validation loss evolution @ walker\_benchmark  $10^{1}$ memory augmented transformer differentiable neural computer unitary rnn matrix exponential unitary rnn transformer recurrent\_network\_attention\_transformer  $6 \times 10^{0}$ recurrent\_network\_augmented\_transformer gru  $5 \times 10^{0}$ neural circuit policies mean squared error ct rnn ct\_gru  $4 \times 10^{0}$ ode Istm unitary ncp  $3 \times 10^{0}$  $2 \times 10^{0}$ 10 20 30 50 60 70 40 epochs