

A GAP potential for carbon from randomly selected, uncurated structures: comparison with GAP-20

C. D. Chalkidis¹, G. Fanourgakis¹

¹Laboratory of Quantum and Computational Chemistry, School of Chemistry, AUTH · chalchri@chem.auth.gr

Topic 5 · Computational Chemistry and Artificial Intelligence

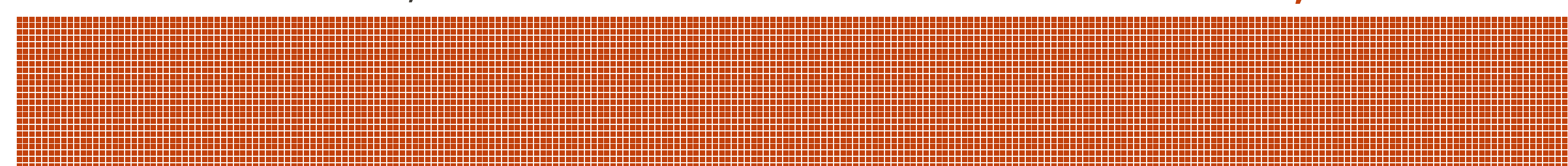
Poster code:

What accuracy does a machine-learning potential reach when its training structures are drawn at random?

With **1,000 random, uncurated structures**, a GAP potential reaches the accuracy of **GAP-20**, which was trained on **6,088 curated ones**.

