All-atom Diffusion Transformers

Unified generative modelling of molecules & materials

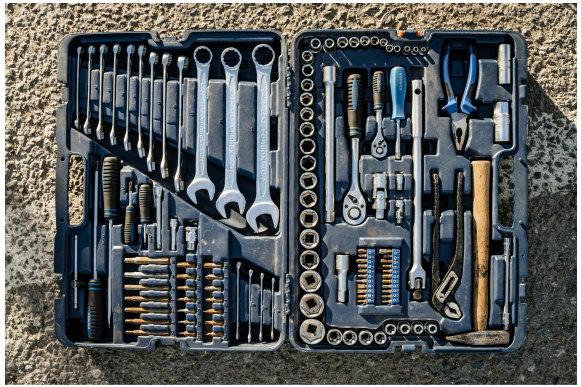
Chaitanya K. Joshi
with Xiang Fu, Yi-Lun Liao, Vahe Gharekhanyan, Benjamin Miller, Anuroop Sriram*, and Zachary W. Ulissi*

📄**Paper:** arxiv.org/abs/2503.03965

💻**Code:** github.com/facebookresearch/all-atom-diffusion-transformer

About me 👋

Final year PhD student; my website: chaitjo.com



Deep learning for molecular modelling

GNNs, Transformers, Diffusion

Theoretical foundations

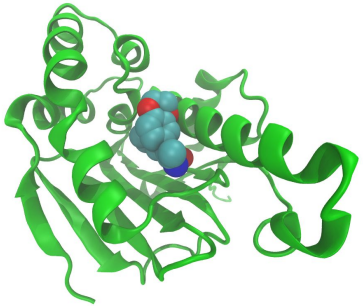
Foundation models

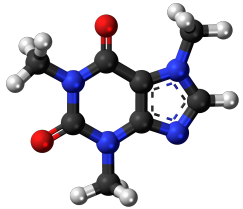
Applications in RNA design

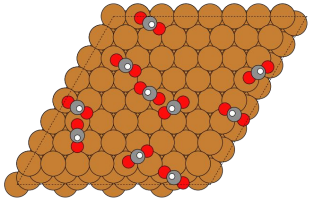
Structure prediction & inverse design

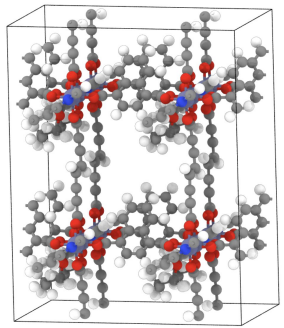
Ribozymes and riboswitches

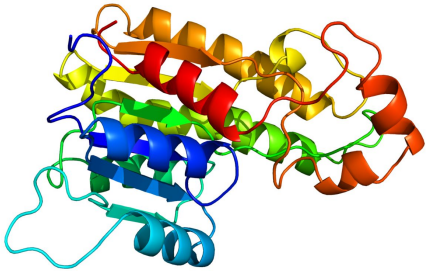
Foundation models for Chemistry?

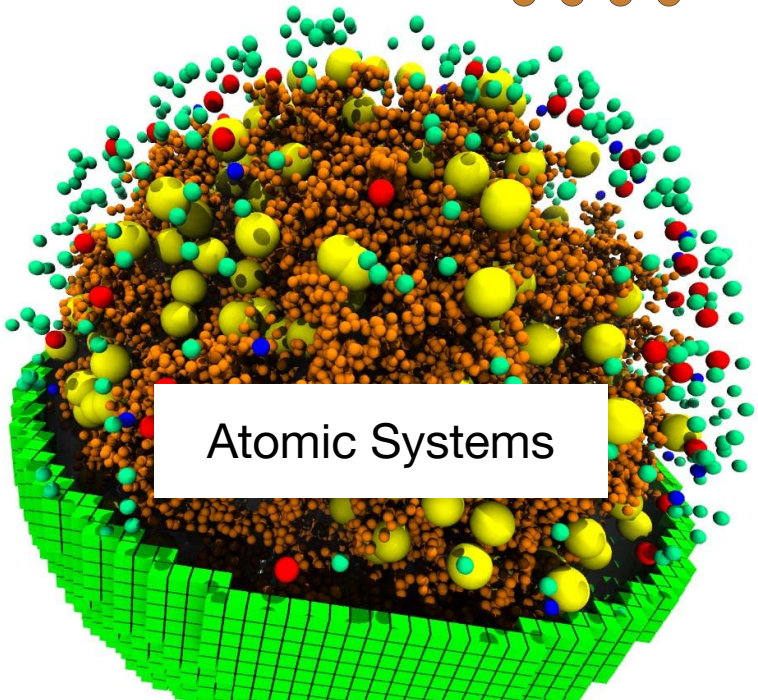




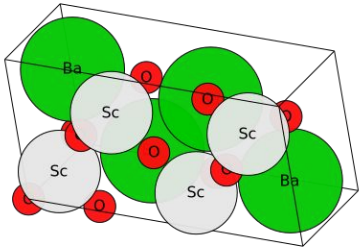








Atomic Systems



Foundation models for Chemistry

Vision: Unified generative models for all types of atomic systems

Large and **diverse datasets** where we treat everything as atoms in 3D space.

⚛️

Learn **generalizable representations** of chemistry and atomic interactions.

⭐

Useful for novel, **data scarce** design tasks.

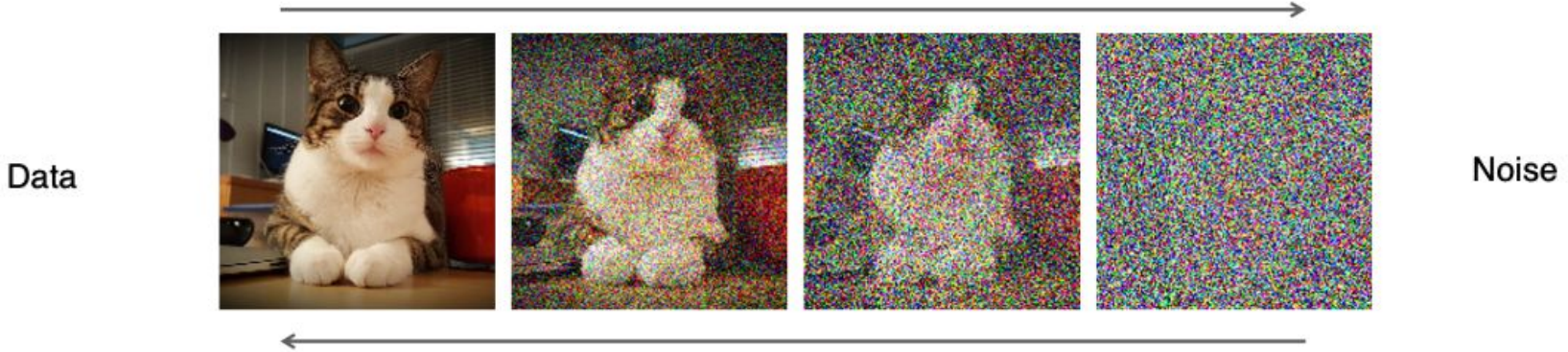
- Molecule + Slab → Catalysis system
- Molecule + Protein → Drug or enzyme

