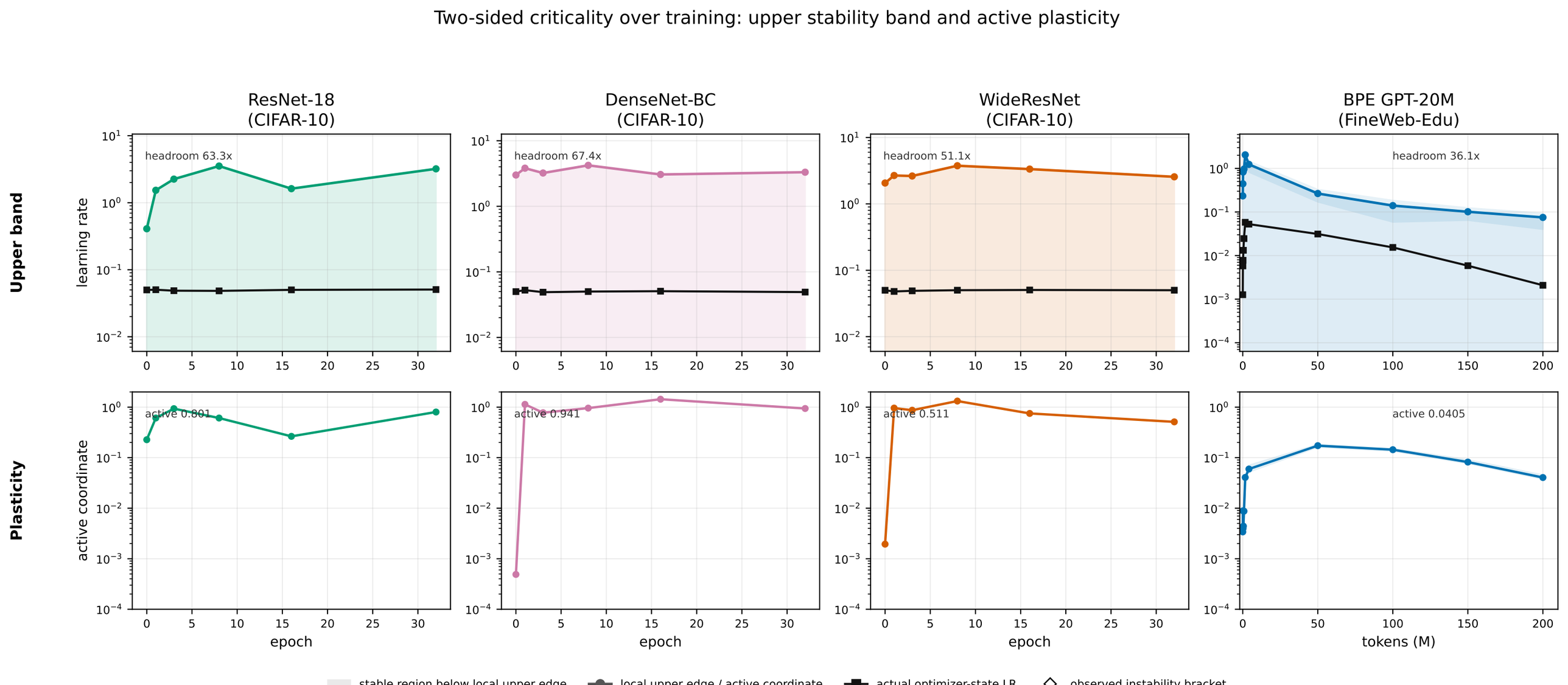
Two-sided criticality over training: upper stability band and active plasticity
ResNet-18
(CIFAR-10)
DenseNet-BC
(CIFAR-10)
WideResNet
(CIFAR-10)
BPE GPT-20M
(FineWeb-Edu)
Upper band
Plasticity
learning rate
active coordinate
headroom 63.3x
headroom 67.4x
headroom 51.1x
headroom 36.1x
active 0.801
active 0.941
active 0.511
active 0.0405
epoch
epoch
epoch
tokens (M)
stable region below local upper edge
local upper edge / active coordinate
actual optimizer-state LR
observed instability bracket