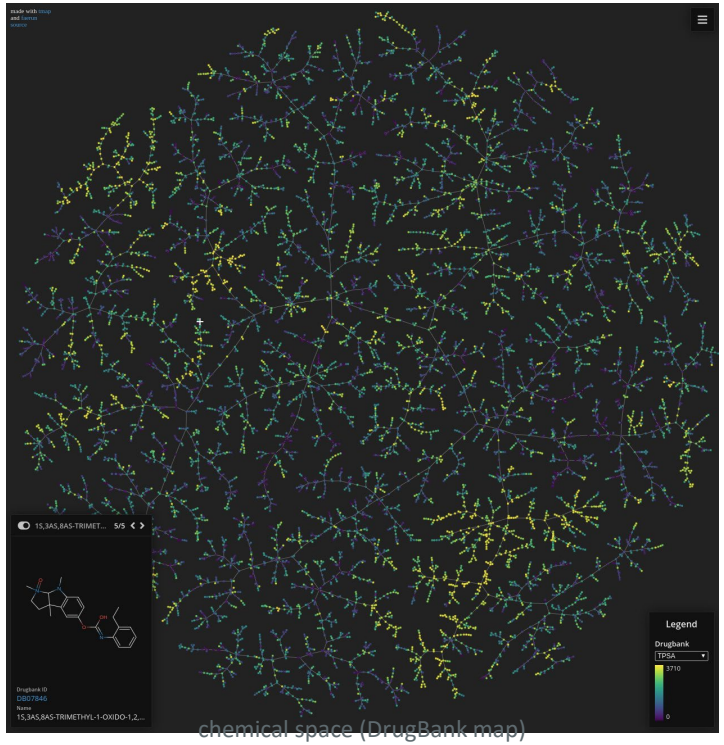
KCDS SUMMER SCHOOL 2026 · SESSION 2 OF 3

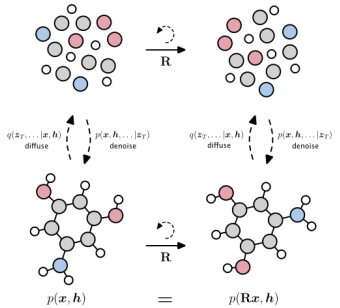
Generating molecules and materials

Generative models, reinforcement learning and foundation models

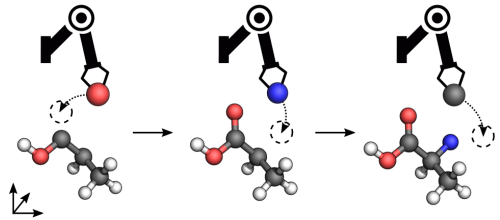
Geometric GNNs and Generative AI for Materials Discovery
Arghya Bhowmik · DTU
KIT Karlsruhe · 25 September 2026



chemical space (DrugBank map)



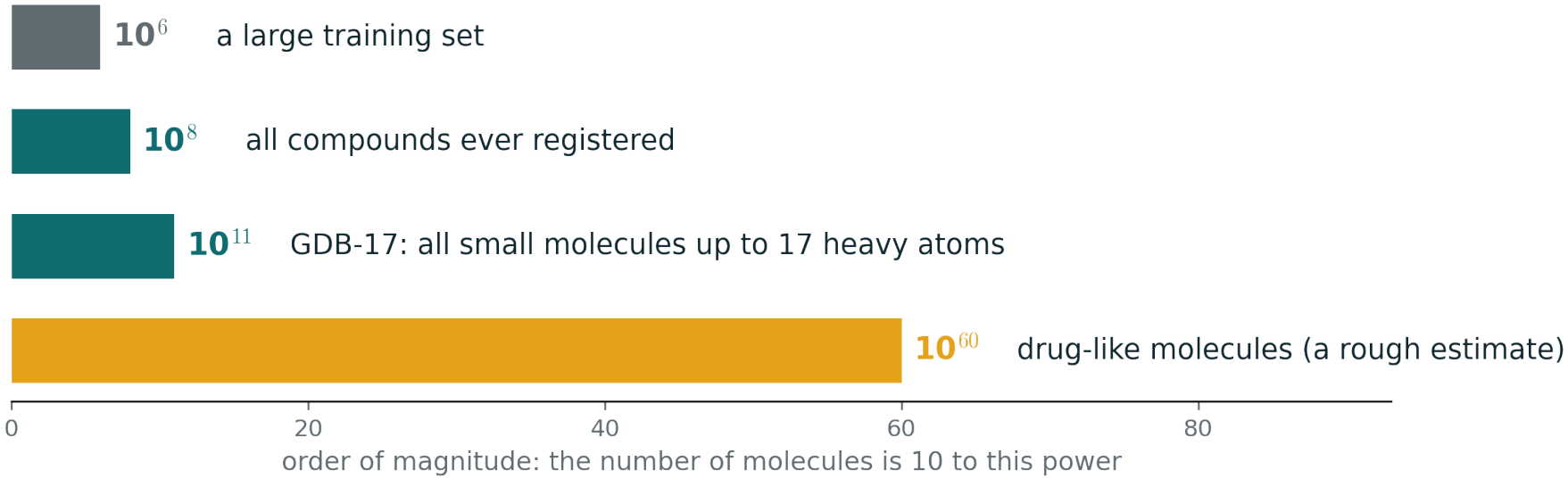
a generative model (EDM)

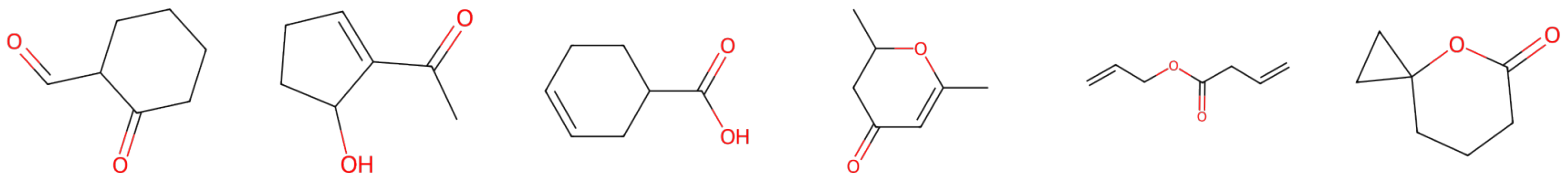


an RL agent (MolGym)

Images: TMAP (Probst & Reymond 2020), EDM (Hoogeboom et al. 2022), MolGym (Simm et al. 2020); MIT-licensed code repositories

Molecular space is huge

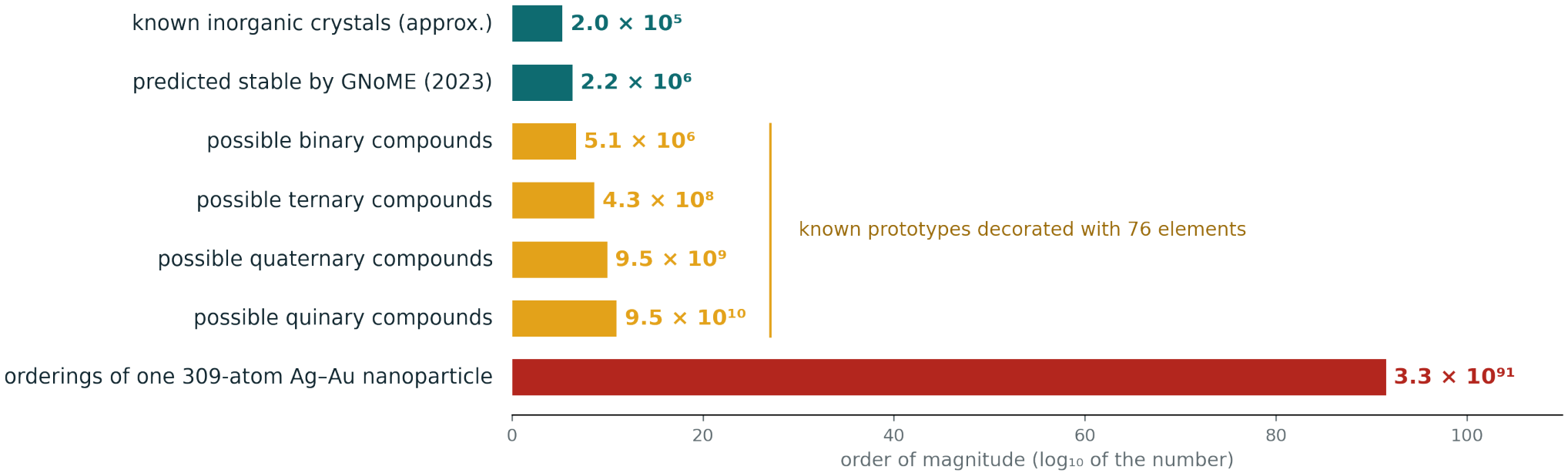


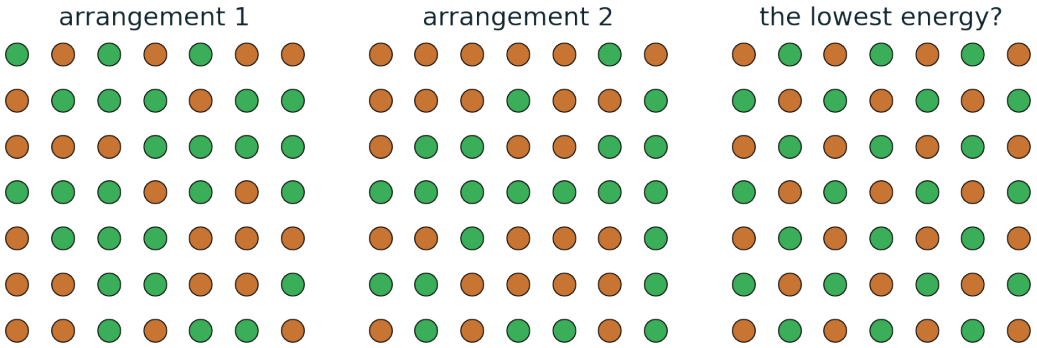


Six of the 6,095 molecules with the formula $C_7H_{10}O_2$ in QM9

We can never test them all. We need models that **suggest** good ones.

Crystal space is huge too



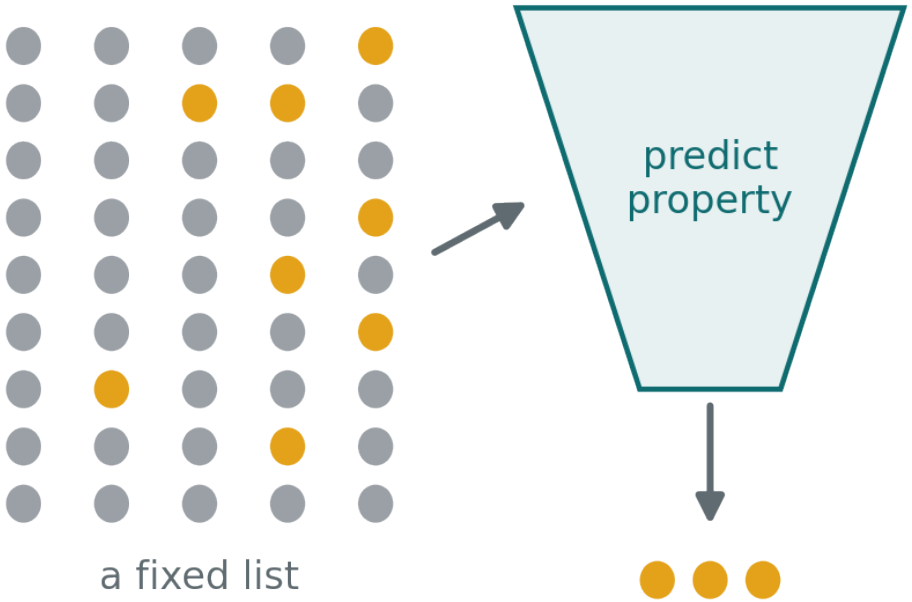


Just decorating **known prototypes** with 76 elements gives about **10^{11}** possible quinary compounds. Then each one can be ordered in many ways.

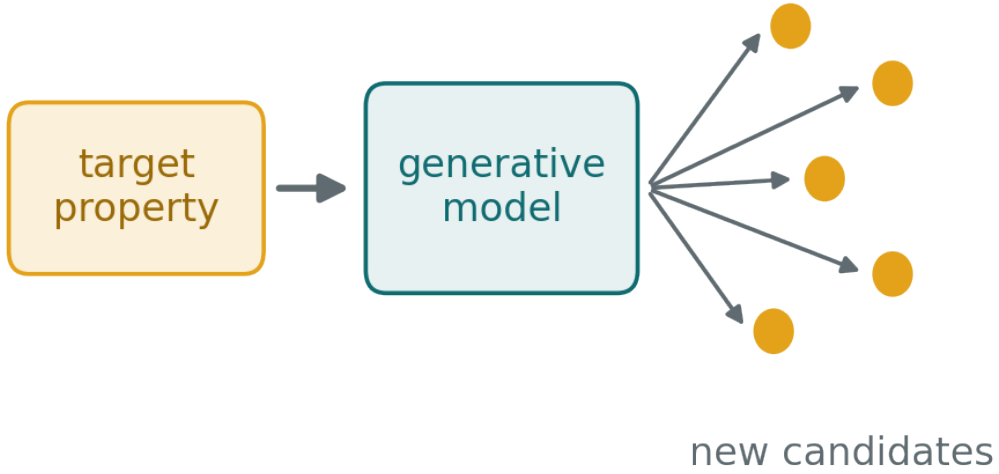
Compound counts: Sci. Adv., doi:10.1126/sciadv.adq1431. Nanoparticle count: Elsborg, Hovmand & Bhowmik, ACS Mater. Au 6, 696 (2026). Known and GNoME counts approximate.

From screening to inverse design

Screening: check every item on a list



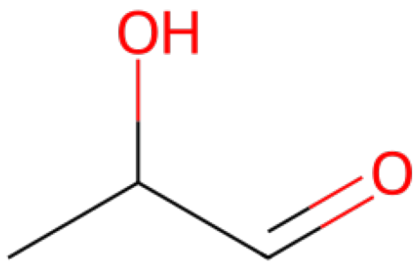
Inverse design: ask for what you want



Both need fast property predictions. GNNs give us those, and they also sit **inside** the generative models.

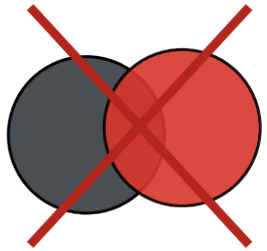
What makes a generated structure good?

chemically valid



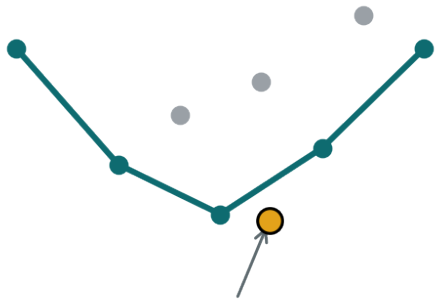
every atom has
the right number of bonds

physically valid



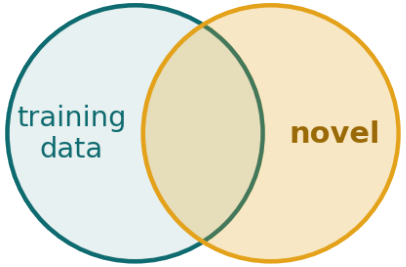
no atoms on top
of each other

stable



new, on or below
the hull: stable

unique and novel



not a copy of the
training data or of each other

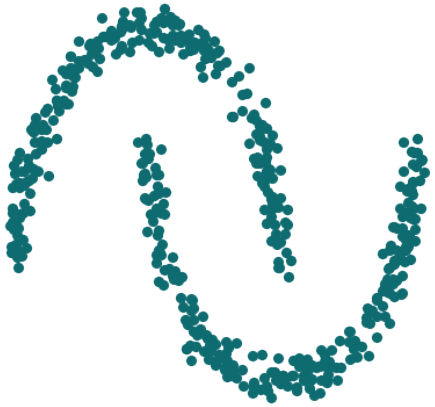
Stable + unique + novel = S.U.N. An example from our battery-materials search:

32,600	generated	→	**27,227**	near the hull after relaxation	→
12,550	stable, unique and novel				

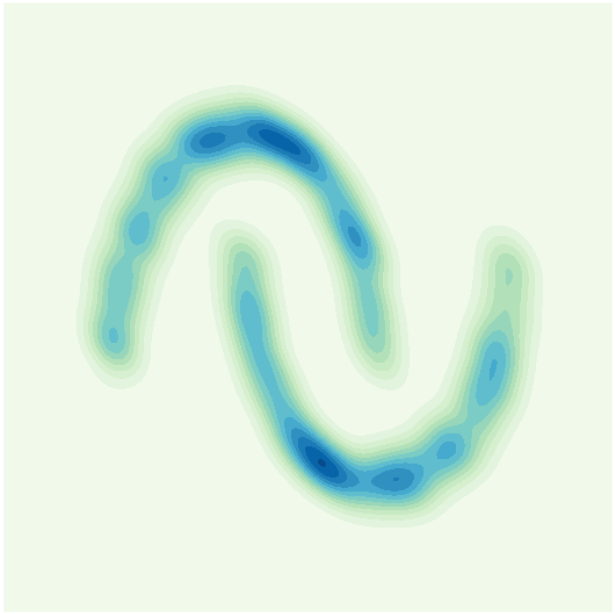
Parida, Roy, García Lastra & Bhowmik, Batteries & Supercaps (2025).

What a generative model does

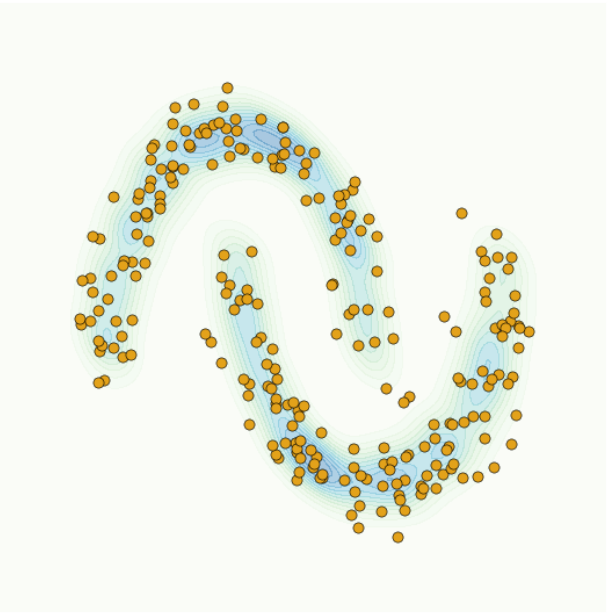
1. training data $x \sim p_{\text{data}}$



2. learn a model $p_\theta(x)$



3. sample new $x \sim p_\theta$



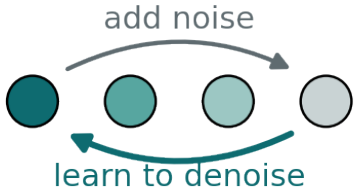
Inverse design needs one more step: learn p(x | y), the structures that have a property y, and sample from that.

Five ways to generate

VAE

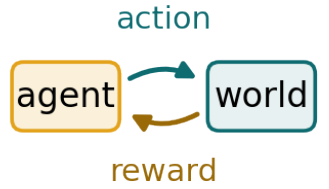
normalising flow

diffusion / flow matching



autoregressive / LM

reinforcement learning



VAE: squeeze into a smooth map, then decode

Language model: write it token by token

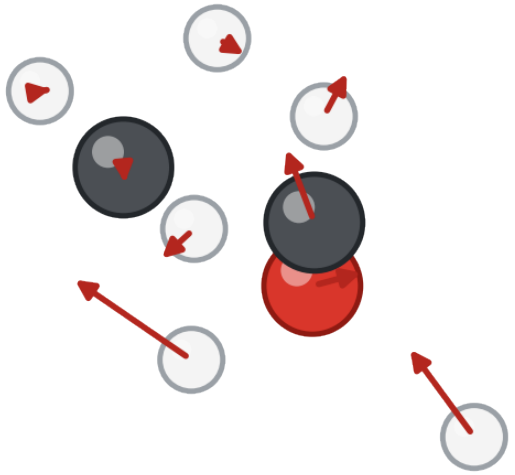
Flow: an exact, reversible transformation

RL: build it step by step for a reward

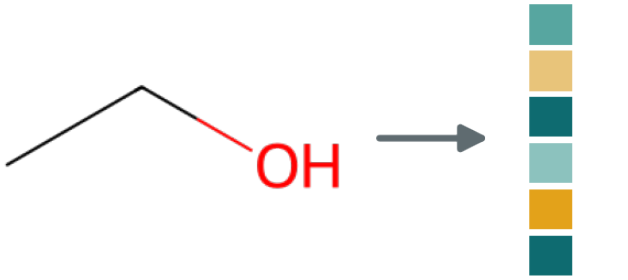
Diffusion / flow matching: start from noise, learn to clean it

Where GNNs sit inside generative models

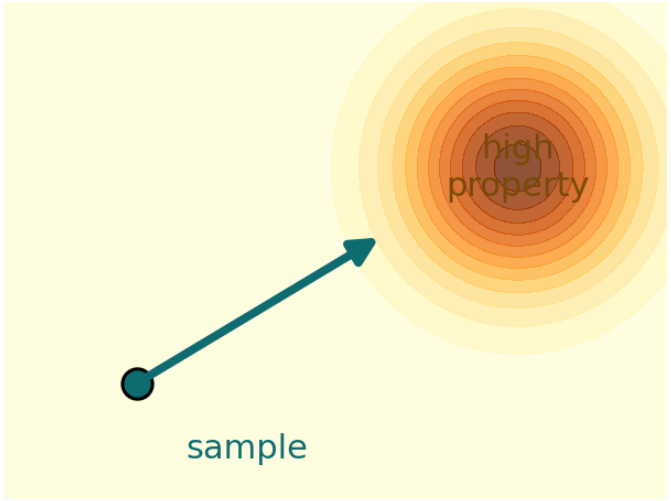
denoiser: predict the noise on every atom



encoder: turn a structure into a vector



guide: a property model pushes samples where we want them



The same layers as in Lecture 1: invariant outputs for atom types, equivariant outputs for positions.