GAP-20 · Rowe et al., 2020

6,088 curated



Large model · this work

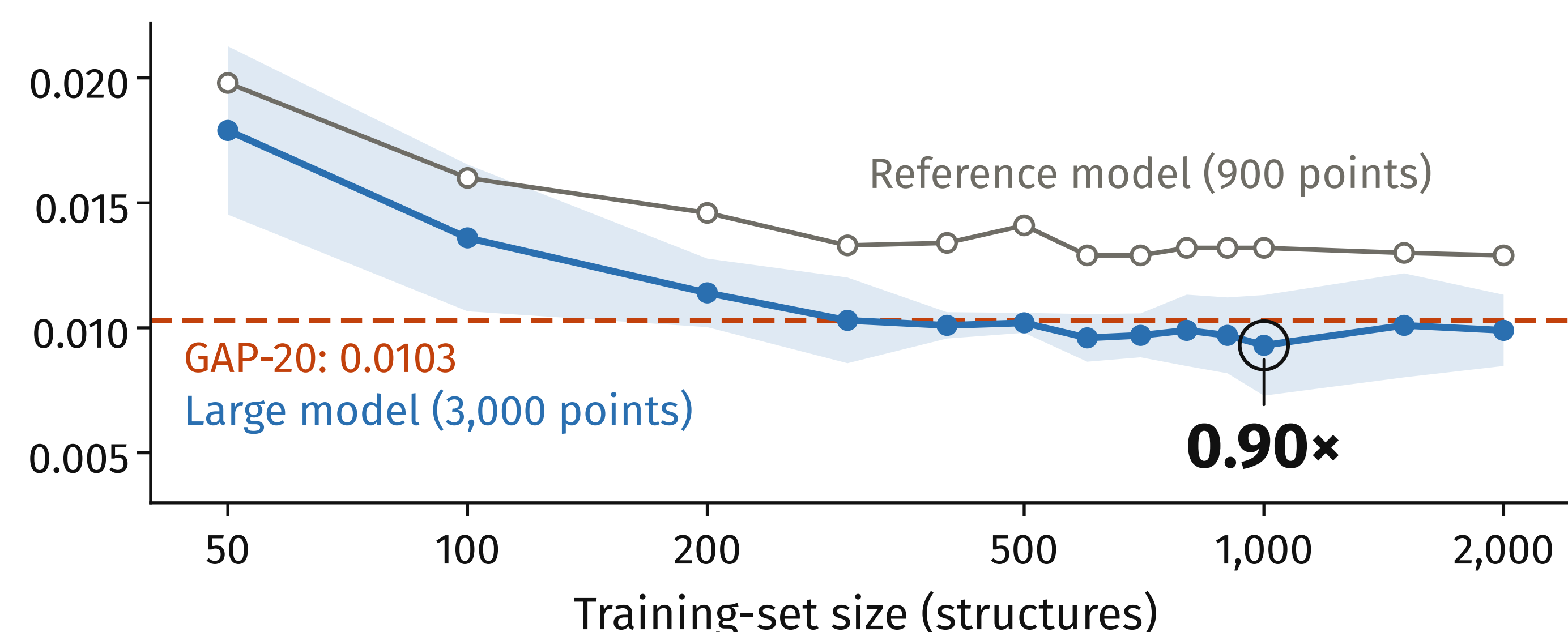
1,000 random

6× fewer structures, with no curation at all



Each square: one training structure. Error: median (eV/atom) on the 116 structures that neither model had seen; ratio 0.90.

Median energy error (eV/atom) on the 116 structures

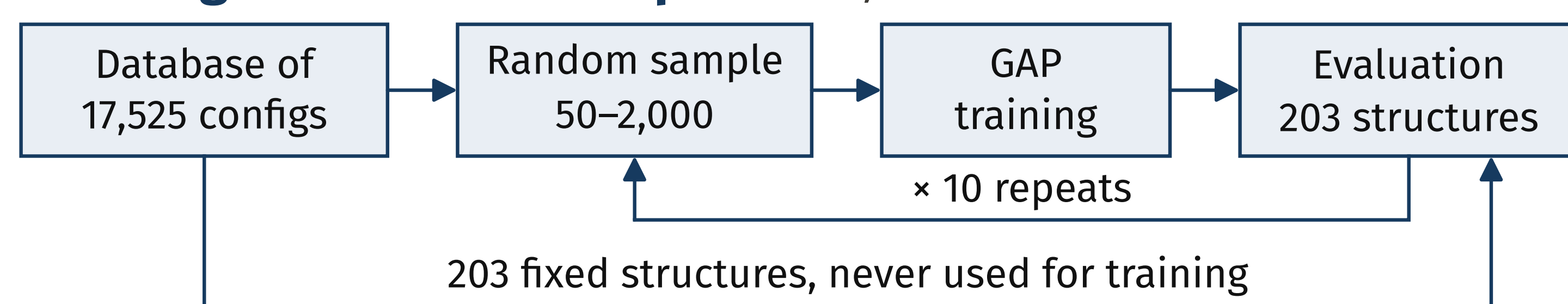


More accurate than GAP-20 at 9 of the 13 sizes; shading: ± 1 standard deviation.

1 1,040 potentials on random samples of 50–2,000 structures

Machine-learning potentials (here GAP with the SOAP descriptor) give DFT energies at a cost orders of magnitude lower; their accuracy is set by the training structures.

8 settings $\times$ 13 sizes $\times$ 10 repeats = 1,040



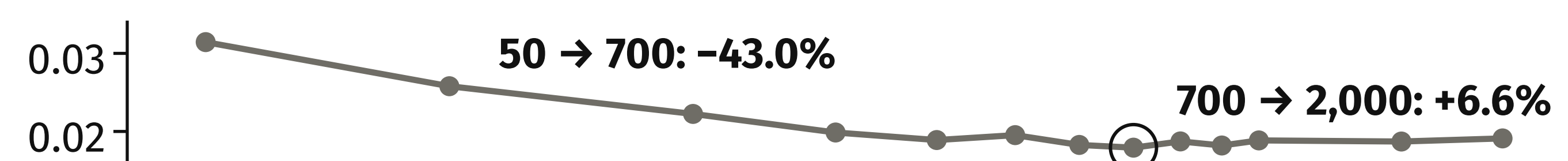
Each setting changes one parameter, e.g. the “capacity”: 300–3,000 representative points. ~316 h on AUTH’s “Aristotle” cluster ($\approx 10,100$ CPU hours).



