# 3D Infomax improves GNNs for Molecular Property prediction

Hannes Stärk*, Dominique Beaini, Gabriele Corso, Christian Dallago, Prudencio Tossou, Pietro Liò, Stephan Günnemann

Technical University of Munich

Valence

UNIVERSITY OF CAMBRIDGE

## Molecular property predictions are better when additionally informed by 3D geometry

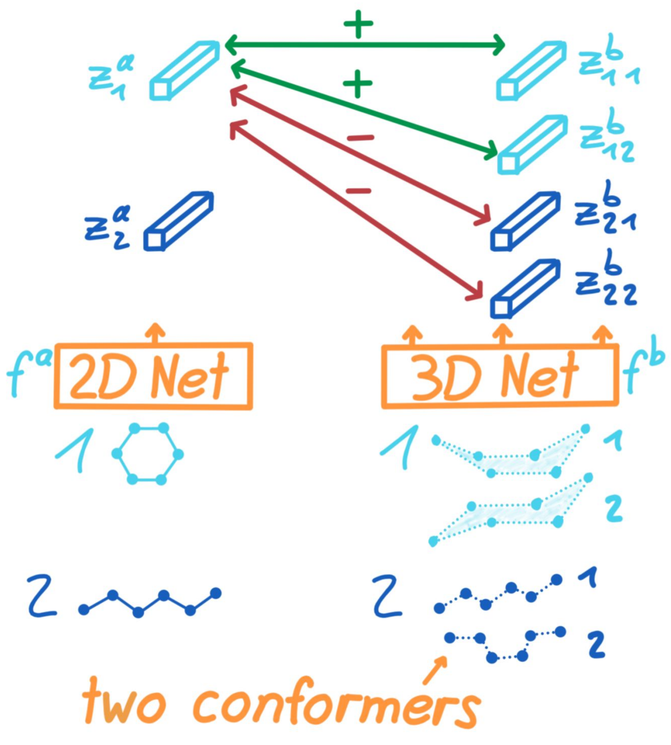
However, 3D information is often unavailable. We pre-train a GNN to generate implicit 3D information from 2D inputs which it can use to inform downstream molecular property predictions.

