香港中文大學
The Chinese University of Hong Kong

UMA: A Family of Universal Models for Atoms

Yu Xie

25 August 2025

https://doi.org/10.48550/arXiv.2506.23971

Content

1. Background
2. Method
3. Results
4. Discussion

Background

- DFT's computational expense limits its usage – $O(n^3)$
- Machine Learning Interatomic Potentials (MLIPs) have the potential to accurately approximate DFT while being dramatically faster – $O(n)$
- Current MLIPs are mostly trained on smaller problem-specific datasets – limited by computational cost (unlike language and vision models - generalised across diverse data distributions and tasks)
- This paper:
  - A family of Universal Models for Atoms (UMA) – a single generalised model
  - Training datasets: 500 million atomic systems, 30 billion atoms
  - Covering materials, catalysis, molecules, molecular crystals and Metal Organic Frameworks (MOFs)
  - Empirical scaling laws relating compute, data and model size
  - Mixture of Linear Experts (MoLE) architecture to improve compute efficiency

Method – UMA Models

Summary of UMA models:

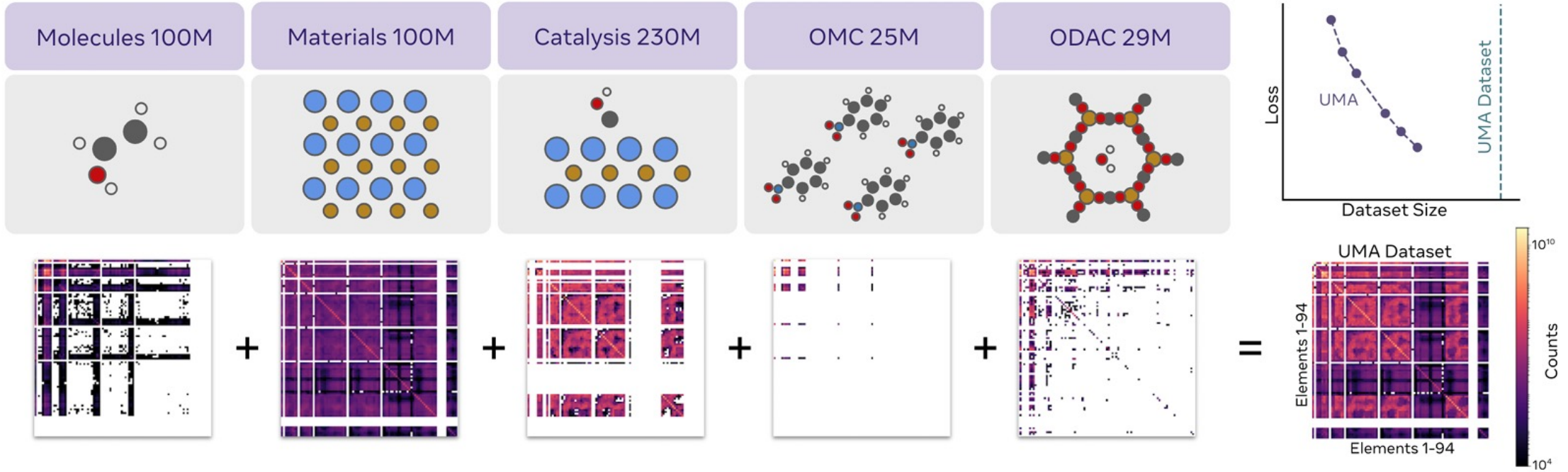
- Small (S): suitable for computationally intensive applications
- Medium (M): most general-purpose model, more accurate than UMA-S
- Large (L): highly accurate, helps understand scaling behaviour

Model	Total Parameters	Active Parameters	Inferences per second for 1k Atoms	Max Atoms per 80GB GPU	Conservative
UMA-S	150M	6M	16	100k+	✓
UMA-M	1.4B	50M	3	10k+	✓
UMA-L	700M	700M	1.6	1k+	✗

Inference speed and max atoms measured on Nvidia H100 with a periodic system that has ≈ 50 neighbors per atom within 6Å, see Appendix D

