Enhancing material property predictions by leveraging transfer learning techniques

24th International Conference on Solid State Ionics
18th July 2023 – London, UK

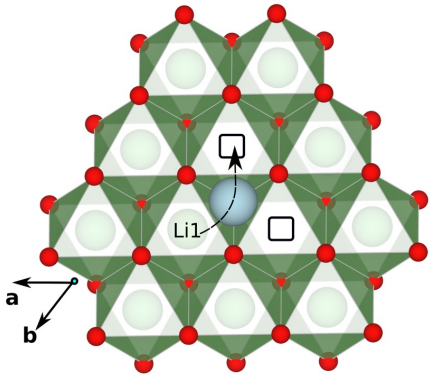
Presented by
Reshma Devi P
IV-Year Ph.D.
Department of Materials Engineering
Indian Institute of Science
India

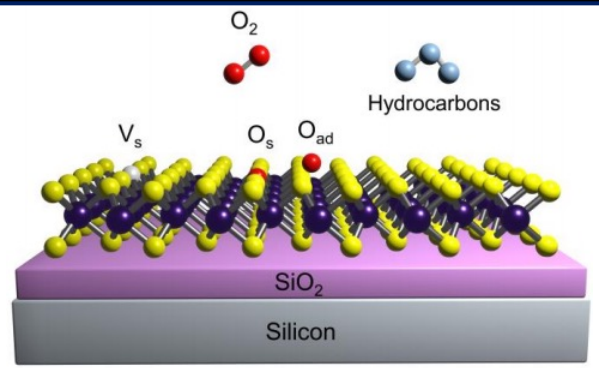
Under the Guidance of
Dr. Keith T. Butler
Professor
Department of Chemistry
University College London
United Kingdom

Under the Guidance of
Dr. Sai Gautam Gopalakrishnan
Assistant Professor
Department of Materials Engineering
Indian Institute of Science
India

How to handle data inadequacy in materials science?

- The accuracy of a Machine Learning (ML) model depends on
  - ❑ Quality of data
  - ❑ **Quantity of data** → **Challenging aspect to meet for specific material properties**
  - ❑ Model framework
  - ❑ Training algorithm
- Complex models like Graph Neural Networks (GNNs) perform better at datapoints > 10^4





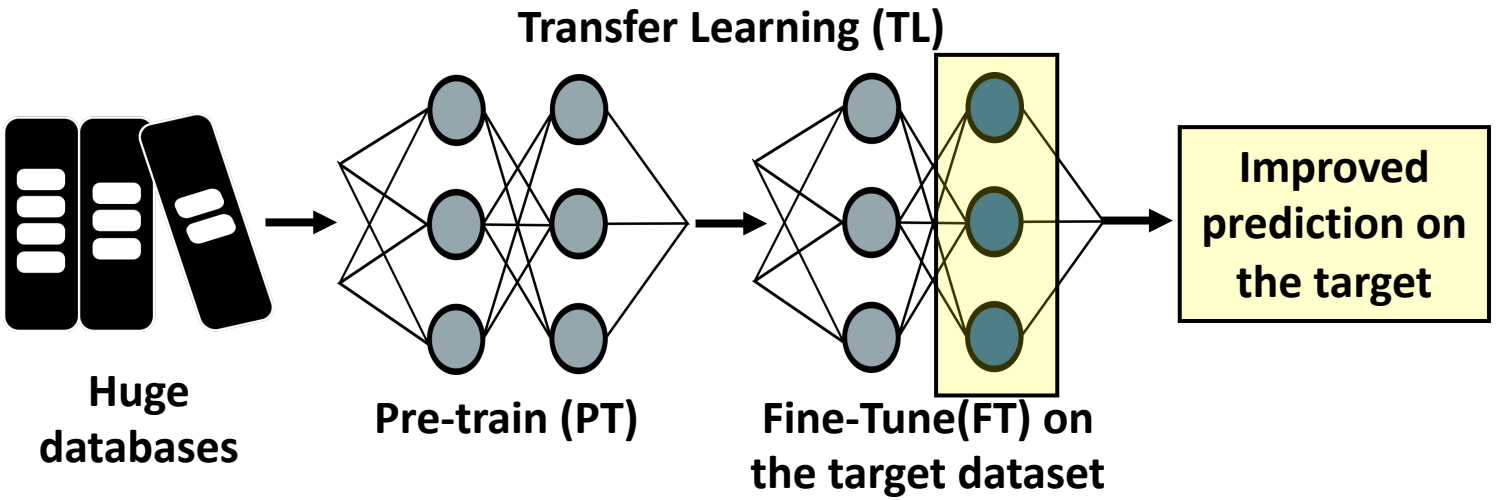
Construct new models that have less variance when trained on small datasets

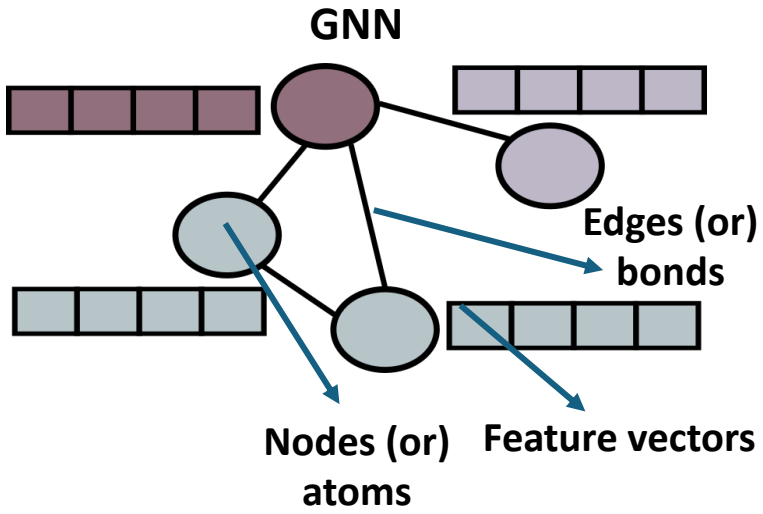
1. Kumar et al. 2024. 2D Mater.

How to handle data inadequacy in materials science?

- The accuracy of a Machine Learning (ML) model depends on
  - ❑ Quality of data
  - ❑ **Quantity of data** → **Challenging aspect to meet for specific material properties**
  - ❑ Model framework
  - ❑ Training algorithm
- Complex models like Graph Neural Networks (GNNs) perform better at datapoints > 10^4

Construct new models that have less variance when trained on small datasets



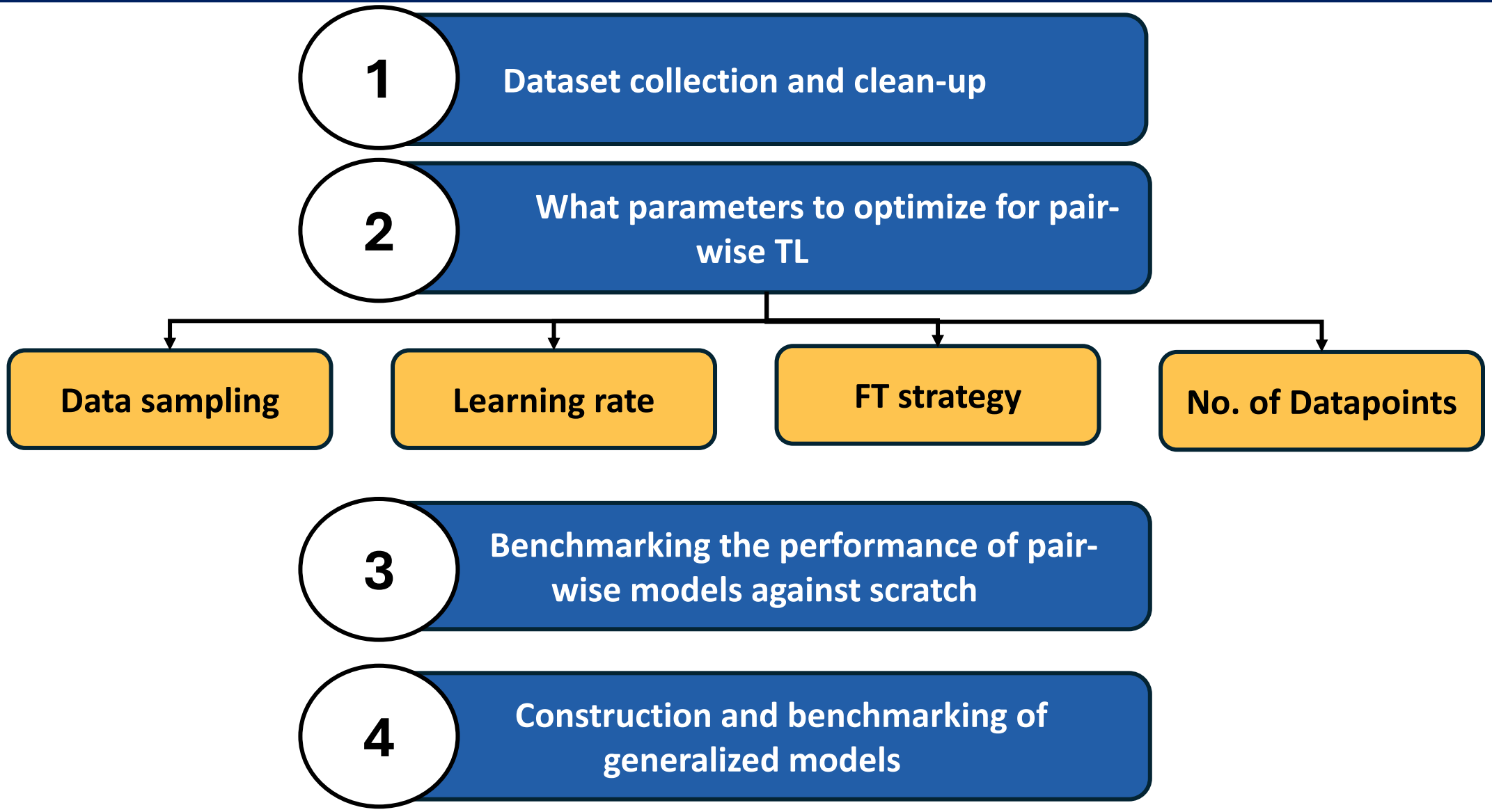


Objectives

- What is the best way to do pair-wise TL?
- Is there a strategy to create generalized models that can learn on multiple properties simultaneously?