How diffusion works

forward: $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$ (fixed, no learning) →

data x_0 noise x_T

← **reverse: a network $\epsilon_\theta(x_t, t)$ removes the noise step by step (learned)**

forward: $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$

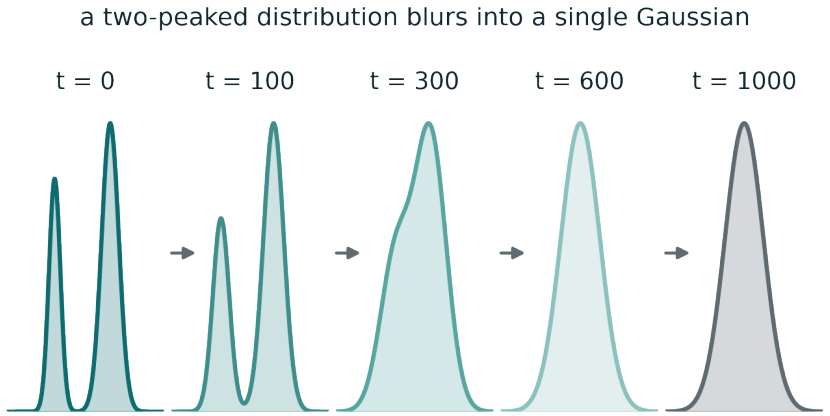
training: $\mathcal{L} = \mathbb{E}_{t, x_0, \epsilon} \| \epsilon - \epsilon_\theta(x_t, t) \|^2$

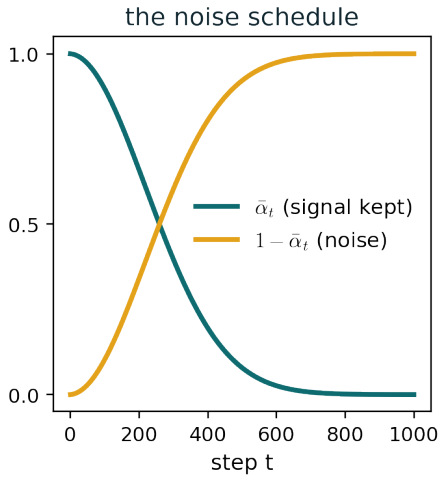
link to the score: $\nabla_{x_t} \log q(x_t|x_0) = -\epsilon / \sqrt{1-\bar{\alpha}_t}$

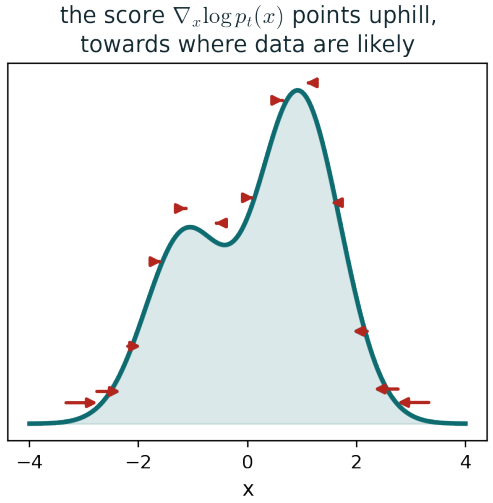
- Destroying data is easy: add noise.
- Learn the reverse: predict the noise.
- Predicting the noise = learning the score.

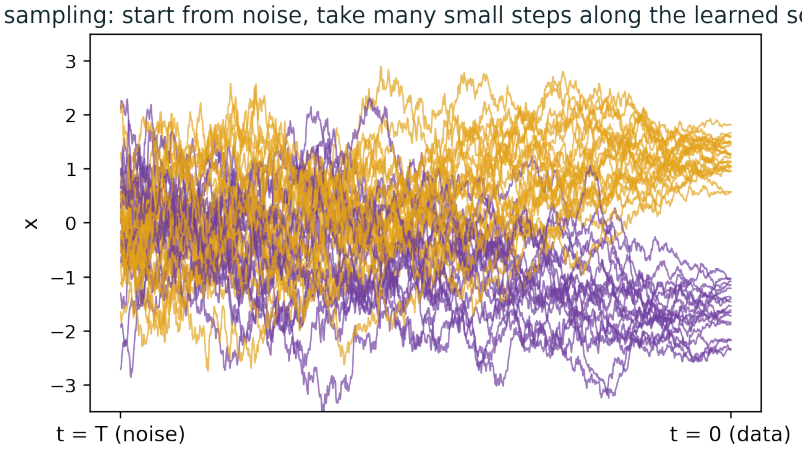
Ho, Jain & Abbeel, NeurIPS 2020 (DDPM); Song et al., ICLR 2021. Toy example computed for this lecture.

Diffusion runs in two directions









1 · Forward (fixed)

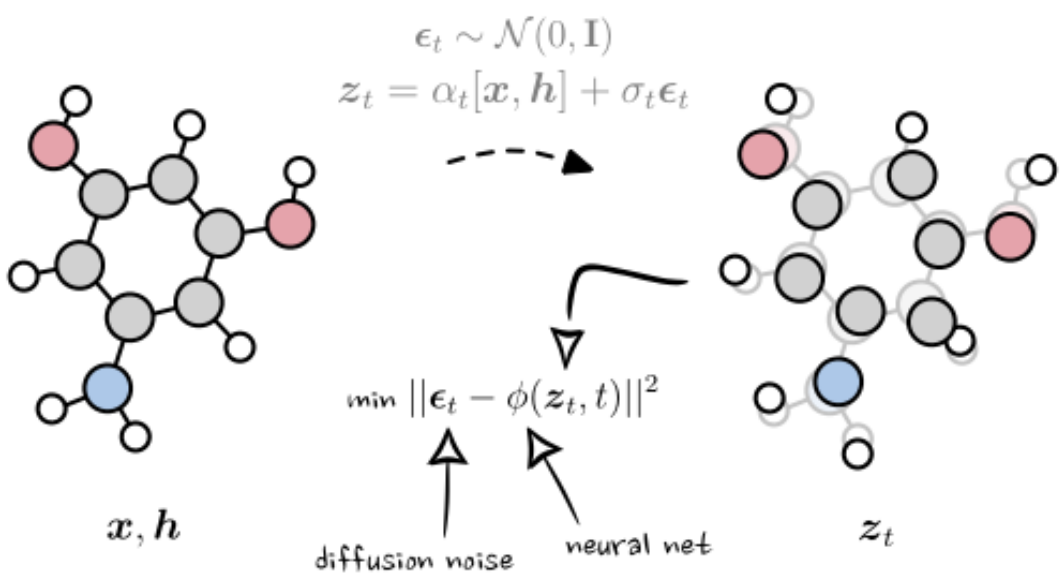
- Shrink the signal, add a little noise, again and again.
- After ~1,000 steps: pure noise.

2 · Reverse (learned)

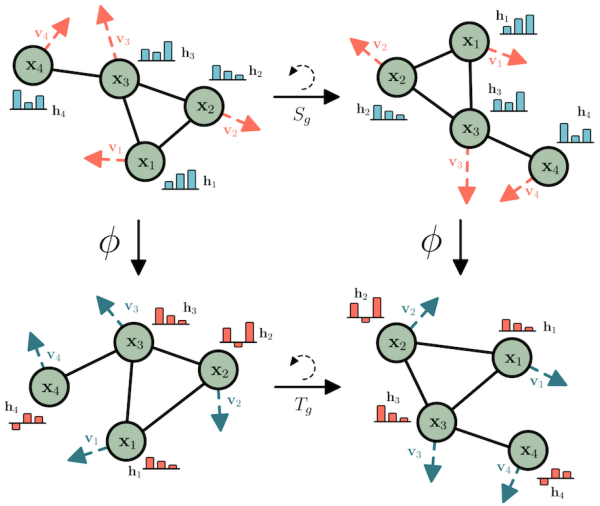
- The score points to where data are likely.
- Start from noise, follow it in small steps.

Ho, Jain & Abbeel, NeurIPS 2020; Song et al., ICLR 2021. Toy mixture with its exact score, computed for this lecture.

For 3D molecules, the denoiser is an equivariant GNN



EDM training: noise atom positions x and types h together



$$m_{ij} = \phi_e(h_i, h_j, \|x_i - x_j\|^2)$$

$$x_i \leftarrow x_i + C\sum_j (x_i - x_j)\,\phi_x(m_{ij})$$

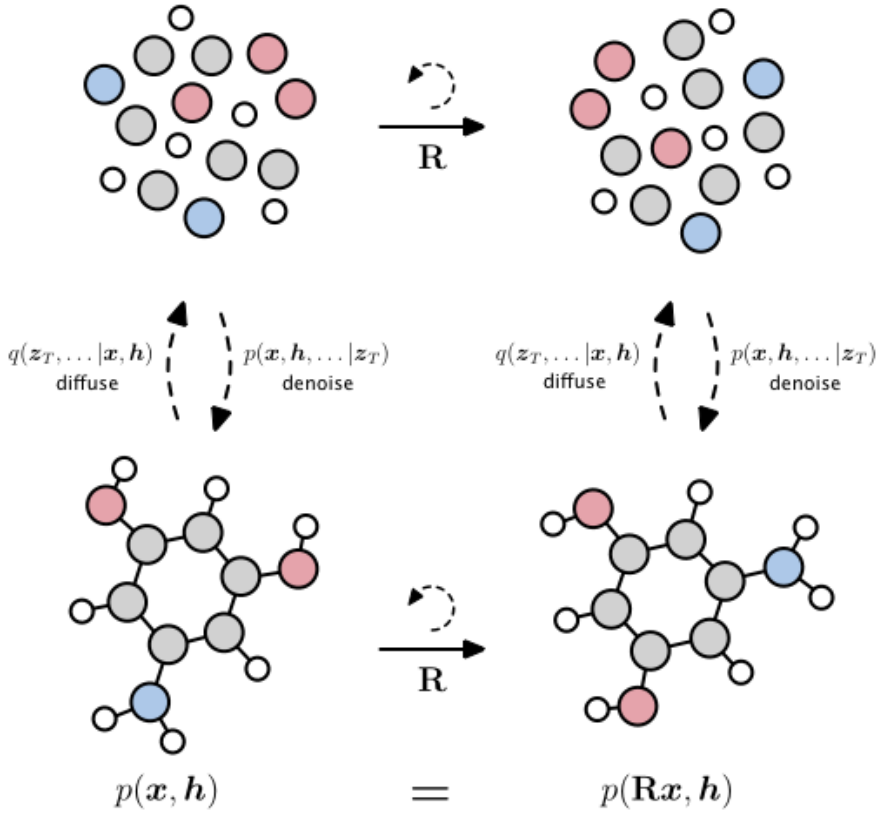
$$h_i \leftarrow \phi_h\left(h_i, \sum_j m_{ij}\right)$$

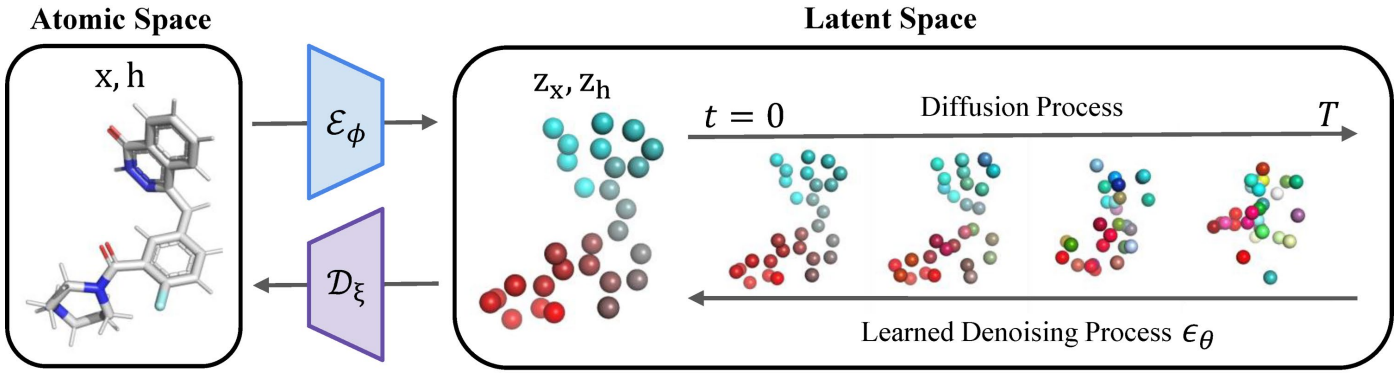
EGNN: rotate the input, the output rotates

- Positions **x** and atom types **h** are noised together.
- Messages use **invariant** inputs (distances, atom types); positions then move along the **vectors between atoms**, weighted by those messages.
- So the noise predicted on positions **rotates with the molecule**, and the noise on atom types stays invariant.
- Distances alone would make an invariant, SchNet-like model; the vector update is what makes EGNN equivariant.

Figures from the EDM and EGNN code repositories (MIT): Hoogeboom et al., ICML 2022; Satorras, Hoogeboom & Welling, ICML 2021.

Why equivariance matters, and diffusion in a latent space





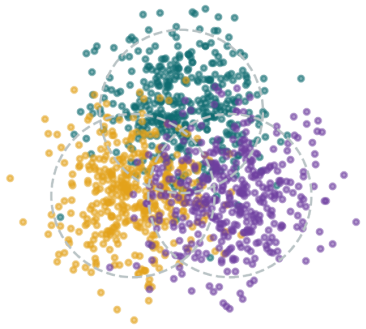
GeoLDM: encode first, then diffuse in a smaller latent space

- Remove the centre of mass, and use a rotation-invariant noise prior.
- With an equivariant denoiser, every rotation of a molecule gets the **same probability**.
- Latent diffusion trades exact 3D noise for speed and quality.

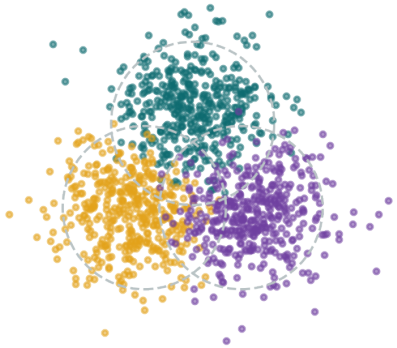
Figures from the EDM and GeoLDM code repositories (MIT): Hoogeboom et al., ICML 2022; Xu et al., ICML 2023.

Steering generation towards a property

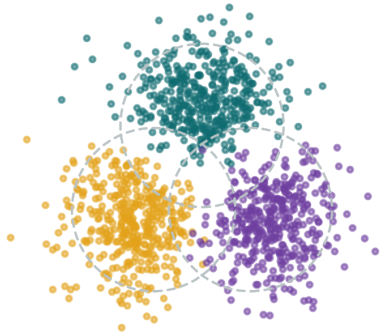
guidance weight w = 0



guidance weight w = 1



guidance weight w = 4



Three classes in 2D. More guidance: samples move away from the other classes

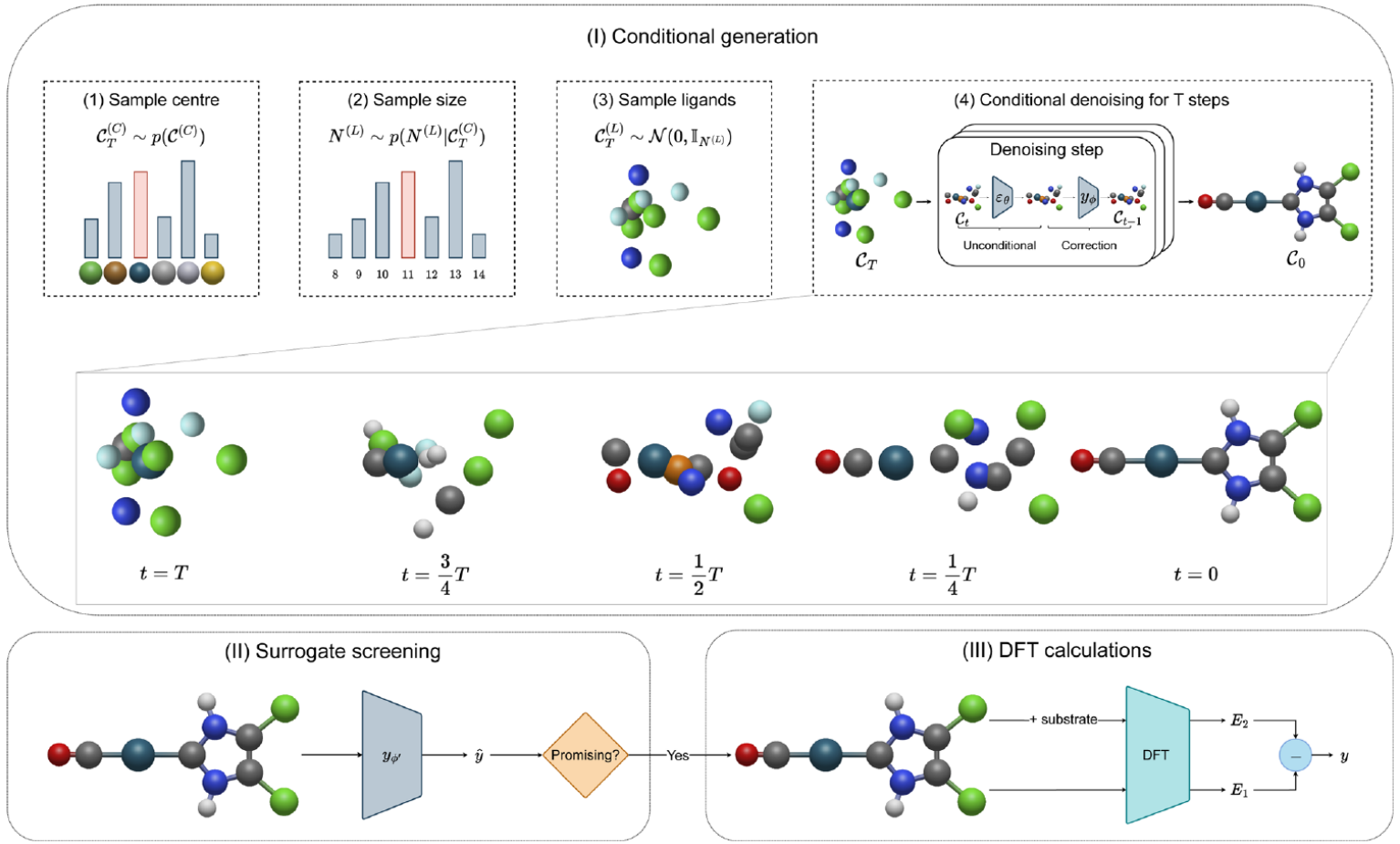
- **Classifier guidance:** add the gradient of a property predictor.
- **Classifier-free:** mix conditional and unconditional predictions.
- Too little ignores the target; too much loses diversity.

Bayes: $\nabla \log p(x \mid y) = \nabla \log p(x) + \nabla \log p(y \mid x)$

classifier-free: $\tilde{\epsilon} = (1 + w)\, \epsilon_\theta(x_t, t, y) - w\, \epsilon_\theta(x_t, t, \varnothing)$

Dhariwal & Nichol, NeurIPS 2021; Ho & Salimans, arXiv:2207.12598 (toy example adapted from their figure and computed for this lecture).

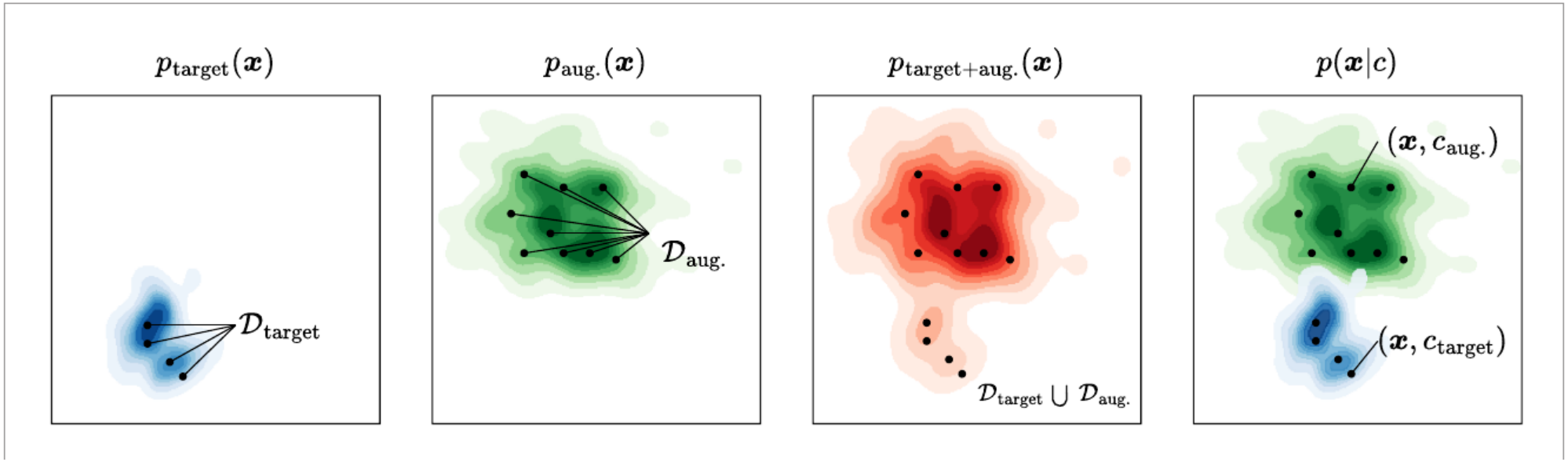
OM-Diff designs catalysts with guided diffusion



- Generates all atoms in 3D around a chosen metal.
- A property model guides it to a target binding energy.
- Screen fast, then check the best with DFT.

