

1.82

1.81

1.80

1.79

1.78

1.77

LLaMA 13B (width 5120 depth 40)

other $lpha_0$ 0.15 1.829 1.769 1.790 1.767 1.770

1.818 1.781 1.776 1.779 1.791

0.2

0.25 1.825 1.787 1.790 1.791 1.781

attention block α_0

0.2 0.4 0.6 8.0 1.0