State of Generative models for 3D atomic systems today

Diffusion models

Current state-of-the-art for continuous data

Images, video, audio, …



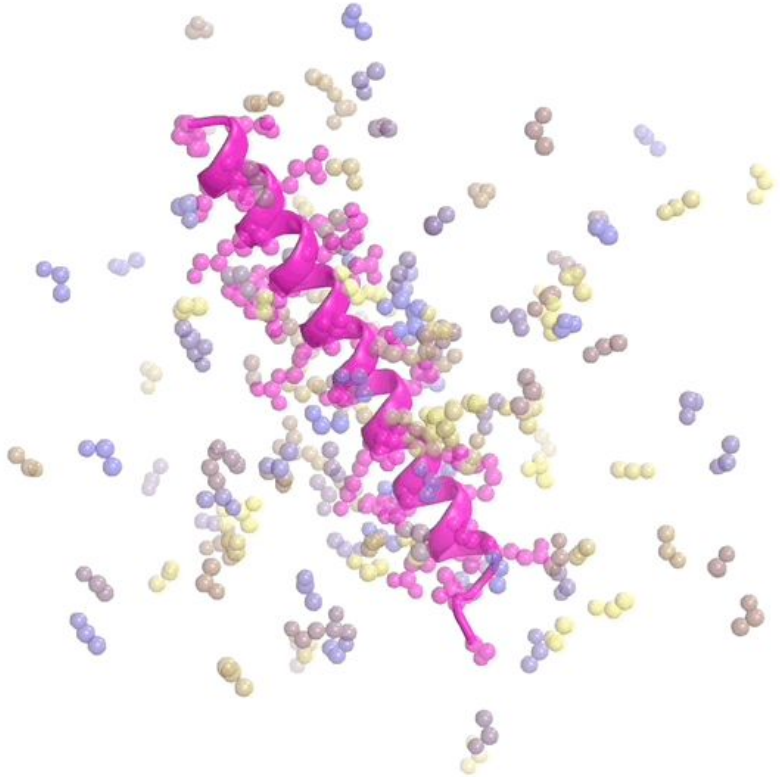
Train a neural network to iteratively transform a **noisy** sample to **clean** or real data.

GIF: Arash Vahdat, NVIDIA

Diffusion models

Current state-of-the-art for continuous data

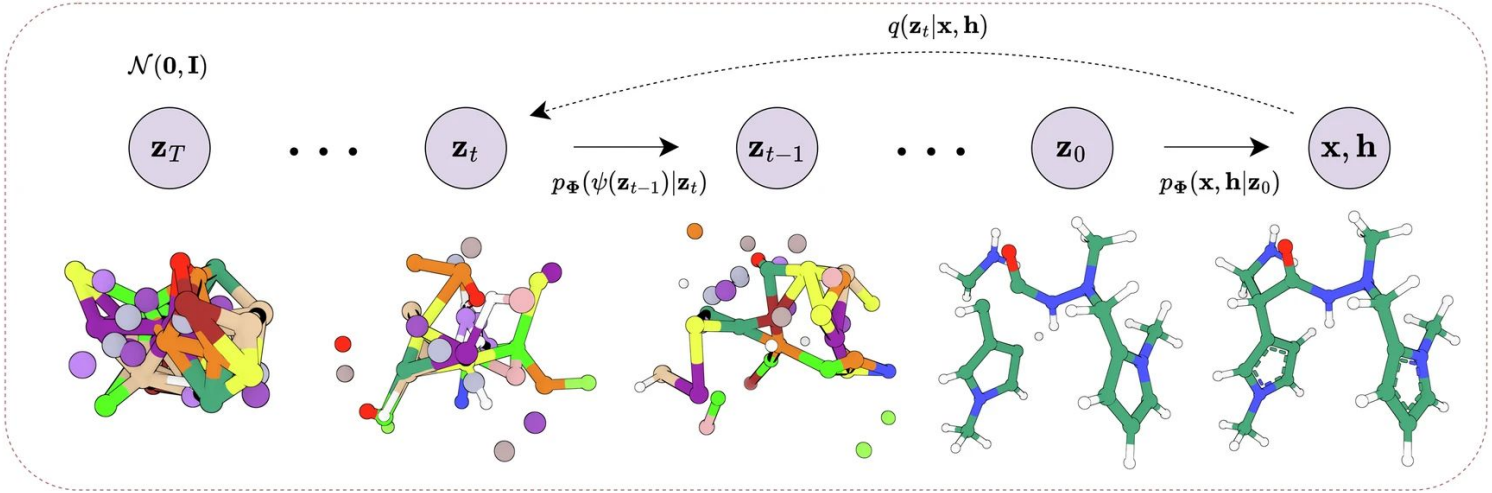
…and molecules



GIF: Institute for Protein Design

Generating small molecules

Diffusion on a complex product manifold

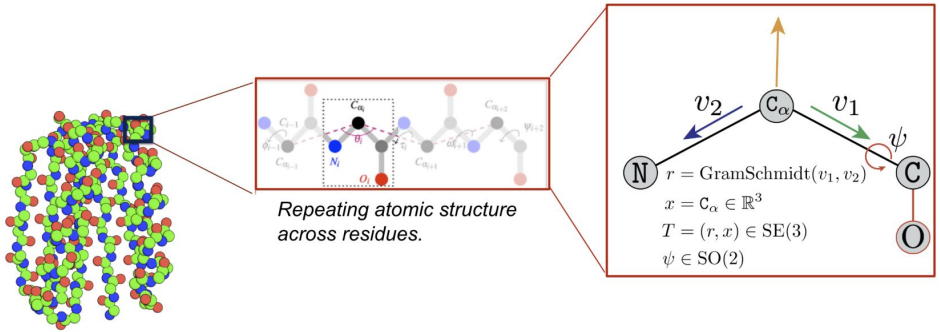


Jointly diffuse **atom types** (categorical) and **3D coordinates** (continuous).

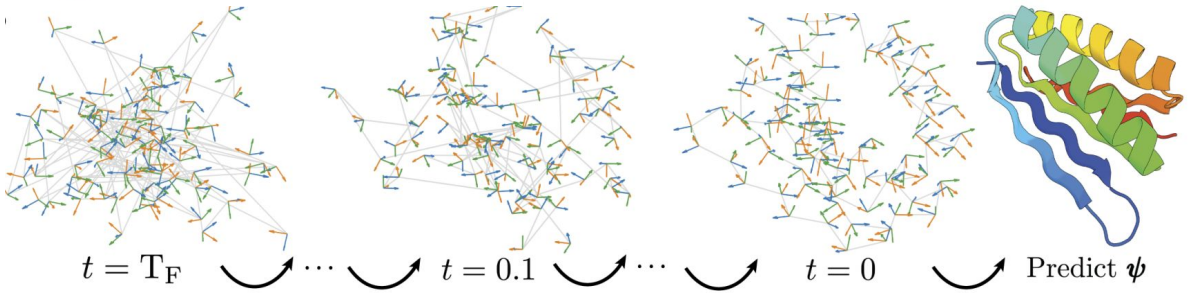
Hoogeboom et al. Equivariant Diffusion Model for Molecule Generation in 3D. ICML 2022.