Cornet, Benediktsson, Hastrup, Schmidt & Bhowmik, Digital Discovery 3, 1793 (2024), Fig. 1.

Little data? Borrow from related data

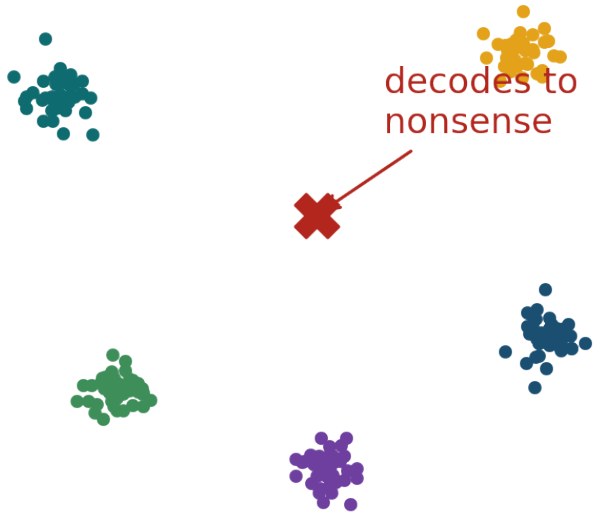


- The target data are few and expensive (blue).
- Add many related, unlabelled structures (green).
- A label keeps them apart, so we can still ask for the target.

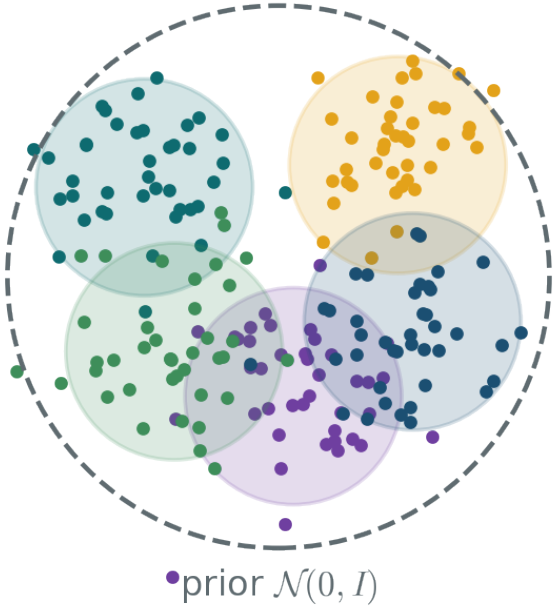
Cornet et al., Mach. Learn.: Sci. Technol. 6, 025057 (2025), Fig. 2.

Why a plain autoencoder is not enough

autoencoder: tight clusters with empty gaps

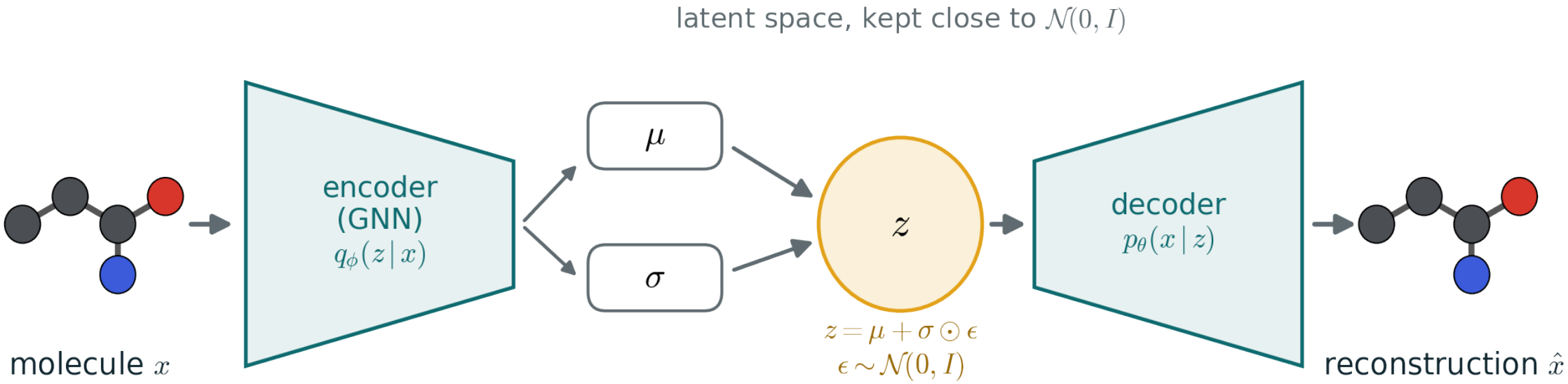


VAE: overlapping clouds fill the space



- A plain autoencoder compresses well, but its map has gaps: points between clusters decode to nonsense.
- A VAE encodes each molecule as a small cloud and pulls all clouds towards one Gaussian.
- So we can pick any point, or walk between two molecules, and still decode something sensible.

How a variational autoencoder works



$$\log p_\theta(x) \geq \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \mathrm{KL}(q_\phi(z|x) \,\|\, \mathcal{N}(0, I))$$

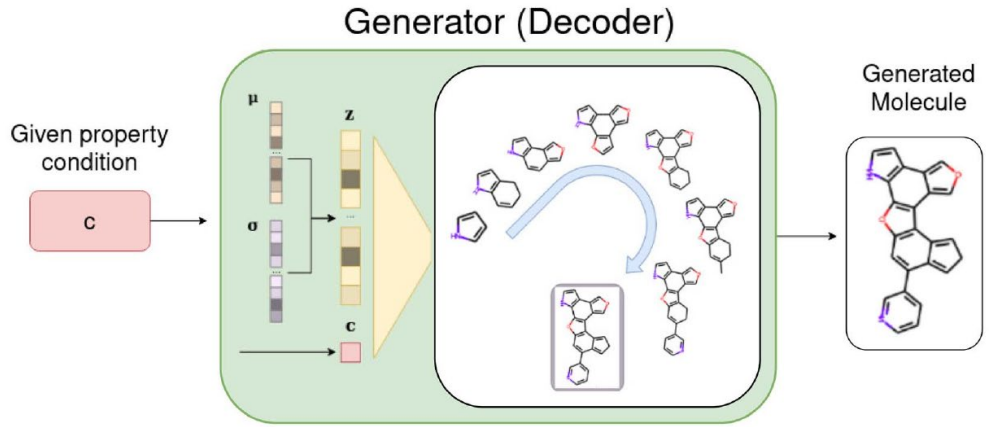
first term: reconstruct the molecule; second term: keep the map smooth

reparameterisation: $z = \mu_\phi(x) + \sigma_\phi(x) \odot \epsilon$

- Encode to a small region, not a point.
- The KL term keeps the map filled and smooth.
- Any point on the map decodes to something sensible.

Kingma & Welling, ICLR 2014; Kingma & Welling, An Introduction to Variational Autoencoders, arXiv:1906.02691 (2019). Diagram drawn for this lecture.

A graph VAE that decodes towards a target

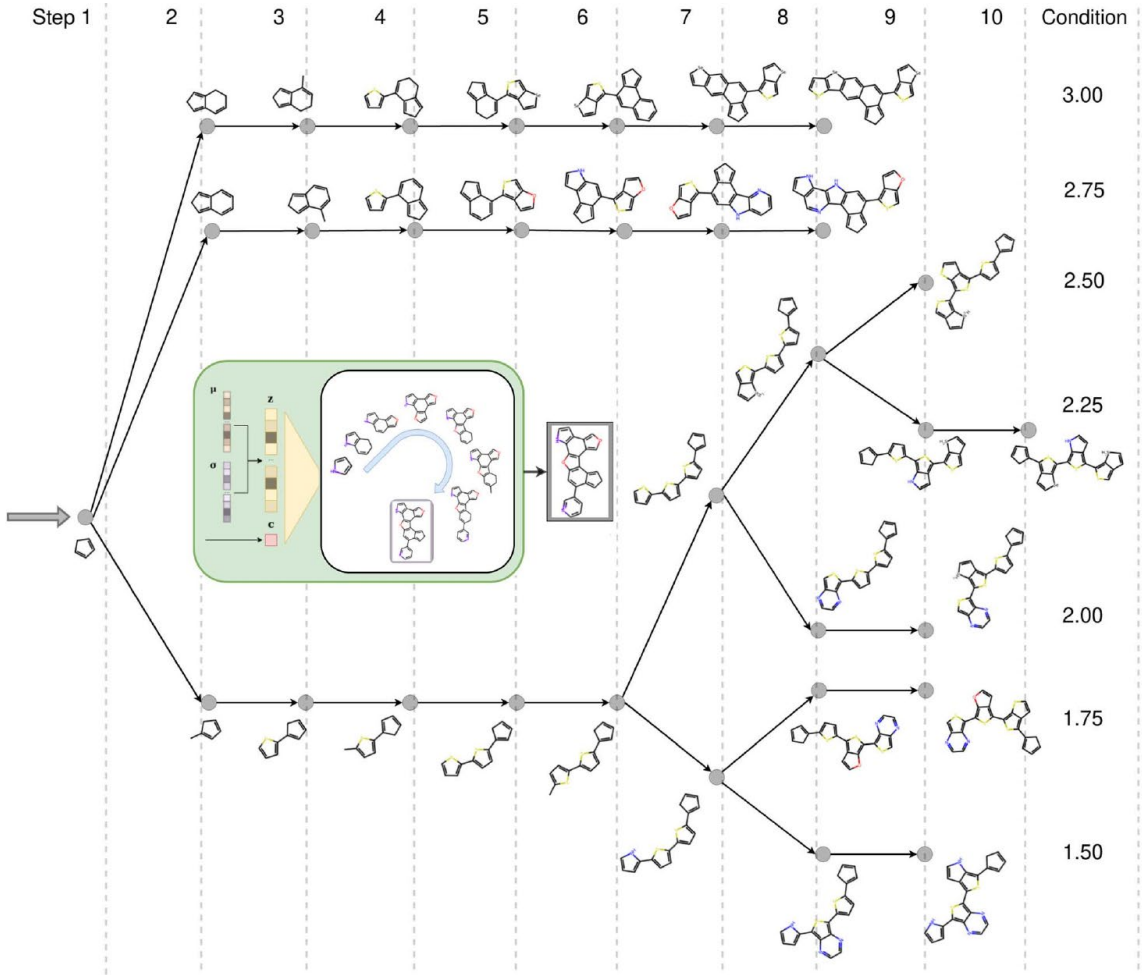


1. **Sample** a latent vector z from the learned distribution (μ, σ).

2. **Append the target**, here the HOMO–LUMO gap c: the decoder reads [z; c].

3. **Grow the graph** one motif per step: pick the next motif and where it attaches.

4. **Same z, different c:** every branch starts from the same motif, then grows different motifs to hit each target gap.

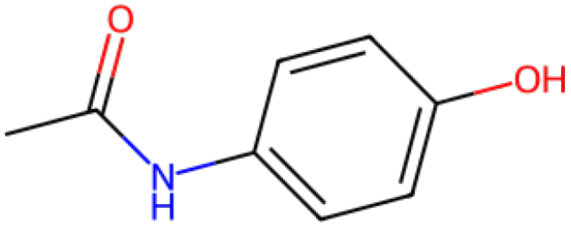


Ortega Ochoa, Benediktsson, Sechi, Jørgensen & Bhowmik, J. Mater. Chem. A 11, 26551 (2023), Fig. 3 (HierVAE-based conditional decoder).

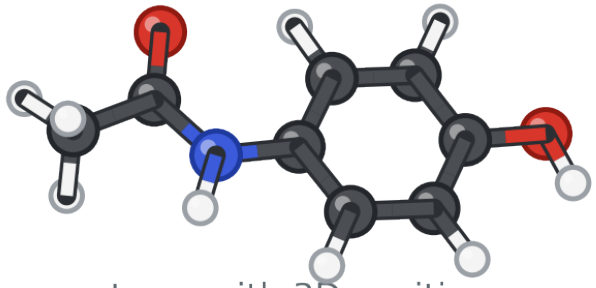
Three ways to generate a molecule

`CC(=O)Nc1ccc(O)cc1`

a string, one token at a time



a graph of atoms and bonds



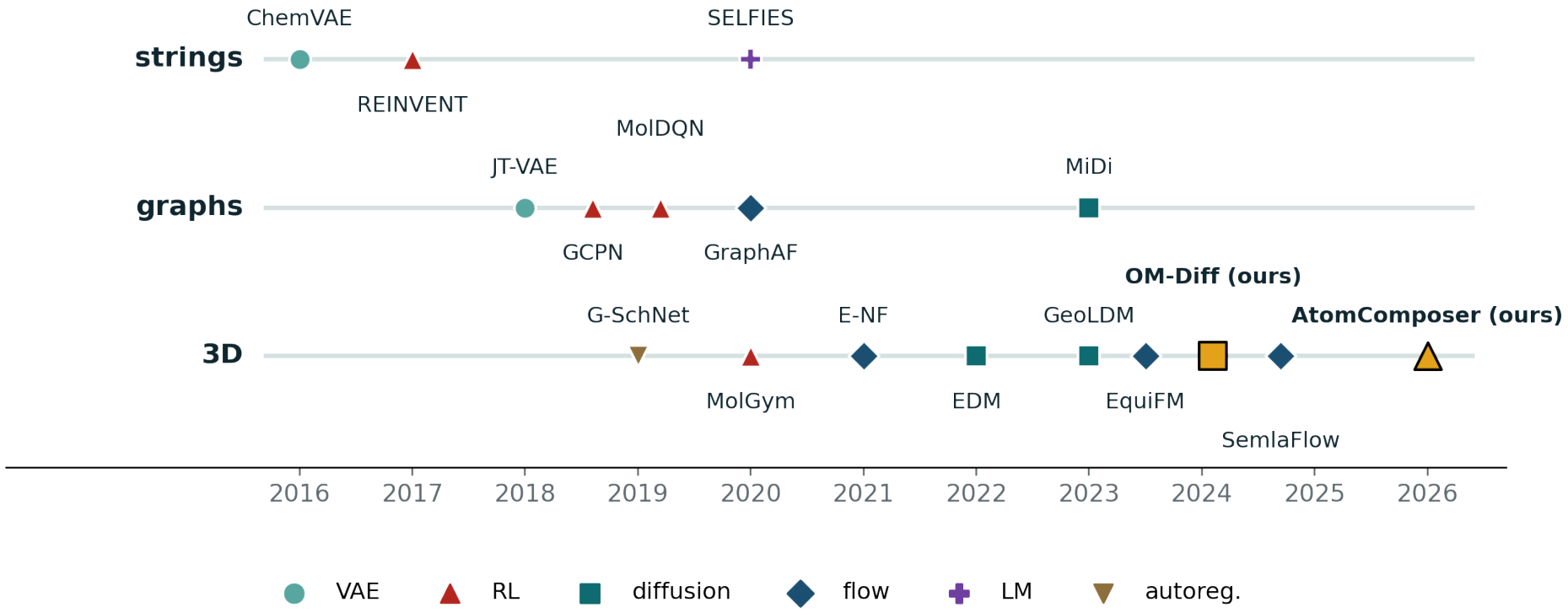
atoms with 3D positions

Strings: fast, huge libraries; no 3D

Graphs: valid bonds by design; no 3D

3D: needed for shape: binding, catalysis

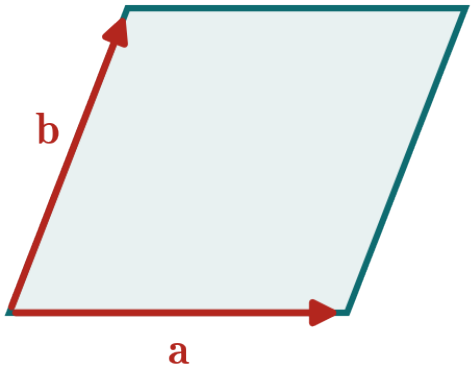
Generative models for molecules, over time



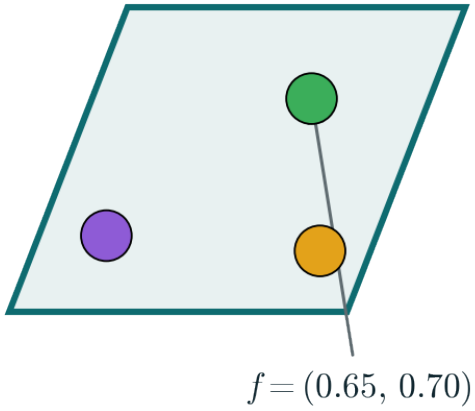
Data: **ZINC**, **ChEMBL** (strings, graphs); **QM9**, **GEOM** (3D). Benchmarks: GuacaMol, MOSES.

What a crystal is, for a generative model

1. the box: lattice L

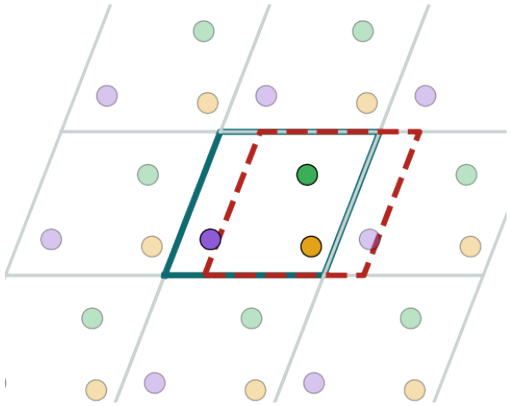


2. atoms A at fractional positions $F \in [0,1)^3$

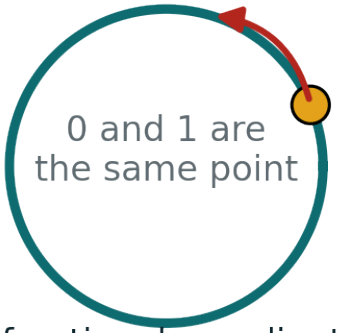


$f = (0.65,\ 0.70)$

3. repeats forever: a shifted box (red) is the same crystal



4. positions wrap around



a fractional coordinate lives on a circle (torus)

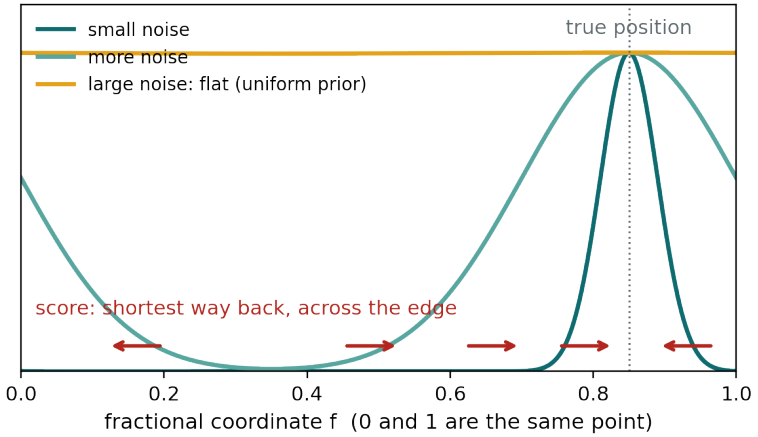
$\mathcal{M} = (A,\ F,\ L)$: atom types, fractional coordinates $F \in [0,1)^{N\times 3}$, lattice $L \in \mathbb{R}^{3\times 3}$

invariant to: atom order, rotating L, shifting $F \to (F+t) \bmod 1$, choice of cell

- Plus **symmetry**: 230 space groups.
- And the atom count varies.

