         -1.605000000       2.779941546       0.000000000
          0.000000000       0.000000000      18.000000000
           3
  1    42       0.666653133       0.333346911       0.140507043
  2    16       0.333270956       0.666729059       0.051680557
  2    16       0.333280310       0.666719644       0.229337766
```

Optimized parameters from siesta.out：

```
outcoor: Relaxed atomic coordinates (fractional):
    0.66665313    0.33334691    0.14050704   1       1  Mo
    0.33327096    0.66672906    0.05168056   2       2  S
    0.33328031    0.66671964    0.22933777   2       3  S
```

Atomic coordinates in input i.fdf:

```
%block AtomicCoordinatesAndAtomicSpecies
  0.666744471         0.333255500         0.141528144     1  1  Mo
  0.333312690         0.666687310         0.052778751     2  2  S
  0.333303303         0.666696727         0.230252653     2  3  S
%endblock AtomicCoordinatesAndAtomicSpecies
```

Upon comparison, it is evident that there is also a very minor change in the structural parameters.

The bond lengths in MoS2, as extracted from the MoS2.BONDS and MoS2.BONDS_FINAL files, show a subtle change from 2.4466 Å before optimization to 2.4476 Å after optimization, visualized in Material Studio.

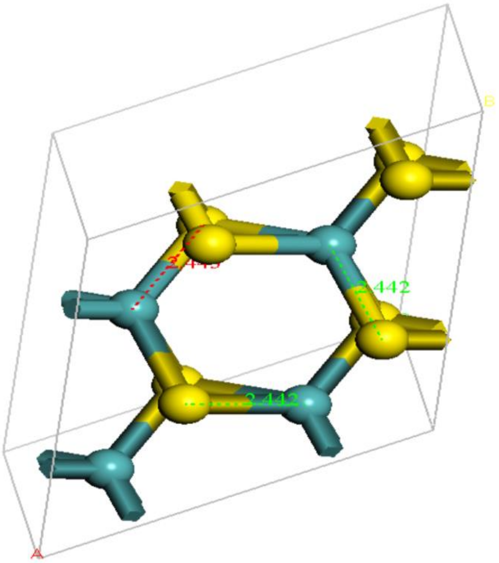


Figure 7 Bondlength in hBN Lattice

3.2.3 Bandstructure

The calculated band structure of MoS2 is shown here, with the Fermi level adjusted to 0 eV. The band gap is 1.68 eV.

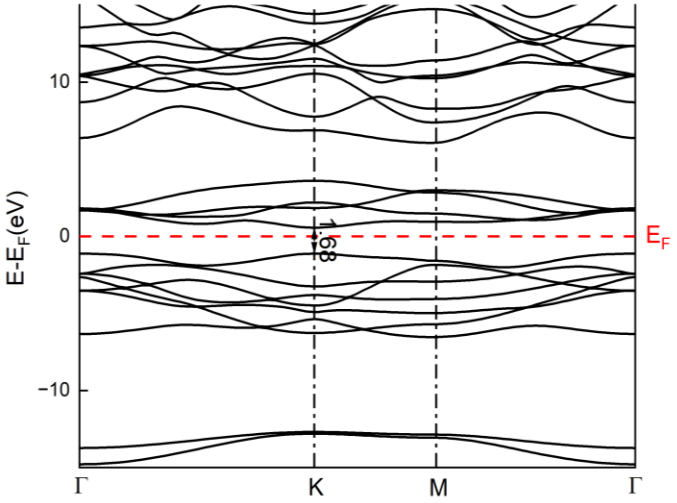


Figure 8 MoS2 calculated band structure

4. Conclusion

Isolating monolayer hBN by exfoliation from bulk hBN has proved challenging, although Elias *et al.* have recently succeeded in growing atomically thin samples of hBN on graphite substrates[7]. Their reflectance and photoluminescence measurements indicate a direct gap of **6.1 eV** for hBN on graphite. Pena Roman et al[8]. directly measured the density of states of a single monolayer of h-BN epitaxially grown on highly oriented pyrolytic graphite, by performing low temperature scanning tunneling microscopy (LT-STM) and spectroscopy (STS). The observed h-BN electronic band gap on defect-free regions is **(6.8 ± 0.2) eV**. Using optical spectroscopy to obtain the h-BN optical band gap, the exciton binding energy is determined as being of **(0.7 ± 0.2) eV**. In addition, the locally excited cathodoluminescence and photoluminescence show complex spectra that are typically associated to intragap states related to carbon defects. Moreover, in some regions of the monolayer h-BN we identify, using STM, point defects which have intragap electronic levels around 2.0 eV below the Fermi level.

