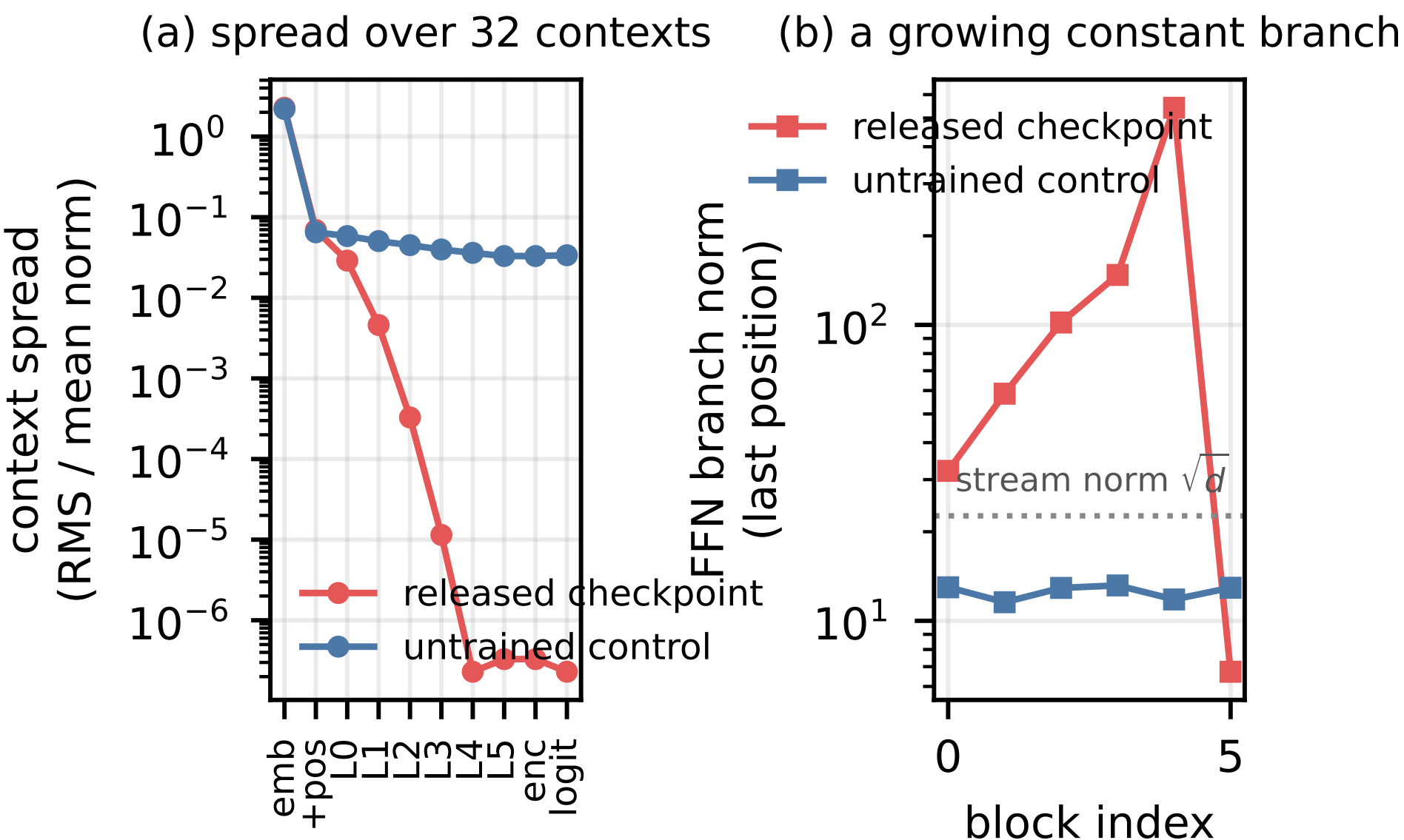

(a) spread over 32 contexts
context spread
(RMS / mean norm)
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
released checkpoint
untrained control
emb
+pos
L0
L1
L2
L3
L4
L5
enc
logit
(b) a growing constant branch
FFN branch norm
(last position)
released checkpoint
untrained control
10^{2}
10^{1}
stream norm $\sqrt{d}$
0
5
block index