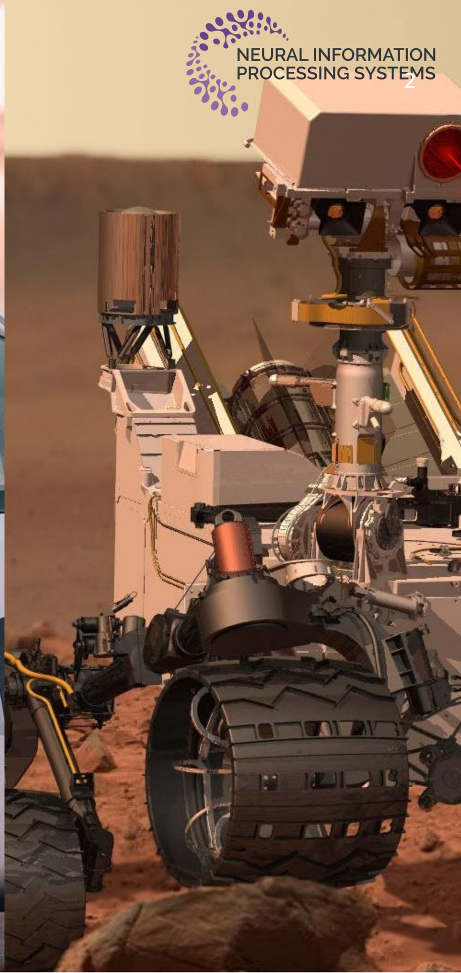
Adaptive Surrogate Gradients for Sequential Reinforcement Learning in Spiking Neural Networks

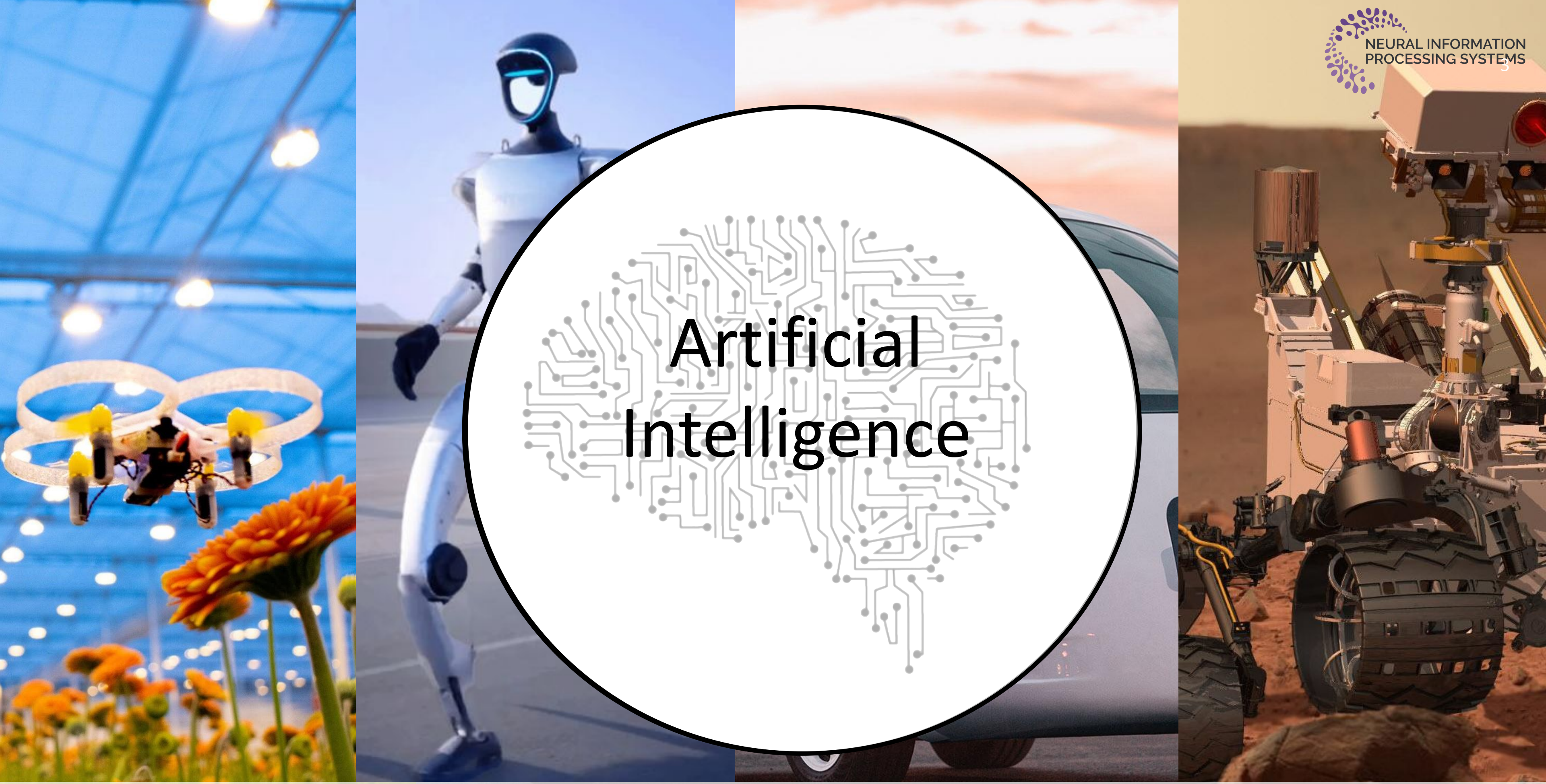
Korneel Van den Berghe[1,2], Stein Stroobants[1],
Vijay Janapa Reddi[2], G.C.H.E. de Croon[1]