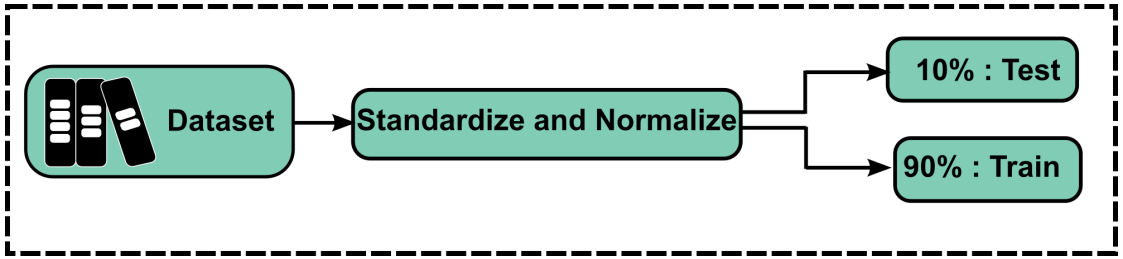
1. Kumar et al. 2024. 2D Mater.

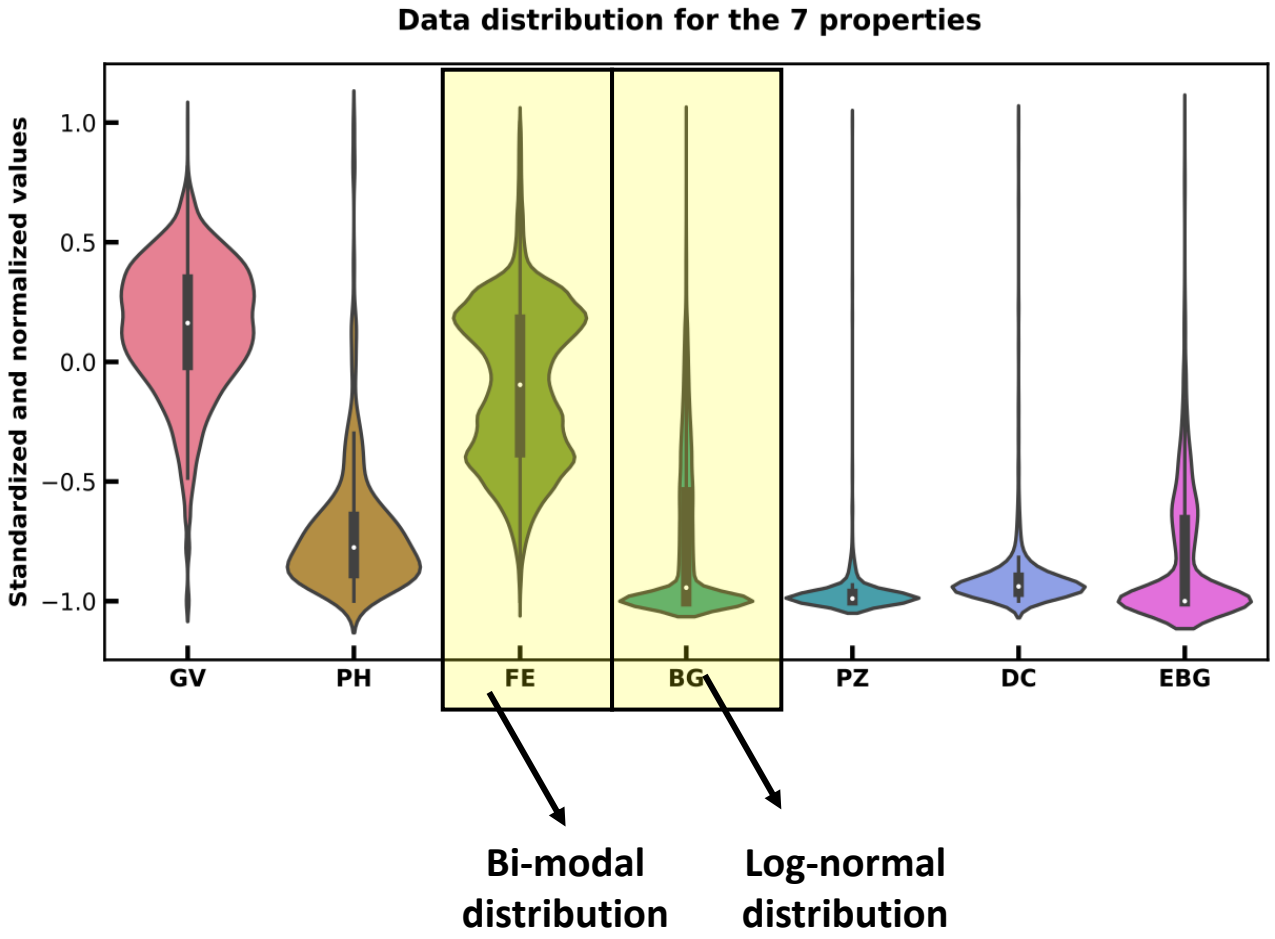
General workflow



7 datasets spanning different distribution

Datasets from Matminer[1]	# of Datapoints
Piezoelectric modulus (PZ)	941
Dielectric constant (DC)	1056
Phonons (PH)	1265
Experimental Band gap (EBG)	2481
GVRH (GV)	10987
Band gap (BG)	106113
Formation energy (FE)	132752