Due to the strong exciton binding (~1 eV), the optical bandgap has been obtained to be about **1.8 eV** from the photoluminescence (PL) and optical absorption experiments[9]. Since the exciton binding energy is large, the measured optical bandgaps are not accurate representation of semiconductor bandgap, determined by the energy difference between valence and conduction band edges. With intercalated potassium (K) in a bulk MoS2, a quasi ML MoS2 has been fabricated from a bulk MoS2, and a direct bandgap of **1.86 eV** at the *K* valley has been measured using angle-resolved photo-emission spectroscopy (ARPES)[10]. For a chemical vapor deposition (CVD) grown ML MoS2 on a Au(111) substrate, the ARPES bandgap of about 1.39 eV has been measured, which is very small. In scanning tunneling spectroscopy (STS) measurements, the bandgap of a ML MoS2 on graphite substrate has been measured to be **2.15 eV**[11].

Notably, the electronic properties of an isolated hBN and MoS2 monolayer (i.e., a freely suspended sample) are at present only accessible through theoretical calculations[12].

In the context of the experimental data mentioned earlier, this section also introduces computational results from the C2DB database (Computational 2D Materials Database), comparing them with results obtained using the SIESTA computational method in our project.

The computational outcomes align closely with the Standard DFT results in the C2DB, indicating the success of our calculations. However, it's important to note that Standard DFT tends to underestimate electronic bandgaps. By applying the Heyd-Scuseria-Ernzerhof 2006 (HSE06) functional, we achieved conclusions consistent with experimental observations for 2d-MoS2, but this was not the case for 2d-BN. Conversely, the GW method yielded results for 2d-BN that closely matched experimental data, but it tended to overestimate the bandgaps in 2d-MoS2.

Material	SIESTA	Standard DFT	HSE06	GW	Experiment
2d-BN	4.54eV	4.67eV	5.68eV	7.12eV	6.1/(6.8 - 7.1) eV
2d-MoS2	1.68eV	1.58eV	2.09eV	2.53eV	1.8 - 2.15 eV

Table 2 Band Gaps Calculated by Different Methods and characterized by experiments. The band gap values for 2d-BN and 2d-MoS2 are presented. Calculations using the SIESTA method were performed in this project, while values from standard DFT, HSE06, and GW methods were obtained from the C2DB database. The 'Experiment' column summarizes band gap data from various experimental studies previously mentioned."

Supplementary Information

i.fdf

i.fdf for hBN

