On the expressiveness and spectral bias of KANs

Yixuan Wang*, Jonathan W. Siegel, Ziming Liu, Thomas Y. Hou

***PhD candidate, Caltech**

Based on arXiv: 2410.01803

KA representation theorem

[From wikipedia]

If f is a multivariate continuous function, then f can be written as a finite composition of continuous functions of a single variable and the binary operation of addition.

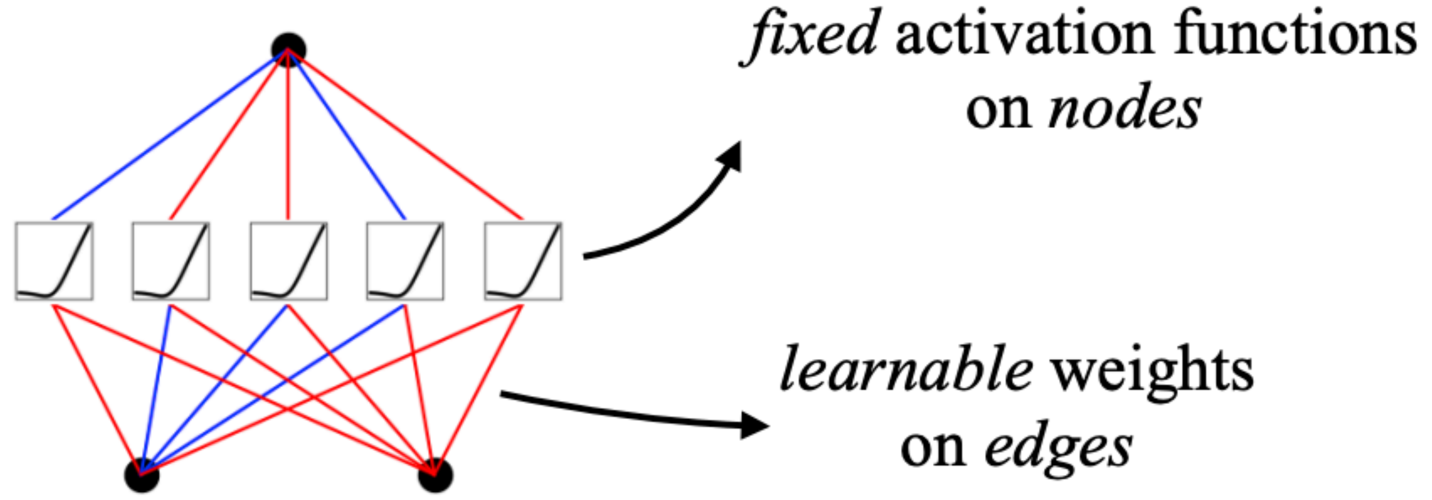