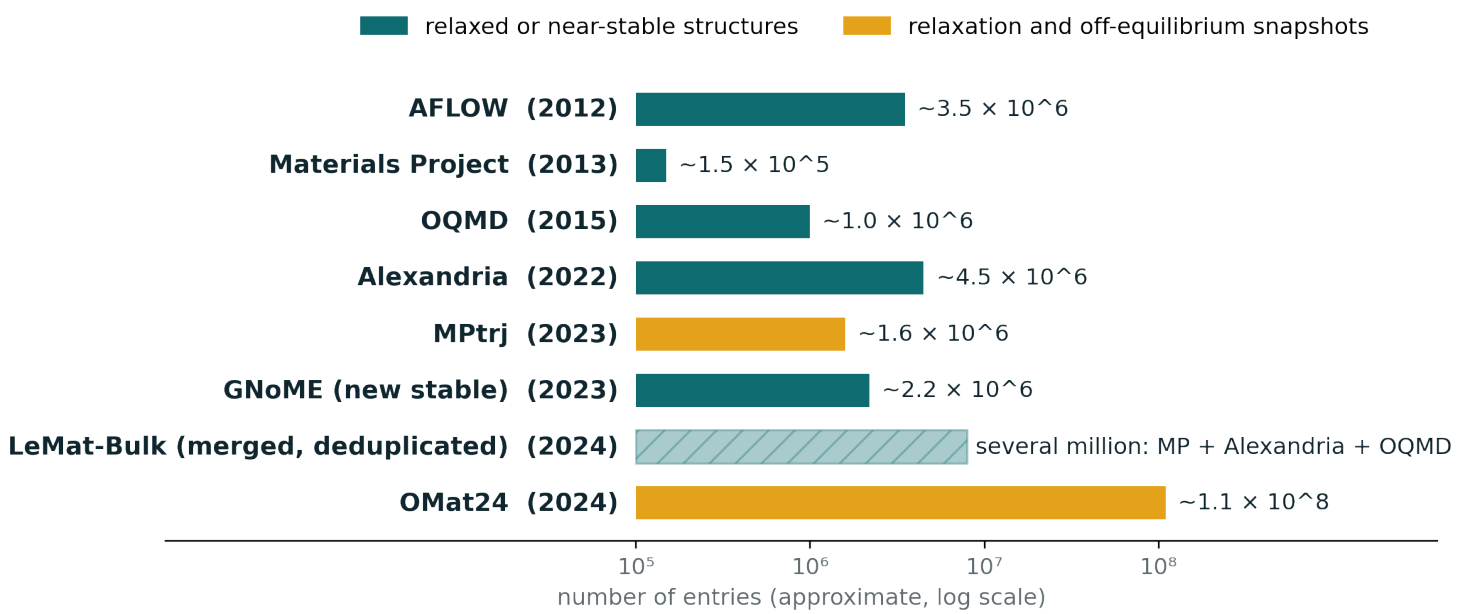
What changes from molecules to crystals

Top: molecule. Bottom: crystal (atoms wrap, the cell changes, types are masked)



	Molecule	Crystal
Positions	Cartesian; centre of mass removed; Gaussian prior	fractional, on a torus; wrapped noise; uniform prior
The cell	none	lattice L is noised and denoised too
Atom types	continuous noise on one-hot types	often discrete: types are masked
Denoiser graph	edges within the molecule	edges into neighbouring cells (offsets n); distances from the noisy cell
Must not matter	rotation, translation, atom order	the same, plus shifting all atoms and the choice of cell

Chemical-space-wide data make foundation generators possible



- From **~150,000** Materials Project entries to **millions** across the periodic table (Alexandria, GNoME, LeMat-Bulk).
- MPtrj and OMat24 add relaxation and off-equilibrium snapshots, mainly for potentials.
- **MatterGen** trains on ~600,000 stable crystals (Alex-MP-20), over ten times the ~45,000 in MP-20.
- Pre-trained on broad data, small **adapters** then steer it towards a property.

Novelty is judged against an even bigger reference set, which includes experimental structures from the ICSD.

Sizes approximate, from the MP, OQMD, AFLOW, Alexandria, MPtrj, GNoME, LeMat-Bulk and OMat24 papers; MatterGen: Zeni et al., Nature 2025.

Two ways to use a crystal generator

Crystal structure prediction (CSP)

input: Na_2Cl_2

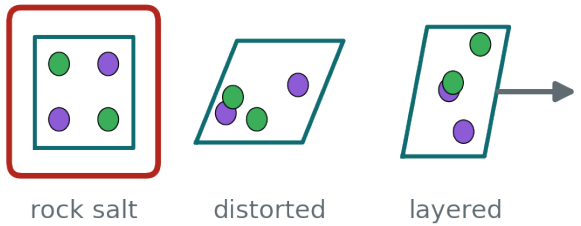
2 Na and 2 Cl per cell

→

generator

types fixed; noise only on F and L

→



rock salt | distorted | layered

relax (ML potential, DFT)

lowest energy = prediction

De novo generation (DNG)

input: nothing

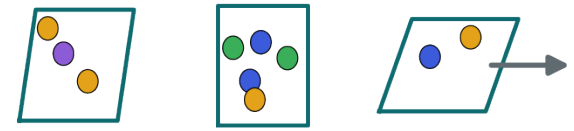
or a target: "stable, band gap ≈ 2 eV"

→

generator

noise on types A, F and L

→



new ternary | new quaternary | new binary

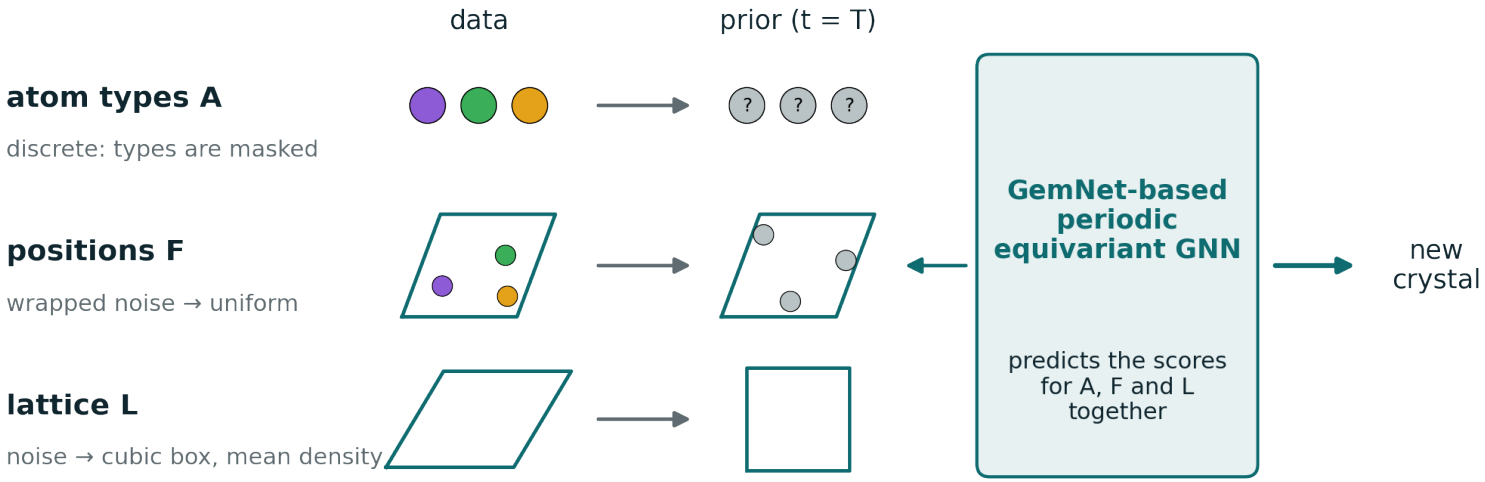
relax, then filter:

stable, unique, novel (S.U.N.)

CSP: structure prediction. Given a formula with a fixed number of atoms of each type per cell, generate only the cell and the positions. Relax many candidates and keep the lowest energy. Scored by **match rate** against known structures.

DNG: de novo generation. Generate atom types, positions and the cell together, optionally towards a target property. Relax, then keep what is **stable, unique and novel**. Scored by the **S.U.N. rate**.

How MatterGen works, and what it learns from



- Diffuses **types, positions and cell** together.
- Types are masked; positions get wrapped noise; the cell relaxes to a cubic box.
- One GemNet-based periodic network predicts all three.

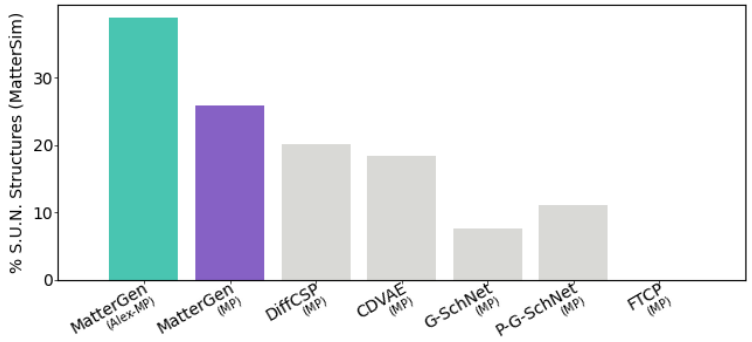
Training data
Alex-MP-20: 607,683 crystals from Alexandria and MP, ≤ 20 atoms, within 0.1 eV/atom of the hull

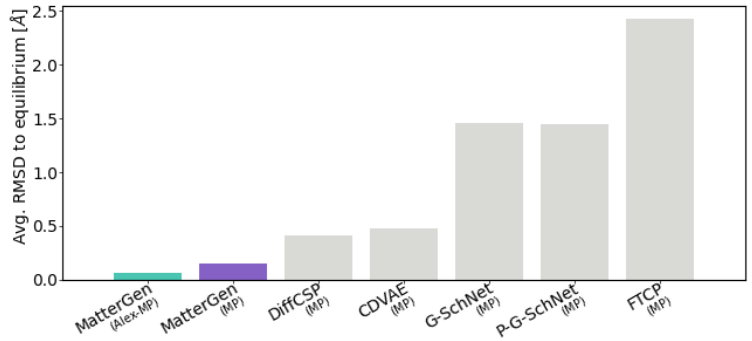
Novelty reference
Alex-MP-ICSD: about 850,000 known structures, including experimental ones from the ICSD

A fair comparison
MatterGen-MP: the same model trained on MP-20 only, to separate model from data

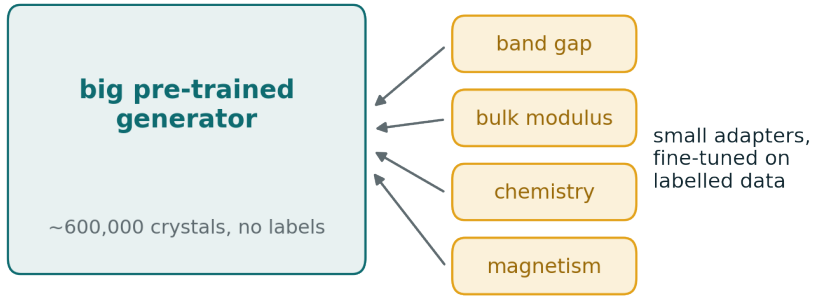
Zeni et al., "A generative model for inorganic materials design", Nature 639, 624 (2025); code and weights: github.com/microsoft/mattergen (MIT).

What MatterGen achieves, and a caveat





Teal: MatterGen. Purple: the same model on MP-20 only



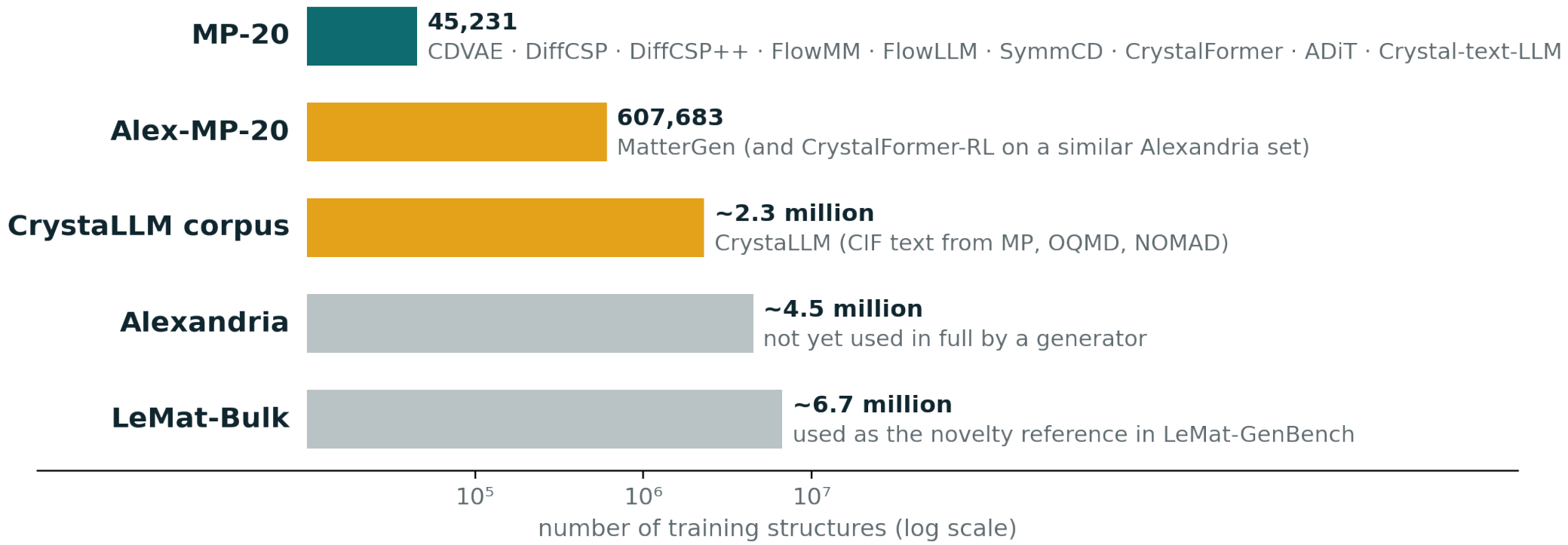
- **2×** the stable, unique and novel rate of earlier models.
- Structures **10× closer** to their DFT minimum.
- Part of the gain is the model; the rest is the **bigger data**.

Fine-tune adapters for:
chemical system
space group
magnetic density
band gap
bulk modulus
supply-chain risk (HHI)

A caveat: $TaCr_2O_6$ was designed for a bulk modulus of 200 GPa and measured at about 169 GPa. But it formed with Ta/Cr disorder, and a disordered phase was already known.

Plots from the MatterGen repository (MIT); Zeni et al., Nature 639, 624 (2025). On novelty and disorder, see also Cheetham & Seshadri, Chem. Mater. (2024).

Most crystal generators learn from very little data

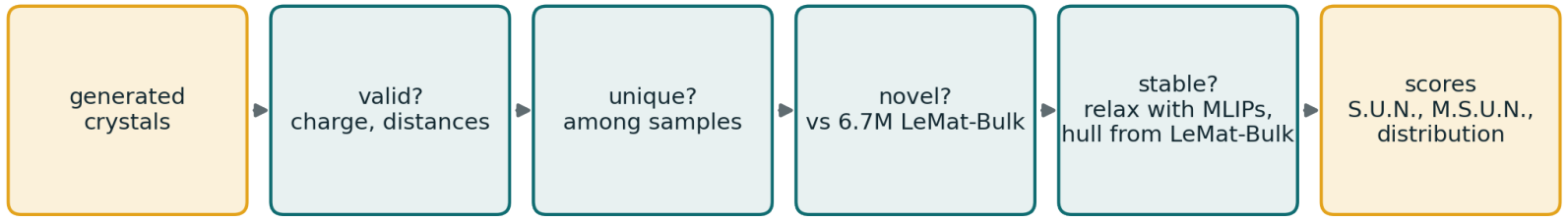


Most generators are trained on **MP-20 alone**. A foundation generator should learn from the **broadest, consistent data** available, and full Alexandria or LeMat-Bulk is still largely unused.

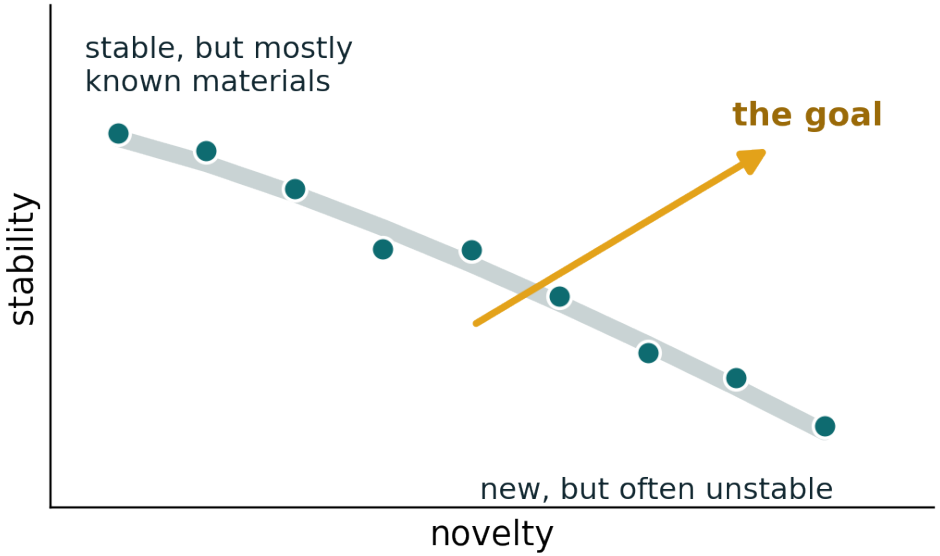
Sizes from the MP-20 (CDVAE), MatterGen, CrystaLLM, Alexandria and LeMat-Bulk papers; approximate where marked ~.

Testing crystal generators fairly, and what they need

LeMat-GenBench: one pipeline for every model



the same checks, the same reference data and the same potentials for every model

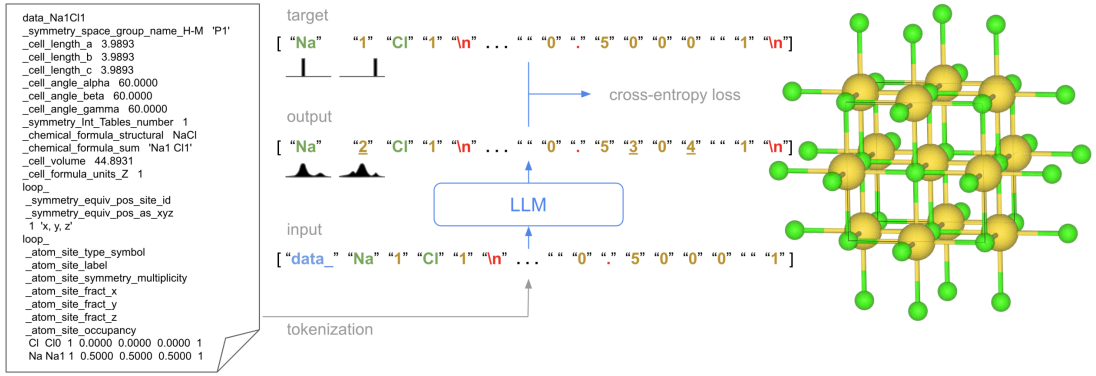


What a universal generator needs

1. broad, consistent data
2. most of the periodic table
3. symmetry built in
4. ways to steer it
5. careful evaluation
6. open and fast

LeMaterial: LeMat-Bulk (2024, ~6.7M entries, CC BY 4.0) and LeMat-GenBench (2025). Trade-off plot is schematic. See also Park, Li & Walsh, Matter 7, 2355 (2024).

Other routes to large-scale crystal generation



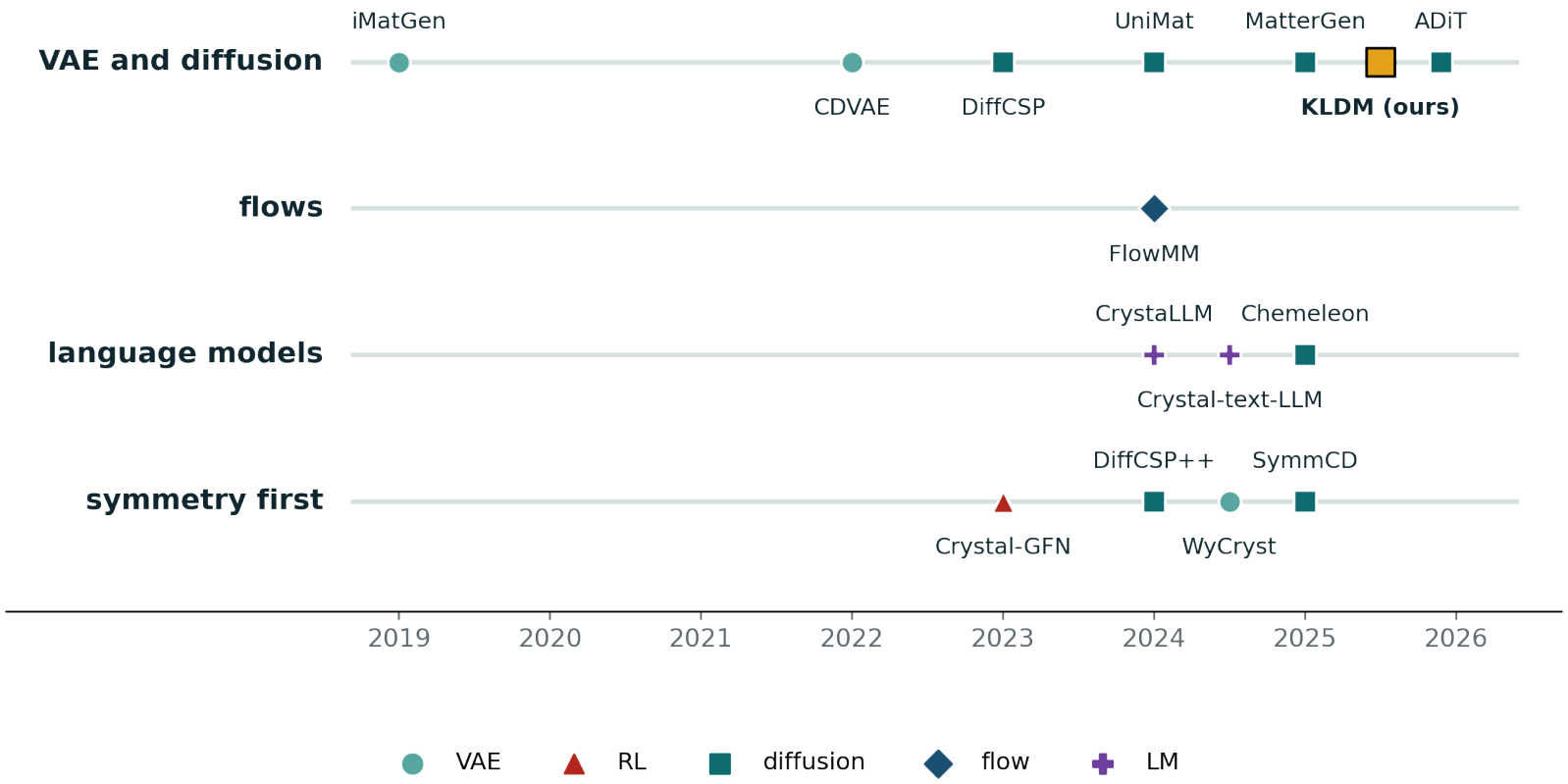
CrystaLLM: a language model that writes CIF files

- Most are trained on MP-20; CrystaLLM and CrystalFormer-RL use broader data.
- Language models learn symmetry from data rather than having it built in.

