

## Neural Network Arena: Investigating Long-Term Dependencies in Deep Models

Hannes Brantner  
Computer Engineering

**TU Wien Informatics**  
Institute of Computer Engineering  
Cyber-Physical Systems Group  
Supervisor: Univ.Prof. Dipl.-Ing. Dr.rer.nat. Radu Grosu  
Contact: e01614466@student.tuwien.ac.at

GitHub repository



### Problem Statement

- implementation of a reusable benchmark suite to compare machine learning models used for sequence modeling
- benchmark suite should test the models for their capabilities to capture long-term dependencies and to model physical systems
- selection of state-of-the-art models should be implemented as well as possible outlined improvements
- thoroughful comparison of all implemented models using the benchmark suite
- all implemented models are Transformer or RNN (Recurrent Neural Network) architectures
- proof-of-concept design and implementation of a continuous-time memory cell architecture based on LTC Networks

### Methodology

- extensive literature review in the domain of sequence modeling
- implementation of the benchmark suite and all the implemented models
- the benchmark suite was invoked three times on all models and the statistics of the invocation output were interpreted

### Motivation

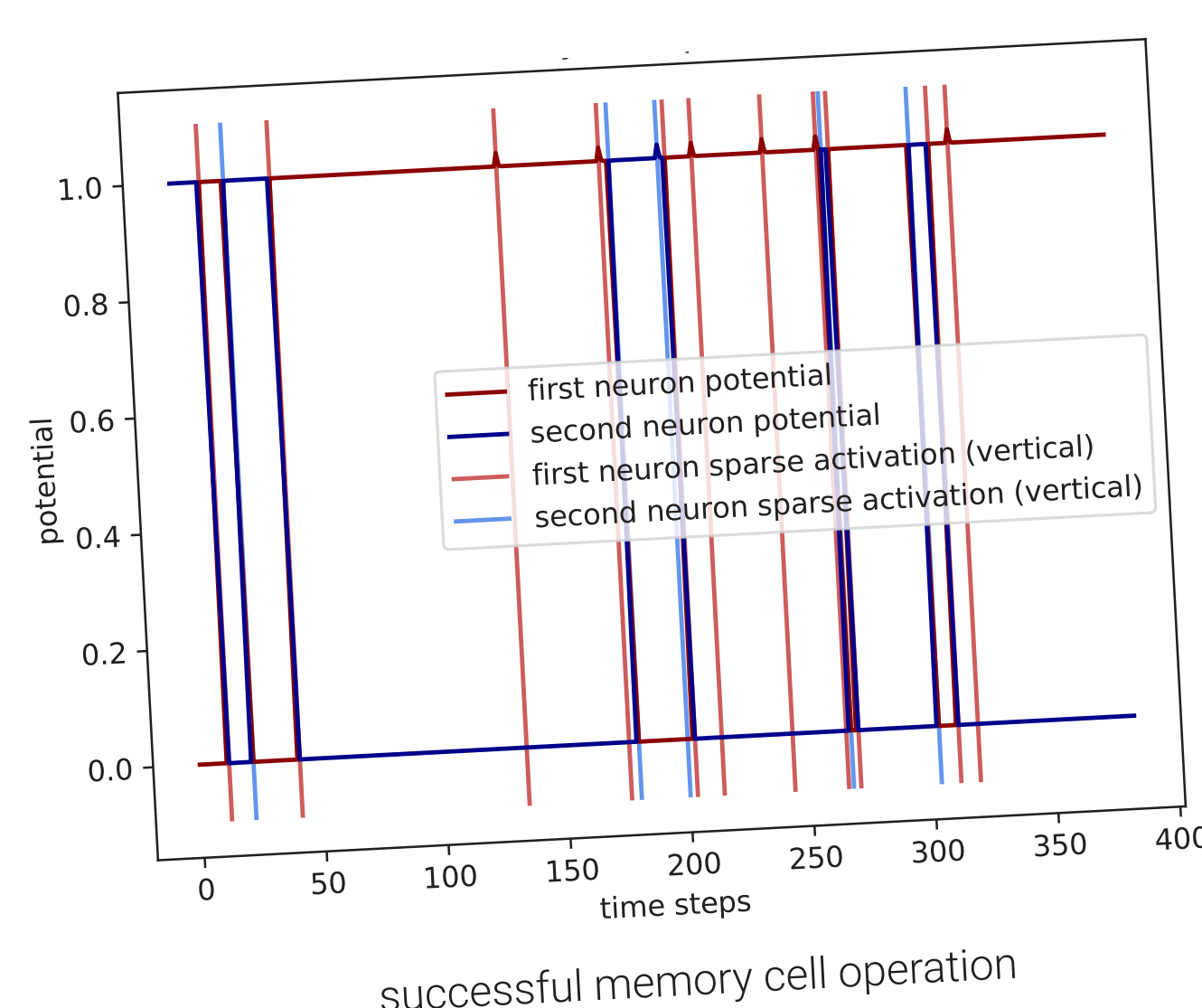
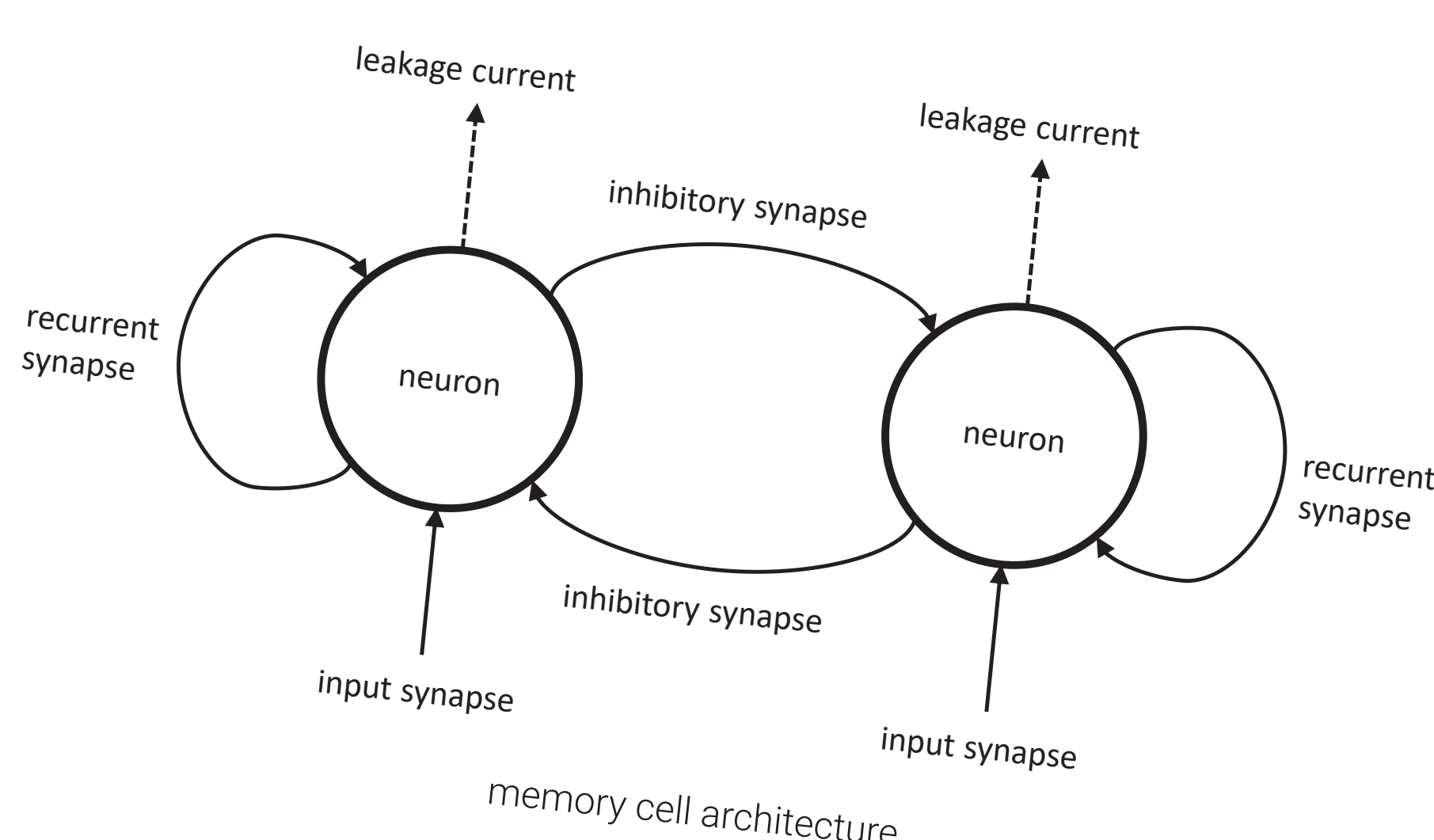
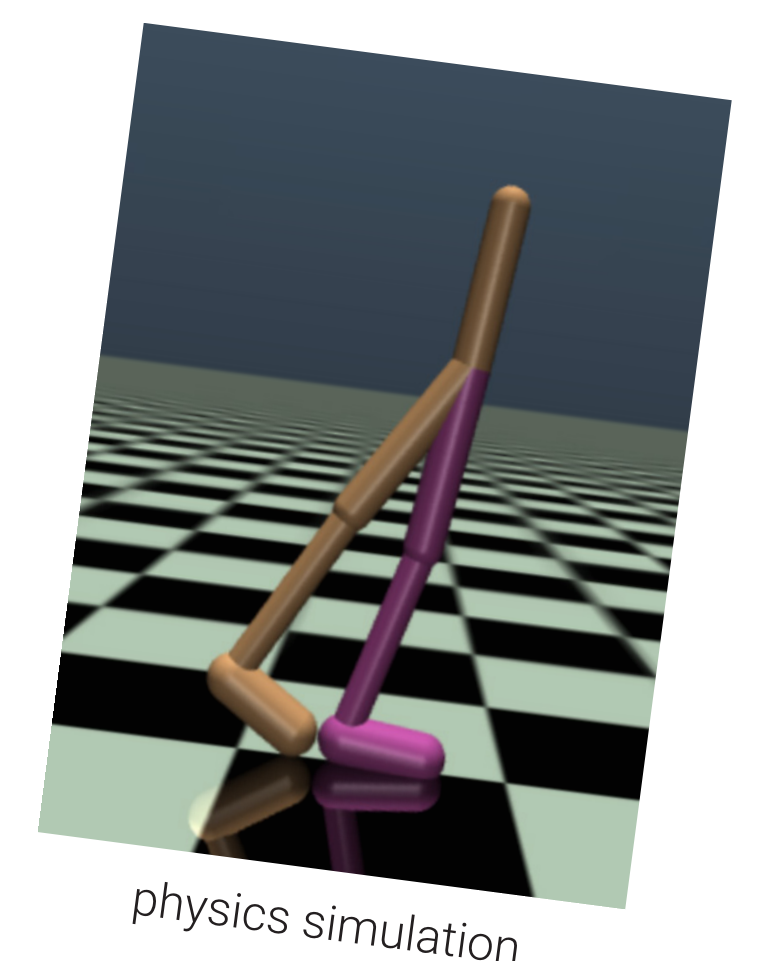
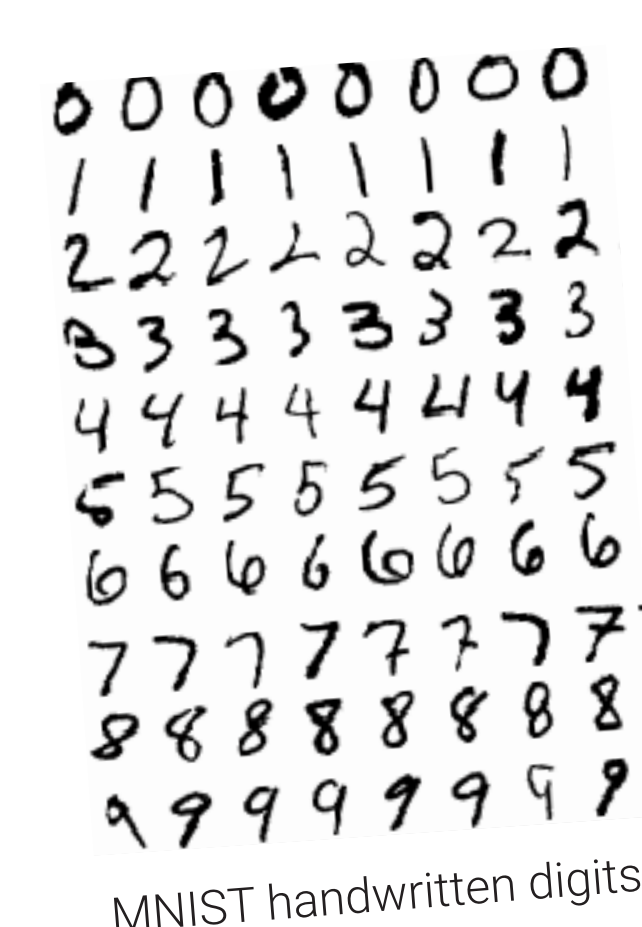
- provide an objective comparison and overview of all implemented models on various sequence modeling tasks
- especially RNN architectures have difficulties of capturing long-term dependencies when being learned by gradient descent
- investigate which mechanism works best in RNN architectures to counteract this difficulty

