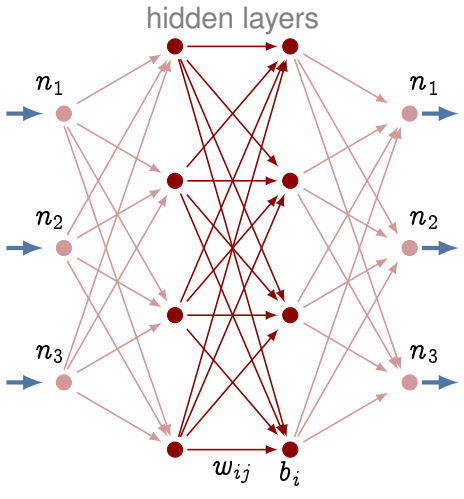
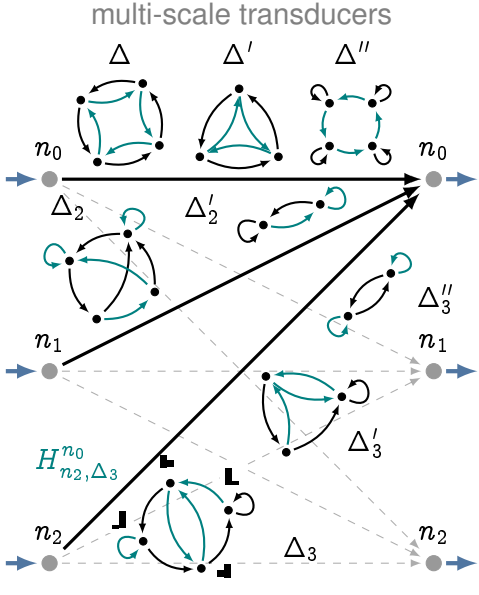
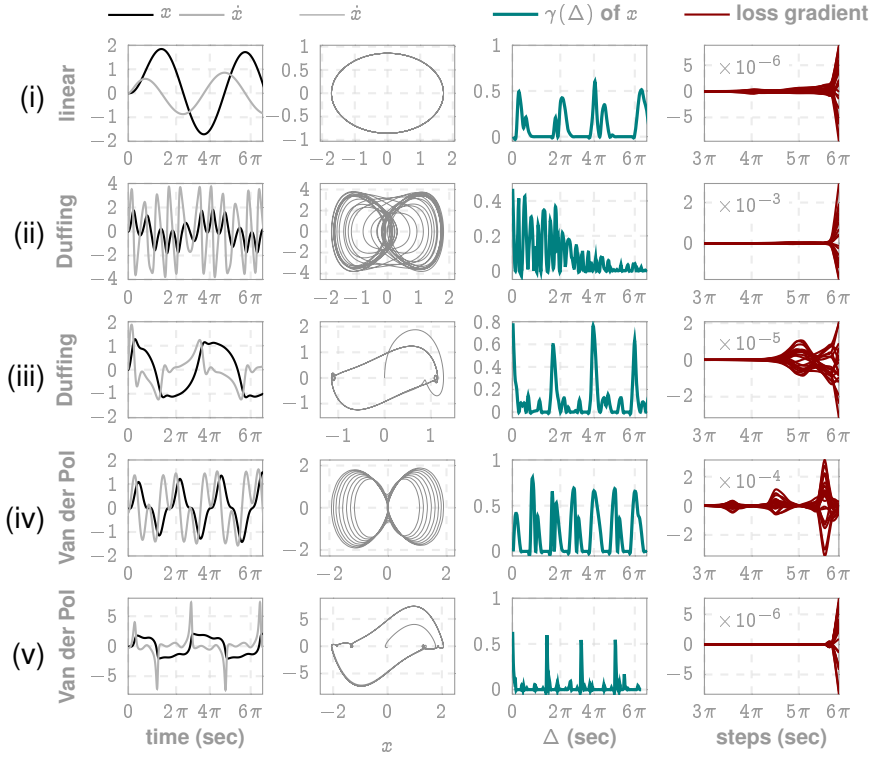
a. Standard Deep Learning Architecture



b. NN-free Deep Learning Architecture



c. Coeff. of Causality (γ) in F-nets vs gradient decay in back-propagation)



d. Linear Combination of Inferred Local Activations with Memory

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \leqq 0}} \omega^{s}_{r,\Delta+j} \mathbb{H}^{s}_{r,\Delta+j} \left(s^{-\infty}_{t-j} \right)$$

Learned activation model from source s to target r with delay $\Delta + j$

target prediction Δ steps from current time

source data upto j steps before current time