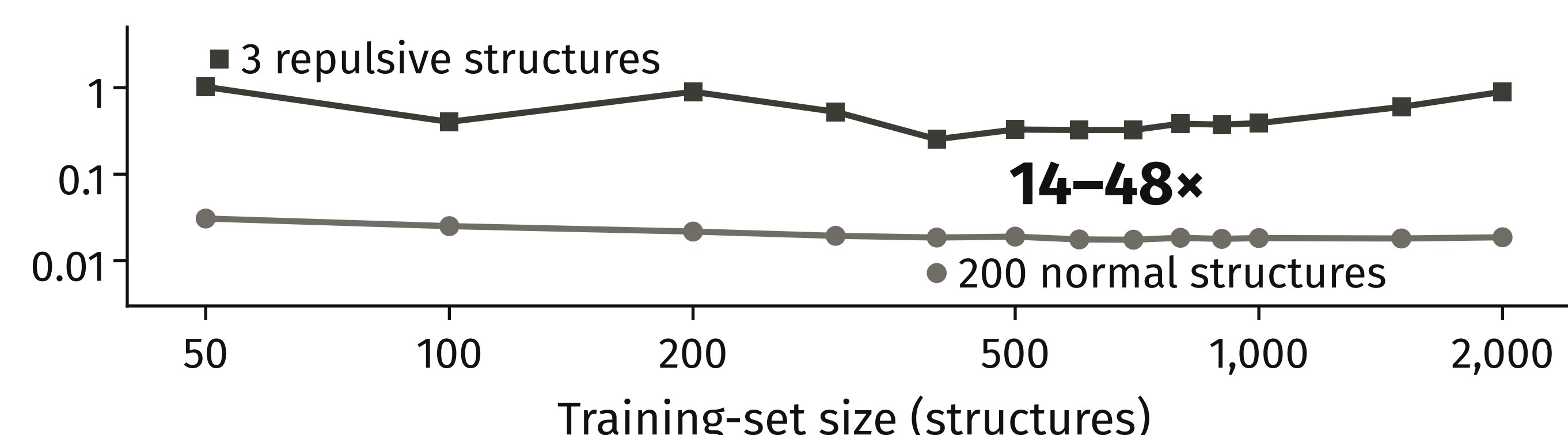
Scan for the poster in Greek and English, the abstract and the references.
9psxmpf.chalkidis.eu

2 The curve hides 3 repulsive structures, 14–48× worse

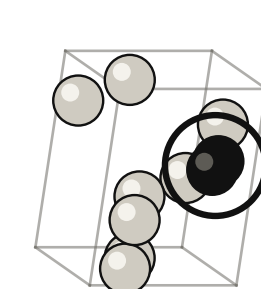
Median energy error (eV/atom) · all 203 structures



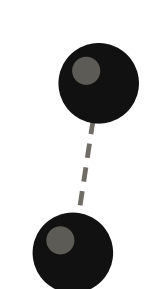
The same, split into the two populations (log scale)



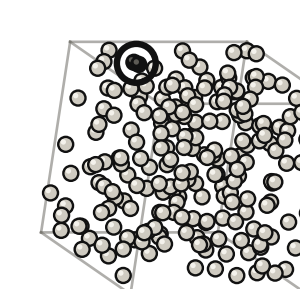
The rise is not overfitting (the training-validation gap closes, $7.6\times \rightarrow 1.1\times$): it comes from a few extreme structures, which are not predicted and which spoil the fit when they enter training.



