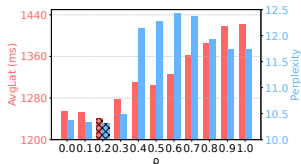
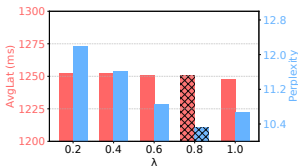


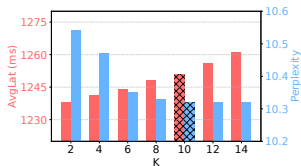
(a) Pruning Compensation Coefficient β



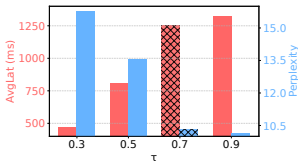
(b) Device Capacity Coefficient ρ



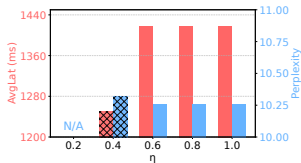
(c) Gradient Smoothing Factor λ



(d) Historical Gradient Length K



(e) Head Pruning Threshold τ



(f) Layer Pruning Threshold η