Generating proteins & biomolecules

Diffusion on an even more complex product manifold



~~Atom type~~
Residue type

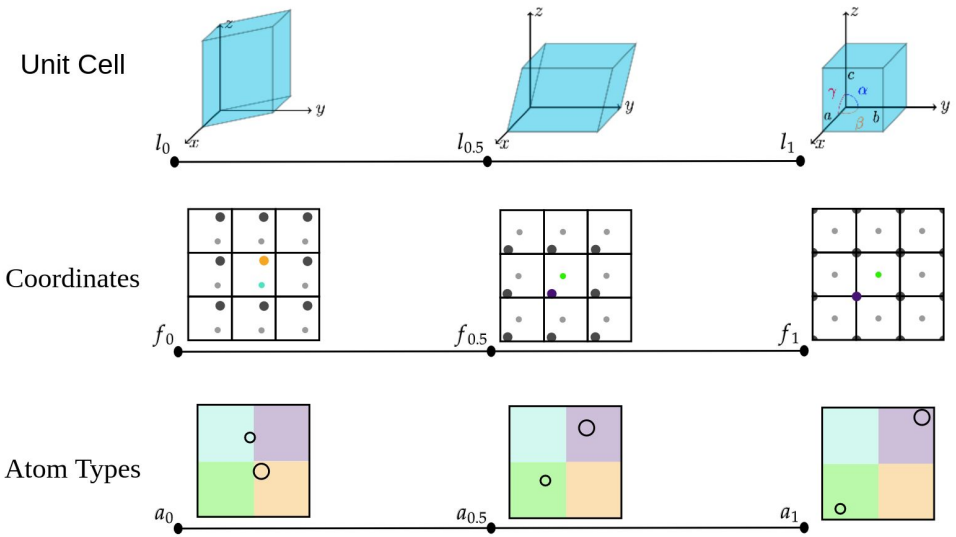


…adds third manifold (**rotations**)

Yim et al. SE(3) diffusion model with application to protein backbone generation. ICML 2023.

Generating crystals & materials

Diffusion on an even more complex product manifold



(1) atom types, (2) fractional coordinates, (3) lattice lengths, (4) lattice angles

Miller et al. FlowMM: Generating Materials with Riemannian Flow Matching. ICML 2024.

Research question:

How can we build a unified generative model for all chemical systems?

💡 Idea 1
It’s all atoms!

Molecular systems are sets of atoms interacting in 3D space

– embed molecules and materials in a shared latent space.

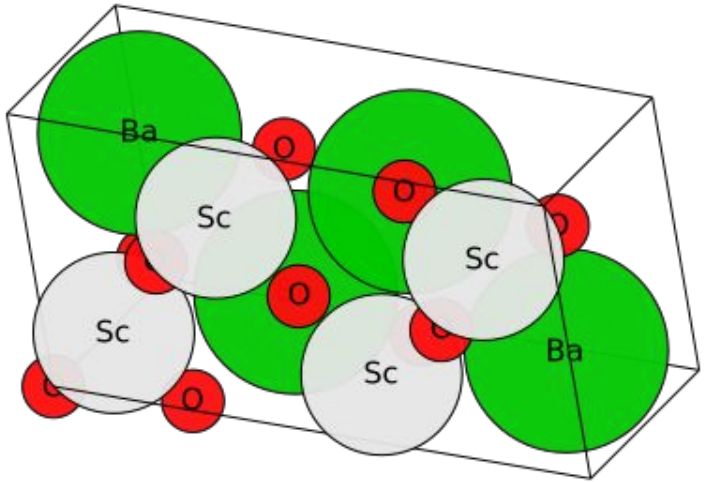
Shogi et al. From Molecules to Materials: Pre-training Large Generalizable Models for Atomic Property Prediction. ICLR 2024.

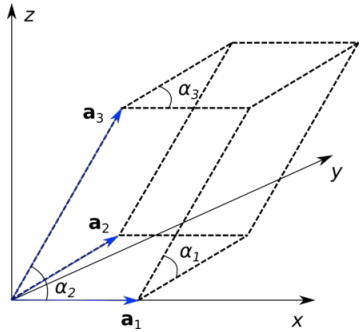
Unified representation of materials

Fractional coordinates = unit_cell^{-1} . pos

For a crystal with N atoms:

- Atom type: (N, 1)
- 3D coordinate: (N, 3)
- Fractional coordinate: (N, 3)
- Lattice parameters:
  - Lengths: (1, 3)
  - Angles: (1, 3)



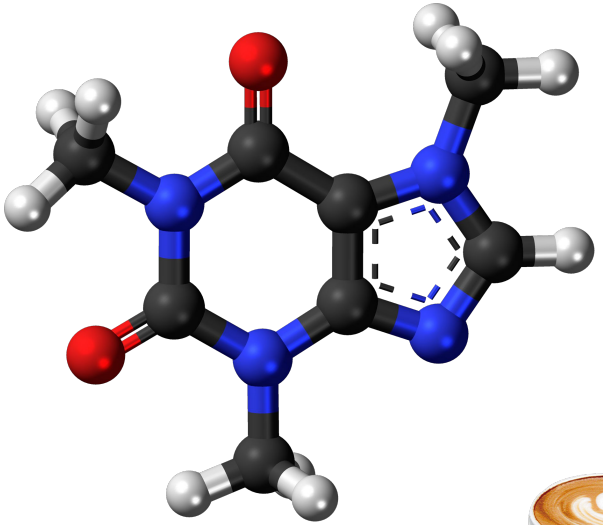


Unified representation of molecules

Set periodic data types to null

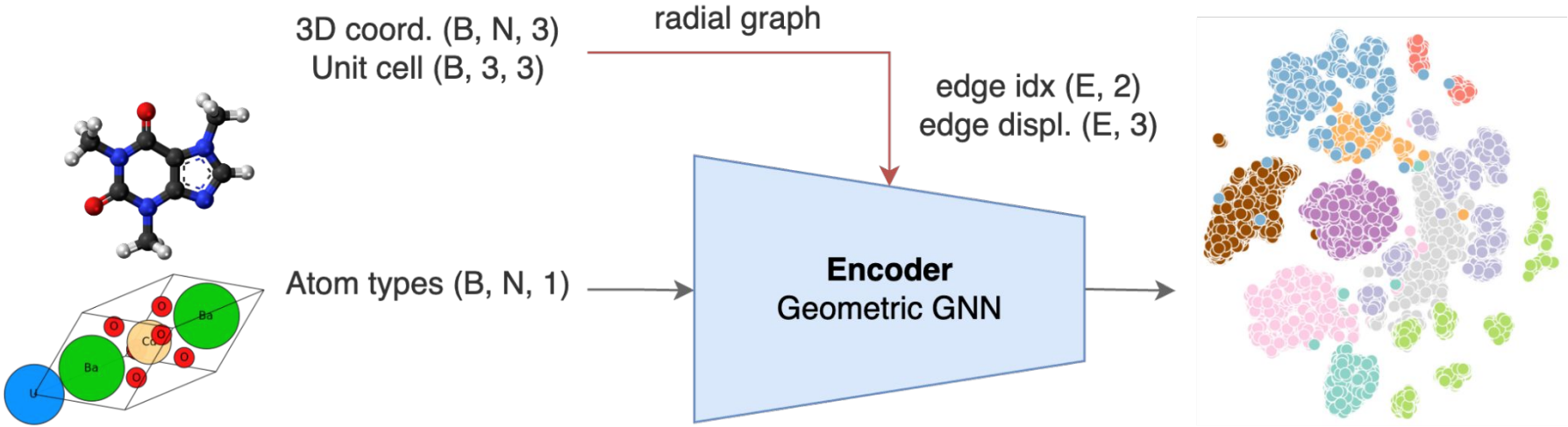
For a molecule with N atoms:

- Atom type: (N, 1)
- 3D coordinate: (N, 3)
- Fractional coordinate: (N, 3) → ∅
- Lattice parameters:
  - Lengths: (1, 3) → ∅
  - Angles: (1, 3) → ∅