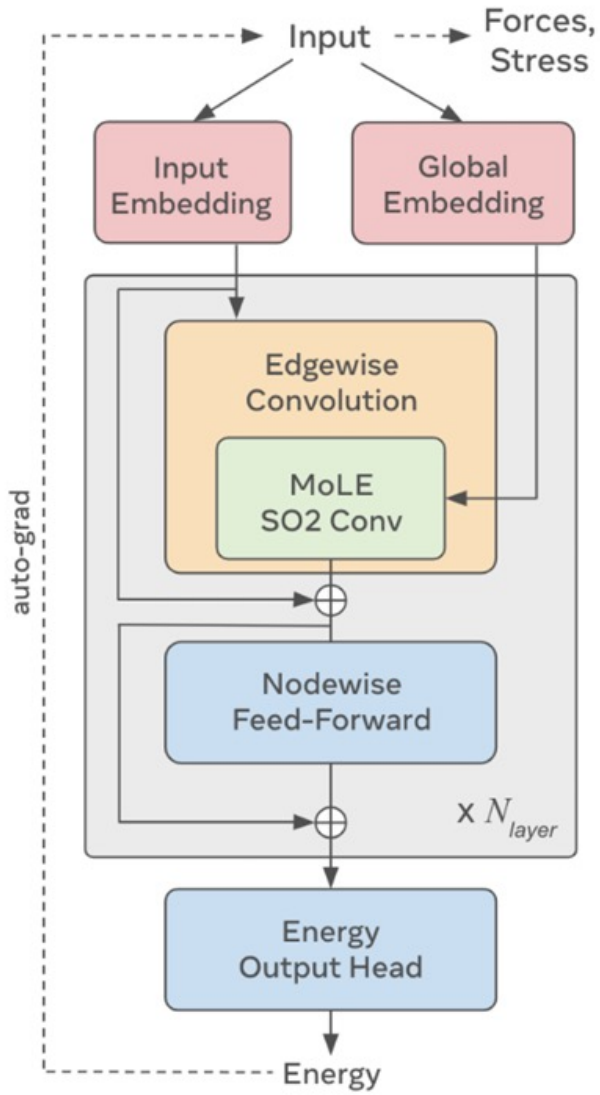
Method – Datasets

- 500 million atomic structures, 30 billion atoms
- Datasets differ in domain-specific DFT settings



Method – Architecture

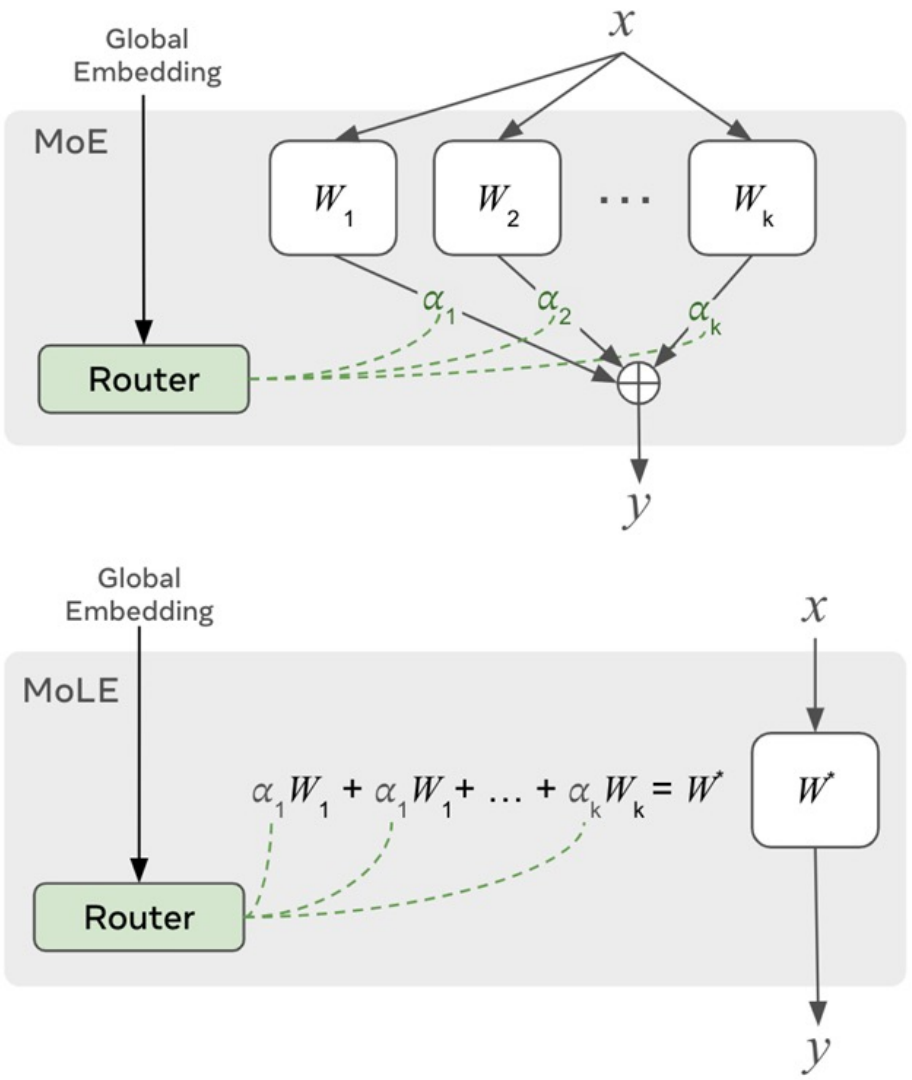
- Based on eSEN – an equivariant graph neural network using spherical-harmonic node embeddings
- Inputs: 3D atomic positions and atomic numbers (handled by eSEN), total charge and spin, DFT task
- Per layer:
  - Edgewise convolution → aggregation from neighbors (≤ 6 Å)
  - Nodewise feed-forward + residual + normalization
- Outputs: total energy, forces, stress