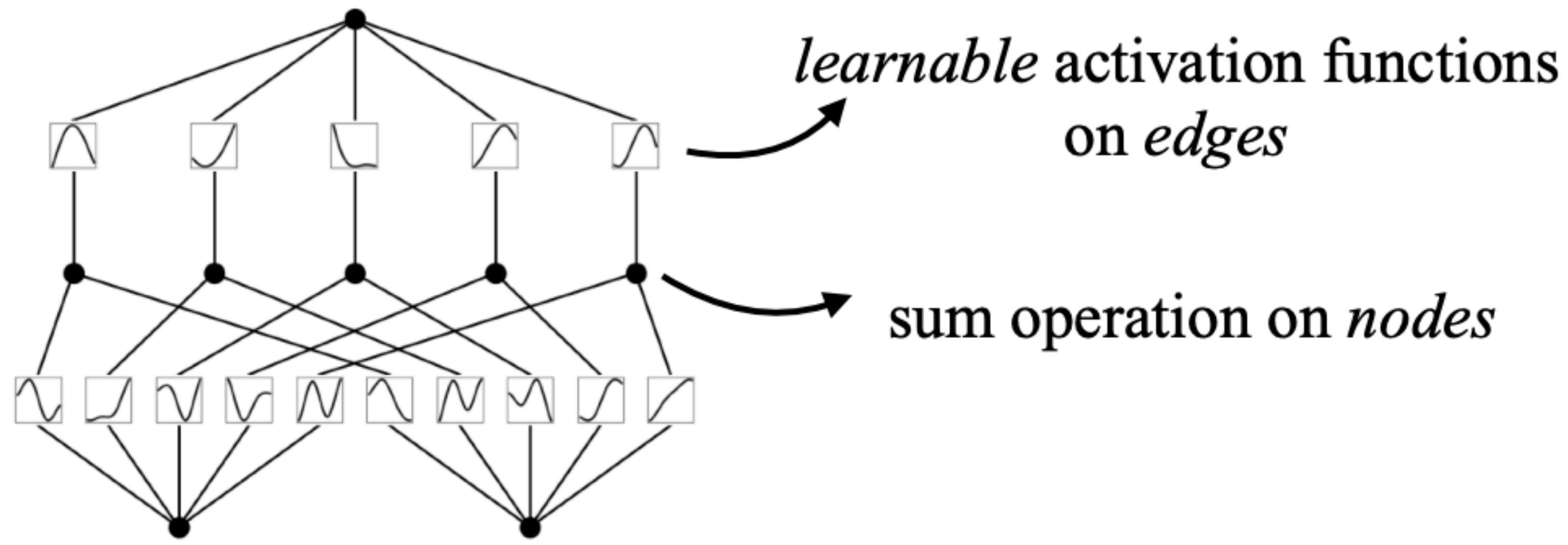
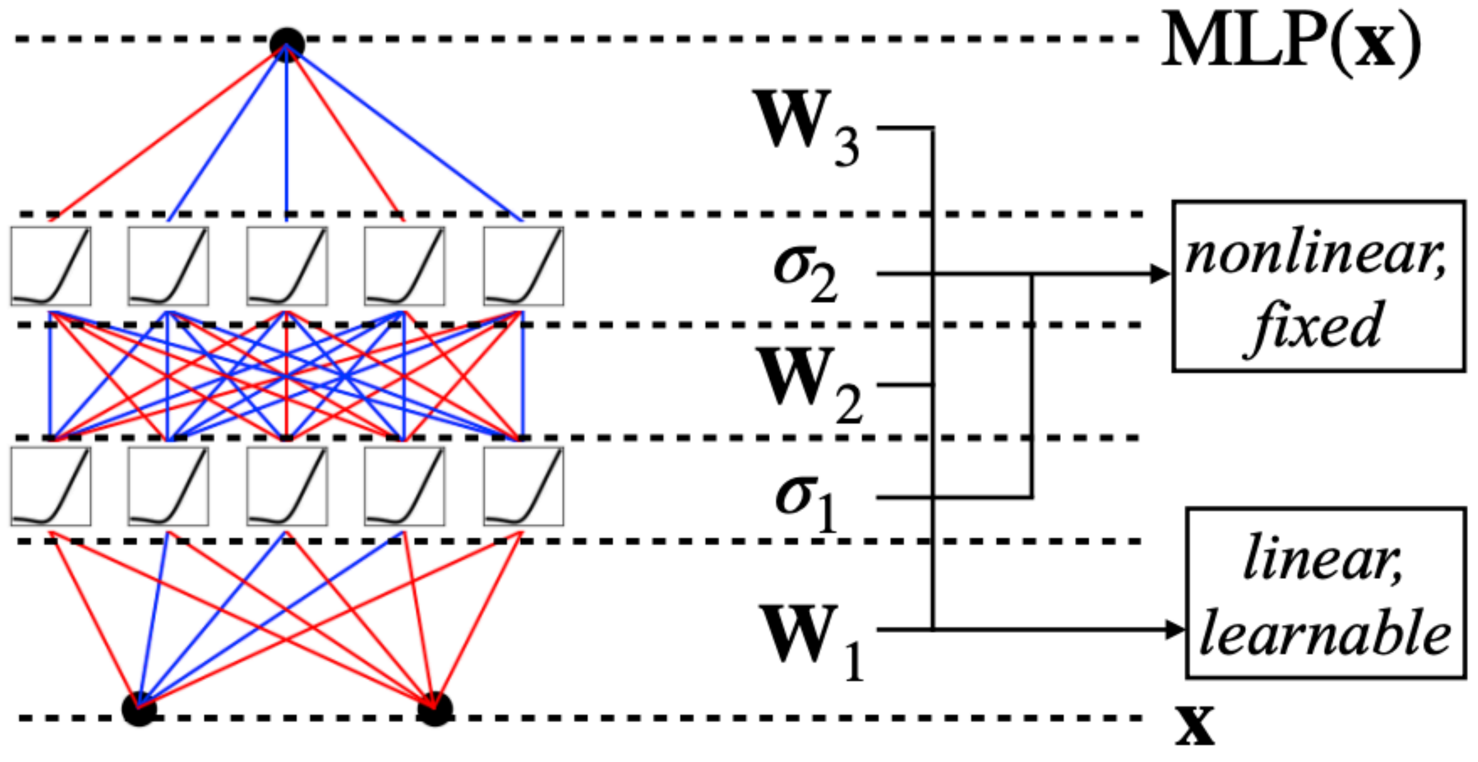
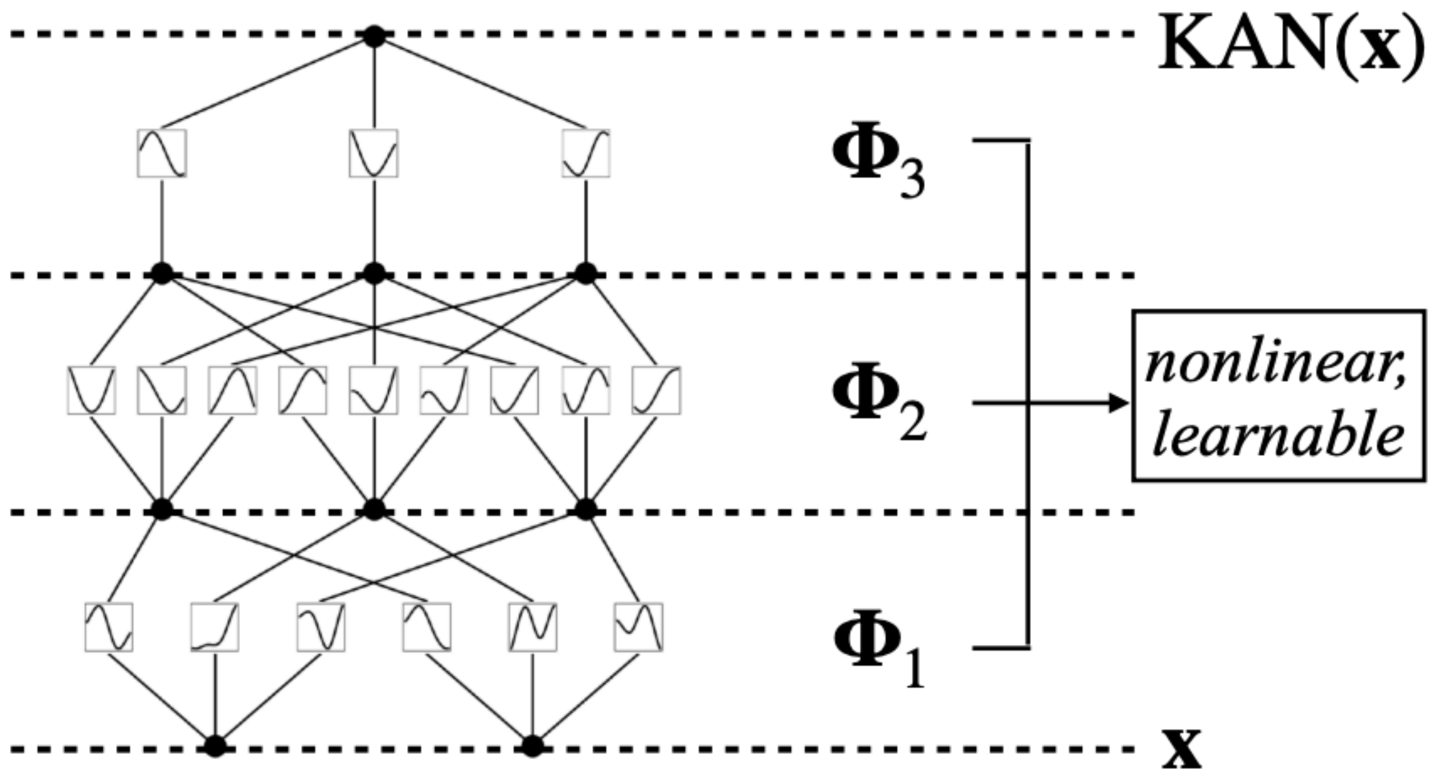
$$f(\mathbf{x}) = f(x_1, \cdots, x_n) = \sum_{q=1}^{2n+1} \Phi_q(\sum_{p=1}^{n} \phi_{q,p}(x_p))$$