CrystaLLM	GPT-style model on ~2.3 million CIF files
CrystalFormer (-RL)	space-group and Wyckoff tokens; RL with ML-potential rewards
Chemeleon	text-guided diffusion; cells up to 40 atoms
ADiT	one latent diffusion transformer for molecules and crystals
FlowMM, FlowLLM	flow matching, optionally seeded by a language model
UniMat	diffusion at GNoME scale, not released

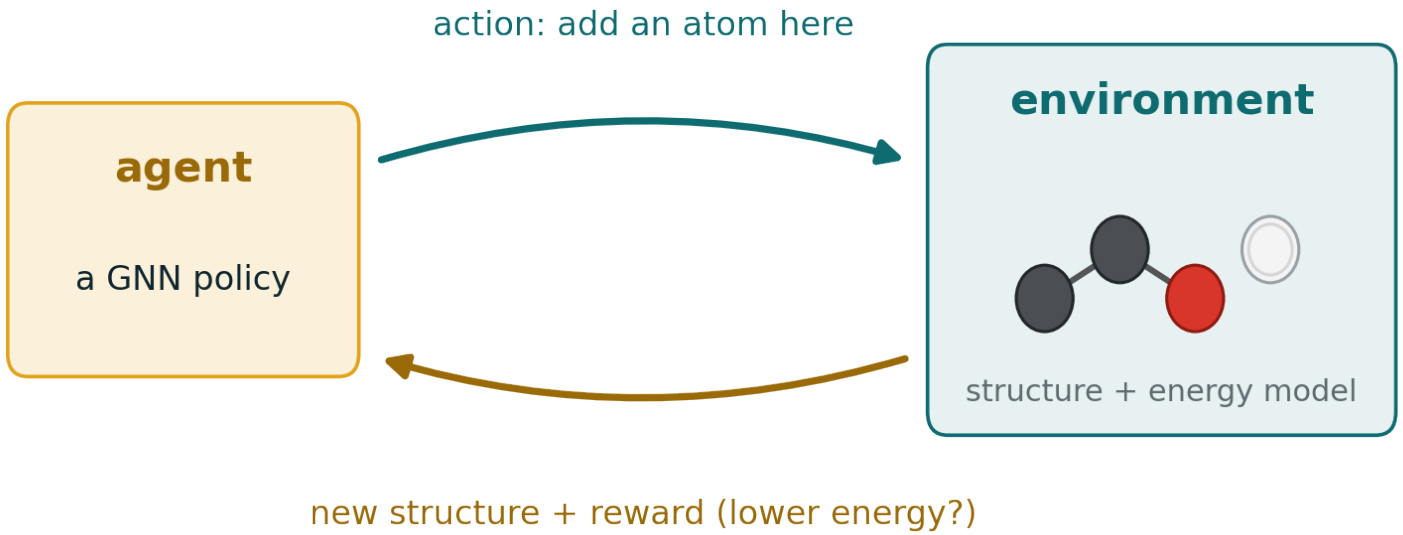
Antunes et al. 2024; Cao & Wang 2024–25; Park et al. 2025; Joshi et al. 2025; Miller et al. 2024; Sriram et al. 2024; Yang et al. 2024. Figure: CrystaLLM repository (MIT).

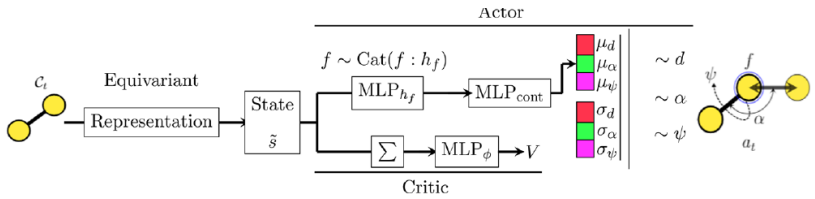
Crystal generators over time



Data: **MP-20**, **Alex-MP-20**. Samples are relaxed with an ML potential, then checked with DFT.

Reinforcement learning works by trial and reward





Our actor–critic agent (2023)

goal: $\max_\theta \mathbb{E}_{\tau\sim\pi_\theta}[R(\tau)]$

REINFORCE: $\nabla_\theta J = \mathbb{E}\left[R(\tau)\sum_t \nabla_\theta \log \pi_\theta(a_t|s_t)\right]$

REINVENT: $\log P_{\mathrm{aug}} = \log P_{\mathrm{prior}}(S) + \sigma\,\mathrm{score}(S)$

No training data needed, only an energy or property function. A trained agent can be **reused**.

Figure right: Elsborg & Bhowmik, J. Chem. Inf. Model. 63, 3731 (2023), Fig. 2.

Three ways to use RL

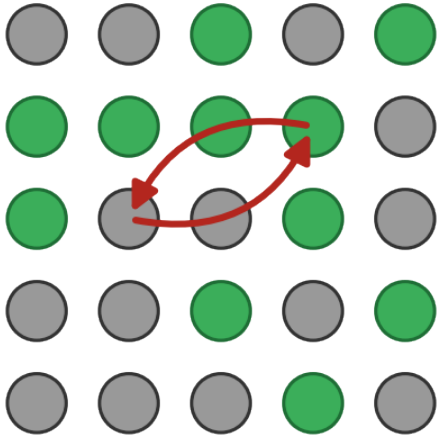
CCO

↓

CC(=O)O

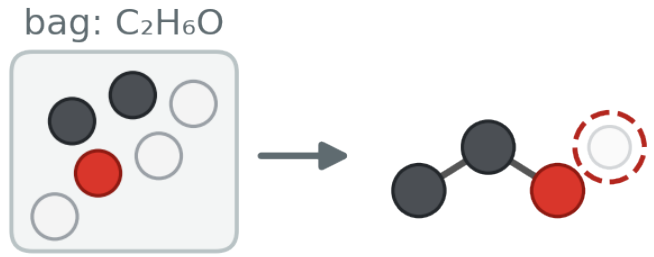
edit a SMILES string

e.g. REINVENT, MolDQN



swap atoms in a structure

e.g. ArtiSAN, alloy nanoparticles (ours)

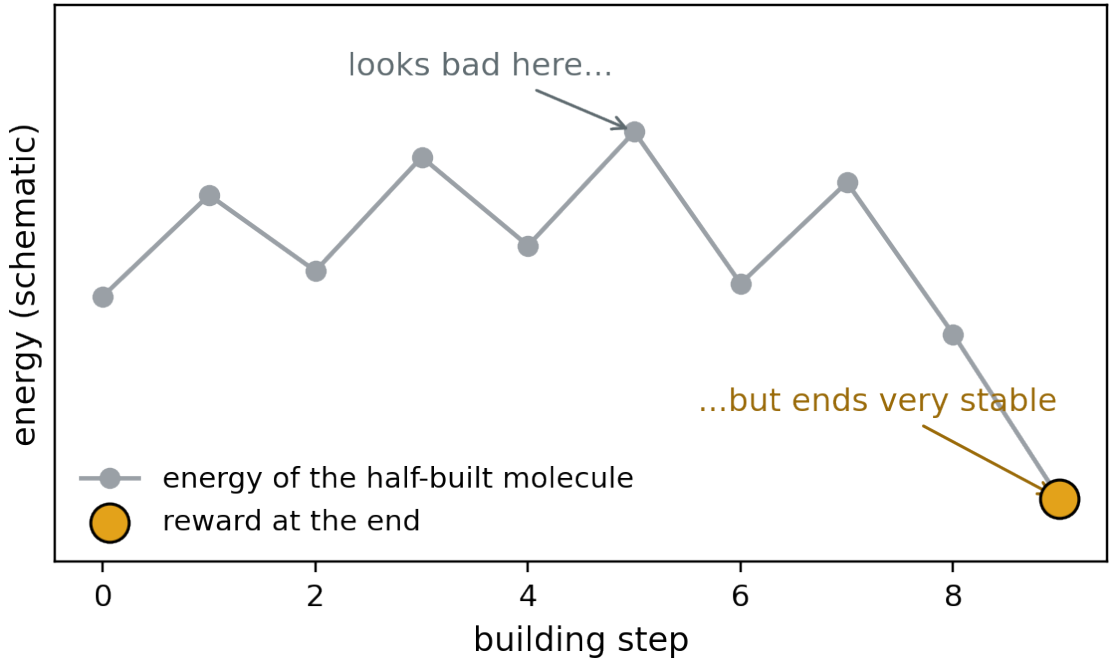


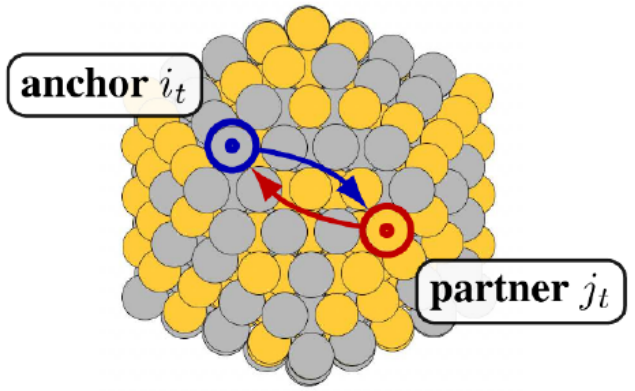
build from a formula, atom by atom

e.g. MolGym, AtomComposer (ours)

Building from a formula is like crystal structure prediction, but one formula allows a huge number of molecules.

What the agent is rewarded for, and when





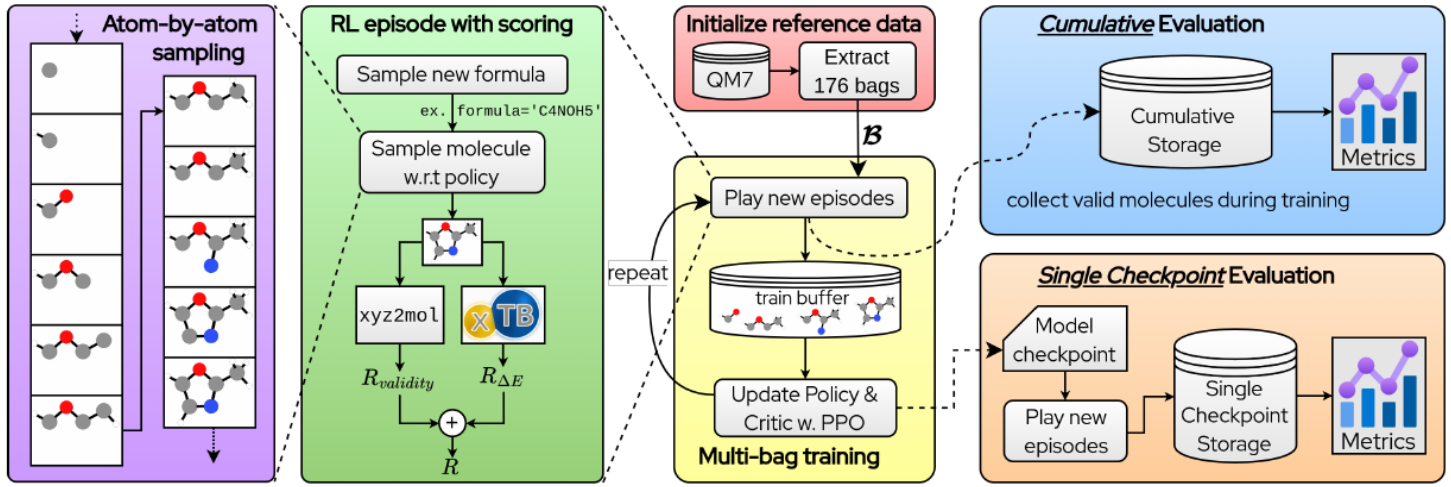
swap (i_t, j_t) → **relax** → **reward** $r_t = E(s_t) - E(s_{t+1})$

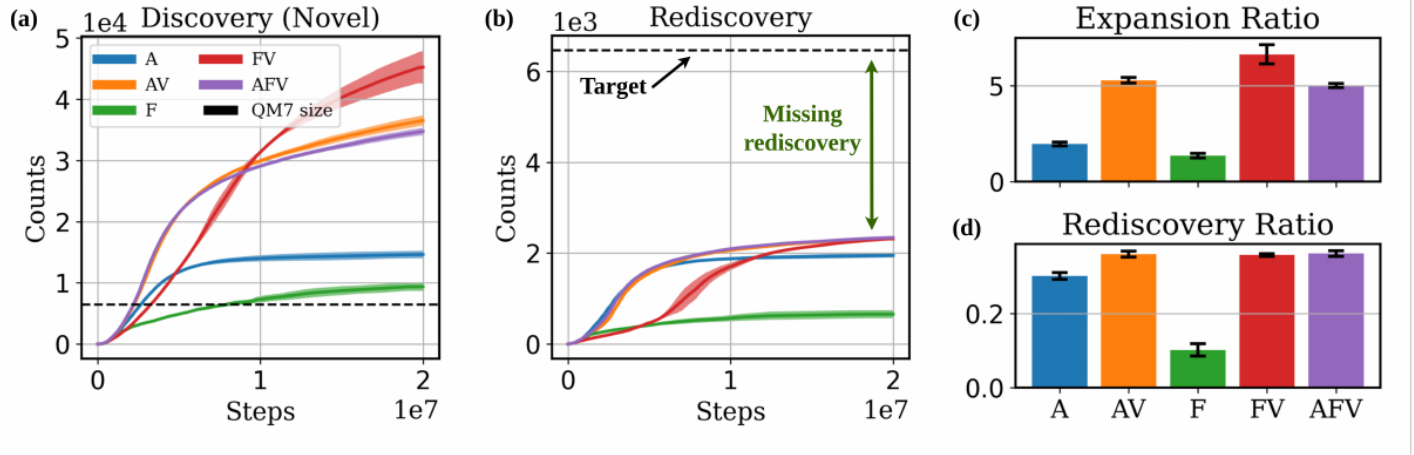
Swaps keep the structure complete, so a reward each step works

Building from scratch: reward **at the end**. Swapping atoms: reward **each step**. Train once, reuse many times.

Right: Elsborg, Hovmand & Bhowmik, ACS Mater. Au 6, 696 (2026), Fig. 1 (CC BY-NC-ND 4.0). Left: schematic.

AtomComposer builds molecules from a formula alone

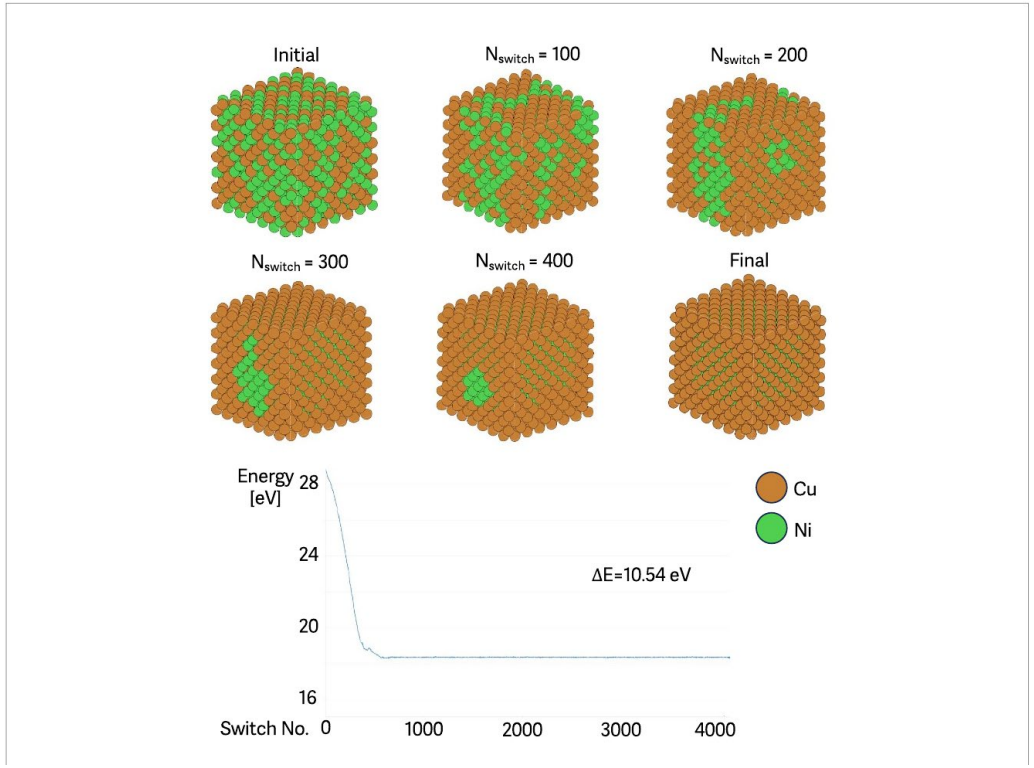




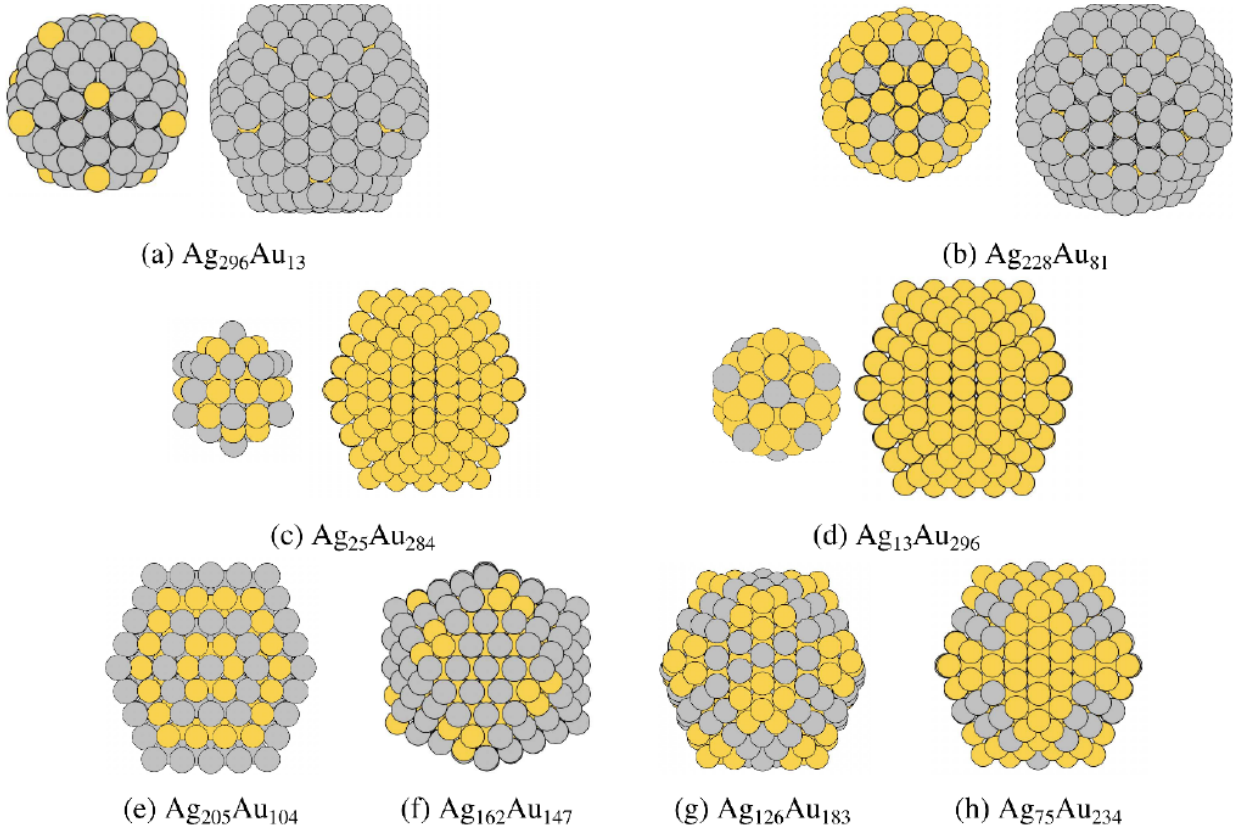
- No training data: rewards from xTB and a validity check.
- Trained on many formulas at once.
- Up to 10× more valid molecules on new formulas than earlier RL.