Embedding atomic systems

Geometric GNNs enforce physical symmetries

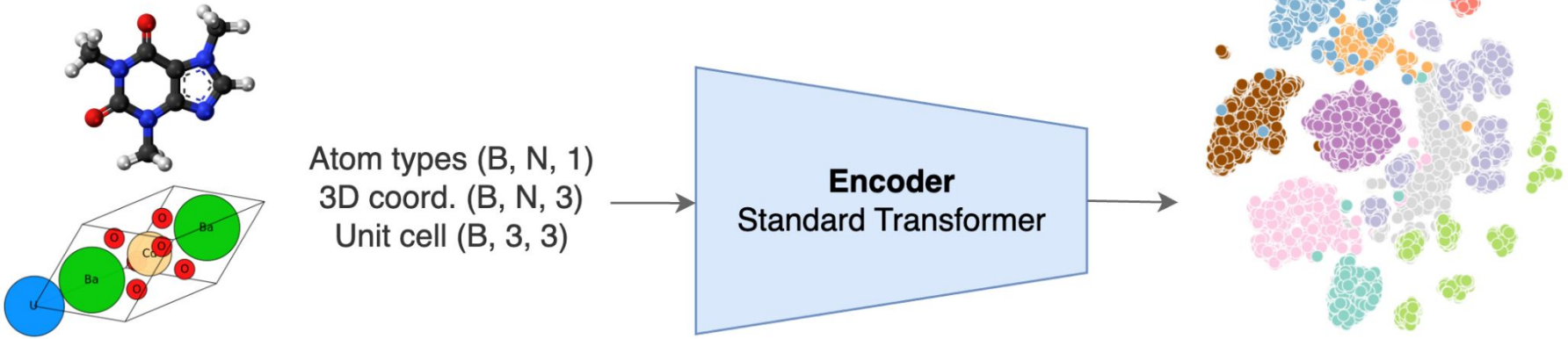


Duval*, Mathis*, Joshi*, Schmidt*, et al. A Hitchhiker's Guide to Geometric GNNs for 3D Atomic Systems. 2023.
Joshi*, Bodnar*, et al. On the Expressive Power of Geometric GNNs. ICML 2024.

Embedding atomic systems

Transformers <u>learn</u> symmetries via data augmentation

(Keen to discuss tradeoffs at the end)



B = batch size dimension

💡 Idea 2
Diffusion in latent space

Do generative modelling in the latent space of a joint encoder

– reconstruct to molecules or materials.

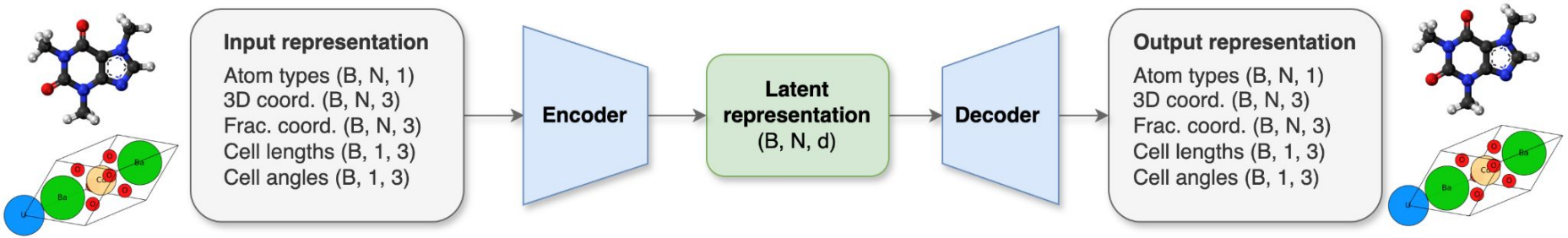
Rombach et al. High-Resolution Image Synthesis with Latent Diffusion Models. CVPR 2022.

Introducing the All-atom Diffusion Transformer

ADiT

Latent diffusion unifies molecules and materials

Stage 1: train Autoencoder for all-atom reconstruction

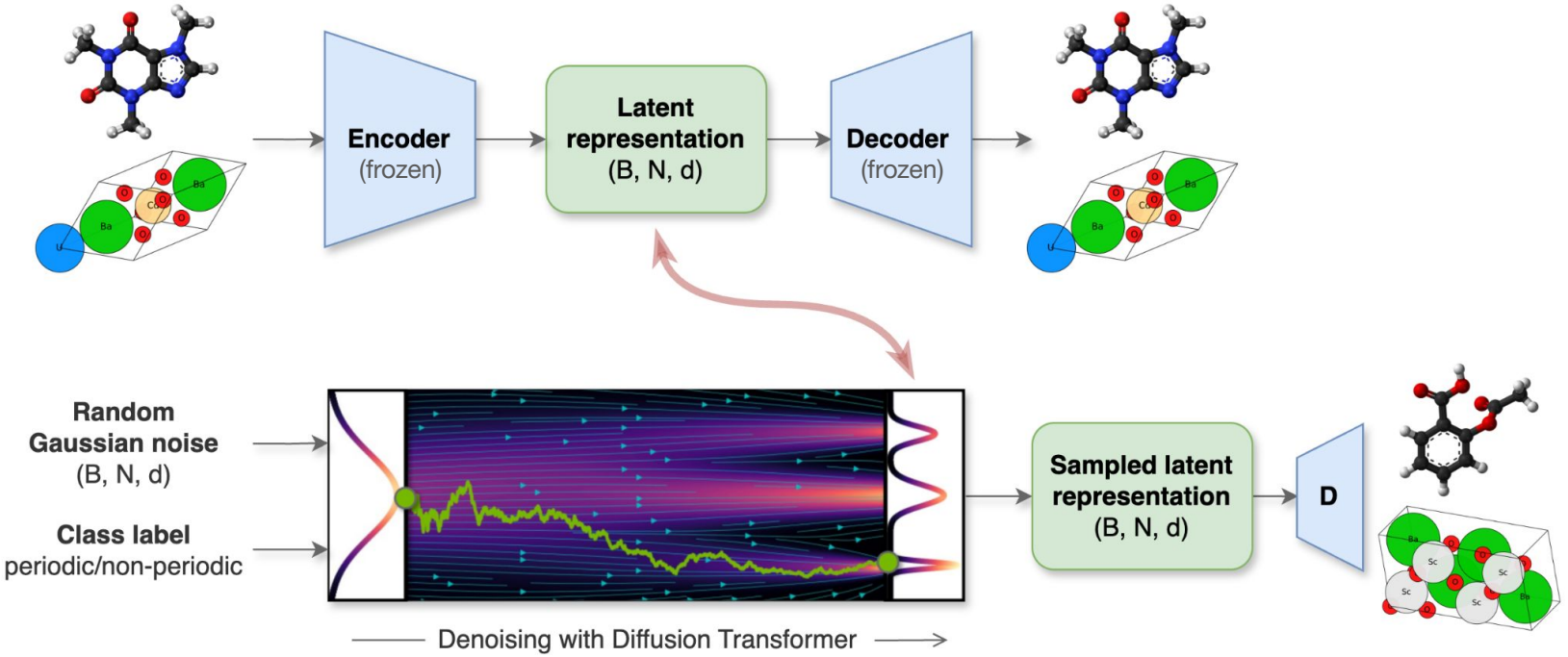


Autoencoder latents can be roto-translationally **equivariant**,
or learn symmetries via **data augmentation**.

Kingma and Welling. Auto-Encoding Variational Bayes. NeurIPS 2013.