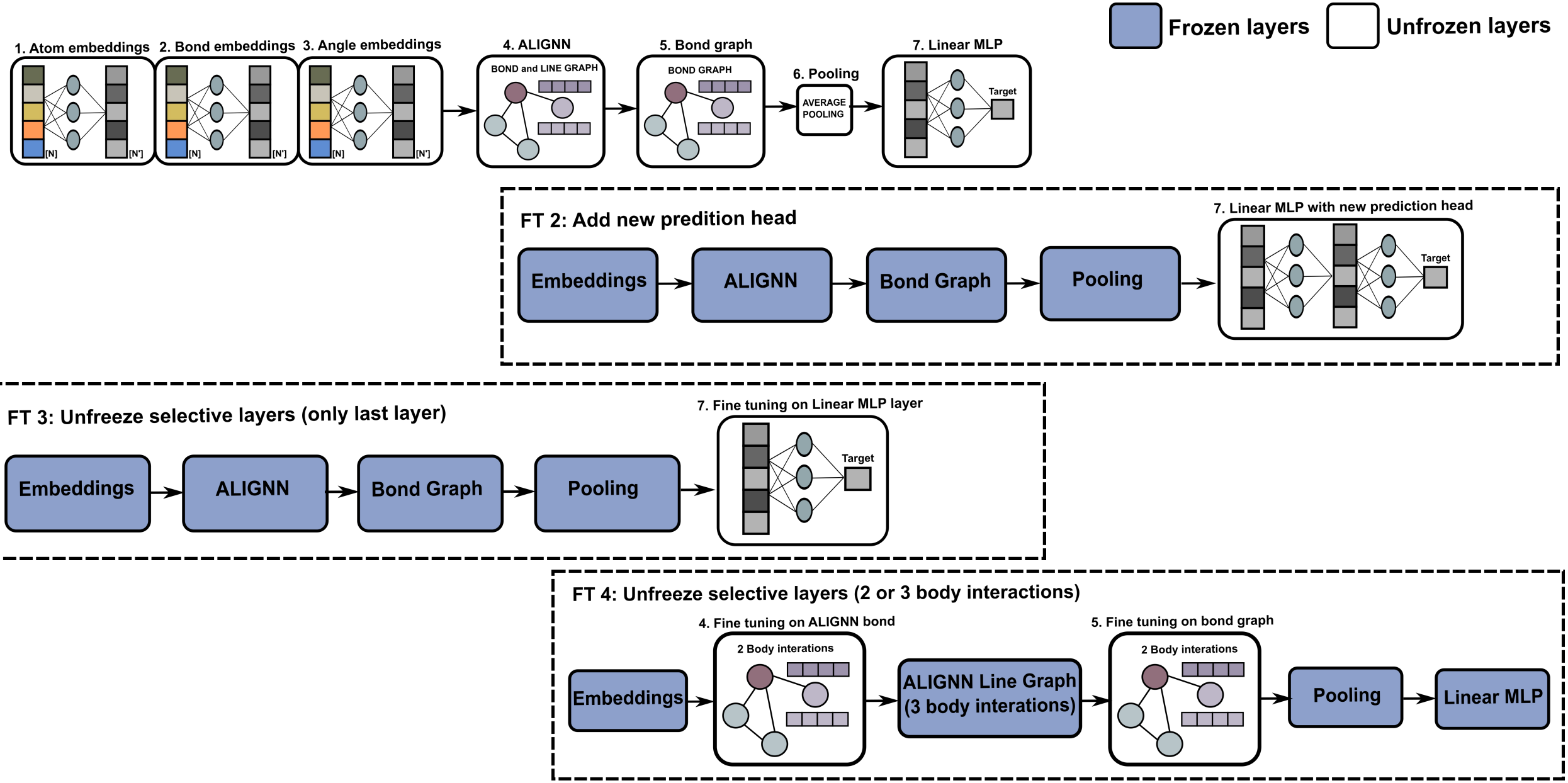




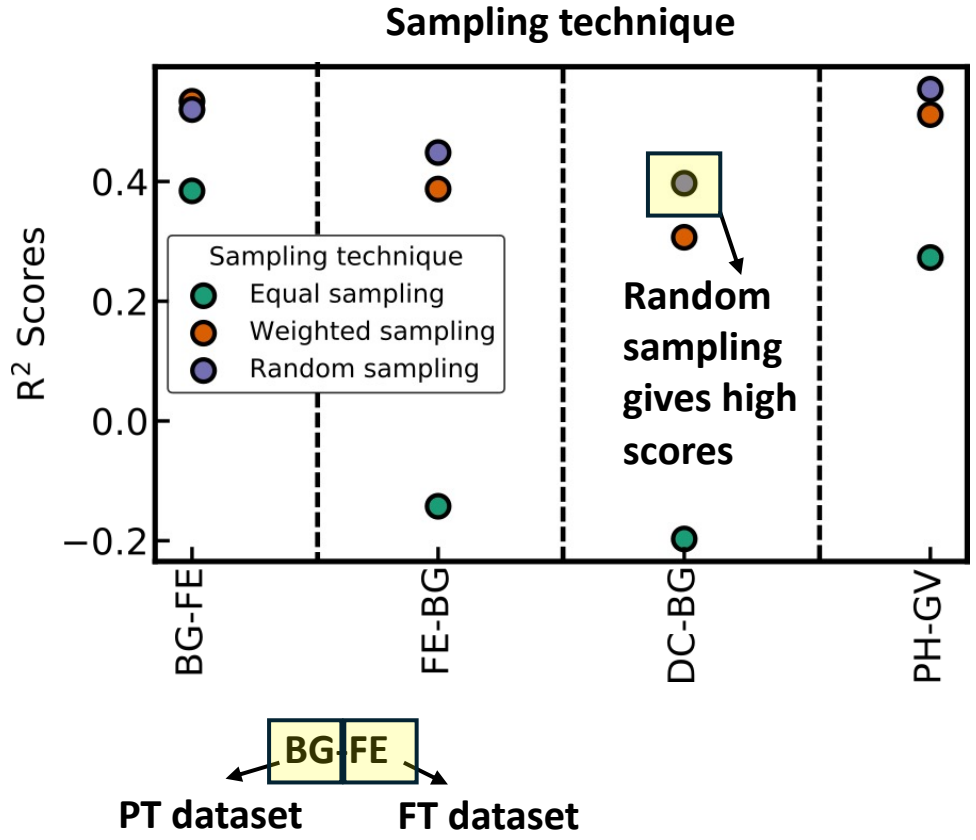
- The test dataset was never used in any of the PT or FT stages
- We report only the test scores in all our results

1. Ward et al. 2018. Comput. Mater. Sci.

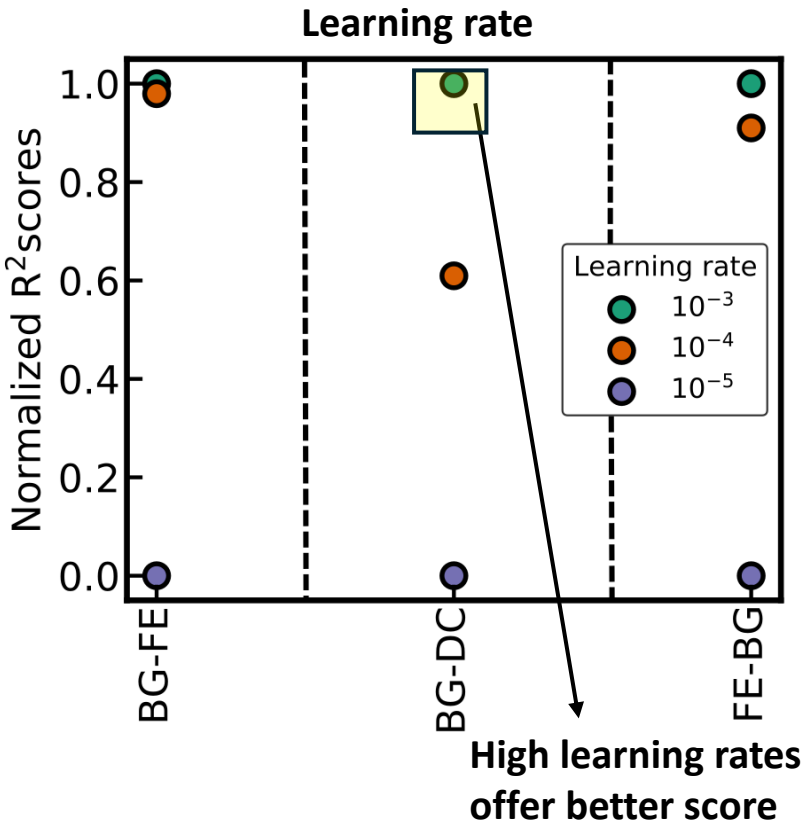
4 Fine-tuning strategies



High learning rates and Random sampling for better R^2 scores



Datapoints: 500
Epochs: 500
Batch size: 16
Learning rate: 10^{-4}
(for sampling technique)

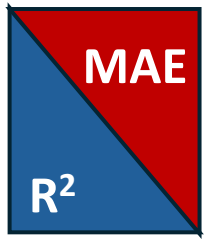


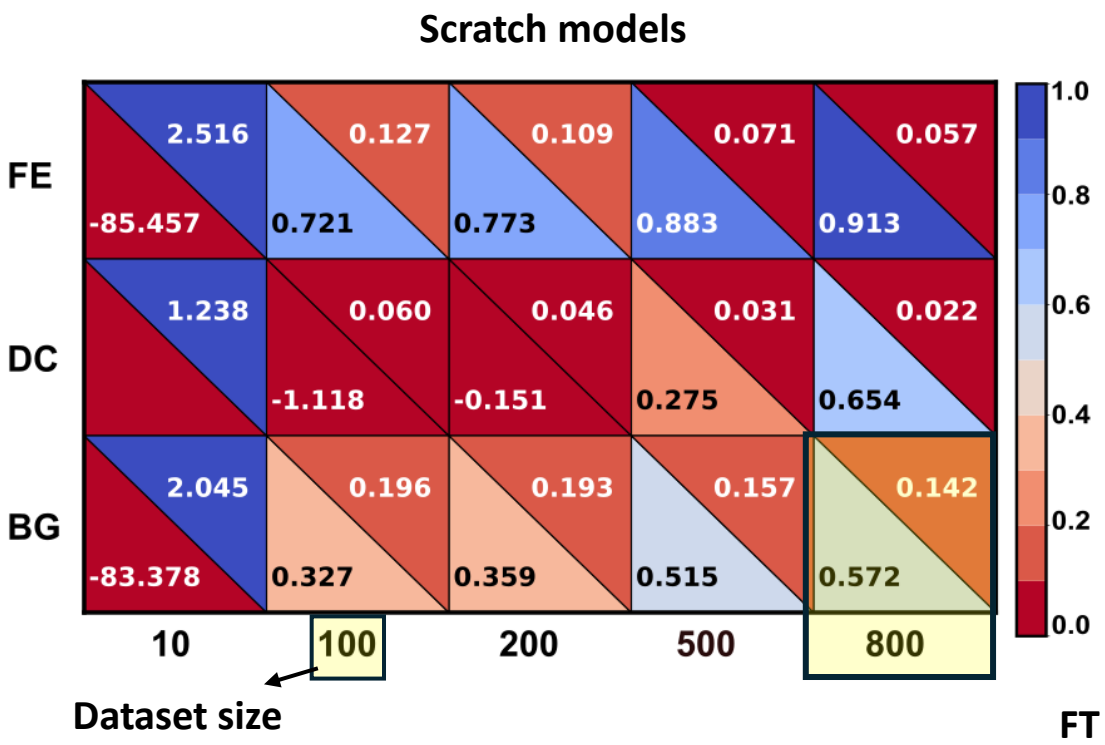
Selective PT-FT pairs were used considering the enormity of the calculation

