

Val Loss vs. GPU VRAM-Hours for DNC

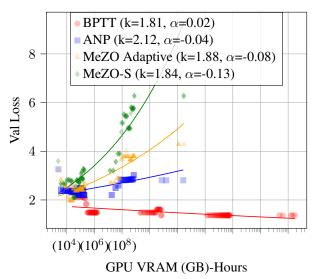


Figure 2. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. Wall Clock Time for DNC

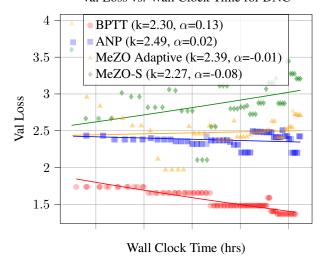


Figure 4. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. GPU VRAM for DNC

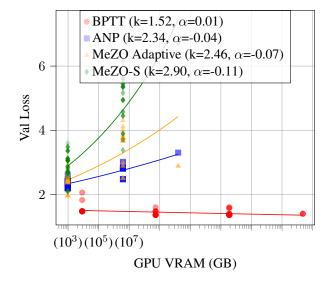


Figure 3. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. GPU VRAM-Hours for NTM

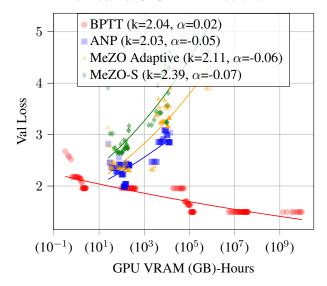


Figure 5. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. GPU VRAM for NTM BPTT (k=1.90, α =0.02) **ANP** (k=2.36, α =-0.05) MeZO Adaptive (k=2.54, α =-0.08) MeZO-S (k=3.18, α =-0.06) Val Loss (10^3) (10^7) (10^{-1}) (10^1) (10^5) GPU VRAM (GB)

Figure 6. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. GPU VRAM-Hours for LSTM

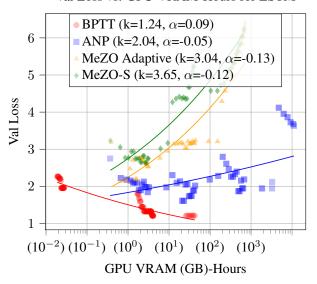


Figure 8. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. Wall Clock Time for NTM

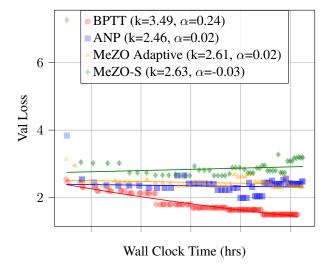


Figure 7. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. GPU VRAM for LSTM

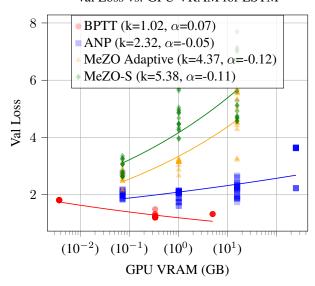


Figure 9. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. Wall Clock Time for LSTM

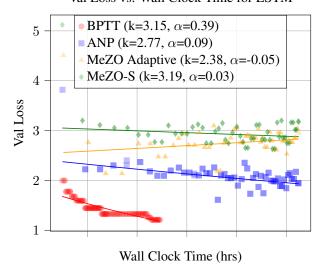


Figure 10. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. GPU VRAM for MAMBA

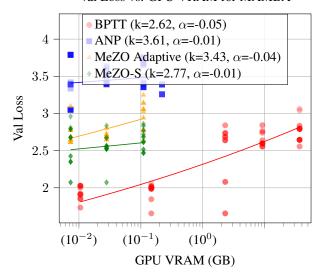


Figure 12. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. GPU VRAM-Hours for MAMBA

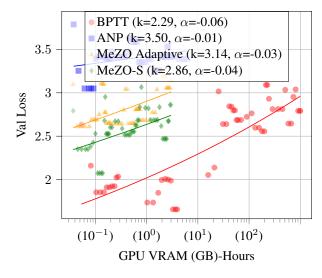


Figure 11. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. Wall Clock Time for MAMBA

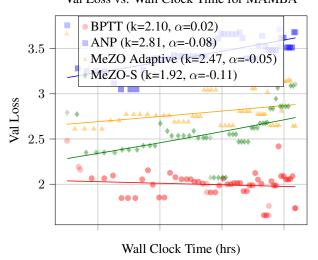
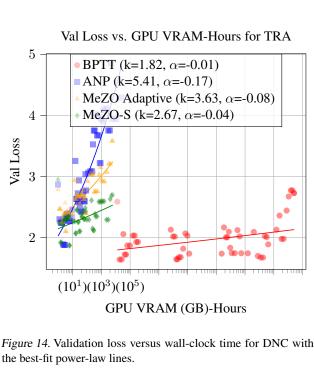


Figure 13. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.



the best-fit power-law lines.

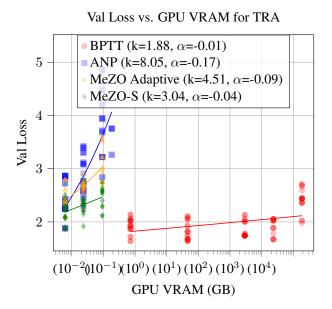


Figure 15. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.

Val Loss vs. Wall Clock Time for TRA

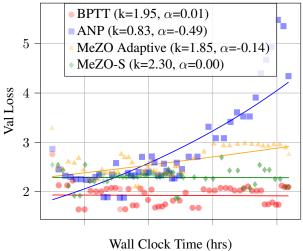


Figure 16. Validation loss versus wall-clock time for DNC with the best-fit power-law lines.