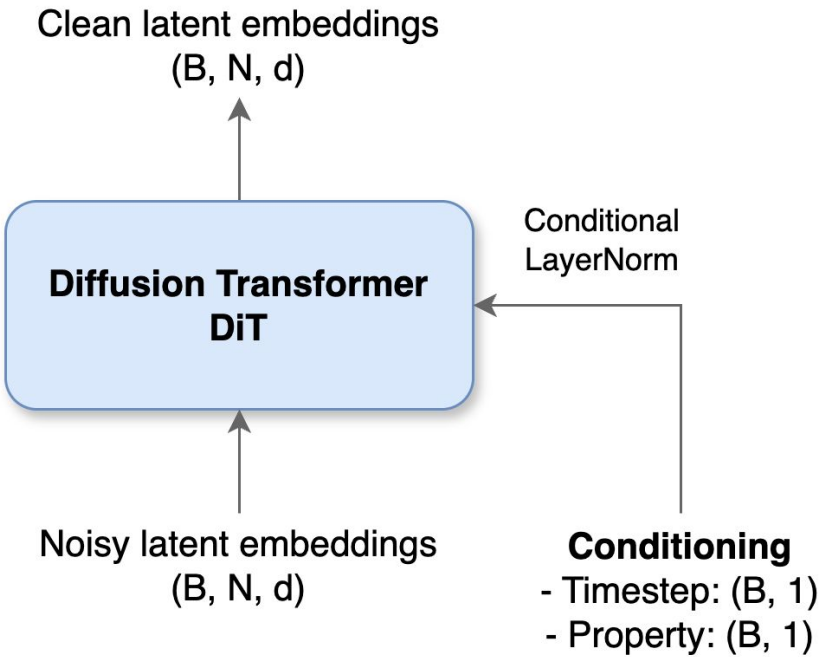
Latent diffusion unifies molecules and materials

Stage 2: train Diffusion model on distribution of AE latent embeddings



Denoiser architecture: Diffusion Transformer

Standard Transformer with conditioning



Property:

- Periodic/non-periodic token
- Any other molecular properties…

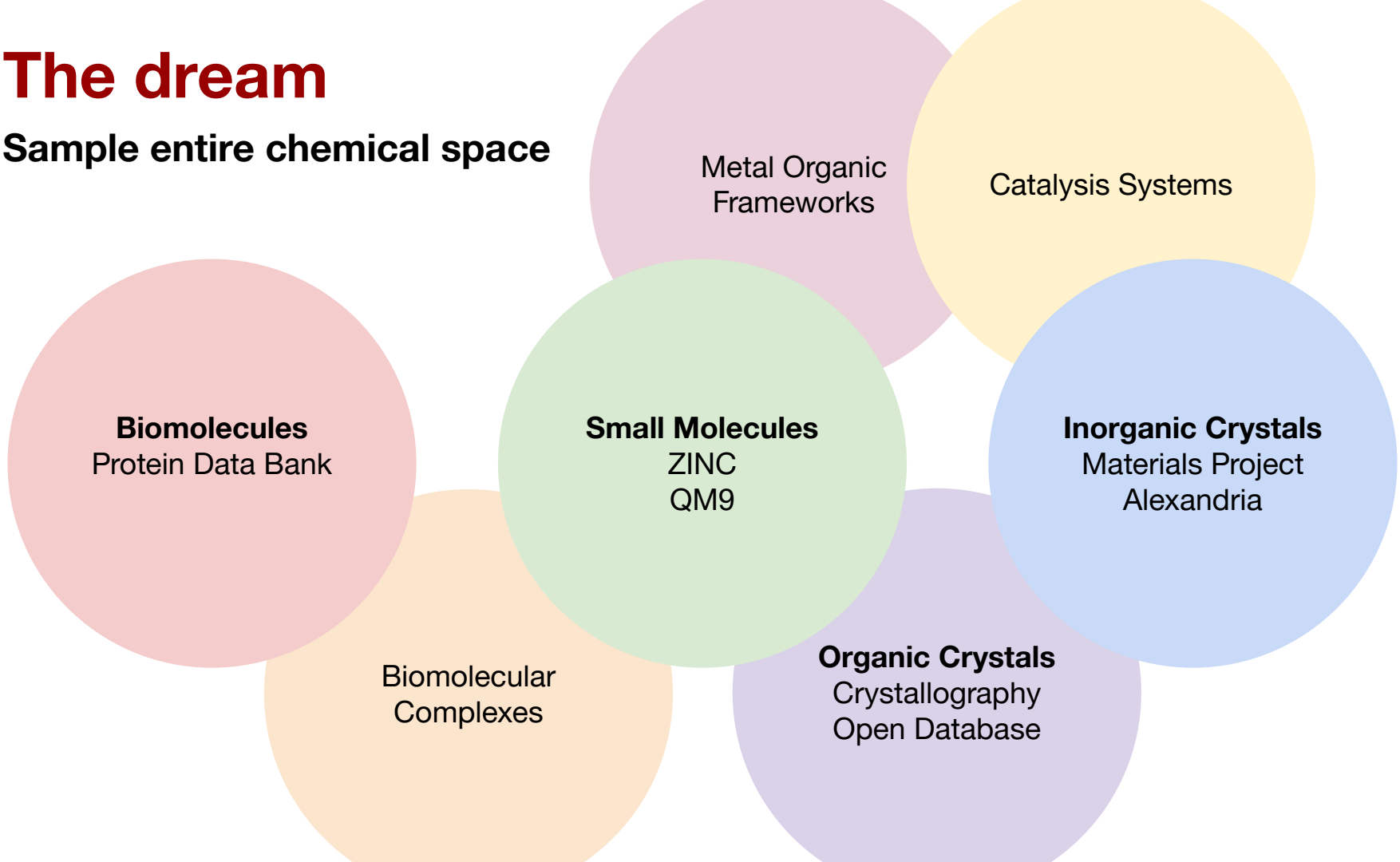
Allows controllability during sampling.

Peebles and Xie. Scalable Diffusion Models with Transformers. ICCV 2023.