Method – Mixture of Linear Experts (MoLEs)

- Mixture of Experts (MoEs – work well for LLMs):
  - Outputs of a block calculated by a set of experts, each with their own individual set of weights
  - Gating function selects which experts to use – sparse activation saves compute
- Mixture of Linear Experts (MoLEs): $y = \sum_k \alpha_k \left(W_k x\right)$
  - Simple linear maps – efficiency
  - Maintains rotational equivariance when used with eSCN convolution – important for force calculations
  - Network weights may be precomputed before running simulations in some cases – much shorter inference times

$$y = W^* x \quad \text{where} \quad W^* = \left(\sum_k \alpha_k W_k\right)$$

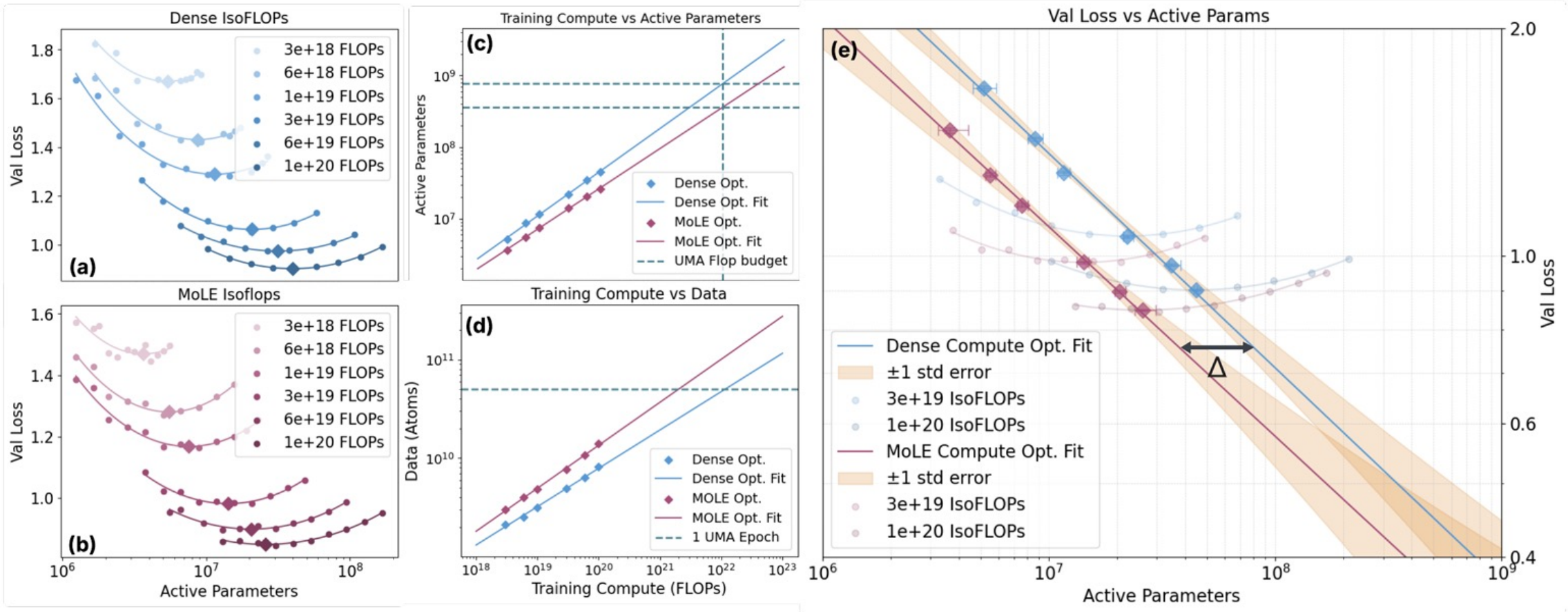


Method – Training Procedure

- Two-stage approach (adopted by UMA-S and UMA-M)
  - First stage: directly predict forces (faster training)
  - Second stage: remove the force head, fine-tune to predict conserving forces and stresses using auto-grad (provide energy conservation and smooth potential energy landscapes)
- Pre-training with BF16 numerical format + fine tuning with FP32 improves accuracy

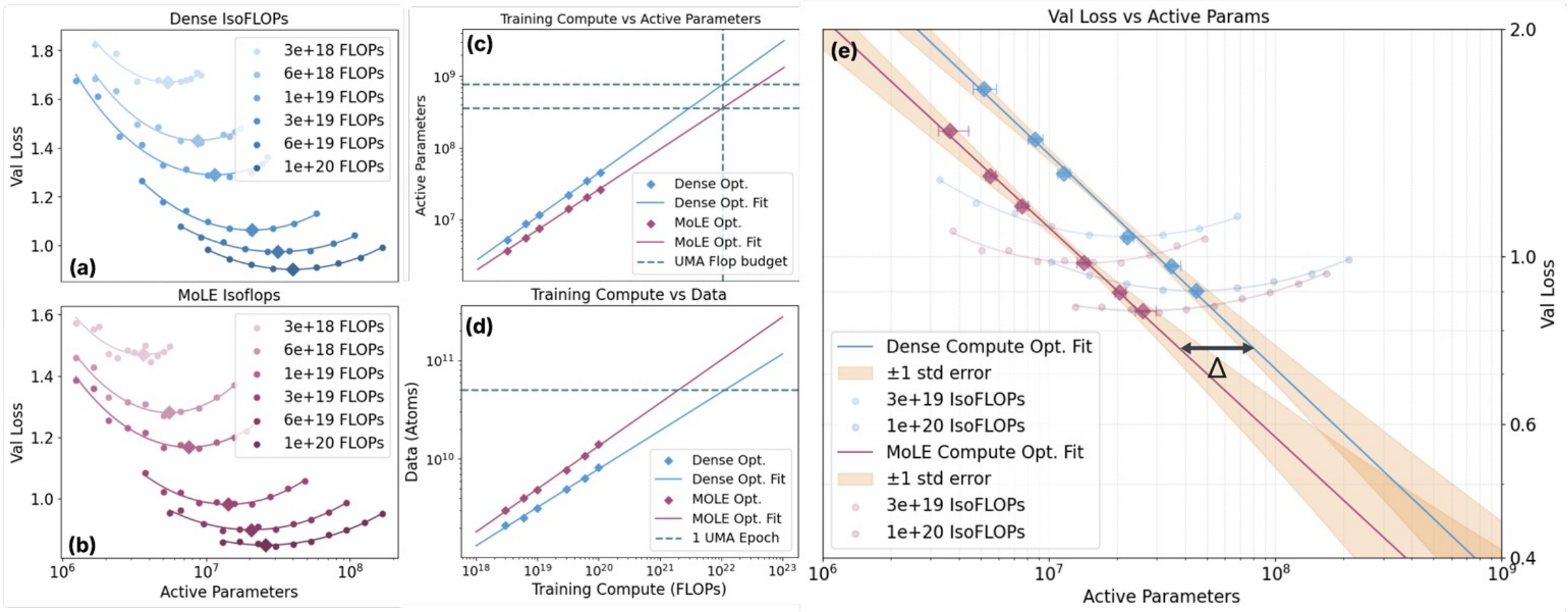
Results

- Iso-FLOPs
  - Varying the amount of training data for each model
  - Mimima represent optimal model at given level of training compute