```
SystemName      hBN
SystemLabel     hBN
```

```
LatticeConstant 2.51 Ang
%block LatticeVectors
  1.0000000000000000  0.0000000000000000  0.0000000000000000
 -0.5000000000000000  0.8660254037844386  0.0000000000000000
    0.000000000000000  0.000000000000000  4.7808764940239043
%endblock LatticeVectors

NumberOfSpecies   2
NumberOfAtoms     2
%block ChemicalSpeciesLabel
  1  5  B
  2  7  N
%endblock ChemicalSpeciesLabel

AtomicCoordinatesFormat Fractional
%block AtomicCoordinatesAndAtomicSpecies
    0.00000   0.00000   0.16667     1        1    B
    0.66667   0.33333   0.16667     2        2    N
%endblock AtomicCoordinatesAndAtomicSpecies


XC.functional   GGA                #exchange-correlation functional
XC.authors      PBE

SolutionMethod   diagon            #default diagonalization solution of the KS Hamilton
ian

#---Options for k-Points Sampling----------------------------------------------

WriteKpoints                 F     #default false, not writing the coordinates of k vec
tors for ksampling
%block kgrid_Monkhorst_Pack
    4    0   0   0.5                # K-sampling kpoints 8x8x1
    0    4   0   0.5
    0    0   1   0.0
%endblock kgrid_Monkhorst_Pack

#---Options for Basis Set------------------------------------------------------

PAO.BasisType   split    # Default split, split-valence scheme for multiple-zeta, based
 on different radii
PAO.BasisSize   DZP      # Default DZP double zeta with polarization orbitals basis set
PAO.SplitNorm   0.15     # Default 0.15
PAO.EnergyShift 0.02 Ry  # Default 0.02 Ry  a standard for orbital-confining cutoff rad
ii


#---Options for Initialisation and SCF Loop------------------------------------

ElectronicTemperature        300 K
OccupationFunction           FD          # function determining the occupation of the e
```

```
lectronic states
MeshCutOff                   200 Ry      # default 100 Ry, plane wave energy cutoff


DM.UseSaveDM                  .True.               # whether or not to read previous de
nsity matrix
DM.MixingWeight               0.02                 # Default 0.25
DM.NumberPulay                5                    # Default 0
DM.Tolerance                  1.d-5                # Default convergence criteria for D
M
DM.RequireEnergyConvergence   .True.
DM.EnergyTolerance            1.d-4 eV             # Energy convergency criteria
MaxSCFIterations              500                  # Maximum number of SCF iterations

#---Options for Structural Relaxations-----------------------------------------

MD.TypeOfRun                 CG                  # structure optimization option
MD.NumCGSteps                500
MD.VariableCell              .False.             #default, no lattice relaxation
MD.ConstantVolume            .False.             # default, no change in lattice volume
MD.MaxForceTol               0.02 eV/Ang         # Default 0.04 eV/Ang, force tolerance
MD.MaxCGDispl                0.03 Bohr           # Default 0.20 Bohr
MD.UseSaveCG                 T
MD.UseSaveXV                 T
#SaveRho                     T
DM.UseSaveDM                 T

WriteBands         F                      # writing band information
BandLinesScale  ReciprocalLatticeVectors
%block BandLines
 1    0.000000   0.000000  0.000000   \Gamma
 30   0.333333   0.333333  0.000000   K
 30   0.500000   0.000000  0.000000    M
 45   0.000000   0.000000  0.000000   \Gamma
%endblock BandLines
```

i.fdf for MoS2, omitting part is the same as hBN i.fdf

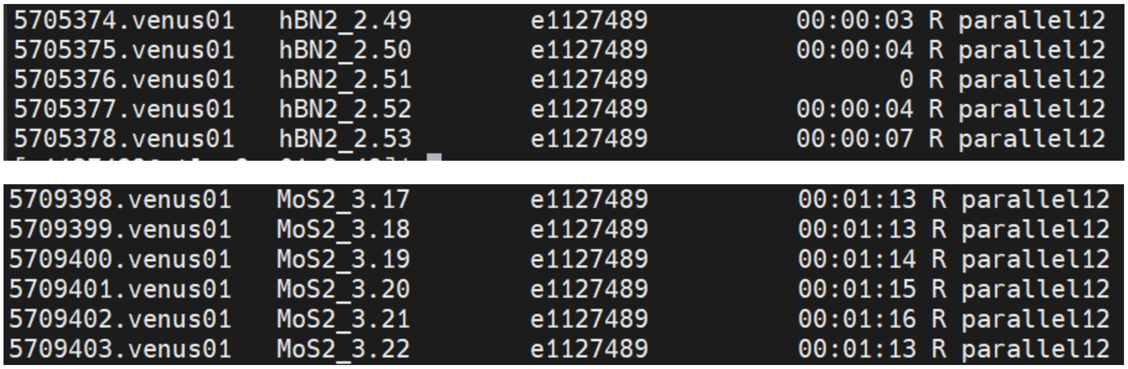
```
SystemName      MoS2
SystemLabel     MoS2