Hastrup, Cornet, Vegge & Bhowmik, arXiv:2605.28287 (2026), Figs 1 and 6.

RL for where atoms sit in alloys and nanoparticles



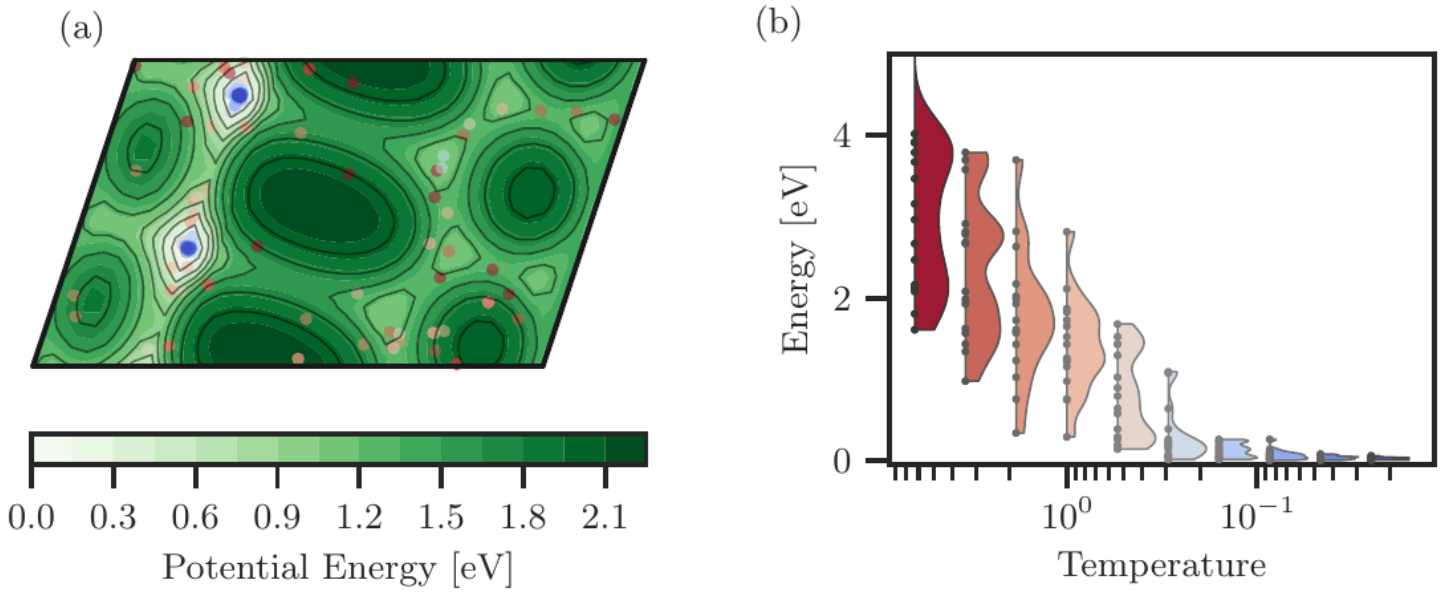
ArtiSAN reorders a 1,372-atom NiCu cell



One agent finds the known best orderings of 309-atom Ag–Au particles

Elsborg & Bhowmik, MLST 5, 035043 (2024), Fig. 4; Elsborg, Hovmand & Bhowmik, ACS Mater. Au 6, 696 (2026), Fig. 3 (CC BY-NC-ND 4.0).

GO-Diff uses diffusion to find low-energy structures

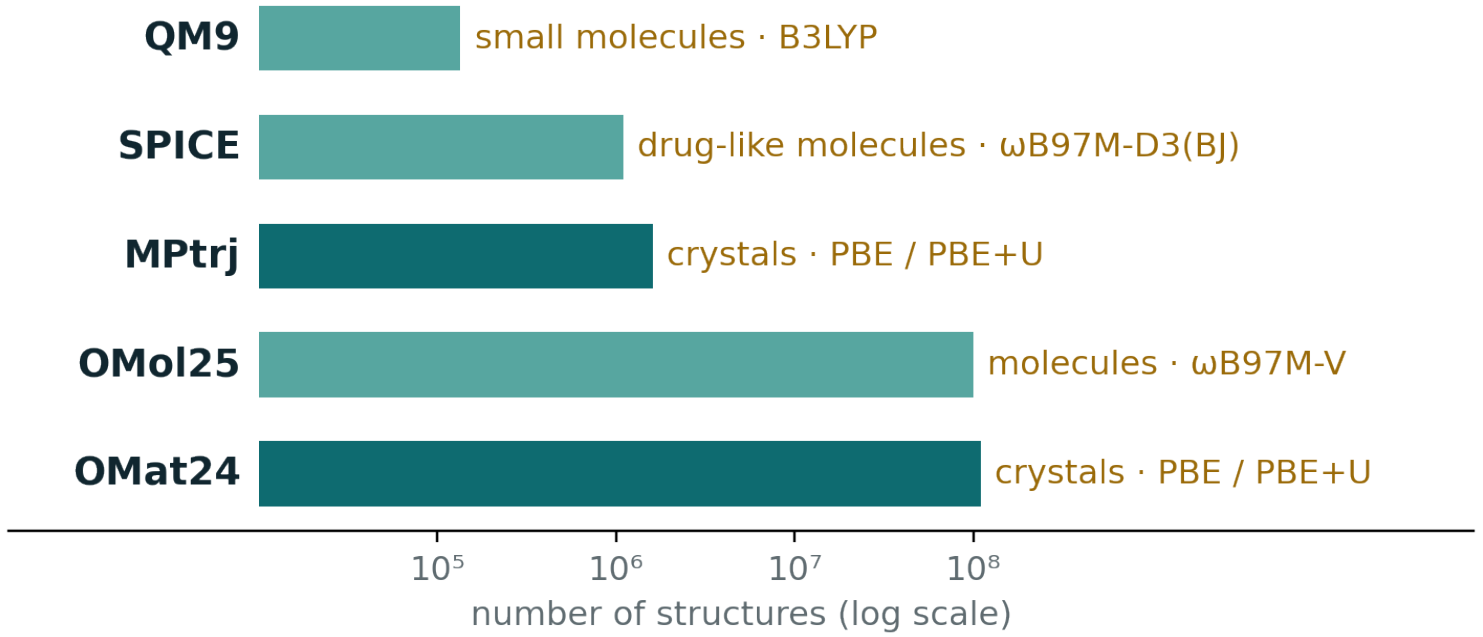


- Trained from scratch, with no data, only the energy.
- Sample, score, refine, cool down, repeat.
- Fewer energy calls, and reusable on related systems.

Related idea: Boltzmann generators sample with the right thermodynamic weights (Noé et al., 2019).

Rønne, Vegge & Bhowmik, arXiv:2510.13448 (2025), Fig. 1: an adatom on a stepped Pt surface.

Foundation models need big, consistent data

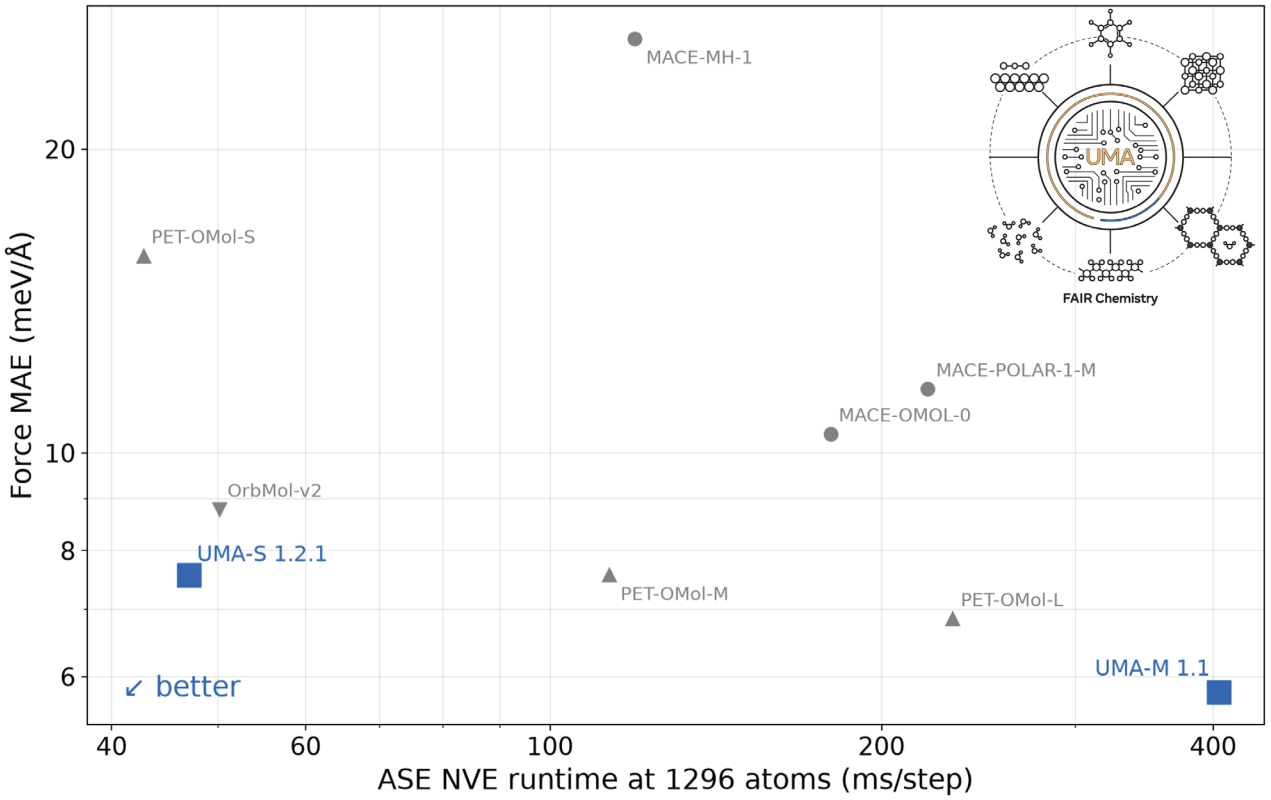


- One level of theory per dataset.
- Mixing levels teaches the model the differences.
- For one chemistry, a focused dataset still wins.

Our examples: a Na-cathode dataset (~113,000 DFT structures) and a halide solid-electrolyte dataset.

Petersen et al., Sci. Data 12, 1440 (2025); Parida et al., MLST (2026), doi:10.1088/2632-2153/aea5d8. Sizes approximate.

Universal ML potentials trade accuracy against speed



Force error vs speed on molecules (OMol25); lower left is better

Crystals

MACE-MP-0 · CHGNet · M3GNet · MatterSim · Orb · SevenNet · eqV2 · UMA

Molecules

MACE-OFF · AIMNet2 · SO3LR · UMA

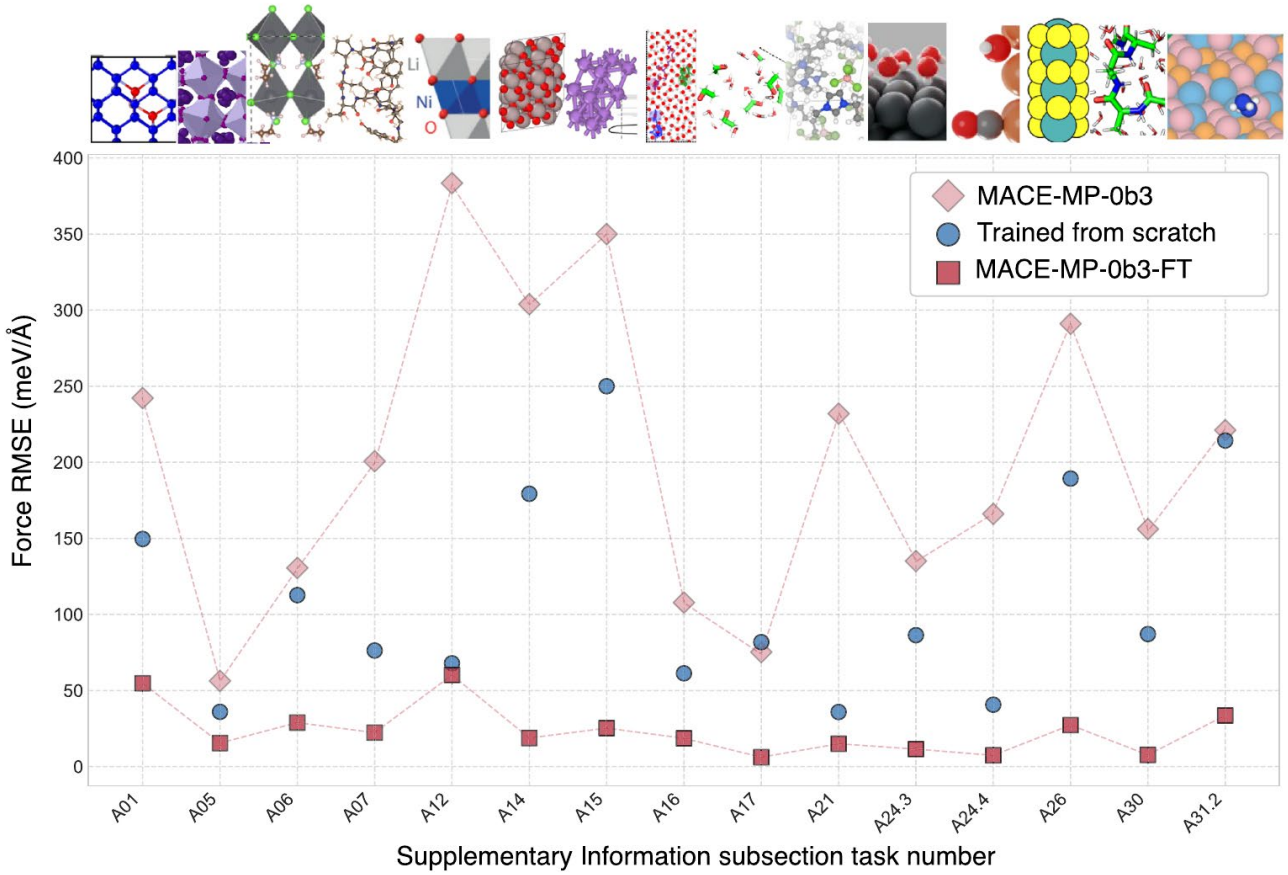
Benchmark

Matbench Discovery: is a new crystal stable?

Leaderboards change every few months: **test on your own problem.**

Figure from the fairchem repository (MIT): Meta FAIR Chemistry, UMA (Wood et al., arXiv:2506.23971, 2025).

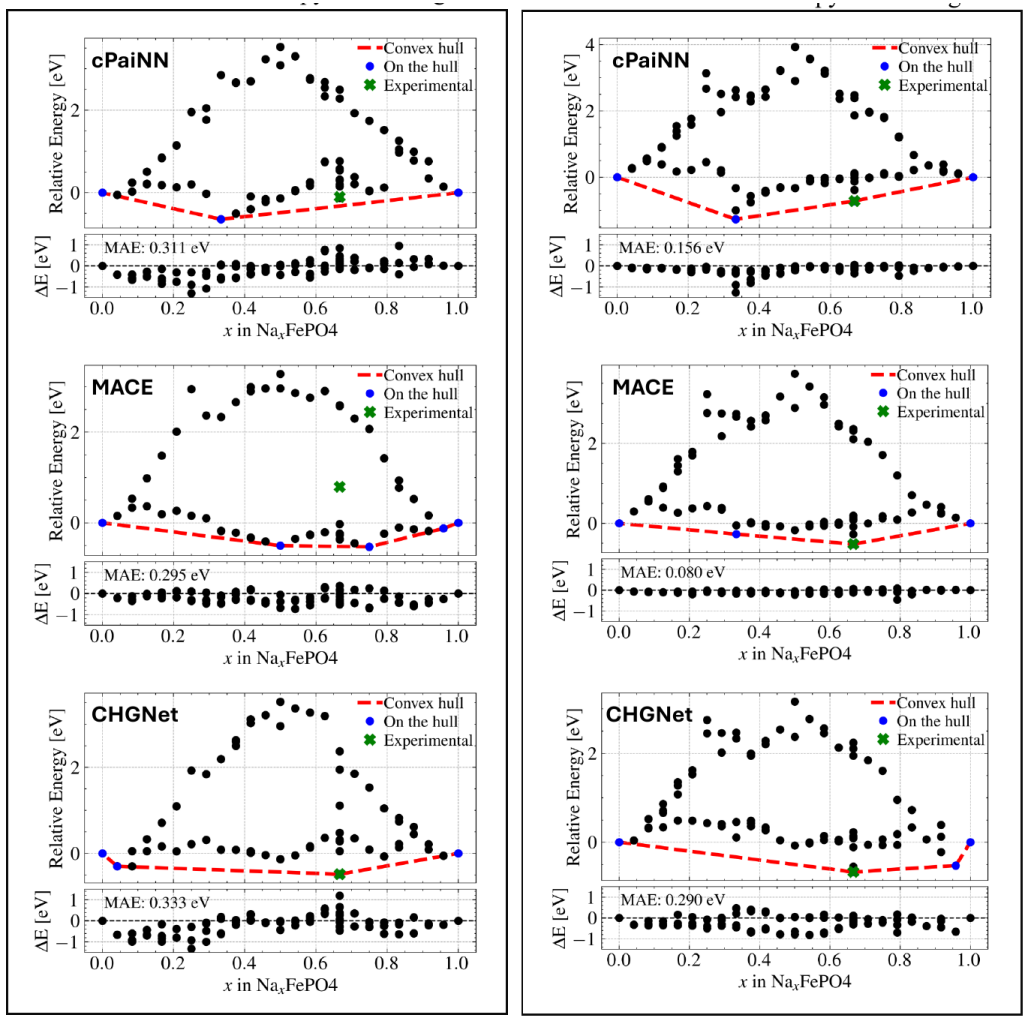
Start from a foundation model, then fine-tune



- Out of the box: good, not great (pink).
- From scratch: needs lots of data (blue).
- Fine-tuned on a little data: best (red).
- Multi-head and multi-task models mix levels of theory safely.

Batatia et al. (incl. Bhowmik), J. Chem. Phys. 163, 184110 (2025), Fig. 4.

Where foundation models still fail



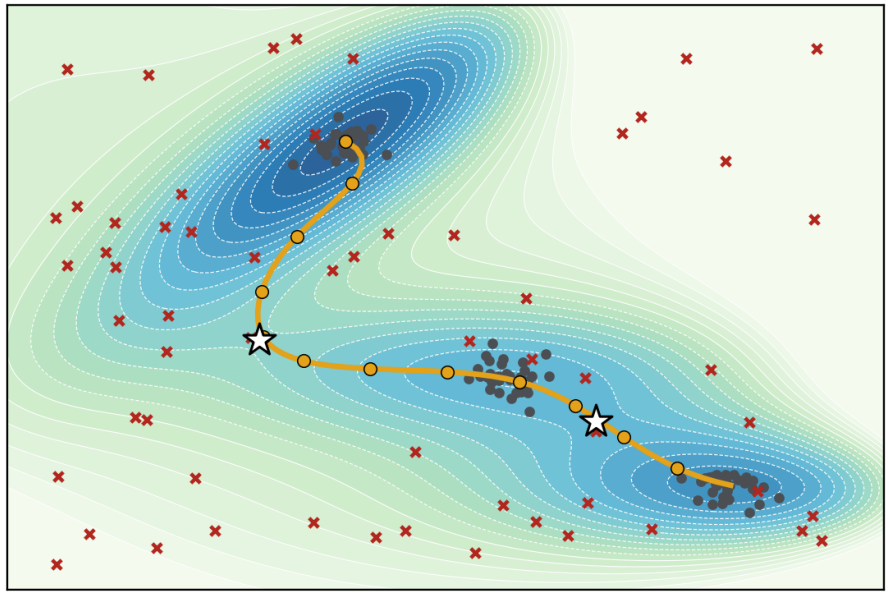
- In $NaFePO_4$, removing Na turns some Fe^{2+} into Fe^{3+}.
- ML potentials put the Fe^{3+} in the wrong places.
- So the convex hull, and the voltage, come out wrong (left).
- Fix: tell the model each Fe's charge state (right).

Other weak spots: surfaces, defects, magnetism, long-range charge, unusual chemistry.

Petersen, Lysgaard, Bhowmik, Hippalgaonkar & García Lastra, arXiv:2603.26471 (2026), Fig. 6.

