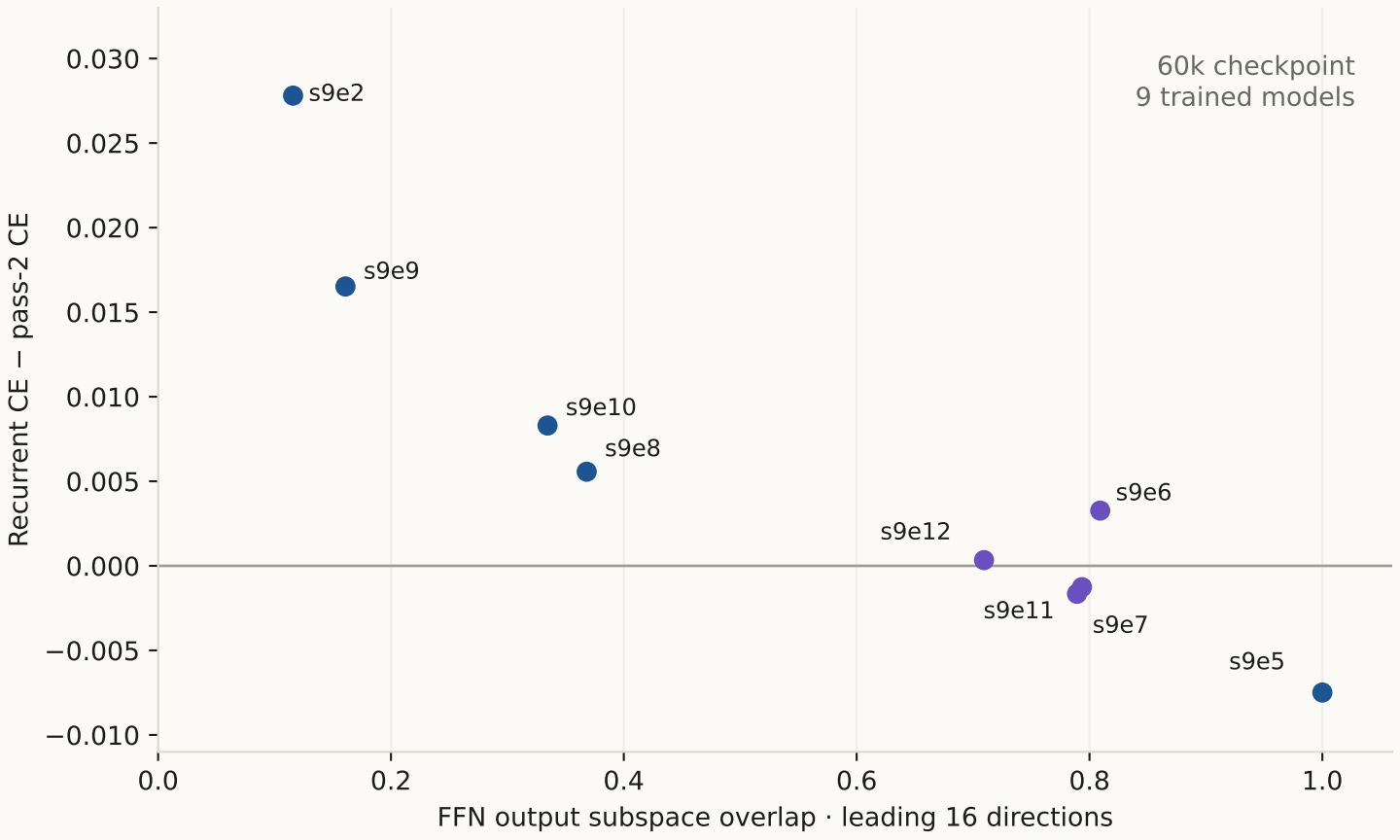

60k checkpoint
9 trained models
s9e2
s9e9
s9e10
s9e8
s9e6
s9e12
s9e11
s9e7
s9e5
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−0.005
−0.010
0.0
0.2
0.4
0.6
0.8
1.0
Recurrent CE − pass-2 CE
FFN output subspace overlap · leading 16 directions