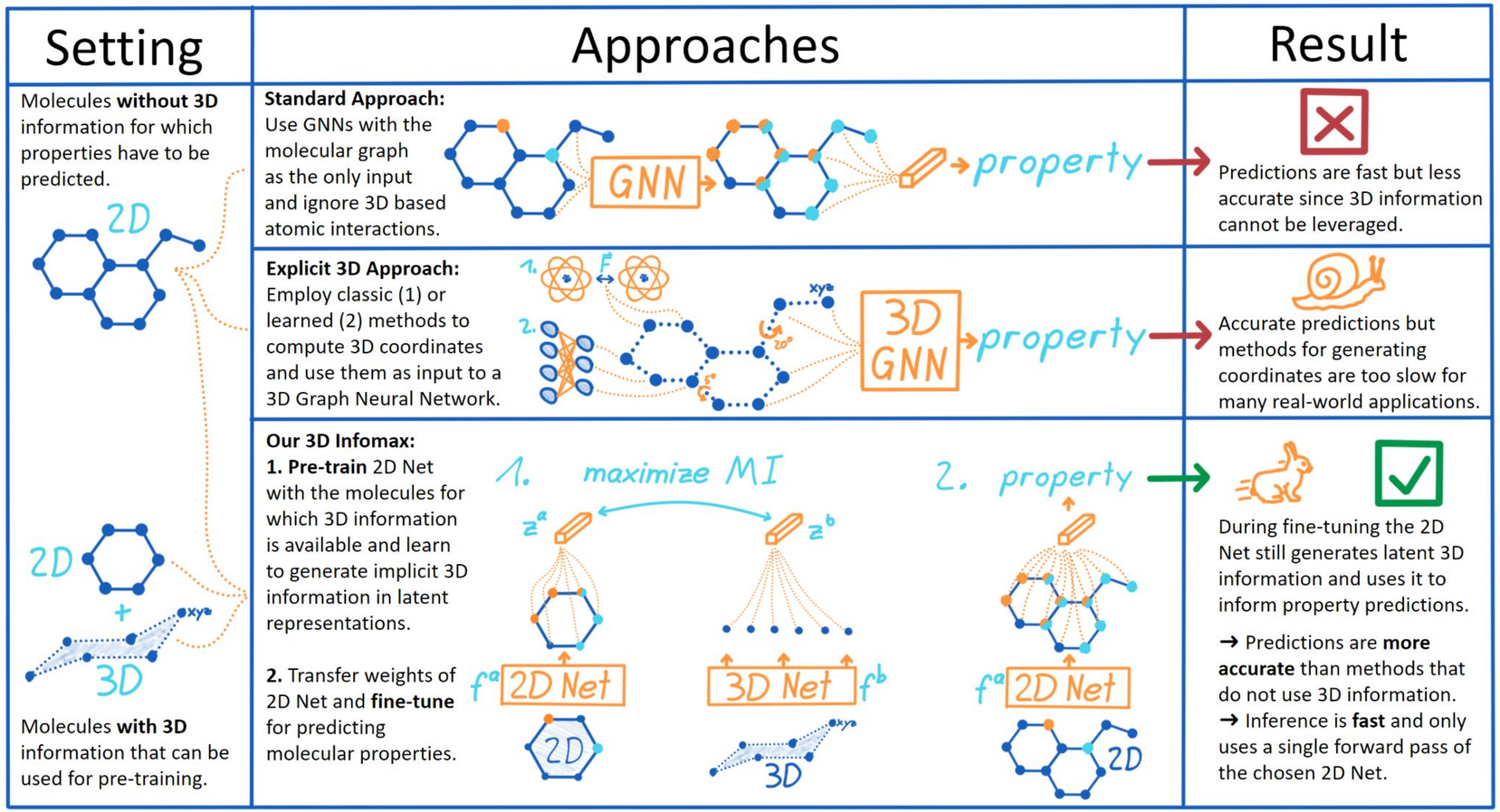
- More accurate than 2D methods but fast
- Highly transferable and robust representations
- Principled SSL: we know that 3D reasoning helps

## Multi-conformer contrastive objective:

$$\log \frac{\sum_{j=1}^{C} e^{sim(\boldsymbol{z}_i^a, \boldsymbol{z}_{i,j}^b)/\tau}}{\sum_{k=1, k\neq i}^{N} \sum_{j=1}^{C} e^{sim(\boldsymbol{z}_i^a, \boldsymbol{z}_{k,j}^b)/\tau}}$$

number of conformers, 2D, 3D




| Target | Rand Init | Pre-training baselines GraphCL | Pre-training baselines PropPred | Our 3D Infomax QM9 | Our 3D Infomax Drugs | Our 3D Infomax QMugs | RDKit SMP | *True 3D SMP* |
|---|---|---|---|---|---|---|---|---|
| $\mu$ | 0.4133±0.003 | 0.3937 | 0.3975 | **0.3507** | **0.3512** | 0.3668 | 0.4344 | 0.0726 |
| $\alpha$ | 0.3972±0.014 | 0.3295 | 0.3732 | 0.3268 | 0.2959 | **0.2807** | 0.3020 | 0.1542 |
| homo | 82.10±0.33 | 79.57 | 93.11 | **68.96** | 70.78 | 70.77 | 82.51 | 56.19 |
| lumo | 85.72±1.62 | 80.81 | 99.84 | **69.51** | 71.38 | 78.10 | 80.36 | 43.58 |
| gap | 123.08±3.98 | 120.08 | 131.99 | **101.71** | 102.59 | 103.85 | 114.24 | 85.10 |
| r2 | 22.14±0.21 | 21.84 | 29.21 | **17.39** | 18.96 | 18.00 | 22.63 | 1.51 |
| ZPVE | 15.08±2.83 | 12.39 | 11.17 | 7.966 | 9.677 | 12.06 | **5.18** | 2.69 |
| $c_v$ | 0.1670±0.004 | 0.1422 | 0.1795 | 0.1306 | 0.1409 | **0.1208** | 0.1419 | 0.0498 |