Where $\phi_{q,p} : [0,1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$.

In a sense, they showed that the **only true multivariate function is the sum**, since every other function can be written using **univariate functions and summing**.

Alleviating COD/ More interp

MLP & KAN are *dual* (Liu 2024a)

Model	**Multi-Layer Perceptron (MLP)**	**Kolmogorov-Arnold Network (KAN)**
Theorem	**Universal Approximation Theorem**	**Kolmogorov-Arnold Representation Theorem**
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$
Model (Shallow)	**(a)** 	**(b)** 
Formula (Deep)	$\mathrm{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\mathrm{KAN}(\mathbf{x}) = (\mathbf{\Phi}_3 \circ \mathbf{\Phi}_2 \circ \mathbf{\Phi}_1)(\mathbf{x})$
Model (Deep)	**(c)** 	**(d)** 

Scaling

$$\ell \propto N^{-\alpha}$$

approximation error — *number of parameters* — *scaling exponent*

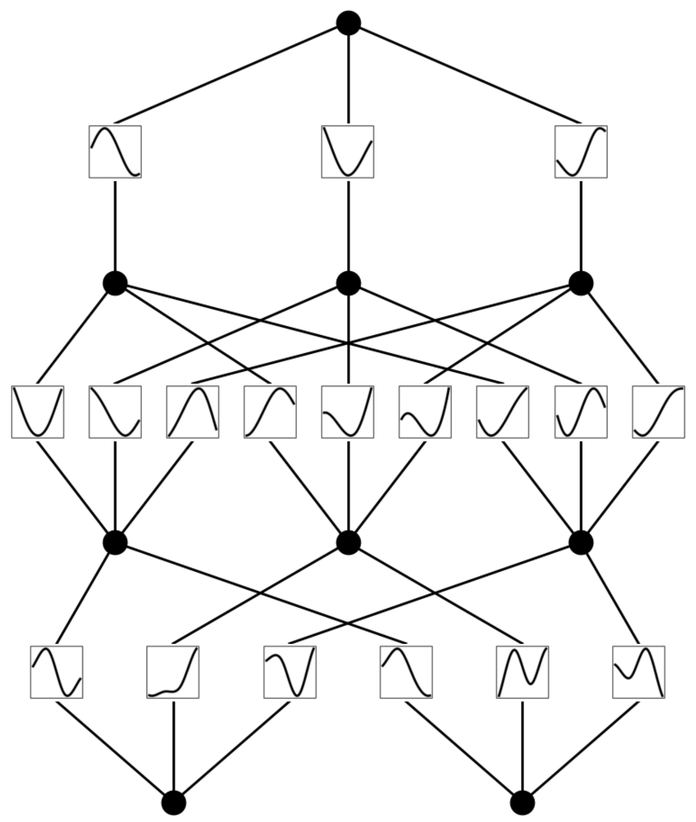
For a function (n dimensions), in general (order k splines on uniform grids):