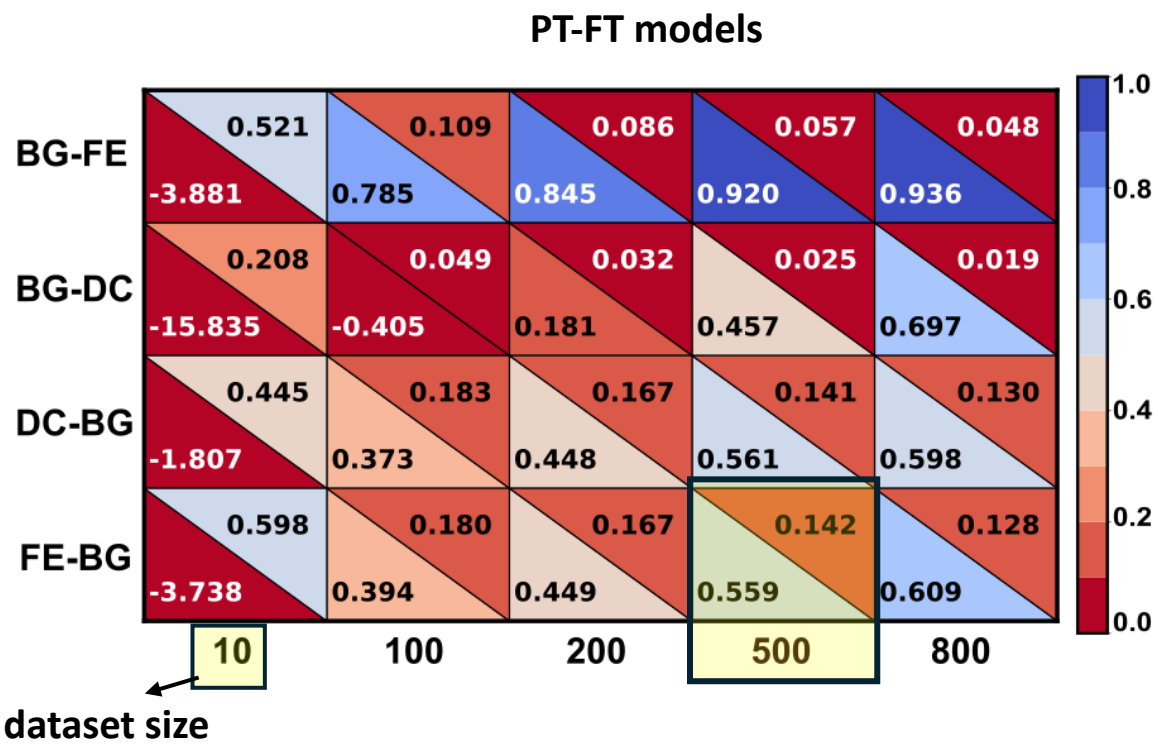
EBG: Experimental Band gap
GV: GVRH
PZ: Piezoelectric modulus
DC: Dielectric constant
BG: Band gap
PH: Phonons
FE: Formation energy

Random sampling, high learning rates, high no. of unfrozen layer and high number of datapoints improvise the performance

Influence of FT size: R^2 scores increase as FT size increases



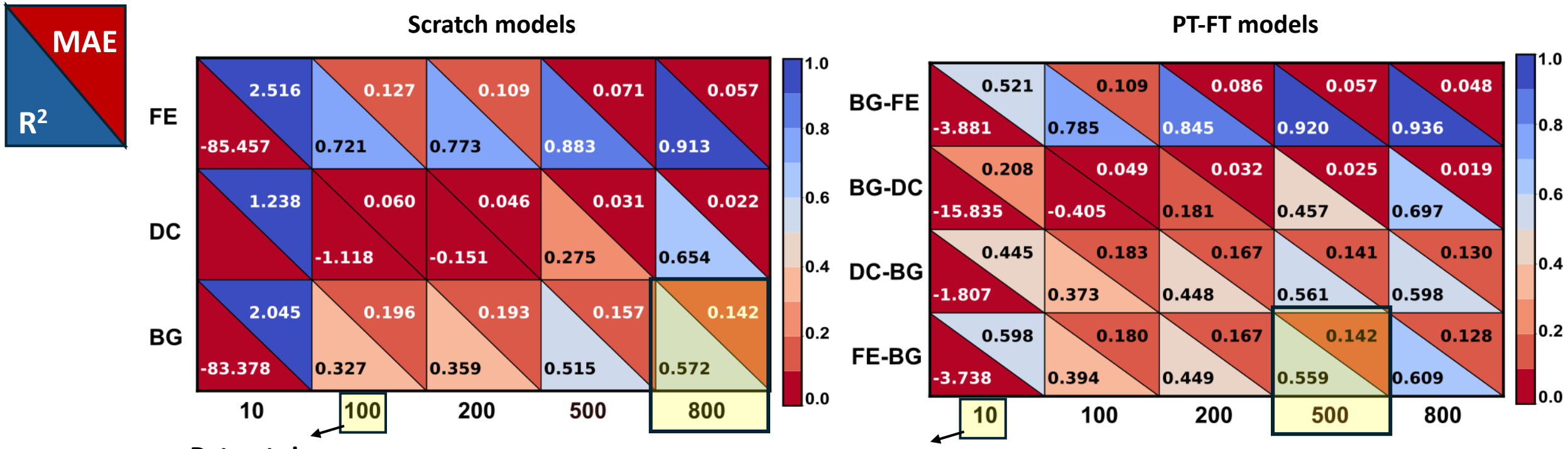




- All PT-FT models perform better than scratch
- R^2 scores and MAEs saturate as FT size increases
- Pair-wise models perform better at smaller dataset sizes than scratch

PT size: 941 (smallest dataset size considered)
FT strategy: Unfreeze all layers
Experiments repeated for 5 different random trials and the mean results are plotted

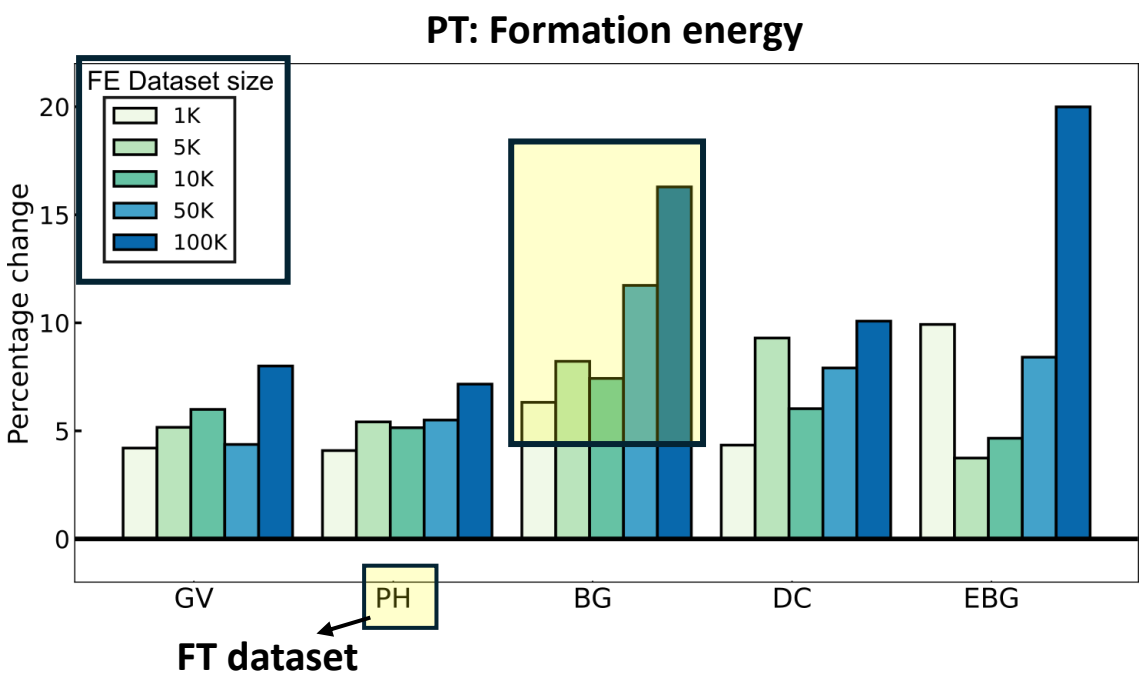
Influence of FT size: R^2 scores increase as FT size increases