Experiments

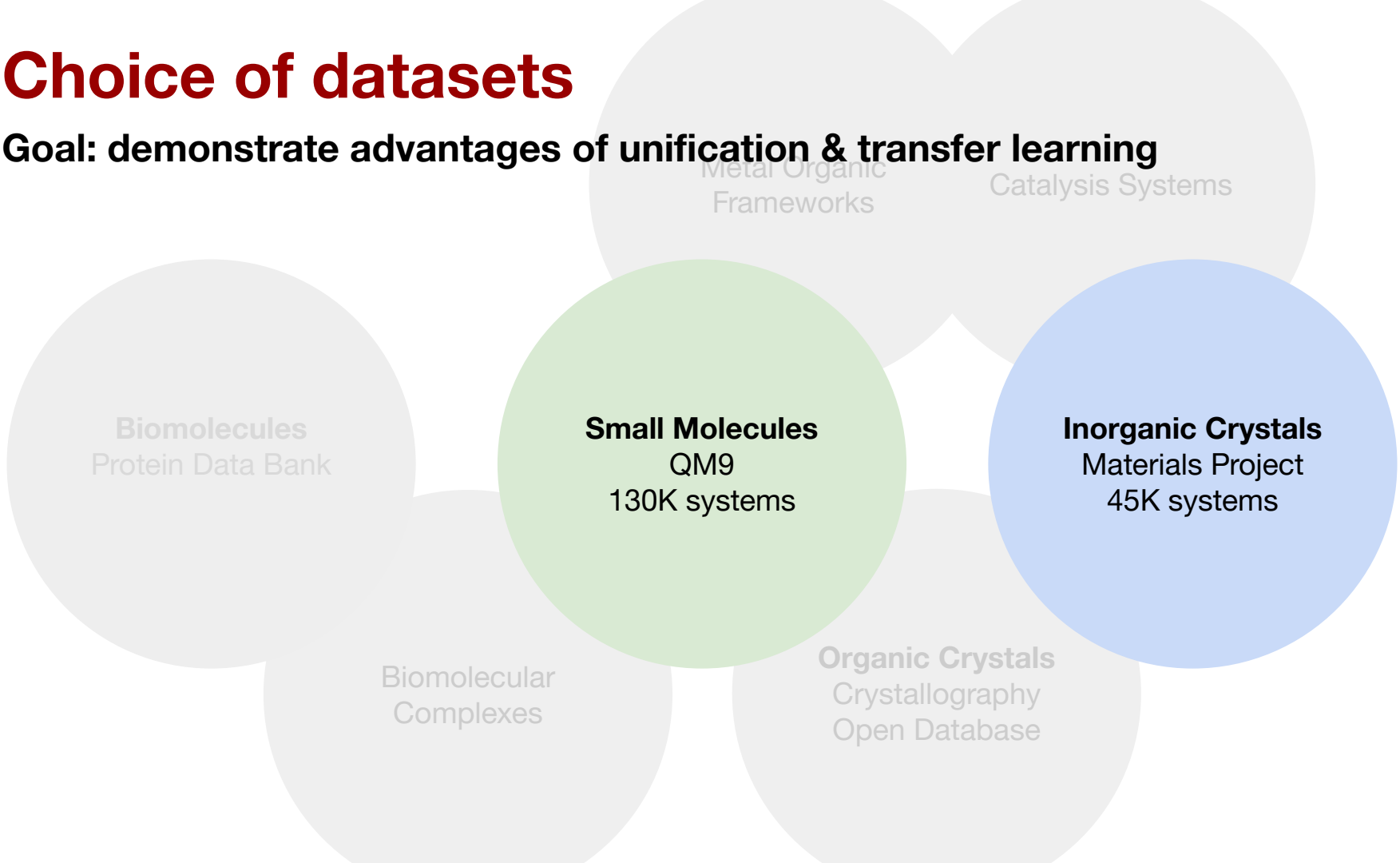
The dream

Sample entire chemical space



Choice of datasets

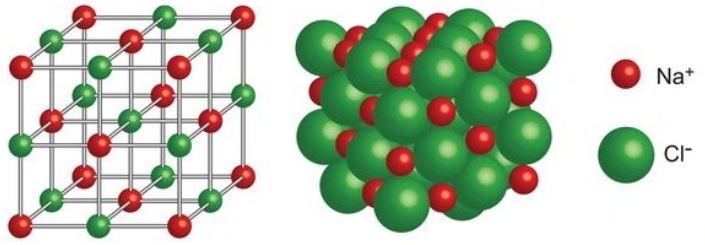
Goal: demonstrate advantages of unification & transfer learning



Crystal Generation

We want to generate physically realistic and stable crystals

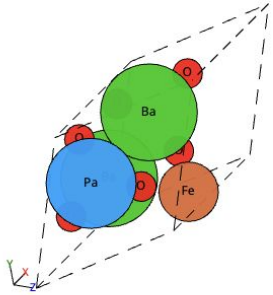
- **Structural validity**: all pairwise distances >= 0.5 & volume >= 0.1
- **Compositional validity** according to SMACT: Semiconducting Materials by Analogy and Chemical Theory
    - Charge neutrality & electronegativity balance
- **Metastability:** DFT Energy above hull from Materials Project < 0.1
    - Thermodynamic stability



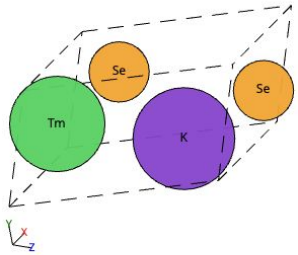
Crystal generation

ADiT improves over equivariant diffusion & benefits from transfer learning

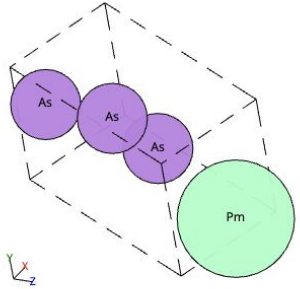
Model	Structural Validity (%) ↑	Compositional Validity (%) ↑	Struct. & Comp. Validity (%) ↑	Metastability Rate (%) ↑
Equivariant diffusion FlowMM, Miller et al.	96.7	83.0	80.3	30.6
Language model FlowLLM, Sriram et al.	99.9	90.8	90.8	66.9
Equivariant ADiT trained on MP20 only	99.2	86.1	85.5	
ADiT trained on MP20 only	99.6	90.5	90.1	**81.6**
ADiT trained on MP20 & QM9	**99.8**	**91.9**	**91.9**	81.0



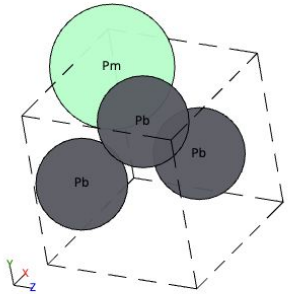
Ba_2 Pa Fe O_6 (F3-3m)



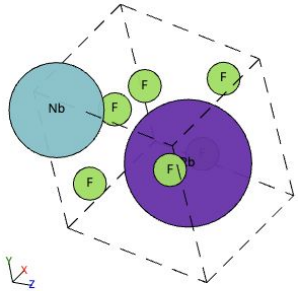
K Tm Se_2 (R-3m)



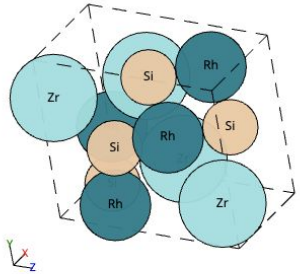
Pm As_3 (Cm)



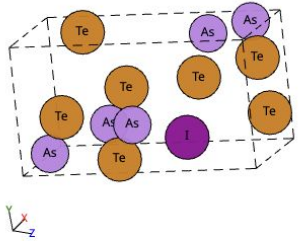
Pm Pb_3 (Pm-3m)



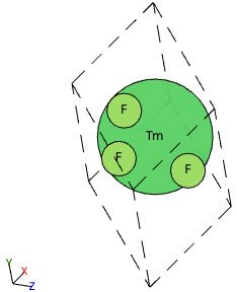
Rb Nb F_6 (P1)



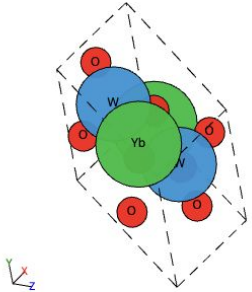
Si_4 Rh_4 Ze_4 (Pnma)



Te_7 As_5 I (P1)



Tm F_3 (Imm2)



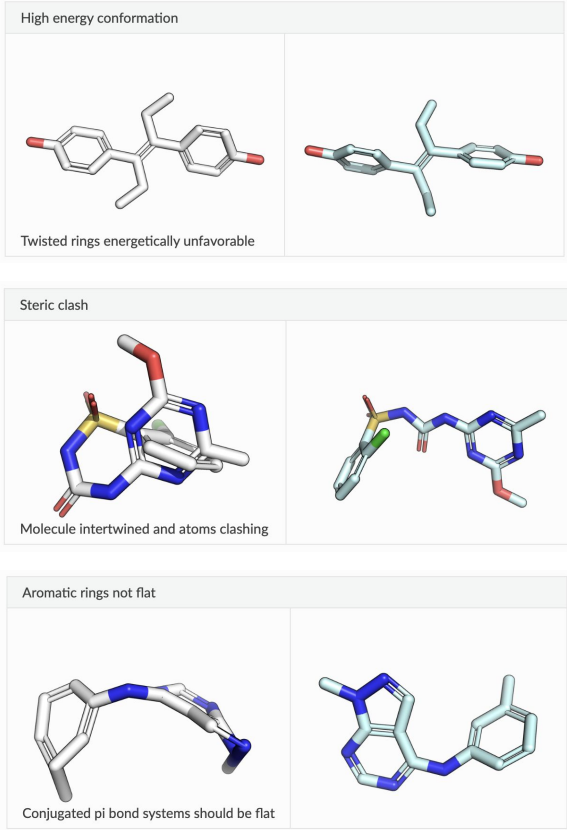
Yb_2 W_2 O_8 (P-1)