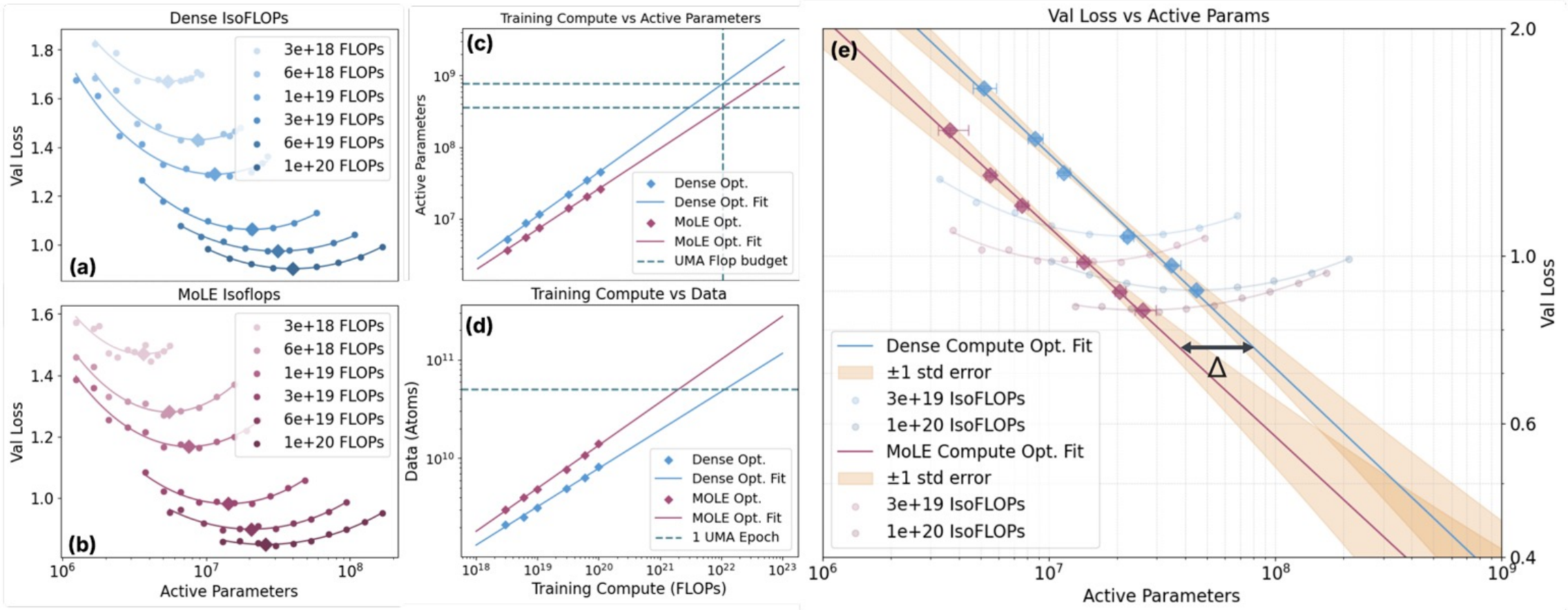
Results

- Log-linear scaling behaviour
- Optimal dense model size – ~700M parameters (UMA-L)