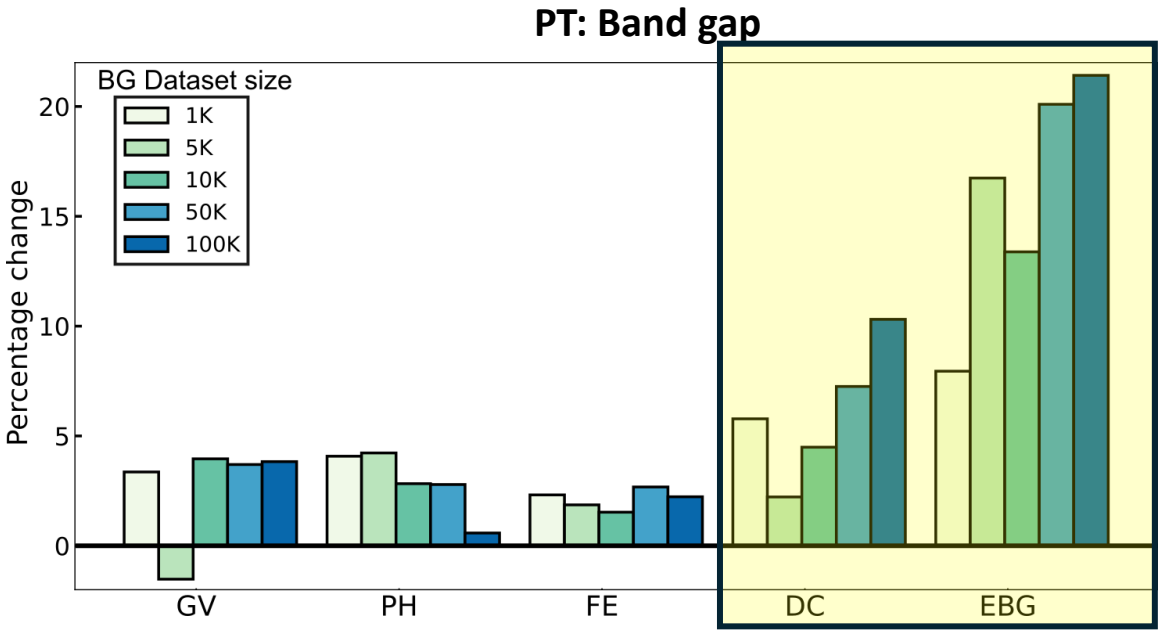


FT size of 800 is identified as optimal and fixed for all following experiments

PT size: 941 (smallest dataset size considered)
FT strategy: Unfreeze all layers
Experiments repeated for 5 different random trials and the mean results are plotted

Influence of PT size: R^2 scores increase as PT size increases





- PT **with FE(100K)** offers the best performance across all FT datasets
- Non-monotonic trend at smaller PT sizes

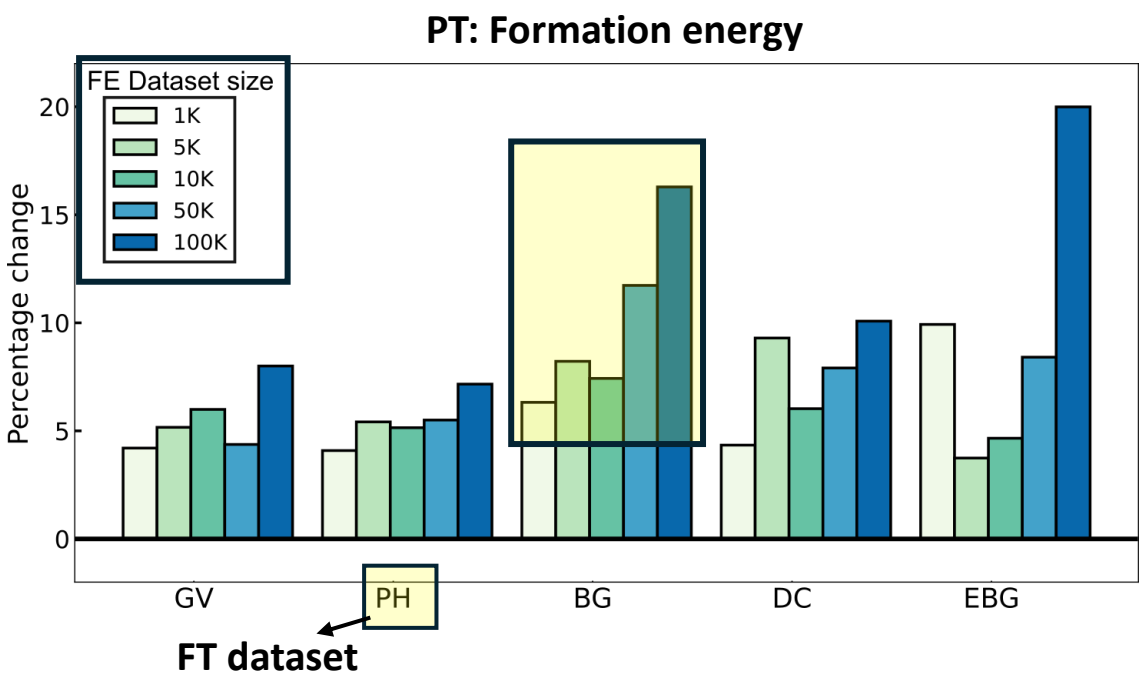
- PT with **BG(50K)** offers the best performance across all FT datasets
- Non-monotonic trend at smaller PT sizes
- BG(100K) Performs specifically better for DC and EBG

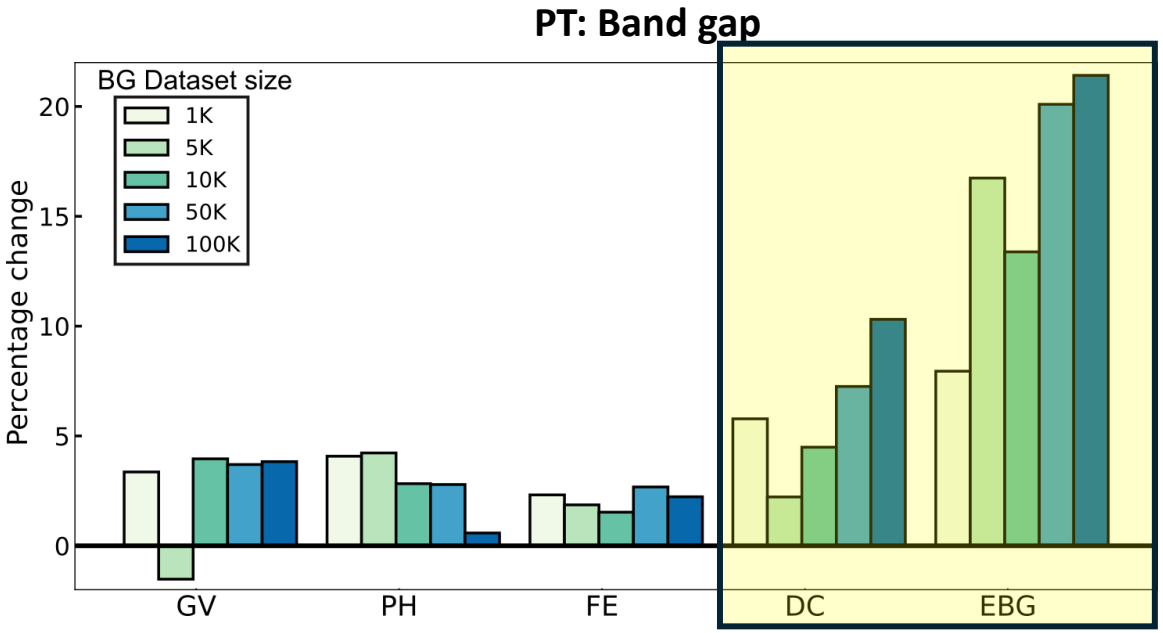
PT size: Largest 2 of the 7 datasets considered – FE and BG
FT strategy: Unfreeze all layers
FT size: 800
Experiments repeated for 5 different random trials and the mean results are plotted

EBG: Experimental Band gap
GV: GVRH
PZ: Piezoelectric modulus
DC: Dielectric constant

BG: Band gap
PH: Phonons
FE: Formation energy

Influence of PT size: R^2 scores increase as PT size increases





- PT with **BG(50K)** offers the best performance across all FT datasets
- Non-monotonic trend at smaller PT sizes

