Train MSE During Training across Architectures n = 64n = 256n = 128 10^{2} 10^{-2} 10^{-6} n = 512n = 1024n = 2048 10^{2} 10^{-2} Marile Contract Child 10^{-6} 30k 45k 60k 30k 60k 15k 15k 45k 15k 30k 45k 60k **Epochs Epochs Epochs** $\sigma = 0$ Linear Layers then ReLU $\sigma = 0.25$ Deep ReLU Shallow