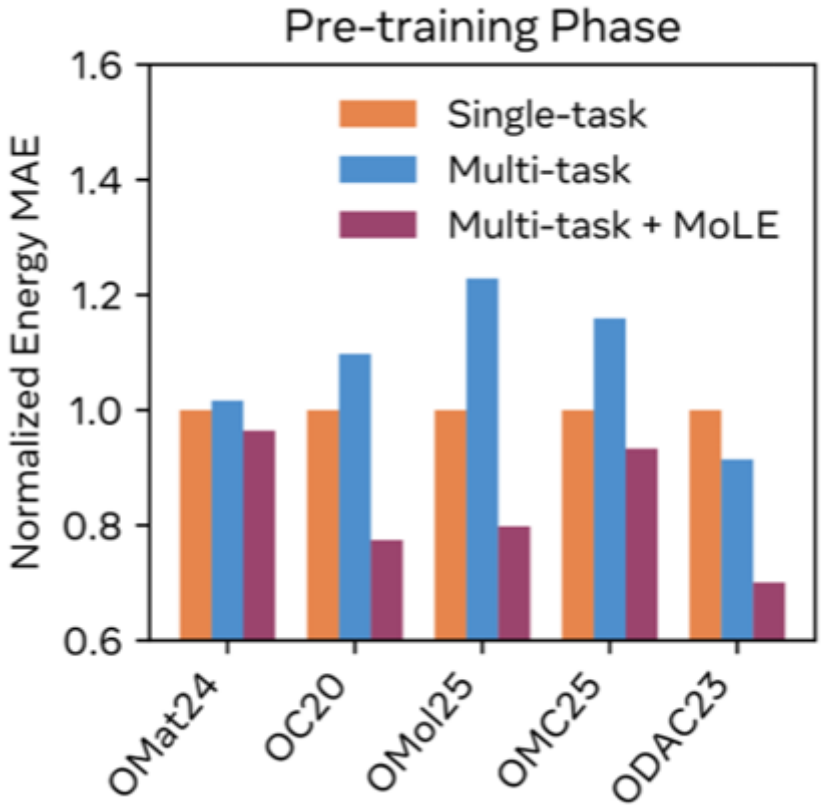
Results

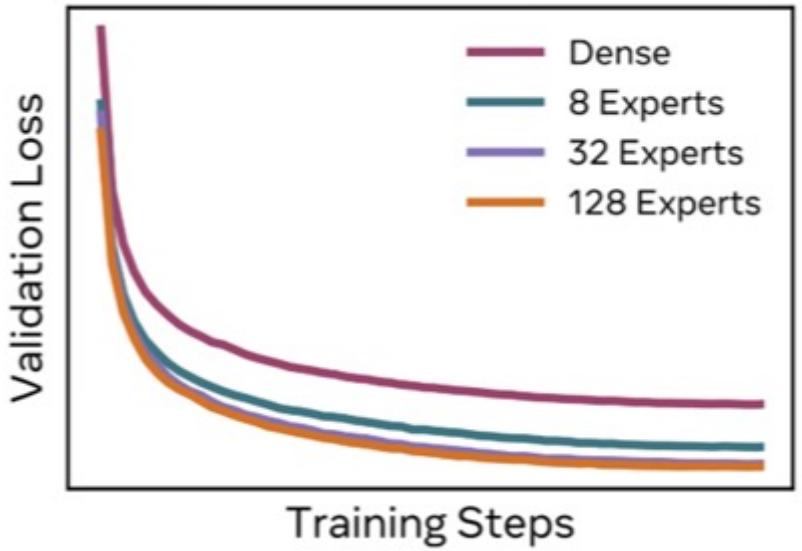
- Dense vs MoLE
  - Smaller MoLE model size is needed to achieve a fixed loss
  - Converge at larger model sizes



Results

- Multi-task vs single-task
- Effects of the number of experts





Results

- Inference efficiency

Atoms	UMA-S (6.6M)	UMA-M (50M)	eSEN-30M-OAM (30M)	Orb-v3 conservative-inf-omat (25M)	MACE-MPA-0 (9M)	MACE-OFF23-L (4.7M)
100	44	21	8	77	38	89
1,000	16	3	1.7	30	24	20
10,000	1.6	0.2	OOM	3.7	2.9	OOM
50,000	0.2	OOM	OOM	OOM	OOM	OOM
100,000	0.1	OOM	OOM	OOM	OOM	OOM

