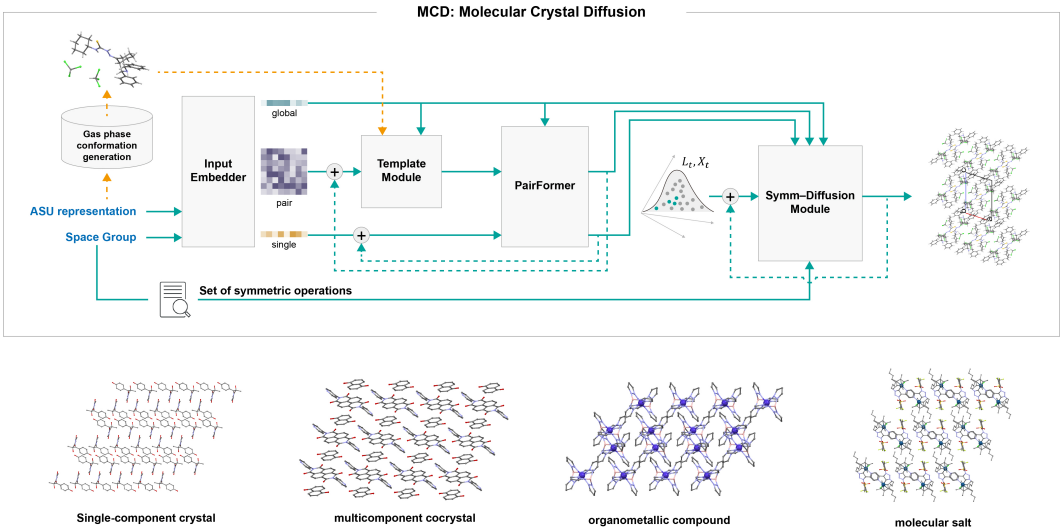
A Unified Diffusion Framework for the End-to-End Generation of Symmetry-Constrained Complex Molecular Crystals

Wendi Cai[1], Fanyang Mo[1]*
[1]Peking University

Abstract

Crystal structure prediction is a central problem in chemistry and materials science. Although recent deep generative models—particularly diffusion models—have achieved substantial progress in crystal structure generation for inorganic materials, generating molecular crystal structures remains challenging. These systems exhibit pronounced complexity, combining intramolecular conformational variability with intricate intermolecular packing interactions (e.g., hydrogen bonding, π-π stacking, and electrostatics) under crystallographic symmetry constraints. Here we present a unified diffusion framework Molecular Crystal Diffusion (MCD) for end-to-end generation of symmetry-constrained complex molecular crystals, built on a Diffusion Transformer (DiT) backbone. Given low-dimensional molecular representations as input, the model directly generates complete crystal structures, including lattice parameters and fractional atomic coordinates under periodic boundary conditions. Symmetry is incorporated as an explicit constraint during generation, enabling the model to sample structures consistent with specified space-group operations while retaining flexibility across diverse compositions and stoichiometries. A single architecture supports multiple crystal categories—single-component molecular crystals, multi-component co-crystals, molecular salts, and metal–organic molecular crystals—without category-specific redesign. Across broad crystal families, the proposed approach produces chemically plausible packings with symmetry-consistent outputs, providing a scalable route toward automated discovery of functional molecular solids.



MCD: Molecular Crystal Diffusion framework and examples of generated crystal structures.

*Corresponding author: fmo@pku.edu.cn