



Figure 1: Illustrations of SWA and SGD with a Preactivation ResNet-164 on CIFAR-100<sup>1</sup>. **Left:** test error surface for three FGE samples and the corresponding SWA solution (averaging in weight space). **Middle and Right:** test error and train loss surfaces showing the weights proposed by SGD (at convergence) and SWA, starting from the same initialization of SGD after 125 training epochs.