Crystalline
10 atoms
-0.96 eV/atom



Dimer
2 atoms · 2.83 Å
+0.06 eV/atom

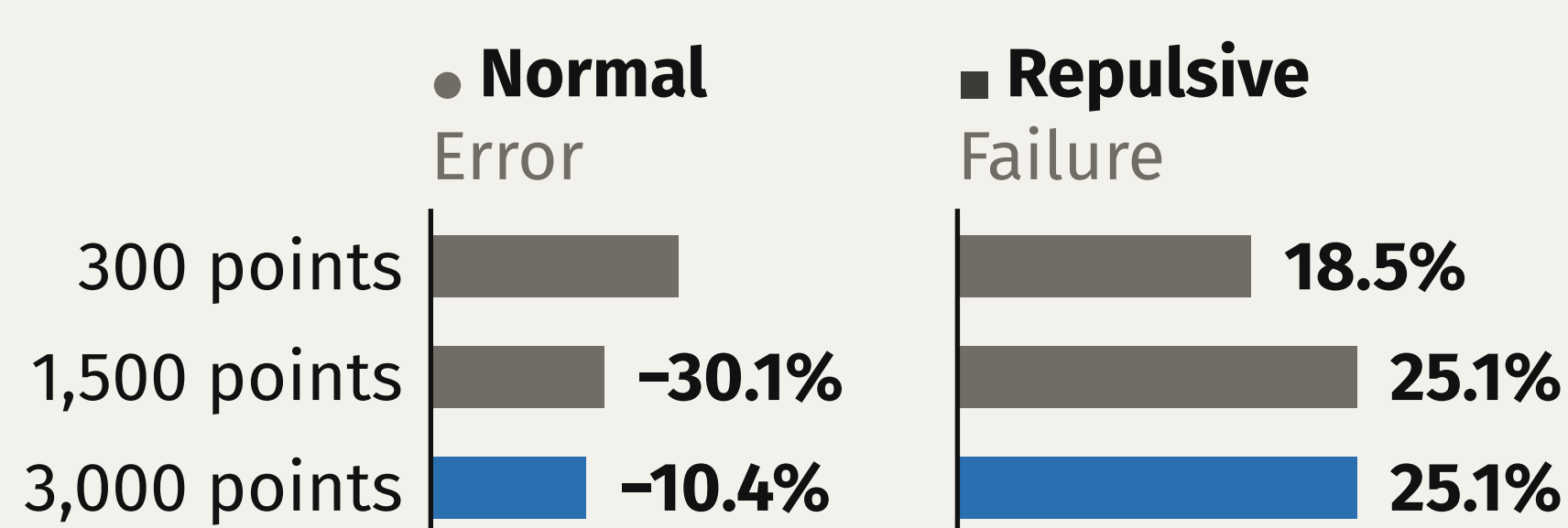


Liquid
216 atoms
+4.05 eV/atom

Repulsive: DFT energy above -4 eV/atom (normal structures: median -7.20).

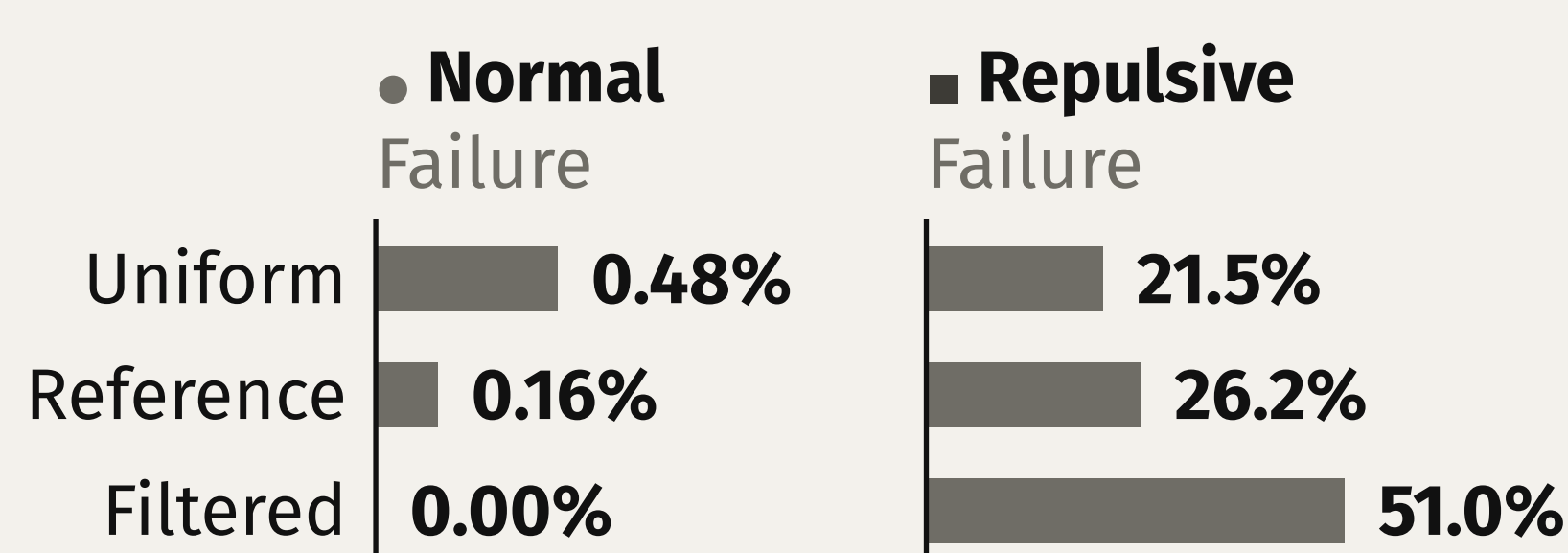
3 Three independent lines of evidence: the limit is in the data