Molecule generation

We want to generate physically realistic and valid molecules

- **Validity:** whether RDKit can assign a SMILES
- **PoseBusters:** physics-based sanity checks for molecular generative models

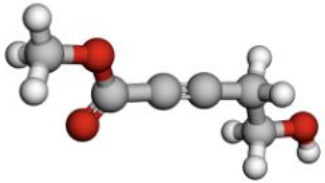
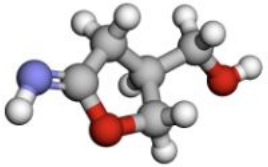
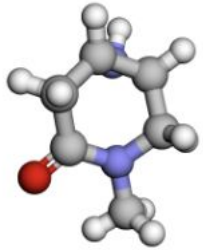
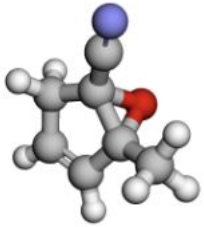
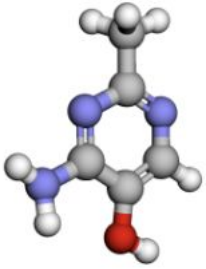
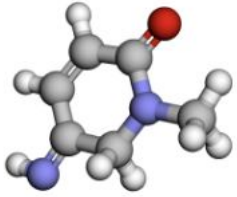
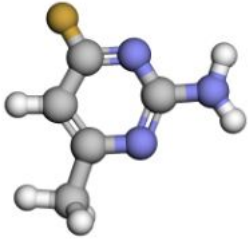
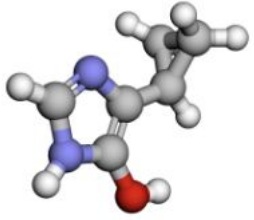
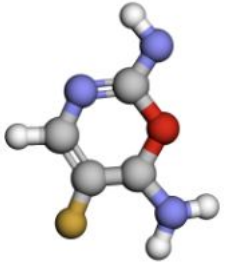
Bond lengths	The bond lengths in the input molecule are within 0.75 of the lower and 1.25 of the upper bounds determined by distance geometry.
Bond angles	The angles in the input molecule are within 0.75 of the lower and 1.25 of the upper bounds determined by distance geometry.
Planar aromatic rings	All atoms in aromatic rings with 5 or 6 members are within 0.25 Å of the closest shared plane.
Planar double bonds	The two carbons of aliphatic carbon-carbon double bonds and their four neighbours are within 0.25 Å of the closest shared plane.
Internal steric clash	The interatomic distance between pairs of non-covalently bound atoms is above 0.8 of the lower bound determined by distance geometry.
Energy ratio	The calculated energy of the input molecule is no more than 100 times the average energy of an ensemble of 50 conformations generated for the input molecule. The energy is calculated using the UFF[32] in RDKit and the conformations are generated with ETKDGv3 followed by force field relaxation using the UFF with up to 200 iterations.



Molecule generation

ADiT improves over equivariant diffusion & benefits from transfer learning

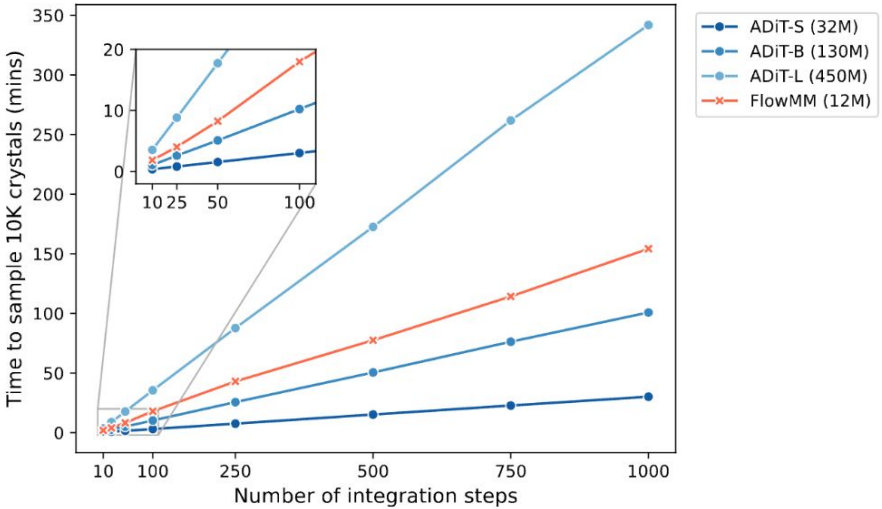
Model	Validity (%) ↑	Uniqueness (%) ↑	No steric clash (%) ↑	Reasonable internal energy (%) ↑
Equivariant diffusion EDM, Hoogeboom et al.	91.9	**98.6**	99.8	94.8
Language model Symphony, Diagavane et al.	83.5	97.9	98.1	95.6
Equivariant ADiT trained on QM9 only	91.5	98.0		
ADiT trained on QM9 only	92.2	97.9	99.9	95.9
ADiT trained on MP20 & QM9	**94.5**	97.8	**99.8**	**95.9**

