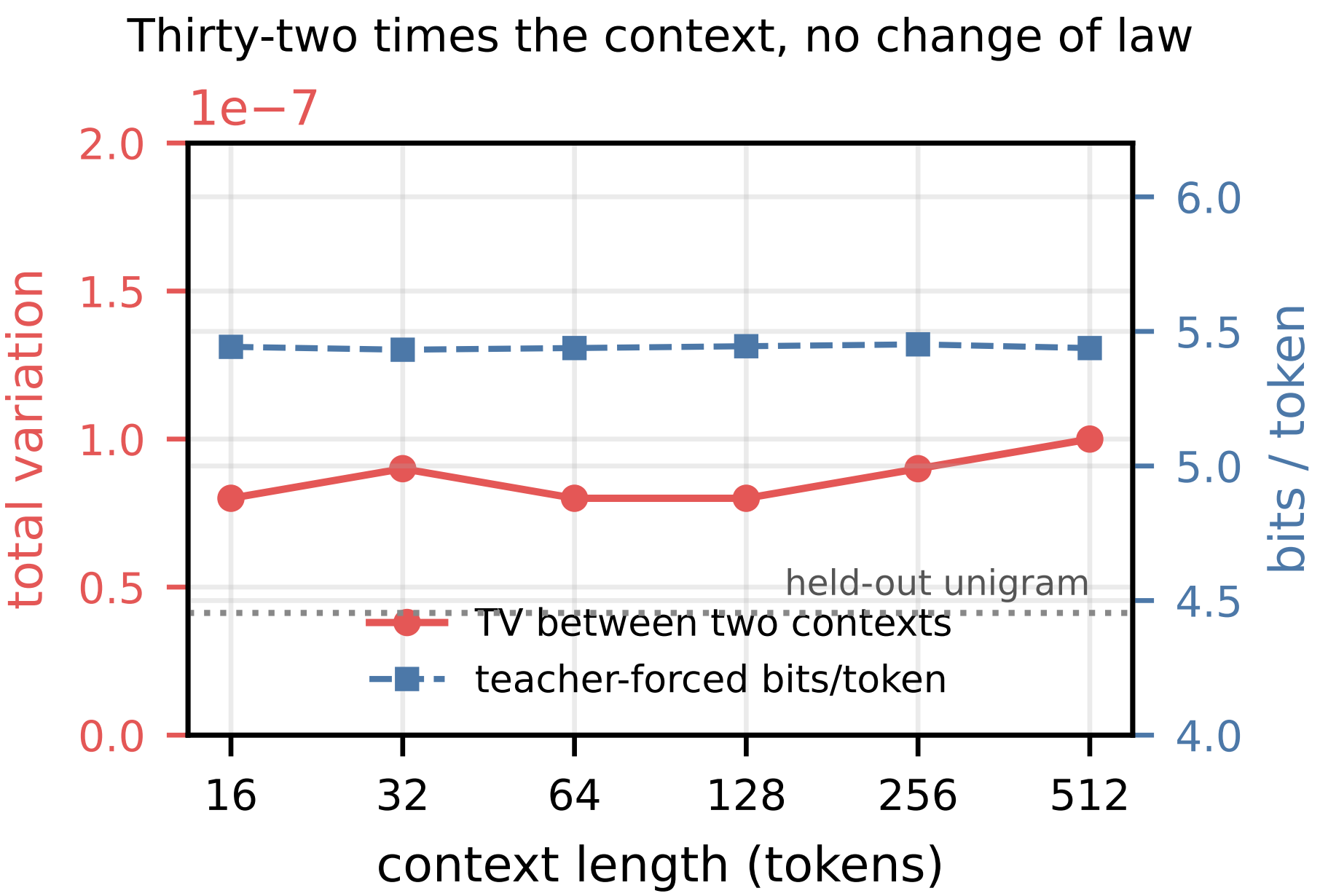

Thirty-two times the context, no change of law
1e−7
2.0
1.5
1.0
0.5
0.0
total variation
6.0
5.5
5.0
4.5
4.0
bits / token
held-out unigram
TV between two contexts
teacher-forced bits/token
16
32
64
128
256
512
context length (tokens)