$$\ell \propto N^{-(k+1)/n}$$

For a function (n dimensions) smoothly represented as a KAN*:

$$\ell \propto \text{poly}(n) N^{-(k+1)}$$

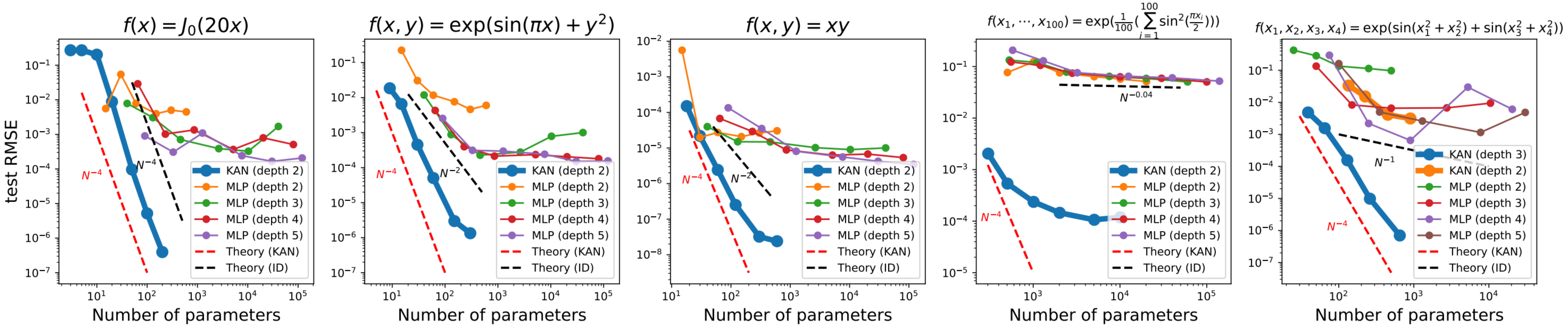
Which is equivalent to $n = 1$, because of KART.



*Informally (Lai-Shen 2021) such functions are dense in C[0,1]

Symbolic formulas

$$\ell \propto N^{-(k+1)}, k = 3$$



KANs and MLPs represent each other

- For KANs with only k-th order B-spline nonlinearity, we have
- ReLU-k MLP with width W, depth L can be represented by k-KAN with width W, depth 2L, grid size 2
- k-KAN with width W, depth L, grid size G can be represented by ReLU-k MLP with width (G+2k+1)W^2, depth 2L
- O(G^2W^4L) parameter count for MLP, O(GW^2L) for KAN in this formulation
- Sharp if we restrict the depth of MLP