Reactions happen where the training data are thin

an energy landscape with three valleys

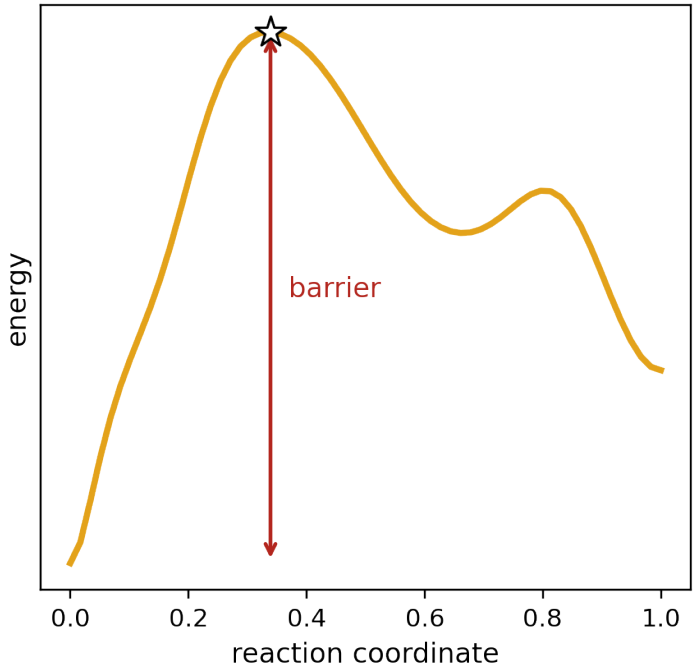


grey dots: equilibrium MD stays in the valleys

red crosses: random or hot sampling, mostly wasted

amber: follow the reaction path (NEB); stars mark the passes

a 1 eV barrier is visited ~10^{-17} times
as often as the valley at room temperature

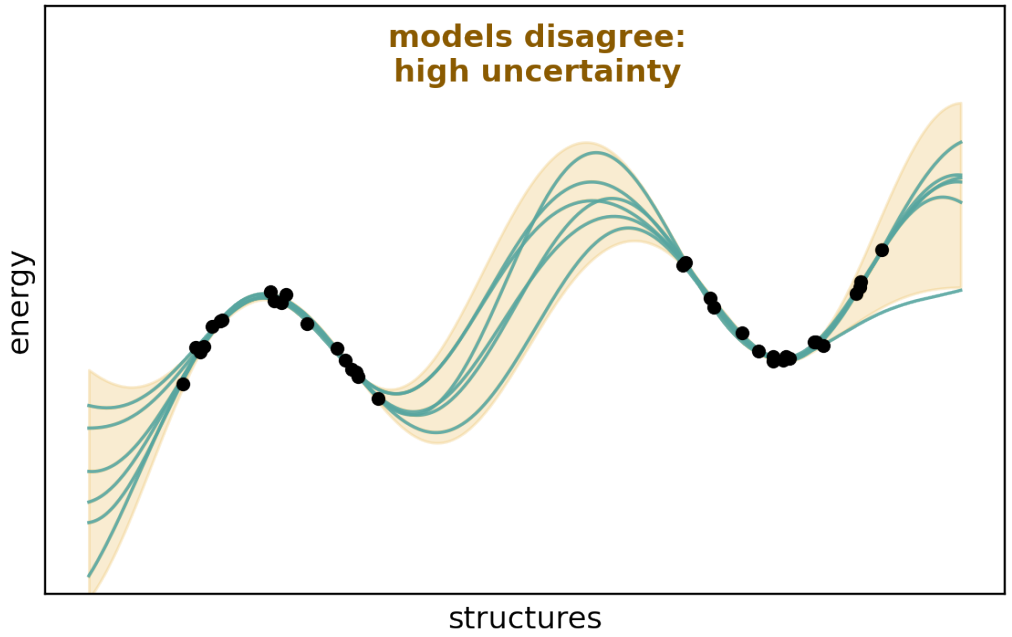


- A potential is only reliable where it has seen data, and reactions happen at the passes between valleys.
- Random or high-temperature sampling spends DFT on useless structures. NEB, the dimer method or metadynamics go straight to the pass.

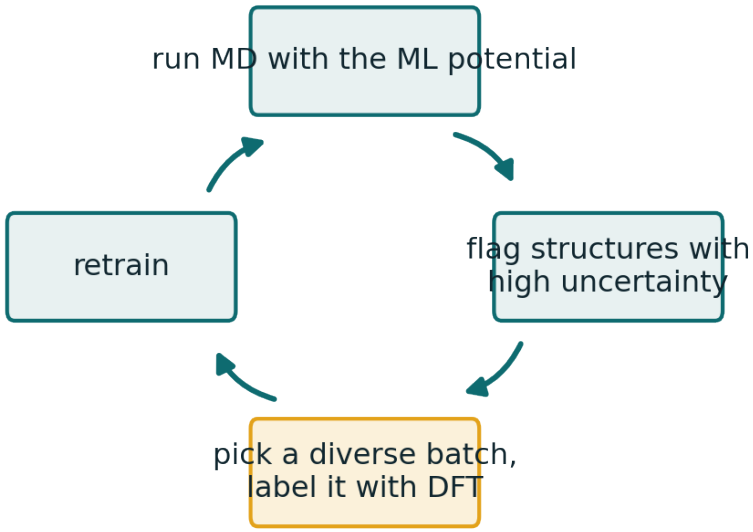
Henkelman, Uberuaga & Jónsson, J. Chem. Phys. 113, 9901 (2000); Laio & Parrinello, PNAS 99, 12562 (2002). Müller–Brown surface computed for this lecture.

Knowing where the model is guessing

an ensemble agrees near the data and fans out away from it



the active-learning loop



Ensembles: the standard; costs several models

Monte Carlo dropout: cheap; often poorly calibrated

Mean–variance output: captures label noise only

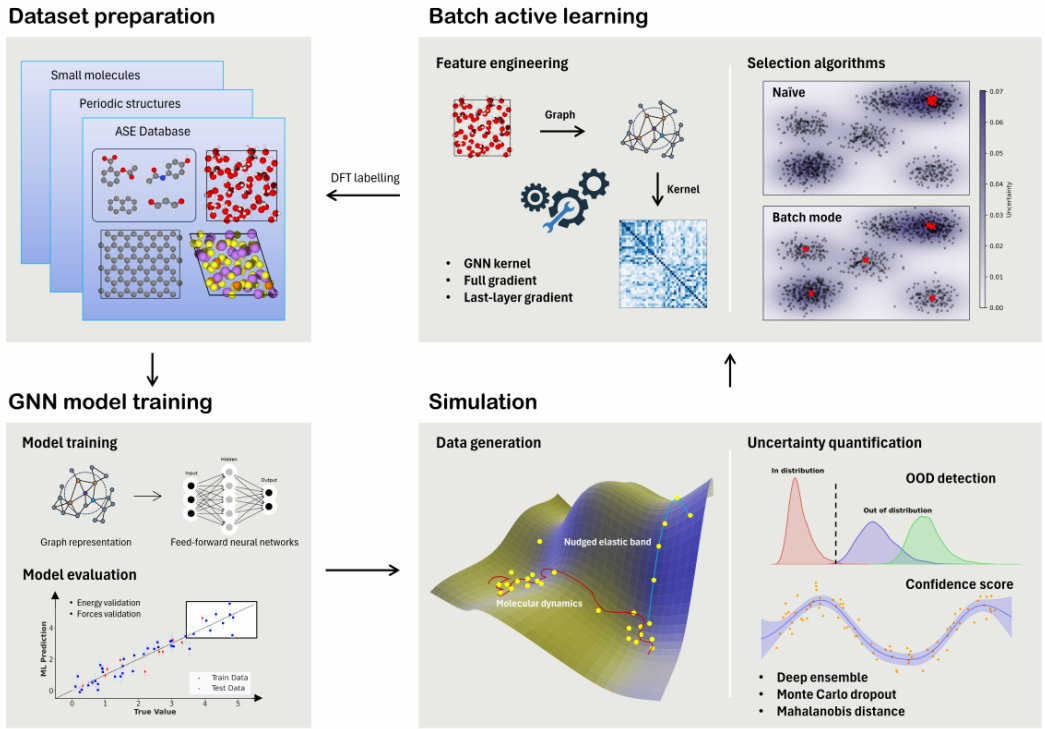
Last-layer Bayesian, GP head: cheap, added after training

Distance in feature or gradient space: good for choosing batches

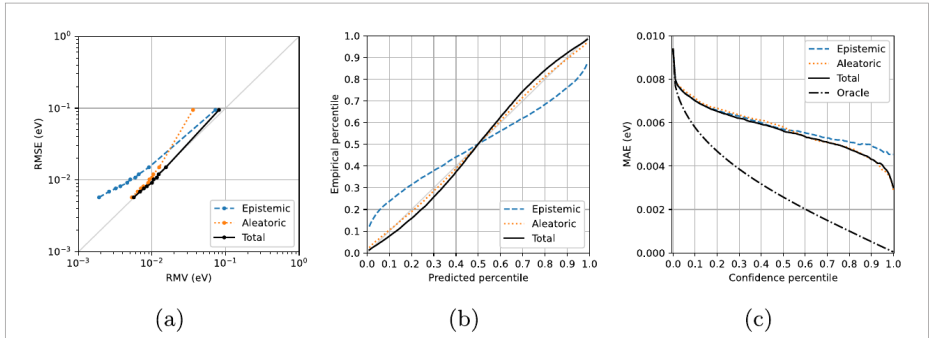
Conformal prediction: calibrated intervals

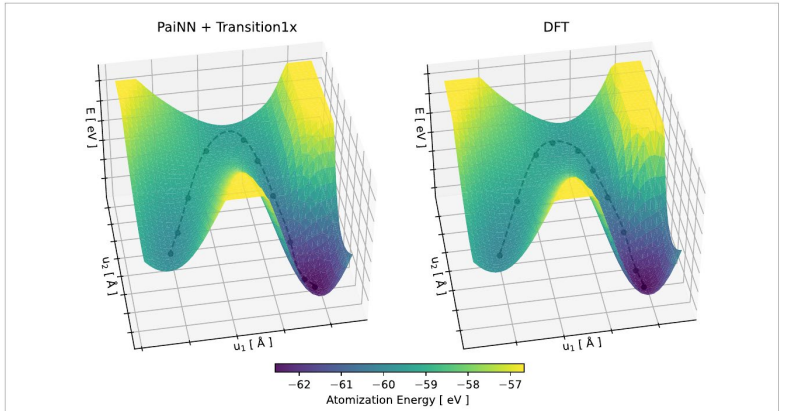
Lakshminarayanan et al., NeurIPS 2017; Gal & Ghahramani, ICML 2016; Daxberger et al., NeurIPS 2021; Holzmüller et al., JMLR 2023; Busk et al., MLST 3, 015012 (2022).

Active learning, error bars and reaction paths in practice



Active learning: add data where the model is unsure

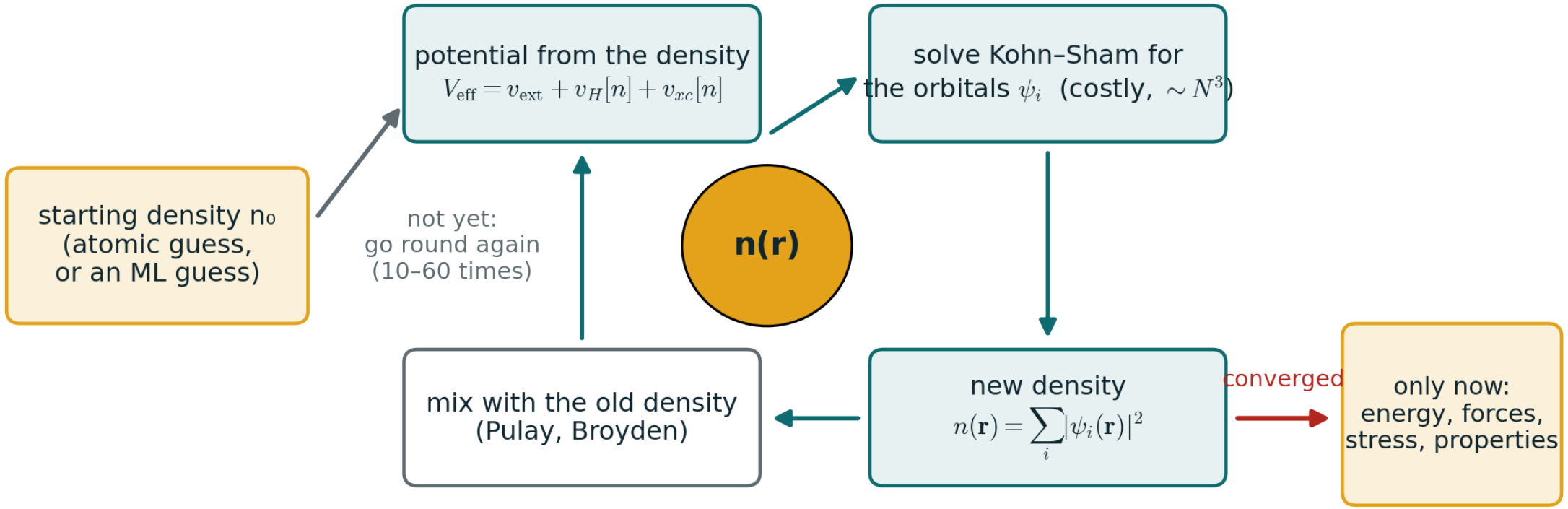




Error bars that match real errors (top); fast reaction paths (bottom)

Yang et al., CURATOR (ChemRxiv 2024), Fig. 1; Busk et al., MLST 3, 015012 (2022), Fig. 2; Schreiner et al., MLST 3, 045022 (2022), Fig. 1.

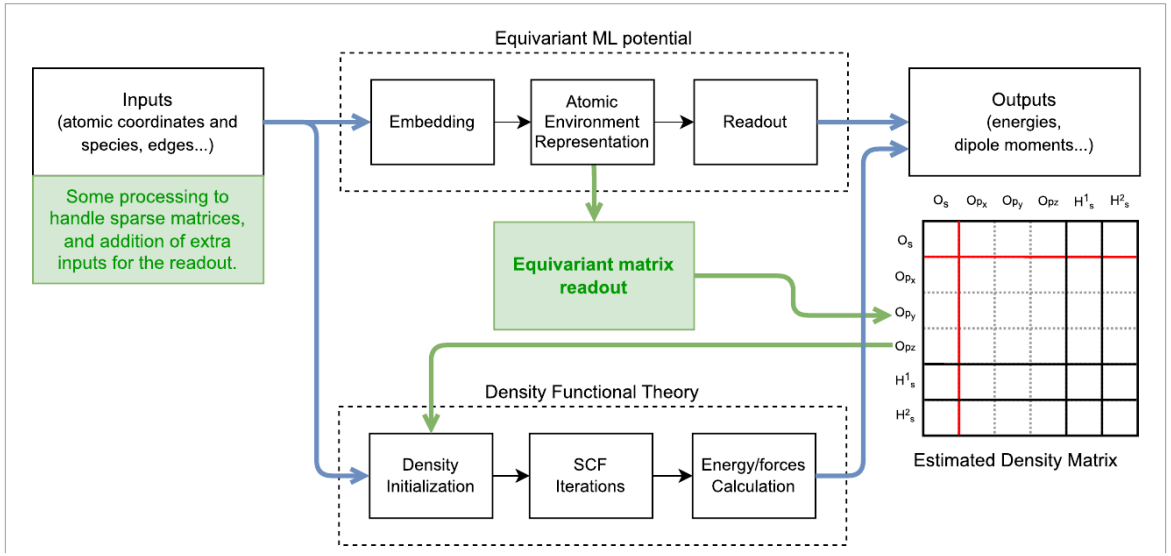
In DFT, the density goes round a loop

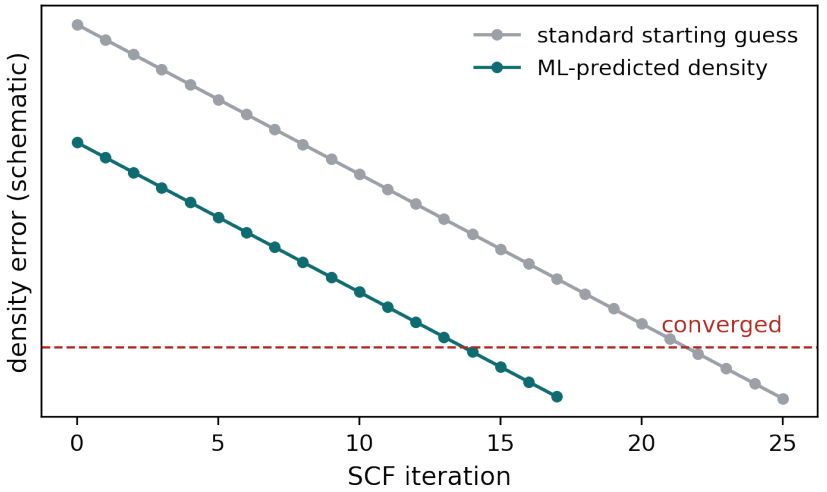


- Energy, forces and other properties come only after the loop has converged.
- Each round solves a large matrix problem (about N^3), so fewer rounds saves real time.
- A better starting density is where machine learning can help.

Hohenberg & Kohn, Phys. Rev. 136, B864 (1964); Kohn & Sham, Phys. Rev. 140, A1133 (1965); Pulay, Chem. Phys. Lett. 73, 393 (1980).

Using GNNs to speed up DFT from the inside



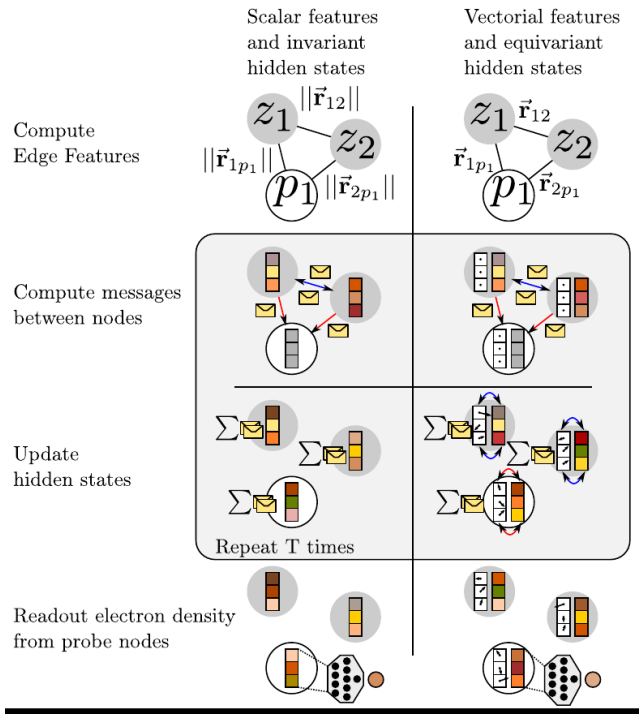


Schematic: a better start converges sooner

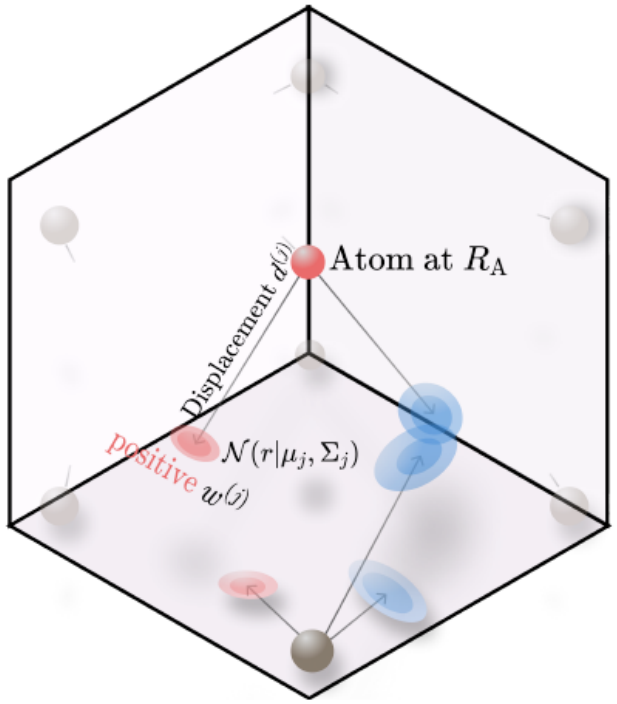
An ML potential **replaces** DFT. Here the GNN **helps** DFT start closer: about 40% fewer SCF steps (Graph2Mat) and up to about 20% less total DFT time (ELECTRA). The answer is still full DFT.

Left: Febrer et al., Mach. Learn.: Sci. Technol. 6, 025013 (2025), Fig. 1.