- PT **with FE(100K)** offers the best performance across
- Non-mo

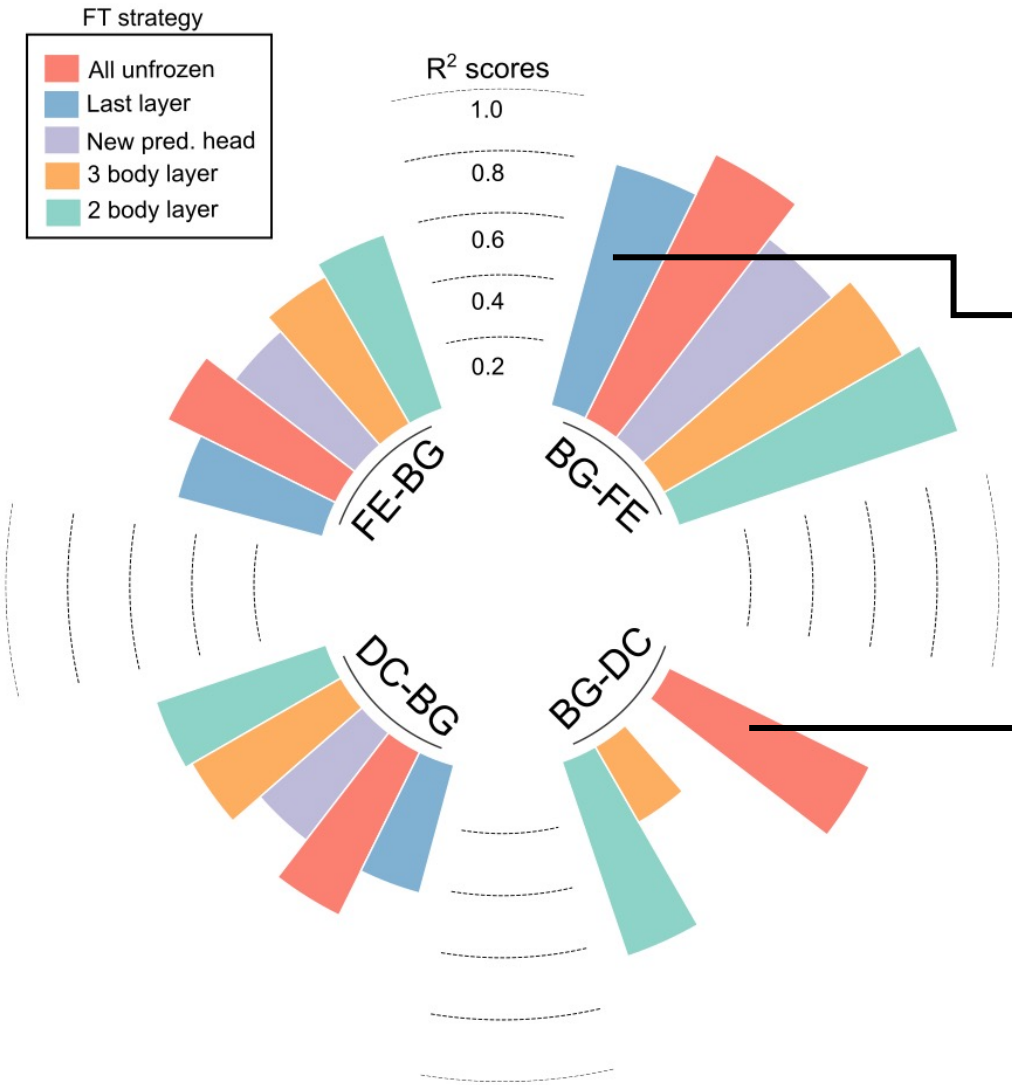
Increase in PT size improves performance if PT dataset has normal distribution or if it is correlated with the FT dataset

PT size: Large
FT strategy: U
FT size: 800
Experiments repeated for 5 different random trials and the mean results are plotted

DC: Dielectric constant

What is the best FT strategy?

PT size: 941
FT size: 800

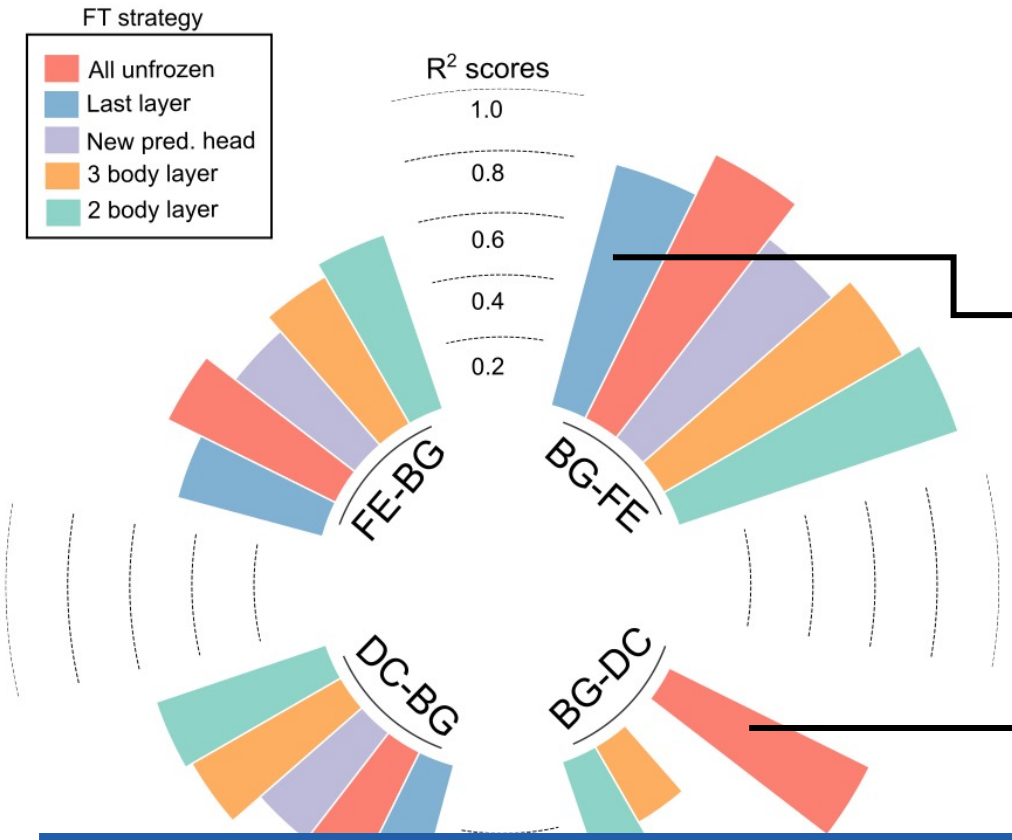


Unfreezing only the last layer offers the least R^2 score

Unfreezing all layers offer the best R^2 score

What is the best FT strategy?

PT size: 941
FT size: 800



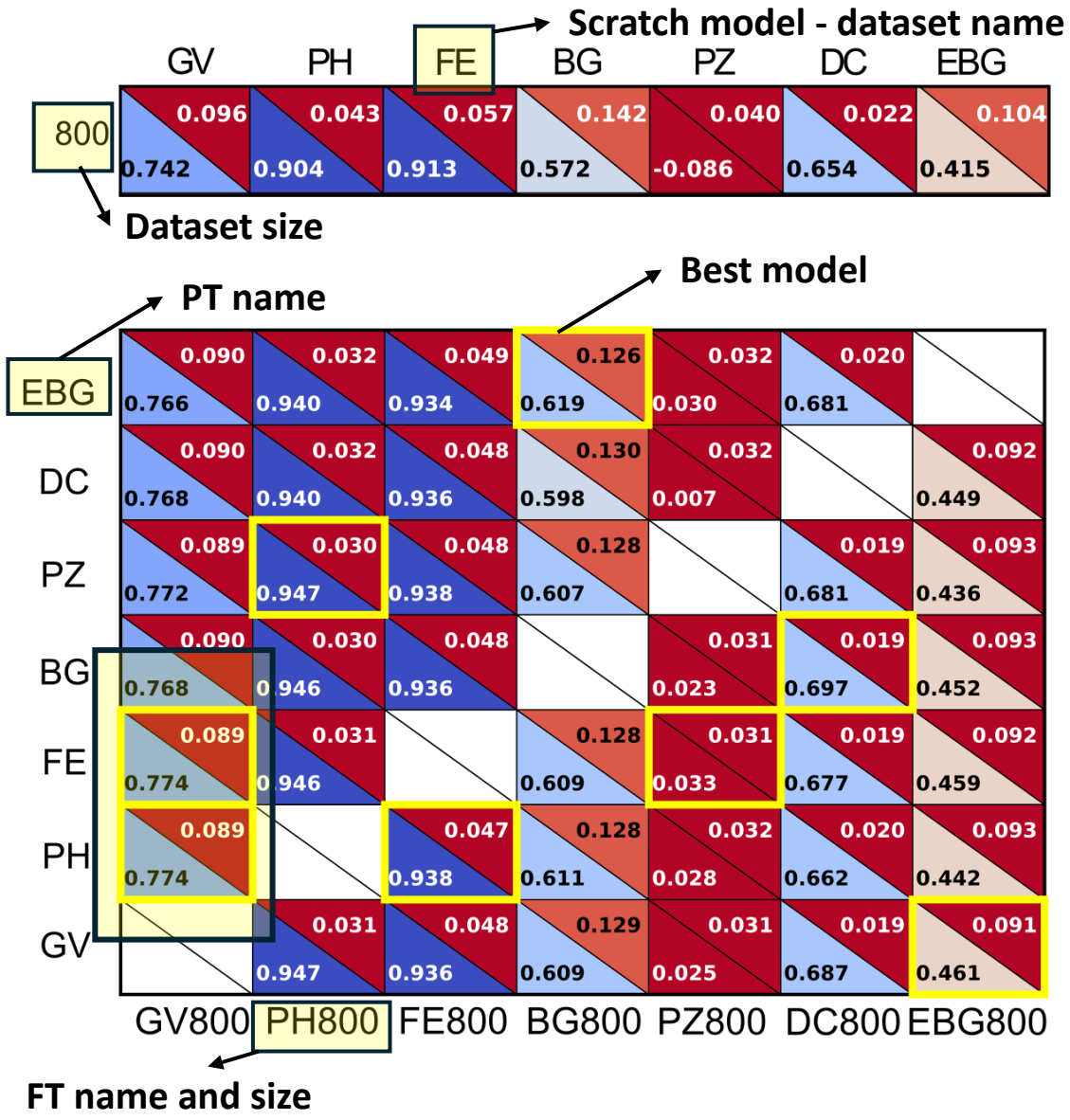
Unfreezing only the last layer offers the least R^2 score