Results

- Test results on test-sets

Model	*Materials*						*Catalysis*				*Molecules*				*Molecular crystals*			*ODAC*	
	WBM *Energy/Atom*	*Forces*	*Stress*	HEA *Energy/Atom*	*Forces*	*Stress*	ID *Ads. Energy*	*Forces*	OOD-Both *Ads. Energy*	*Forces*	OOD-Comp *Energy/Atom*	*Forces*	PDB-TM *Energy/Atom*	*Forces*	OMC-Test *Energy/Atom*	*Forces*	*Stress*	OOD-L/T *Ads. Energy*	*Forces*
UMA																			
UMA-S	20.0	60.8	4.4	22.0	72.8	3.1	52.1	24.3	70.2	30.9	3.64	10.80	0.88	16.12	0.91	4.77	0.97	292.4	16.0
UMA-M	18.1	51.4	4.3	**19.0**	62.2	3.2	33.4	16.0	46.5	21.0	3.26	9.09	**0.69**	10.37	0.82	3.00	0.98	**290.2**	10.7
UMA-L	17.6	**45.5**	3.8	24.8	48.3	2.8	**32.4**	12.2	**43.5**	15.9	**2.33**	**5.19**	0.81	**8.76**	**0.59**	**2.28**	**0.10**	291.1	**6.5**
Literature																			
eSEN-OMat [23]	16.2	49.6	4.1	20.0	59.5	3.2	-	-	-	-	-	-	-	-	-	-	-	-	-
eqV2-OMat [5]	**14.9**	46.3	**3.6**	20.3	**47.0**	**2.7**	-	-	-	-	-	-	-	-	-	-	-	-	-
eqV2-OC20 [45]	-	-	-	-	-	-	149.1	**11.6**	306.5	**15.7**	-	-	-	-	-	-	-	-	-
GemNet-OC20 [24]	-	-	-	-	-	-	163.5	16.3	343.3	23.1	-	-	-	-	-	-	-	-	-
eqv2-ODAC [66]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	316.0	7.2
ST Baselines																			
eSEN-S-OMol	-	-	-	-	-	-	-	-	-	-	3.67	11.56	0.79	14.11	-	-	-	-	-
eSEN-S-OMC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.05	5.39	0.94	-	-
Target																			
Practical Utility	10-20	-	-	10-20	-	-	100	-	100	-	1-3	-	1-3	-	1-3	-	-	100	-

Results

- Test results on important benchmarks

Model	Materials									Catalysis	Molecules					Molecular Crystals		
	Matbench [58] *F1*	*RMSD*	*MAE [eV/atom]*	κ_{srme} [55]	Phonons [46] ω_{max} *[K]*	*Free Energy [kJ/mol]*	Elasticity [16, 35] G_{vrh} *[GPa]*	K_{vrh} *[GPa]*	NVE MD [23] *Conserve*	AdsorbML [43] *Success Rate*	OMol25 [44] *Ligand-strain [meV]*	*PDB-pocket [meV]*	*Dist-SR [meV]*	*Dist-LR [meV]*	NVE MD [23] *Conserve*	CSP Targets [30] *Lattice Energy [kJ/mol]*	*Kendall Rank*	*RMSD [Å]*
UMA																		
UMA-S	0.916	0.064	0.020	0.203	17.59	5.00	8.54	4.96	✓	68.35%	4.39	150.3	67.6	432.1	✓	2.695	0.82	**0.12**
UMA-M	**0.930**	**0.061**	**0.018**	0.195	**13.91**	**3.39**	**8.40**	**4.76**	✓	71.12%	**2.45**	89.7	41.6	588.7	✓	2.664	0.81	0.13
UMA-L	0.928	0.065	0.018	0.671	78.50	18.20	20.56	14.48	✗	**74.41%**	3.37	**71.7**	**16.6**	**246.1**	✗	**2.488**	**0.84**	**0.12**
Literature																		
eSEN-30M-OAM [23]	0.925	0.061	0.018	**0.170**	15.00	4.00	9.13	5.73	✓	-	-	-	-	-	-	-	-	-
ORB v3 [57]	0.905	0.075	0.024	0.210	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SevenNet-MF-ompa [37]	0.901	0.064	0.021	0.317	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GRACE-2L-OAM [8]	0.880	0.067	0.023	0.294	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MACE-MPA-0 [6]	0.852	0.073	0.028	0.412	-	-	-	-	-	-	-	-	-	-	-	-	-	-
eqv2-OC20 [43]	-	-	-	-	-	-	-	-	-	60.80%	-	-	-	-	-	-	-	-
GemNet-OC20 [43]	-	-	-	-	-	-	-	-	-	54.88%	-	-	-	-	-	-	-	-
ST Baselines																		
eSEN-S-OMol	-	-	-	-	-	-	-	-	-	-	5.15	154.48	73.02	608.9	✓	-	-	-
eSEN-S-OMC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6.18	0.74	0.18

Discussion

- Limitations
  - Long-range interactions – standard MLIP cutoff distance of 6 Å
  - Separate embedding for each discrete charge or spin – limited in generalizing to unseen spins and charges
- Large training dataset – 500M atomic structures
- Empirical scaling relations
- Mixture of Linear Experts – increasing model capacity while maintaining inference efficiency
- Strong performance in test-sets and benchmarks

Thank You