Hyperparameter	Small	Large
Layers L	2	4
Heads H	4	8
Model Dim d	256	512
MLP Dim d_{ff}	1024	2048
Determinants K	1	2
MCMC walkers N_w	1024	2048
MCMC steps / iter	10	10
Learning rate	2×10^{-4}	1×10^{-4}
Total parameters		