[1]TU Delft, [2]Harvard University



Autonomous robot market: 7.5B$ in 2024 to ~34B$ by 2035.

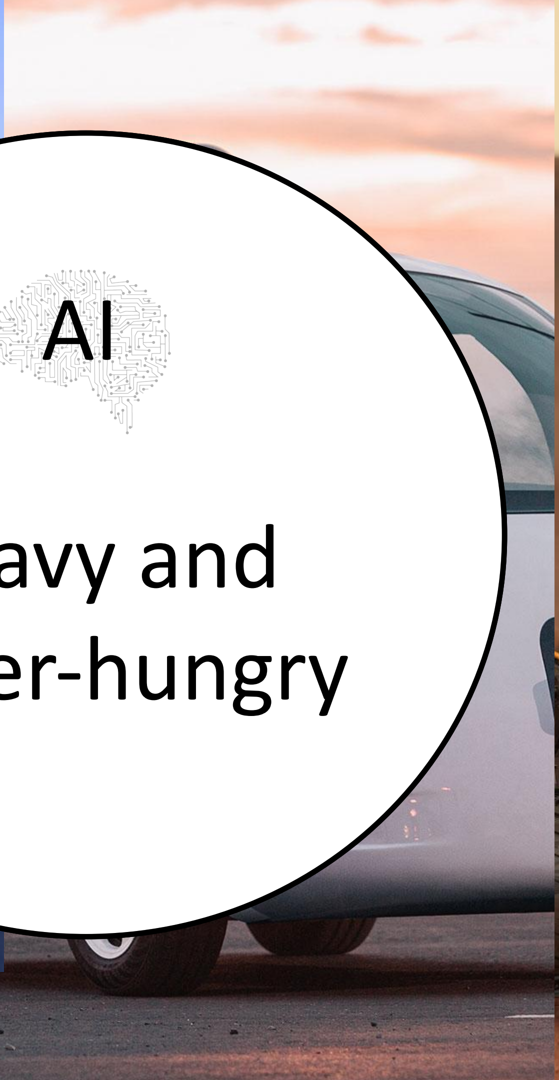
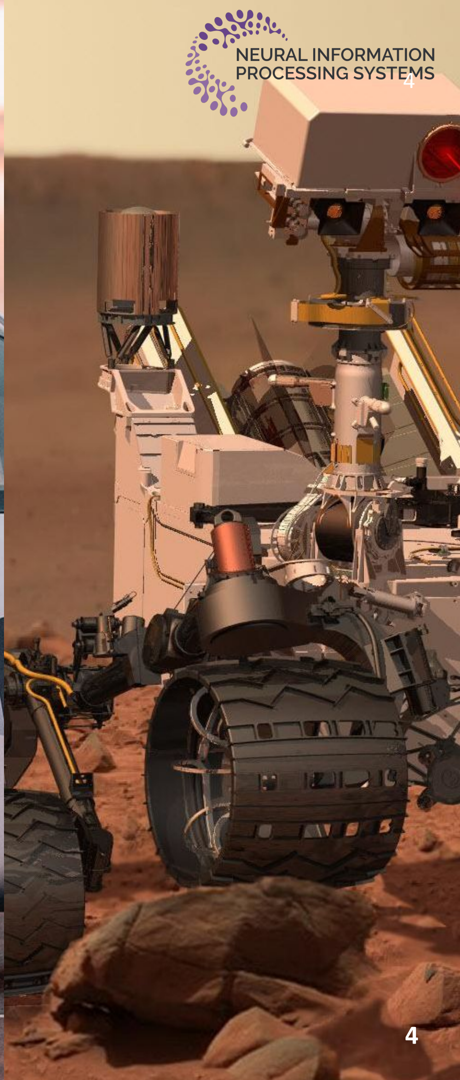
(https://www.marketresearchfuture.com/reports/autonomous-robots-market-6912)



