Transformer test loss vs. learning rate and embedding dimension, at n = 10242.00 Ir 0.0001 Ir 0.0003 1.75 -Ir 0.001 Ir 0.003 1.50 -- Ir 0.01 1.25 -Test loss 1.00 -0.75 -0.50 0.25 0.00 10<sup>2</sup>

embedding dimension