In-Distribution Generalization r=1r=2r=5 $10^{2}$  $10^{0}$ MSE  $10^{-2}$  $10^{-4}$  $10^{2}$  $10^{1}$ MSE  $10^{0}$  $10^{-1}$ 26  $\frac{1}{2^{6}}$  $2^{11}$ 27  $2^{8}$ 29  $2^{10}$   $2^{11}$  $2^{7}$  $2^{6}$  $2^{10}$  $2^8$  $2^{9}$  $2^{i0}$ Number of training samples (n) Number of training samples (*n*) Number of training samples (n)  $\sigma = 0$ with linear layers  $\sigma = 0.25$  $\sigma^2$ , irreducible error

without linear layers