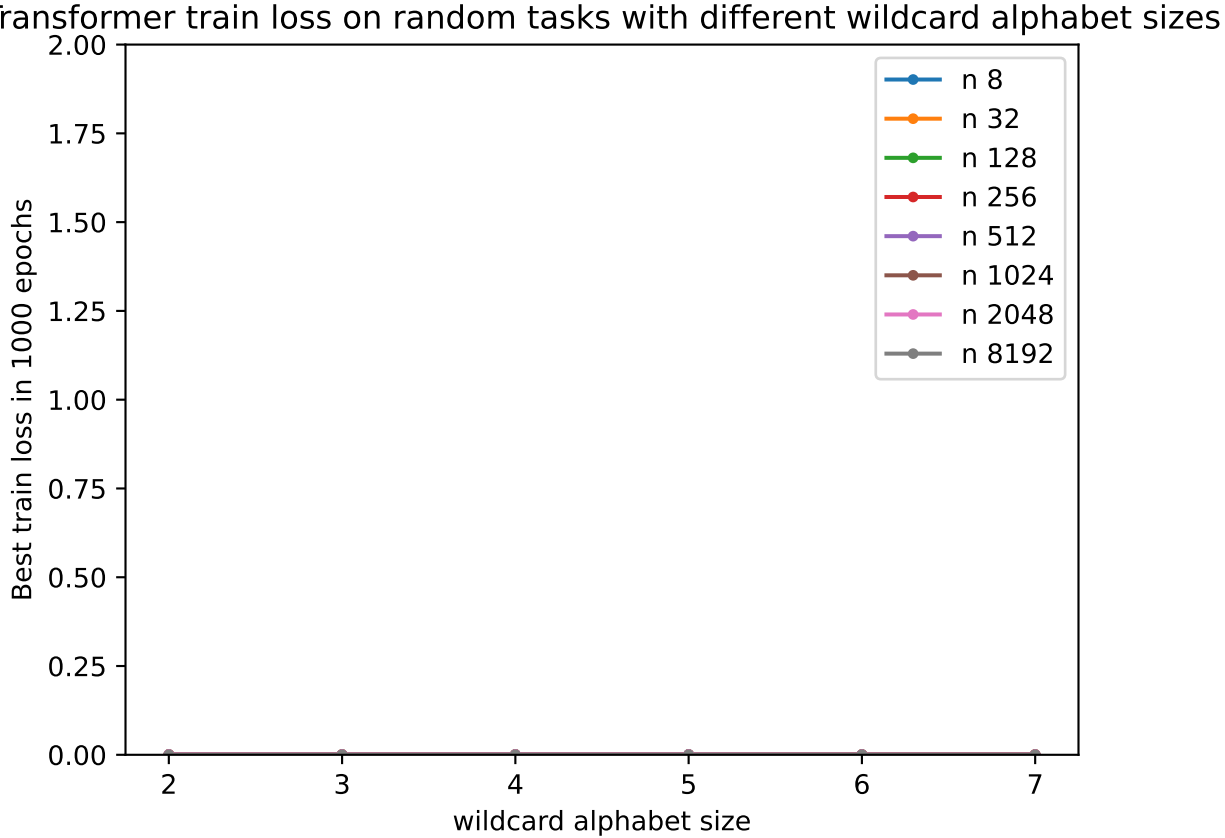
Transformer train loss on random tasks with different wildcard alphabet sizes
Best train loss in 1000 epochs
wildcard alphabet size
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
n 8
n 32
n 128
n 256
n 512
n 1024
n 2048
n 8192