Unfreezing all layers offer the best R^2 score

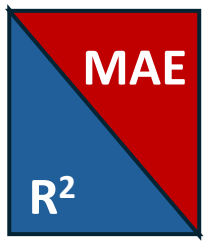
<u>Best strategy: Unfreezing all the layers</u>

Indicates that the PT model requires more re-training to generalize on the FT property

Pair-wise TL on all 7×6 combinations: Better performance at lower datapoints

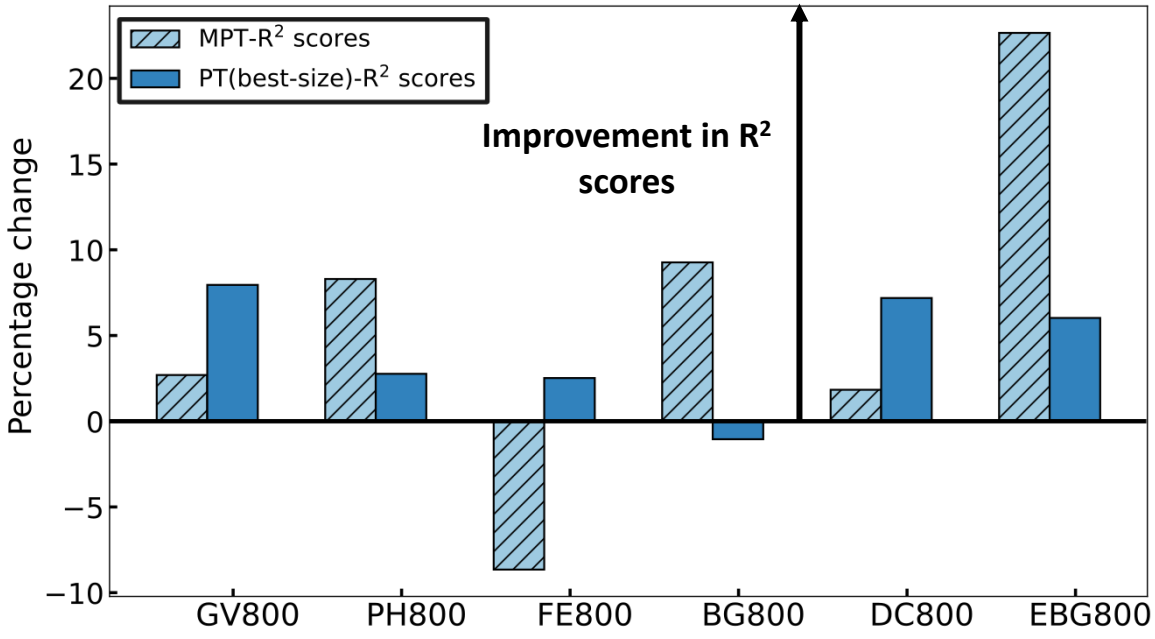


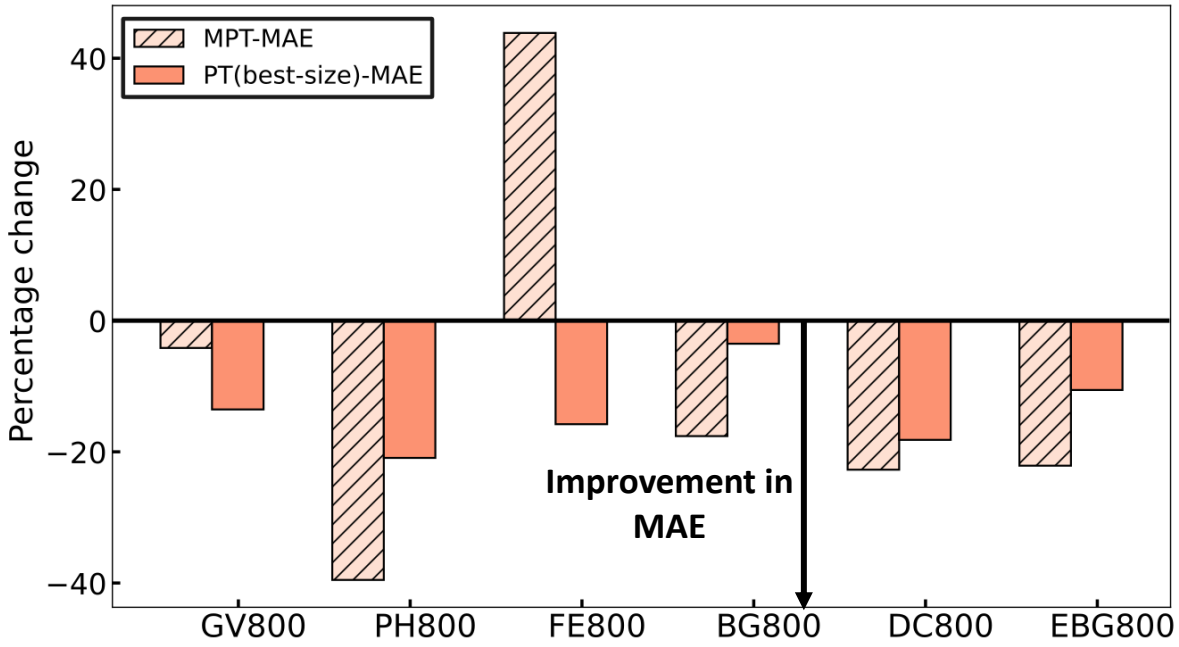
- All PT-FT models perform better than scratch
- Average increase in R^2 score and MAE is 28.4 % and 17.1% respectively
- Good models for GV, PH and FE ($|R^2| > 0.75$)
- Average models for BG, DC, and EBG ($0.4 < |R^2|$ 0.75)
- The specific PT property has little influence on FT when the PT size is capped



PT size: 941
FT strategy: Unfreeze all layers
FT size: 800
Experiments repeated for 5 different random trials and the mean results are plotted

MPT models: Improved R^2scores versus best pair-wise model





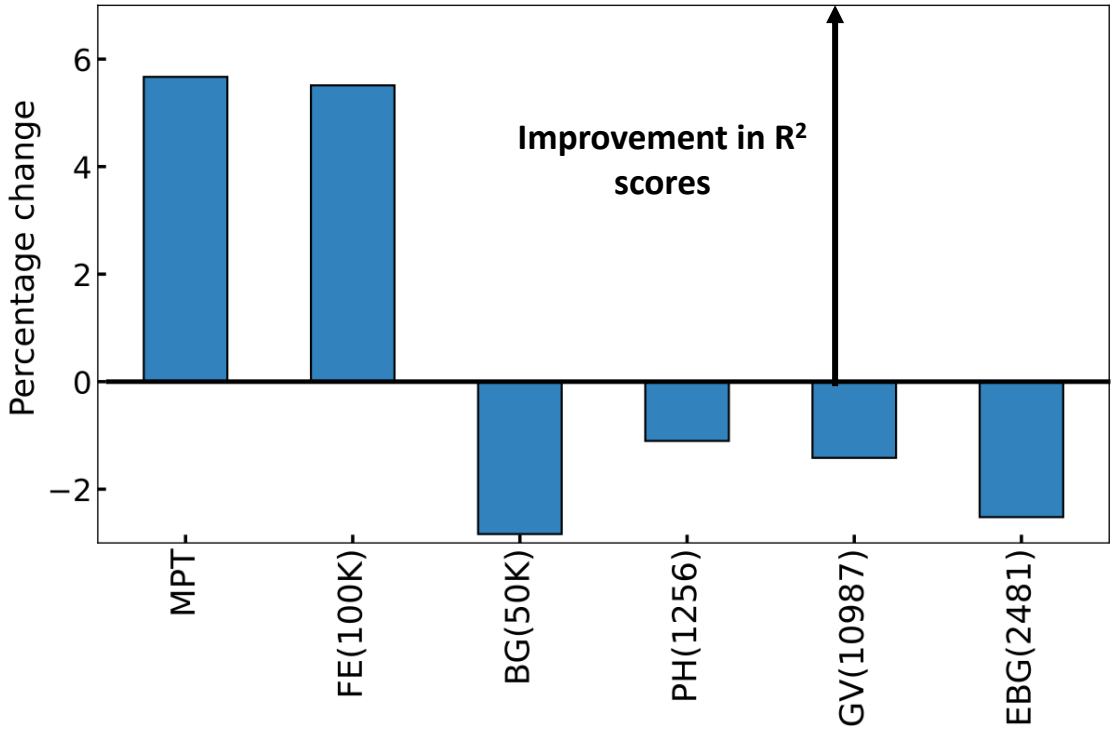
MPT offers best performance in 3/6 and 5/6 cases in terms of R^2scores and MAEs respectively, excluding FE