### Results

- state-of-the-art in efficient unitary matrix parameterization was improved by using an approximated matrix exponential
- continous-time memory cell architecture was successfully trained to store sparse activations
- positional encoding used in Transformer architectures shows deficiencies in tasks where exact positional information is required
- the newly introduced Memory Augmented Transformer architecture shows promising results in some tasks

### Implemented Models

- LSTM (Long Short-Term Memory)
- GRU (Gated Recurrent Unit)
- CT-RNN (Continuous-Time RNN)
- CT-GRU (Continuous-Time GRU)
- ODE-LSTM (Ordinary Differential Equation LSTM)
- NCP (Neural Circuit Policies)
- Unitary RNN
- Matrix Exponential Unitary RNN **new**
- Unitary NCP **new**
- Transformer
- Recurrent Network Augmented Transformer **new**
- Recurrent Network Attention Transformer **new**
- Memory Augmented Transformer **new**
- Differentiable Neural Computer
- Memory Cell **new**



### Benchmark Suite Tasks

- Activity Benchmark - human activity classification of inertial sensor measurement data sequences
- Add Benchmark - adding up two marked numbers in a very long number sequence
- Walker Benchmark - predict the next state of a physics simulation given a sequence of previous simulation states
- Memory Benchmark - store a seen category exactly and recall it after seeing a sequence of irrelevant filler symbols
- MNIST Benchmark - digit classification using a sequence of MNIST handwritten digit image chunks
- Cell Benchmark - validates if sparse activations are correctly stored in the time-continuous memory cell architecture