mol2crystal: Automatic Crystal Structure Generation Tool from Molecular Input

1 Overview

mol2crystal is a Python-based software tool that automatically generates and evaluates candidate molecular crystal structures using space group symmetry. It takes a single molecular structure (in MOL format, specifically named `precursor.mol`) as input and performs the following tasks:

- Exhaustive generation of crystal structures using space groups (1–230)
- Automatic detection of atomic overlaps
- Support for high-accuracy post-processing tools such as DFTB+, xtb, and DFT-D3
- Fully automated structure selection and density evaluation

2 Environment Requirements

- Python 3.8 or higher
- Recommended packages (verified versions):

```
pip install ase==3.22.1 scipy==1.13.0 psutil==7.0.0
pip install pymsym==0.3.4 spglib==2.6.0
```

3 Directory Structure (Post Execution)

- `molecular_files/` : Place input molecular file (e.g., `precursor.mol`)
- `valid_structures/` : Accepted structures (VASP format)
- `optimized_structures_vasp/` : Post-processed optimized structures
- `structure_vs_energy.txt` : Records of density, volume, atom count, etc.

4 Usage Instructions

1. Place the molecular structure in `molecular_files/precursor.mol`
2. Run the command:

```
python3 mol2crystal.py
```

3. Results are saved in `valid_structures/` and density information is recorded in `structure_vs_energy`

5 User Settings (Configurable at Script Start)

- `user_margin` : Margin around molecule (Å), e.g., 1.70
- `user_nmesh` : Number of rotation angle divisions, e.g., 3
- `user_overlap_scale` : Threshold scale for atomic overlap, e.g., 0.80
- `user_max_depth` : Depth of space group expansion (physical similarity), e.g., 1
- `user_skipping_n_molecules` : Skip structures with too many molecules, e.g., 100

6 Post-processing Options

High-accuracy optimization can be performed using the following scripts:

```
python3 postprocess_dftb.py
python3 postprocess_xtb.py
python3 postprocess_siesta.py
```

7 Output Example

```
valid_structures/POSCAR_theta_0_phi_0_sg_62 0.000000 1.234 48 1234.56
```

8 Background and Significance

mol2crystal aims to automate and ensure reproducibility in molecular crystal structure prediction. It contributes to education, research, and industrial applications as a fully open and reproducible tool, in contrast to many commercial or proprietary codes.

9 Code Structure and Key Scripts

- `mol2crystal.py` : Main script for automatic crystal structure generation
- `select_data.py` : Structure selection based on energy and density
- `postprocess_*.py` : High-accuracy optimization using DFTB+, xtb, SIESTA
- `run.sh` : Bash script for orchestrating execution with logging

10 Processing Flow

1. Read and center molecular structure
2. Analyze point group symmetry and extract space group candidates
3. Rotate molecule and define unit cell
4. Apply space group and generate crystal structures
5. Check for atomic overlaps
6. Save structure and evaluate density

11 Key Functions

- `rotate_molecule()` : Rotate molecule by angles θ, ϕ
- `has_overlap_neighborlist()` : Fast detection of atomic overlaps
- `adjust_cellpar_by_spacegroup()` : Adjust unit cell parameters based on space group
- `get_primitive_cell()` : Convert to primitive cell using spglib
- `density_calc()` : Calculate density from mass and volume
- `get_all_subgroups()` : Recursively obtain strict subgroups of point group
- `expand_physical_supergroups()` : Expand physically similar supergroups to specified depth

12 Design Philosophy

- Strict yet flexible handling of point group and space group correspondence
- Emphasis on automation and reproducibility in structure generation
- Designed for seamless integration with post-processing tools