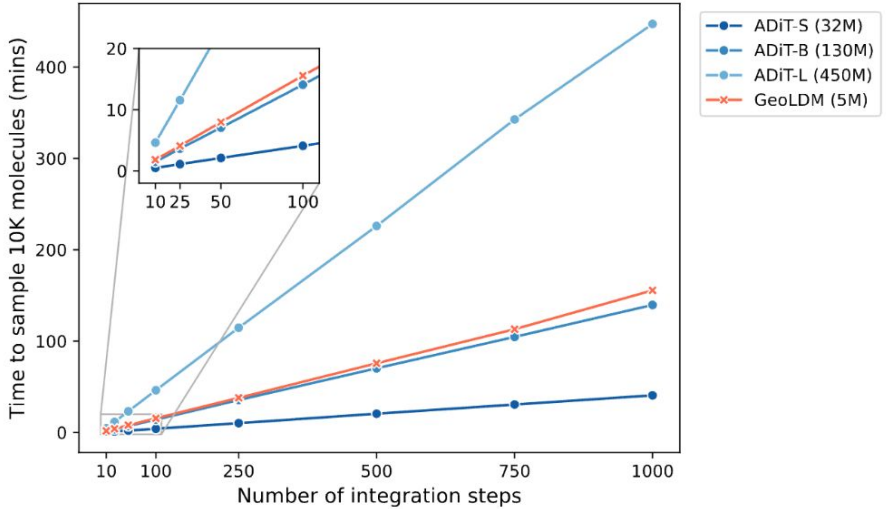
Scaling Results

ADiTs are faster than equivariant diffusion

Time to sample 10K systems vs. number of diffusion timesteps



(a) Crystals – MP20

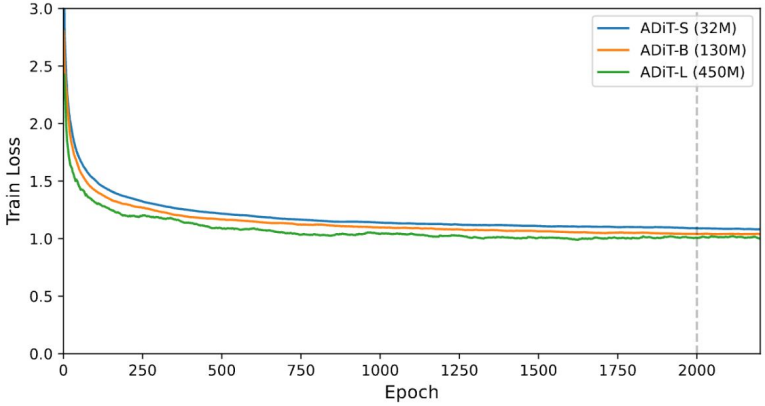


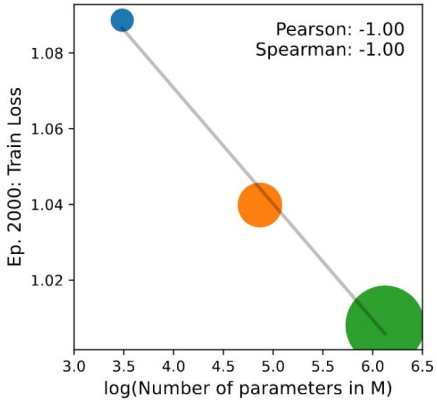
(b) Molecules – QM9

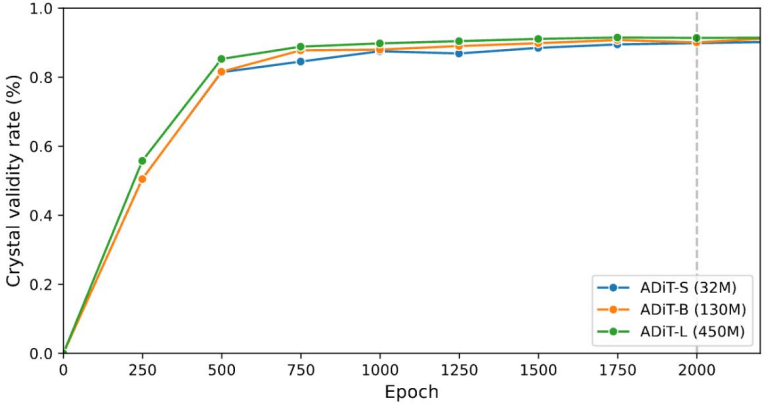
And using **standard Transformers** enables scaling up to **500M parameters** (not possible with expressive Equivariant GNNs).

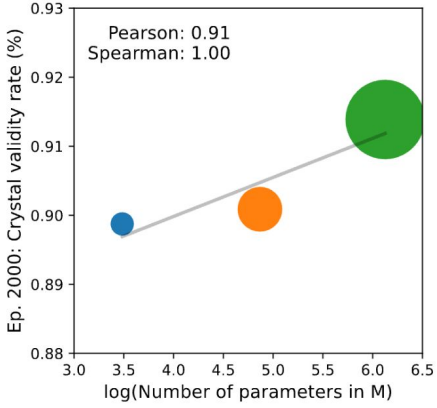
Scaling laws for ADiT

Performance improves predictably with DiT denoiser size



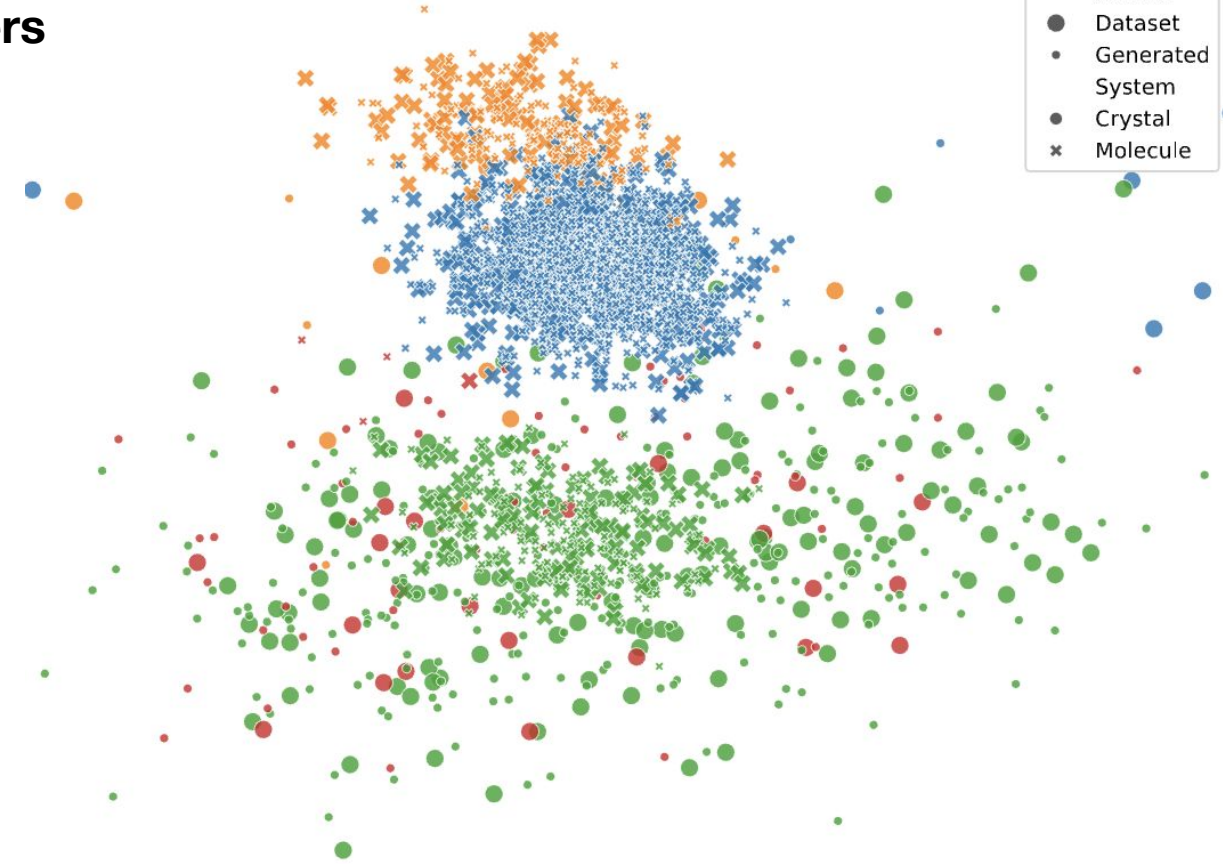






First two PCAs are interpretable

PC 1: Periodic or non-periodic system
PC 2: Atom type clusters



Conclusion

One-slide summary

Generating molecules and materials with latent diffusion models

- **What?** A generative model which can generalize across chemical systems, from periodic crystals to non-periodic molecular systems.
- **How?** Latent diffusion framework:
  a. Variational Autoencoder (VAE) learns a shared latent space by reconstructing all-atom representations of both molecules and materials.
  b. Diffusion Transformer (DiT) samples new latents from the shared distribution, which are decoded to valid molecules or crystals using the VAE.
- **Why?** First step towards broadly generalizable foundation models for designing chemical systems.

Outlook

First step towards generative foundation models for chemistry

- This paper introduced a new architecture. What's next?
- Enabling practical inverse design via conditional training:
  - Conditioning on experimental properties
  - Motif scaffolding and partial infilling
  - **Collaborate with scientists!**
- And scaling to larger and more diverse datasets.
  - Data curation from ZINC, PDB, CCDC, Alexandria, etc.
  - **Good engineering essential!**