LatticeConstant 3.21 Ang
%block LatticeVectors
  1.0000000000000000  0.0000000000000000  0.0000000000000000
 -0.5000000000000000  0.8660254037844386  0.0000000000000000
    0.0000000000000000  0.0000000000000000  5.6074766355140186
%endblock LatticeVectors

NumberOfSpecies   2
NumberOfAtoms     3
%block ChemicalSpeciesLabel
  1  42 Mo
  2  16 S
%endblock ChemicalSpeciesLabel

AtomicCoordinatesFormat Fractional
%block AtomicCoordinatesAndAtomicSpecies
  0.666744471        0.333255500        0.141528144     1  1  Mo
  0.333312690        0.666687310        0.052778751     2  2  S
  0.333303303        0.666696727        0.230252653     2  3  S
%endblock AtomicCoordinatesAndAtomicSpecies
......
PAO.SplitNorm   0.15392     # Default 0.15
......
```

Submission screenshot



```
5705374.venus01   hBN2_2.49        e1127489         00:00:03 R parallel12
5705375.venus01   hBN2_2.50        e1127489         00:00:04 R parallel12
5705376.venus01   hBN2_2.51        e1127489                0 R parallel12
5705377.venus01   hBN2_2.52        e1127489         00:00:04 R parallel12
5705378.venus01   hBN2_2.53        e1127489         00:00:07 R parallel12
```

```
5709398.venus01   MoS2_3.17        e1127489         00:01:13 R parallel12
5709399.venus01   MoS2_3.18        e1127489         00:01:13 R parallel12
5709400.venus01   MoS2_3.19        e1127489         00:01:14 R parallel12
5709401.venus01   MoS2_3.20        e1127489         00:01:15 R parallel12
5709402.venus01   MoS2_3.21        e1127489         00:01:16 R parallel12
5709403.venus01   MoS2_3.22        e1127489         00:01:13 R parallel12
```

Script

LatticeTest.sh:

```
# Start and end lattice constants
start=2.49
end=2.53

# Step size
step=0.01

# Loop over lattice constants
for lattice in $(seq $start $step $end)
do
   # Calculate new z-coordinate value for LatticeVectors
   z_coord=$(echo "scale=16; 12 / $lattice" | bc)

   # Create directory named after the lattice constant
   dir_name=$lattice
   mkdir $dir_name

   # Copy files
   cp template/{i.fdf,B.psf,N.psf,hBN.pbs} $dir_name/

   # Modify the lattice constant in i.fdf file
   sed -i "s/^LatticeConstant.*/LatticeConstant $lattice Ang/" $dir_name/i.fdf
   # Modify the LatticeVectors block in i.fdf file
   sed -i "/%block LatticeVectors/,/%endblock LatticeVectors/s/0\.0000000000000000  0\.
0000000000000000  .*/  0.0000000000000000  0.0000000000000000  $z_coord/" $dir_name/i.f
df

   # Modify the job name in MoS2.pbs file to reflect the lattice constant
   sed -i "s/#PBS -N hBN/#PBS -N hBN2_${lattice}/" $dir_name/hBN.pbs

   # Enter the directory and submit the job
   cd $dir_name
   qsub hBN.pbs
   cd ..
done
```

getdata.sh

```
#!/bin/bash

output_file="energy.dat"

echo -e "LatticeConstant\tTotalEnergy" > $output_file

for lattice in $(seq 2.49 0.01 2.53); do
    energy=$(grep "Total =" $lattice/siesta.out | awk '{print $4}')
    echo -e "$lattice\t$energy" >> $output_file
done