Autonomous robot market: 7.5B$ in 2024 to ~34B$ by 2035.

(https://www.marketresearchfuture.com/reports/autonomous-robots-market-6912)



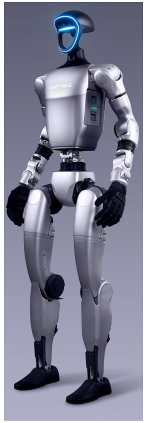
AI

Heavy and power-hungry

Energy and Compute budgets

Humanoids:

Unitree G1



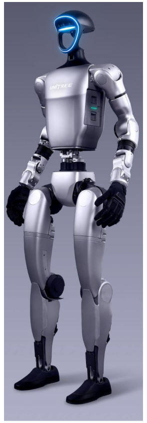
8-core CPU
226.8 Wh battery

https://www.unitree.com/g1

Energy and Compute budgets

Humanoids:

Unitree G1



8-core CPU
226.8 Wh battery

https://www.unitree.com/g1

Quadrupeds:

Boston Dynamics Spot

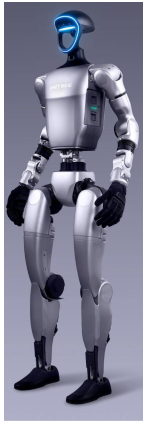
Nvidia Xavier
564Wh

https://bostondynamics.com/products/spot/

Energy and Compute budgets

Humanoids:

Unitree G1



8-core CPU
226.8 Wh battery

https://www.unitree.com/g1

Quadrupeds:

Boston Dynamics Spot

Nvidia Xavier
564Wh

https://bostondynamics.com/products/spot/

AVs:

Tesla FSD

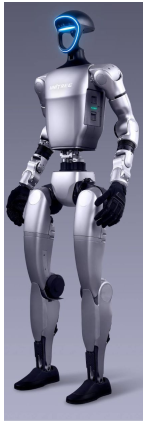
Tesla FSD chip (72 TOPS), similar to Nvidia Orin NX

https://www.tesla.com/

Energy and Compute budgets

Humanoids:

Unitree G1



8-core CPU
226.8 Wh battery

https://www.unitree.com/g1

Quadrupeds:

Boston Dynamics Spot

Nvidia Xavier
564Wh

https://bostondynamics.com/products/spot/

AVs:

Tesla FSD

Tesla FSD chip (72 TOPS), similar to Nvidia Orin NX

https://www.tesla.com/

TUDelft

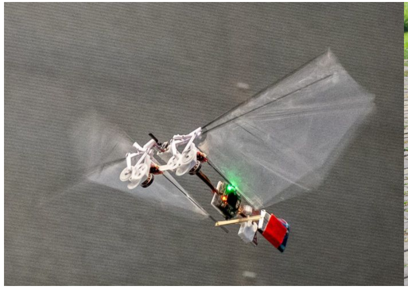
Limited compute and energy available!