What the network predicts



DeepDFT: density at probe points

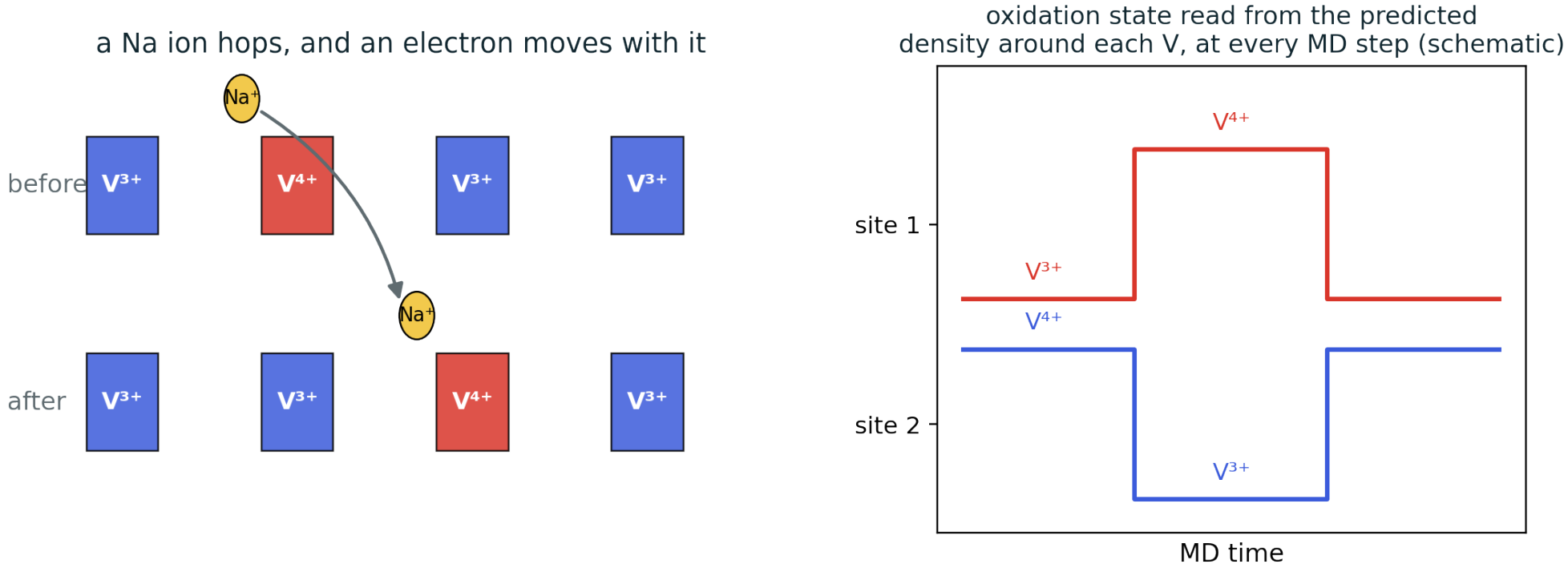


ELECTRA: Gaussians around atoms

- **On a grid:** probe points that only listen (DeepDFT).
- **As Gaussians:** one Fourier transform gives the periodic density (ELECTRA).
- **As a matrix:** blocks split into $l_1 \otimes l_2$ parts, like Lecture 1 (Graph2Mat).

Jørgensen & Bhowmik, npj Comput. Mater. 8, 183 (2022), Fig. 1; Elsborg et al., arXiv:2601.19966 (2026), Fig. 1.

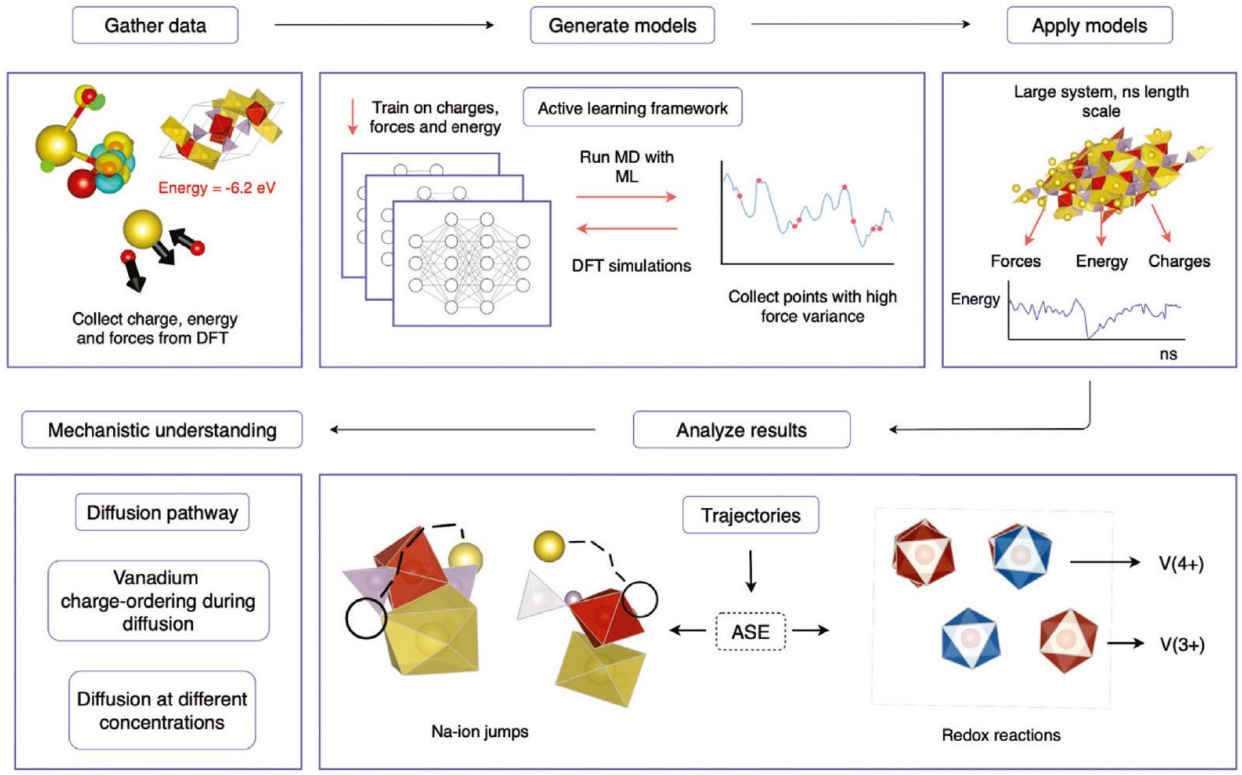
Why battery materials need more than energies



- When a Na^+ ion moves in or out, an electron is added to or taken from a nearby vanadium.
- Which vanadium changes affects the energy, the local structure and how the ions move.
- An energy-only ML potential cannot tell which atom was oxidised. Predicting the density, or magnetic moments as CHGNet does, can.

Schematic drawn for this lecture. CHGNet: Deng et al., Nat. Mach. Intell. 5, 1031 (2023).

Watching vanadium change charge in a Na-ion cathode

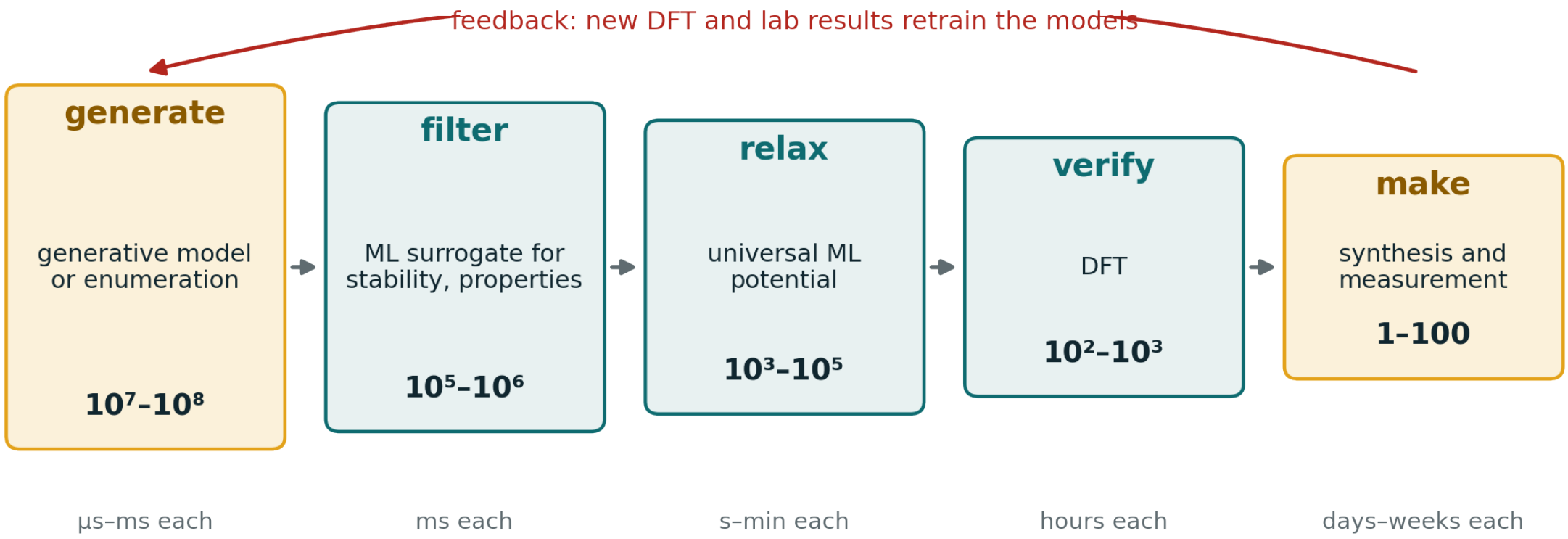


1. Train equivariant GNNs on DFT for **energies, forces and the spin-resolved density**.
2. Use their **uncertainty** to add DFT data where they are unsure.
3. Run MD of about **300 atoms of $Na_3V_2(PO_4)_3$** for **200 million steps**.
4. Integrate the predicted density around each V: **V^{3+} or V^{4+}** at every step. The charge moves as Na hops, and diffusion depends on the Na content.

The ML potential moves the atoms; the density model tells us where the electrons are.

de Blasio, Jørgensen, García Lastra & Bhowmik, Energy Storage Mater. 63, 103023 (2023), Fig. 1.

The recipe behind most discovery campaigns



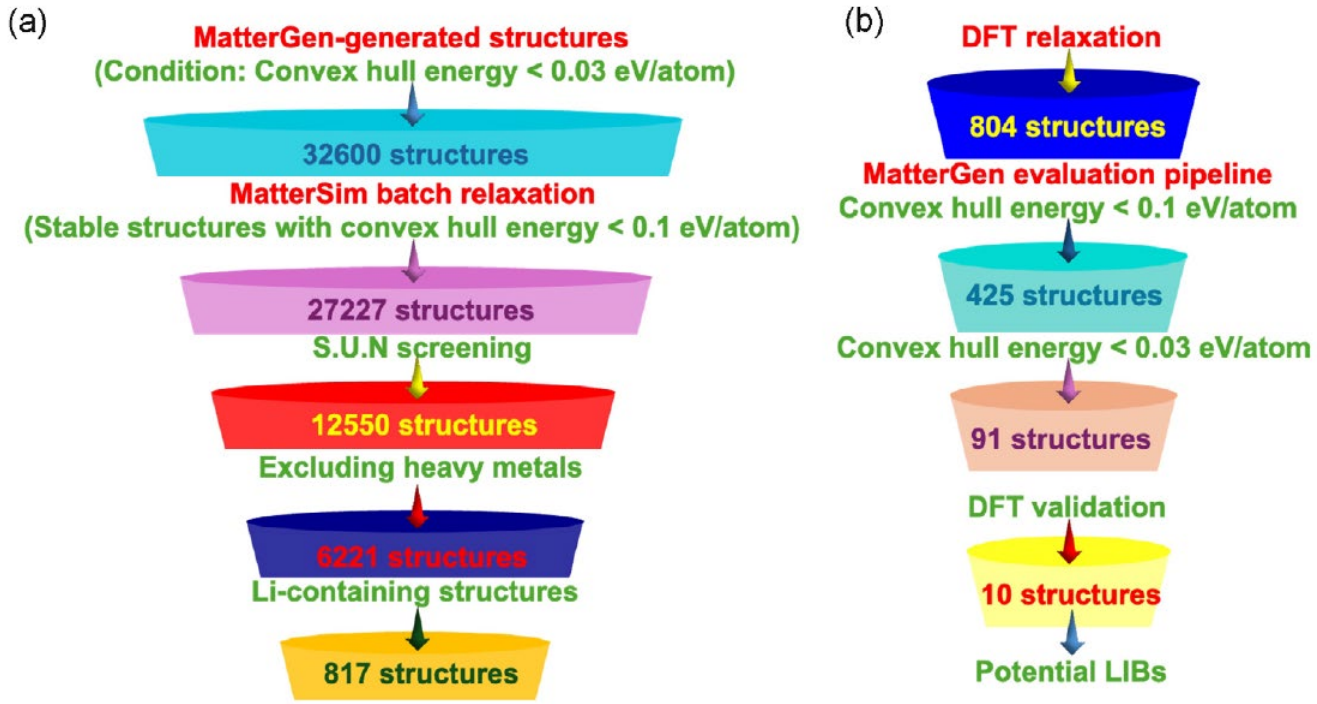
Each step costs 10 to 1,000 times more per candidate than the one before, so each step has to throw away most candidates.

Campaigns from the community

Campaign	What they did	Result	Checked by
GNoME, DeepMind (2023)	GNN filter + DFT in active-learning rounds	2.2 million predicted stable; 381,000 on the new hull	DFT; 736 already made by others
A-Lab, Berkeley (2023)	predicted targets, literature recipes, robot synthesis	41 of 58 targets made in 17 days	automated XRD
Microsoft + PNNL (2024)	~32 million candidates, ML filters and ML-potential MD, then DFT	a new Na–Li–Y chloride electrolyte	synthesis, ionic conductivity
MatterGen, Microsoft (2025)	generator aimed at a bulk modulus, ML potential, DFT	$TaCr_2O_6$: target 200 GPa, measured ~169 GPa	one compound made
Cu–Al catalyst (2020)	ML surrogate + active learning over alloy surfaces	Cu–Al for CO_2 to ethylene	electrochemical tests
Robotic chemist, Liverpool (2020)	Bayesian optimisation driving a mobile robot	688 experiments in 8 days	measured photocatalysis

Merchant et al. 2023; Szymanski et al. 2023; Chen et al. 2024 (preprint); Zeni et al. 2025; Zhong et al. 2020; Burger et al. 2020.

Our own funnel for new Li-battery materials



- **Generate** 32,600 crystals (MatterGen)
- **Relax** with a universal potential (MatterSim)
- **Filter:** S.U.N., no heavy metals, contains Li
- **Check** with DFT: 91 stable, then 10
- **Compute** voltage and capacity: **2 new cathodes**

Parida, Roy, García Lastra & Bhowmik, Batteries & Supercaps (2025), doi:10.1002/batt.202500309, Fig. 1.

What these campaigns teach us

1 Stable is not the same as new, made or useful

A DFT hull at 0 K ignores disorder, entropy, kinetics and precursors.

2 Check disorder and earlier reports

Several "new" ordered compounds turned out to be known disordered phases.

3 Keep humans in the loop

Automated XRD analysis needs expert checks.

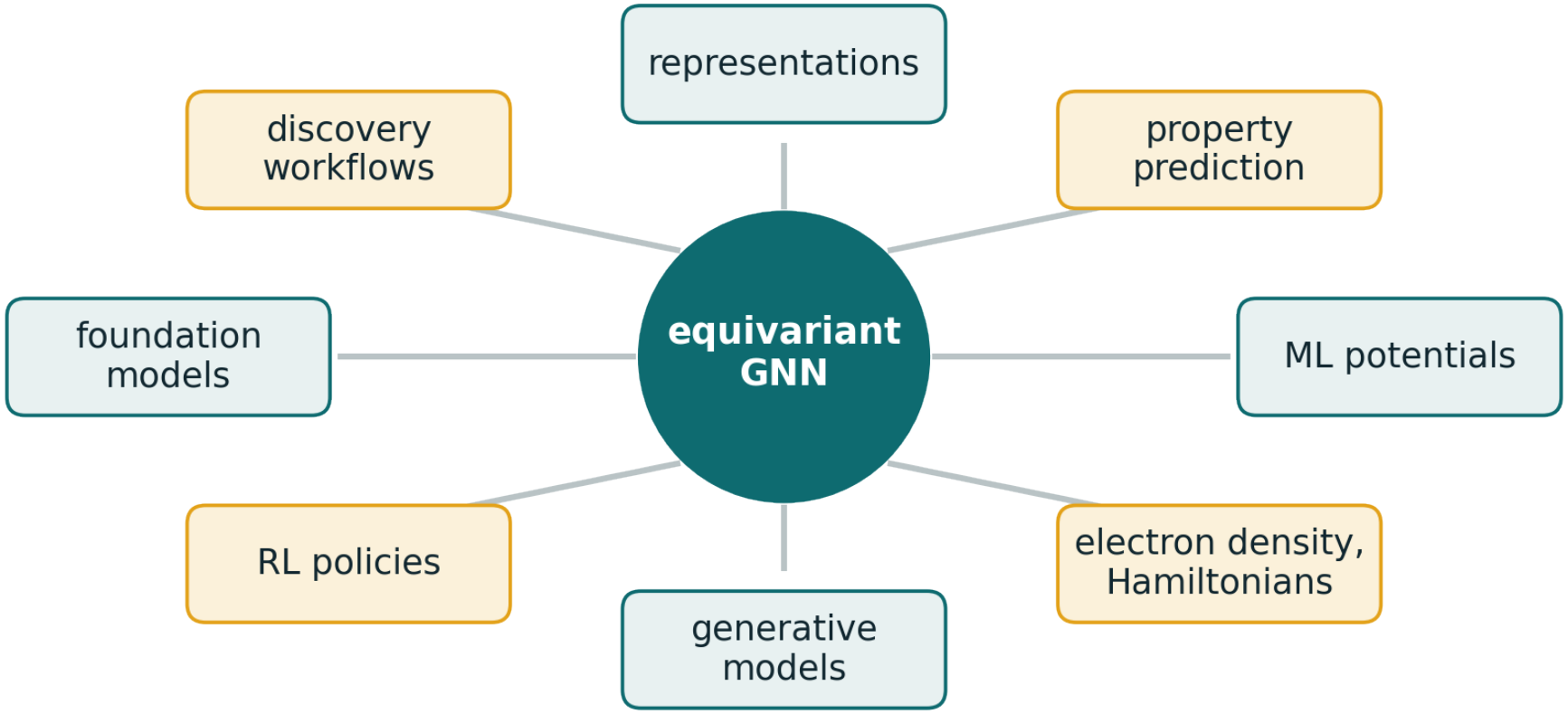
4 Report the whole funnel

Hit rates at every stage, failures included.

ML suggests and filters, physics checks, and experiments decide.

Cheetham & Seshadri, Chem. Mater. 36, 3490 (2024), on GNoME; Leeman et al., PRX Energy 3, 011002 (2024), on A-Lab.

Where GNNs show up in chemistry and materials



Same building blocks every time. What changes is **what they output** and **where they sit**.

HANDS-ON

Notebook 2: foundation models and a small discovery pipeline

1 A foundation potential

Relax molecules and run short MD.

2 A generative model

Generate crystals with MatterGen.

3 Mini pipelines

Filter, relax, check stability, rank.

Pre-generated samples come with it, so everything runs on a CPU.

Papers from our group used today

Generative models

Cornet et al., Digital Discovery 3, 1793 (2024). OM-Diff.

Cornet et al., MLST 6, 025057 (2025). Small-data design.

Cornet, Bergamin et al., arXiv:2507.03602 (2025). KLDM.

Rønne et al., arXiv:2510.13448 (2025). GO-Diff.

Parida et al., Batteries & Supercaps (2025). Battery funnel.

Reinforcement learning

Elsborg & Bhowmik, J. Chem. Inf. Model. 63, 3731 (2023).

Elsborg & Bhowmik, MLST 5, 035043 (2024). ArtiSAN.

Elsborg et al., ACS Mater. Au 6, 696 (2026). Nanoparticles.

Hastrup et al., arXiv:2605.28287 (2026). AtomComposer.

ML potentials and data

Batatia et al., J. Chem. Phys. 163, 184110 (2025). MACE-MP.

Yang et al., CURATOR, ChemRxiv (2024).

Busk et al., MLST 3, 015012 (2022). **Schreiner et al.**, MLST 3, 045022 (2022).

Schreiner et al., Sci. Data 9, 779 (2022). Transition1x.

Petersen et al., Sci. Data 12, 1440 (2025); arXiv:2603.26471 (2026).

GNNs inside DFT

Jørgensen & Bhowmik, npj Comput. Mater. 8, 183 (2022).

Febrer et al., MLST 6, 025013 (2025). Graph2Mat.

Elsborg et al., arXiv:2601.19966 (2026). ELECTRA.

de Blasio et al., Energy Storage Mater. 63, 103023 (2023).

Further papers from the wider field

How the models work

Ho, Jain & Abbeel, NeurIPS (2020). DDPM.

Song et al., ICLR (2021). Score-based SDEs.

Ho & Salimans, arXiv:2207.12598. Classifier-free guidance.

Kingma & Welling, arXiv:1906.02691. VAE introduction.

Lipman et al., ICLR (2023); arXiv:2412.06264. Flow matching.

Molecules

Gómez-Bombarelli et al., ACS Cent. Sci. (2018). ChemVAE.

Jin et al., ICML (2018). JT-VAE.

Hoogeboom et al., ICML (2022). EDM.

Xu et al., ICML (2023). GeoLDM.

Simm et al., ICML (2020). MolGym. **Olivecrona et al.**, J. Cheminform. (2017).

Crystals

Xie et al., ICLR (2022). CDVAE.

Jiao et al., NeurIPS (2023); ICLR (2024). DiffCSP, DiffCSP++.

Miller et al., ICML (2024). FlowMM.

Antunes et al., Nat. Commun. 15, 10570 (2024). CrystaLLM.

Zeni et al., Nature 639, 624 (2025). MatterGen.

Foundation models and discovery

Batatia et al., J. Chem. Phys. 163, 184110 (2025). MACE-MP.

Wood et al., arXiv:2506.23971 (2025). UMA.

Riebesell et al., arXiv:2308.14920. Matbench Discovery.

Merchant et al., Nature (2023). GNoME.

Anstine & Isayev, J. Am. Chem. Soc. 145, 8736 (2023). Review.