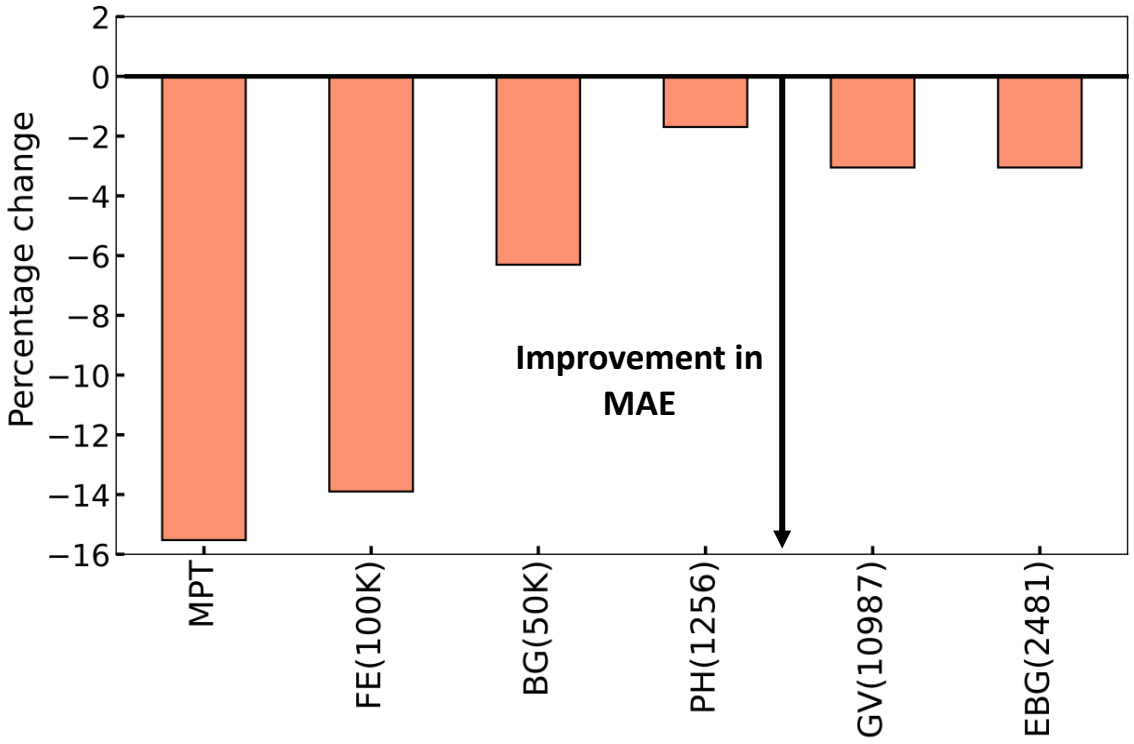
MPT-PT size: 132,270
Pair-wise PT size: Maxed-out best PT dataset
FT size: 800

Dataset size: 1103
PT: MPT model PT on all 7 properties
Pair-wise PT size: : Maxed-out best PT dataset

MPT models: Improved R^2scores versus best pair-wise model

Performance on a completely unrelated dataset: JARVIS 2D band gap



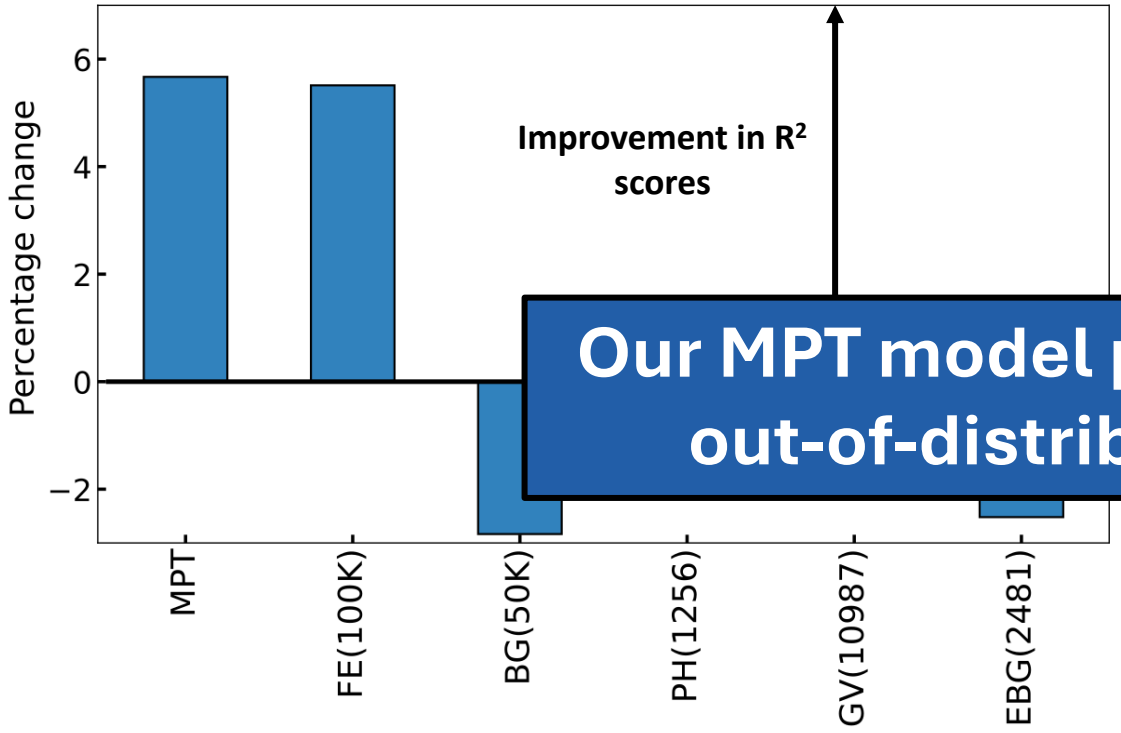


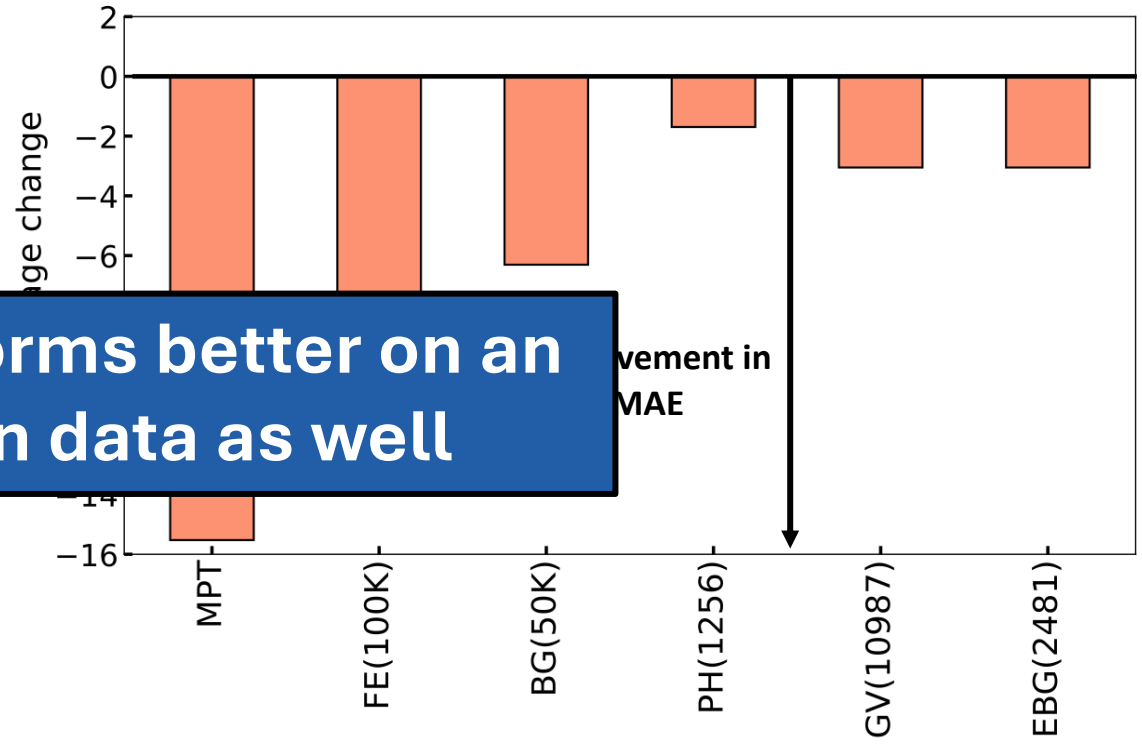
MPT-PT size: 132,270
Pair-wise PT size: Maxed-out best PT dataset
FT size: 800

Dataset size: 1103
PT: MPT model PT on all 7 properties
Pair-wise PT size: : Maxed-out best PT dataset

MPT models: Improved R^2 scores versus best pair-wise model

Performance on a completely unrelated dataset: JARVIS 2D band gap





Our MPT model performs better on an out-of-distribution data as well

MPT-PT size: 132,270
Pair-wise PT size: Maxed-out best PT dataset
FT size: 800

Dataset size: 1103
PT: MPT model PT on all 7 properties
Pair-wise PT size: : Maxed-out best PT dataset

Major takeaways and Acknowledgements

- We provide an improvised TL paradigm for knowledge transfer from source to data insufficient target datasets
- We investigate the impact of FT strategy, PT and FT dataset size on pair-wise TL
  - ❑ Our pair-wise models outperform models from scratch in majority of the cases
- We investigate the MPT model performance, and find it to provide better (or equivalent) performance even in completely unrelated dataset like JARVIS 2D band gap

arXiv version of the paper

Github repository

THE ROYAL SOCIETY

UKRI

SERB INDIA

Dr. Keith T. Butler
UCL, London