a A larger model: better on normal structures, not on repulsive ones



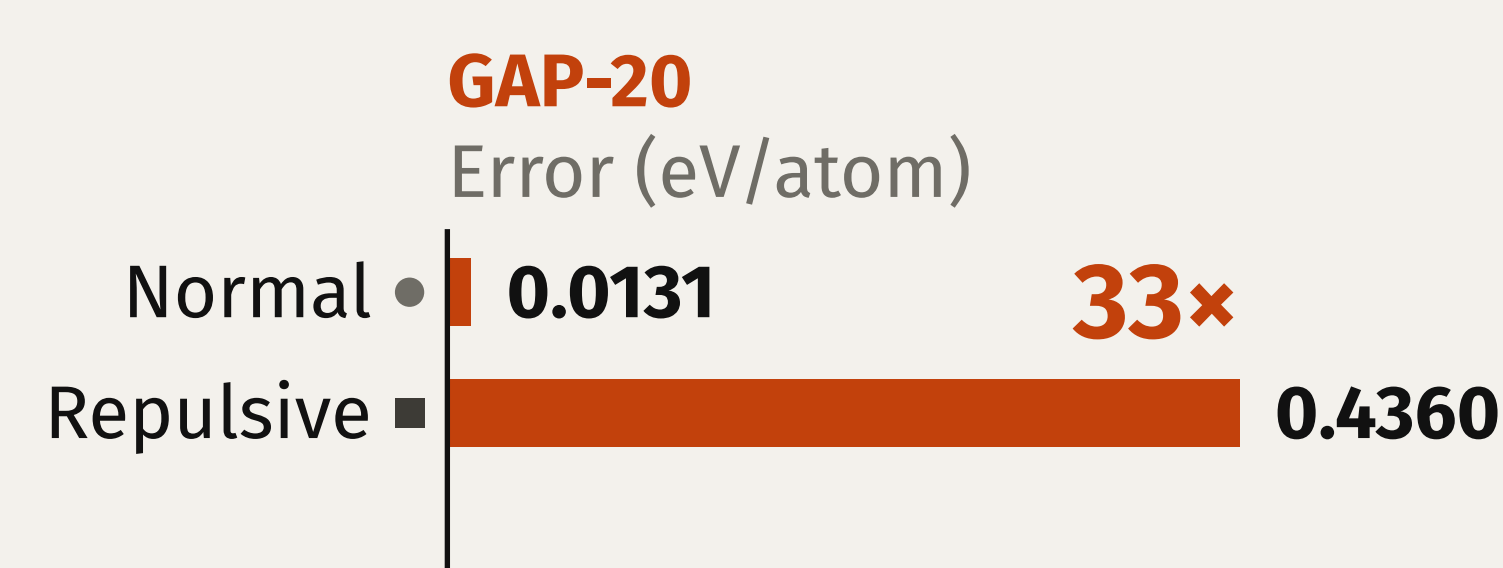
Overall -40.5% for 10× the points. Its error was predicted before training (0.0146–0.0166 eV/atom); the measured value was 0.0152.

b Different training: what normal structures gain, repulsive ones lose



Failure: error > 1 eV/atom. Uniform: one tolerance for all structures; Filtered: no structures with a force > 50 eV/Å.

c GAP-20 fails on the same structures



The same order of magnitude as our 14–48×, despite its curation and its 9,000 points.

On an uncurated database, accuracy is limited by the data, not by the model.

Curation is not a prerequisite, but one of the levers.

Next step: more repulsive structures in training, through active learning driven by the disagreement of a committee of models.

Limitations: only 3 repulsive evaluation structures; the evaluation set comes from the same database as our samples, so the 0.90 is an indication of equivalence.