echo "Data saved to $output_file"
```

Reference

(1) Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. The SIESTA Method for Ab Initio Order-N Materials Simulation. *J. Phys. Condens. Matter* **2002**, *14* (11), 2745. https://doi.org/10.1088/0953-8984/14/11/302.

(2) Haastrup, S.; Strange, M.; Pandey, M.; Deilmann, T.; Schmidt, P. S.; Hinsche, N. F.; Gjerding, M. N.; Torelli, D.; Larsen, P. M.; Riis-Jensen, A. C.; Gath, J.; Jacobsen, K. W.; Mortensen, J. J.; Olsen, T.; Thygesen, K. S. The Computational 2D Materials Database: High-Throughput Modeling and Discovery of Atomically Thin Crystals. *2D Mater.* **2018**, *5* (4), 042002. https://doi.org/10.1088/2053-1583/aacfc1.

(3) Gjerding, M. N.; Taghizadeh, A.; Rasmussen, A.; Ali, S.; Bertoldo, F.; Deilmann, T.; Knøsgaard, N. R.; Kruse, M.; Larsen, A. H.; Manti, S.; Pedersen, T. G.; Petralanda, U.; Skovhus, T.; Svendsen, M. K.; Mortensen, J. J.; Olsen, T.; Thygesen, K. S. Recent Progress of the Computational 2D Materials Database (C2DB). *2D Mater.* **2021**, *8* (4), 044002. https://doi.org/10.1088/2053-1583/ac1059.

(4) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys Rev Lett* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.

(5) Junquera, J.; Paz, Ó.; Sánchez-Portal, D.; Artacho, E. Numerical Atomic Orbitals for Linear-Scaling Calculations. *Phys. Rev. B* **2001**, *64* (23), 235111. https://doi.org/10.1103/PhysRevB.64.235111.

(6) E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García, J. M. Soler. Linear-Scaling Ab-Initio Calculations for Large and Complex Systems. *Phys. Status Solidi B* **1999**, *215* (1), 809–817. https://doi.org/10.1002/(SICI)1521-3951(199909)215:1<809::AID-PSSB809>3.0.CO;2-0.

(7) Elias, C.; Valvin, P.; Pelini, T.; Summerfield, A.; Mellor, C. J.; Cheng, T. S.; Eaves, L.; Foxon, C. T.; Beton, P. H.; Novikov, S. V.; Gil, B.; Cassabois, G. Direct Band-Gap Crossover in Epitaxial Monolayer Boron Nitride. *Nat. Commun.* **2019**, *10* (1), 2639. https://doi.org/10.1038/s41467-019-10610-5.

(8) Román, R. J. P.; Costa, F. J. R. C.; Zobelli, A.; Elias, C.; Valvin, P.; Cassabois, G.; Gil, B.; Summerfield, A.; Cheng, T. S.; Mellor, C. J.; Beton, P. H.; Novikov, S. V.; Zagonel, L. F. Band Gap Measurements of Monolayer H-BN and Insights into Carbon-Related Point Defects. *2D Mater.* **2021**, *8* (4), 044001. https://doi.org/10.1088/2053-1583/ac0d9c.

(9) Splendiani, A.; Sun, L.; Zhang, Y.; Li, T.; Kim, J.; Chim, C.-Y.; Galli, G.; Wang, F. Emerging Photoluminescence in Monolayer MoS2. *Nano Lett.* **2010**, *10* (4), 1271–1275. https://doi.org/10.1021/nl903868w.

(10) Miwa, J. A.; Ulstrup, S.; Sørensen, S. G.; Dendzik, M.; ıfmmode \checkC\else Č\fiabo, A. G. \checks\else š\fiiıfmmode \acutec\else ć\fi; Bianchi, M.; Lauritsen, J. V.; Hofmann, P. Electronic Structure of Epitaxial Single-Layer \mathrmMoS_2. *Phys Rev Lett* **2015**, *114* (4), 046802. https://doi.org/10.1103/PhysRevLett.114.046802.

(11) Zhang, C.; Johnson, A.; Hsu, C.-L.; Li, L.-J.; Shih, C.-K. Direct Imaging of Band Profile in Single Layer MoS2 on Graphite: Quasiparticle Energy Gap, Metallic Edge States, and Edge Band Bending. *Nano Lett.* **2014**, *14* (5), 2443–2447. https://doi.org/10.1021/nl501133c.

(12) Hunt, R. J.; Monserrat, B.; Zólyomi, V.; Drummond, N. D. Diffusion Quantum Monte Carlo and GW Study of the Electronic Properties of Monolayer and Bulk Hexagonal Boron Nitride. *Phys Rev B* **2020**, *101* (20), 205115. https://doi.org/10.1103/PhysRevB.101.205115.