Grid Size matters

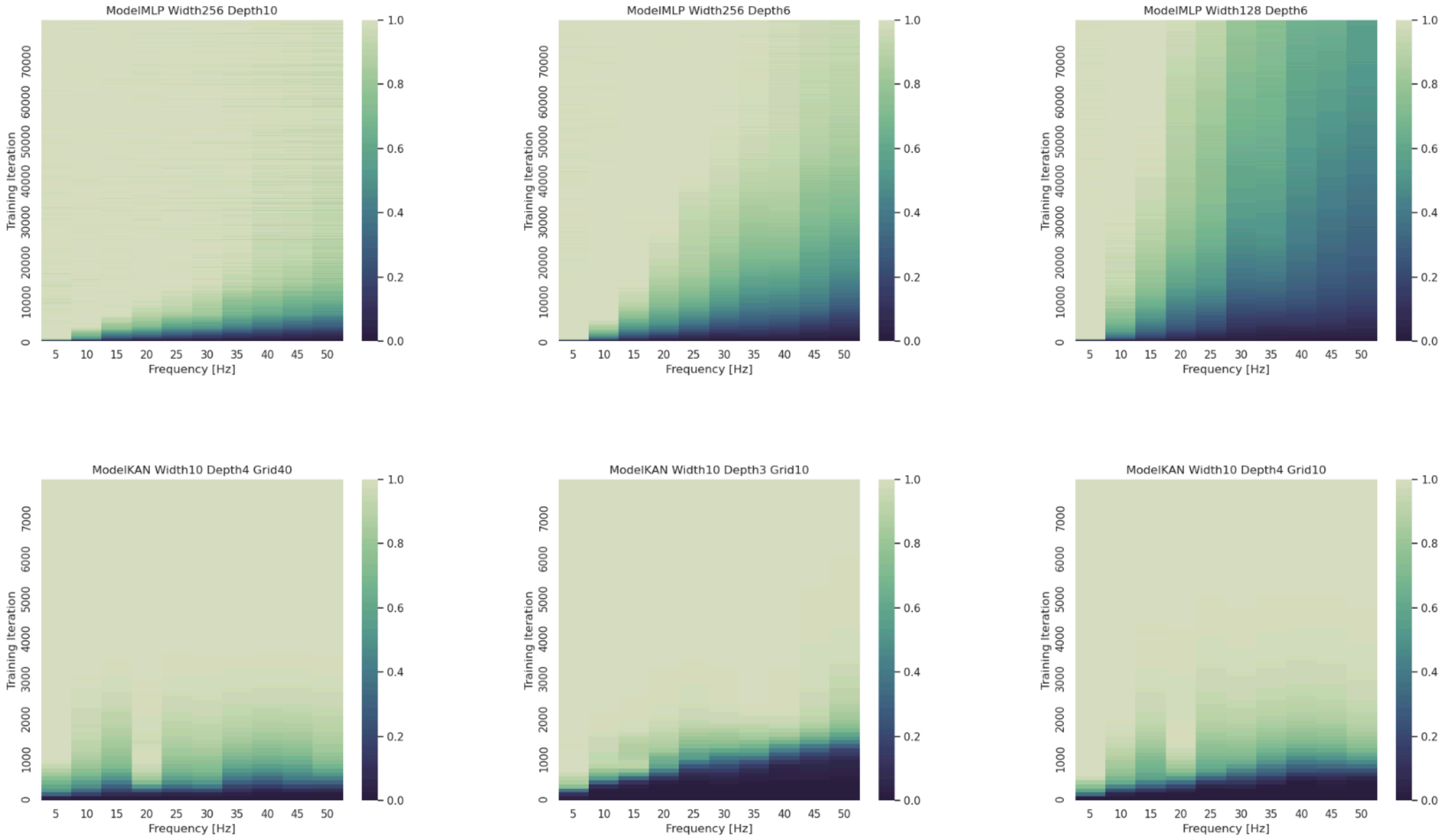
KANs & MLPs for high frequency tasks

- Grid refinement of splines (no spectral bias)
- Compositional structure of nns (spectral bias)
- Example of regression of 1D waves of different frequencies
- Example of regression of Gaussian random field with different scales
- Example of 1D Poisson equation with high frequency by deep Ritz method
- By default MLP: width 256, depth 6; KAN: width 10, depth 2, grid size 20

KANs suffer less from spectral bias, but could overfit!

1D Waves Regression

$$f(x) = \sum A_i \sin\left(2\pi k_i z + \varphi_i\right), \quad k = (5,10,\cdots,45,50)\,.$$



$$-u_{xx} - u_{yy} = f \text{ in } [0,1]^2, \; u = 0 \text{ on } \partial[0,1]^2 .$$

$$f = 2\pi^2 \sin(\pi x)\sin(\pi y) + 2\pi^2 k \sin(k\pi x)\sin(k\pi y),$$

$$u = \sin(\pi x)\sin(\pi y) + \frac{1}{k}\sin(k\pi x)\sin(k\pi y) .$$