NEURAL INFORMATION PROCESSING SYSTEMS

MAVLab: *Autonomous flight of Micro Air Vehicles*

Micro Air Vehicles:

- Limited energy budget
- Limited compute budget
- Limited payload capacity



Delfly

DelftaCopter

Crazyflie

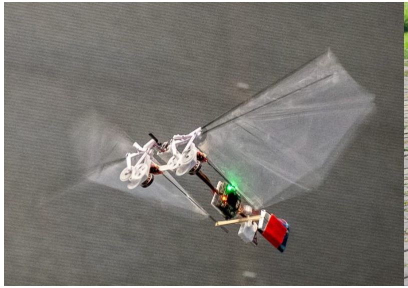
Racing Drone

MAVLab: *Autonomous flight of Micro Air Vehicles*

Micro Air Vehicles:

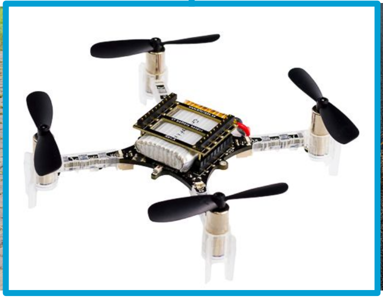
- Limited energy budget
- Limited compute budget
- Limited payload capacity

- 27g
- 0.925Wh
- Cortex-M4, 168MHz
- 192kb RAM



Delfly

DelftaCopter



Crazyflie

Racing Drone

TUDelft

NEURAL INFORMATION PROCESSING SYSTEMS

Autonomous drones

TU Delft wins the AI-human drone race tournament in Abu Dhabi, April 2024, beating three human world-champions in a row in a knockout tournament!

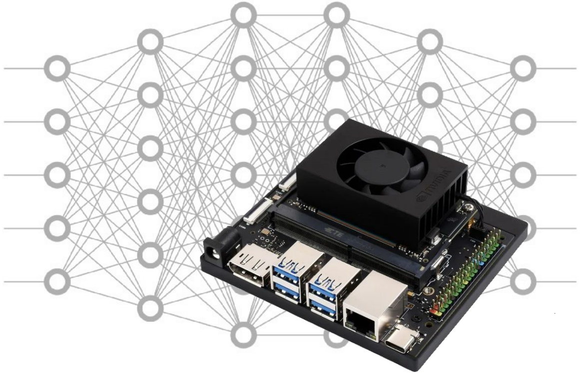
State of the art in AI for drones

30 cm diameter, 700 g (flight 200 W)

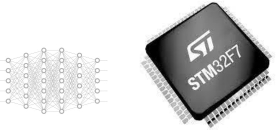
NVidia Jetson Orin NX (25 W, 50 g)

- GPU (1024 CUDA & 32 tensor cores)
- 8-core ARM Cortex CPU
- 16 Gb of memory

Vision



Control



TU Delft

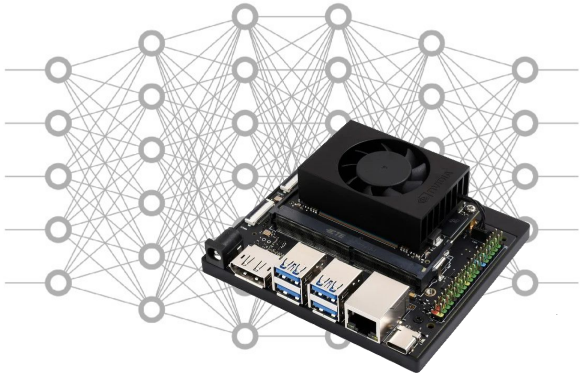
NEURAL INFORMATION PROCESSING SYSTEMS

Processing of High-Dimensional Data

Compute of high-dimensional data is:

- Heavy
- Energy hungry
- Relatively slow

Vision

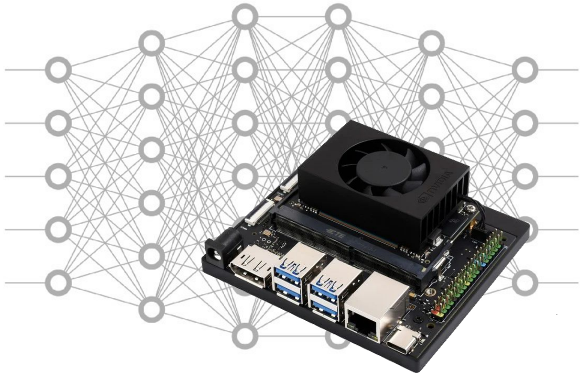


Processing of High-Dimensional Data

Compute of high-dimensional data is:

- Heavy
- Energy hungry
- Relatively slow

Vision

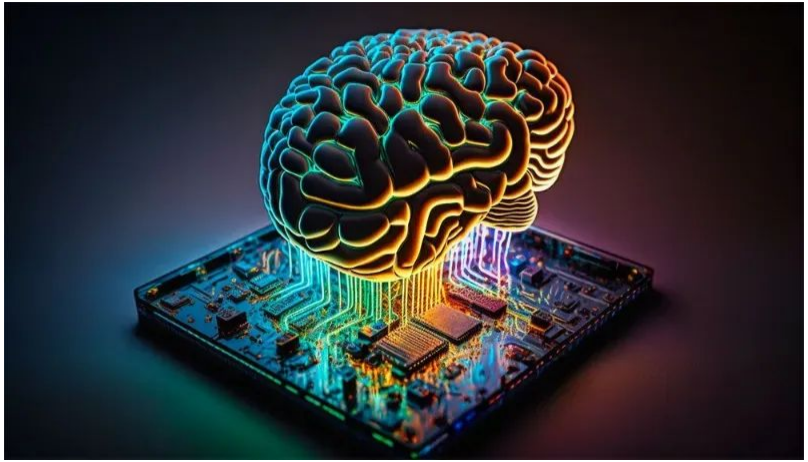


Alternative:

Neuromorphic computing!

Neuromorphic Computing

- Bio-inspired computing paradigm
- Sparse and event-based
- Hardware-software co-design

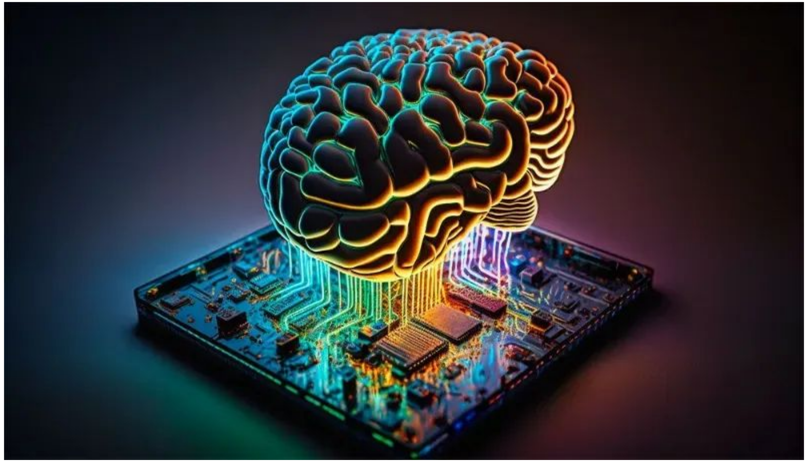


https://www.thedigitalspeaker.com/neuromorphic-computing-hyper-realistic-generative-ai/

Neuromorphic Computing

- Bio-inspired computing paradigm
- Sparse and event-based
- Hardware-software co-design

→ Fast, energy efficient



https://www.thedigitalspeaker.com/neuromorphic-computing-hyper-realistic-generative-ai/

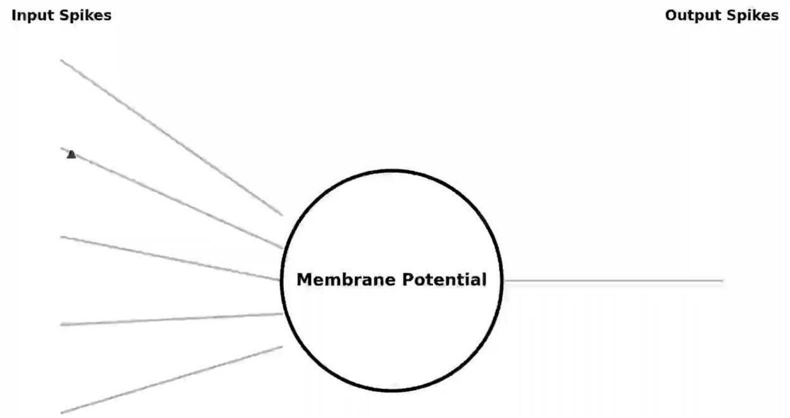
Spiking Neural Networks

Bio-inspired neuron dynamics

- Membrane potential
- Sparse, binary spikes

Non-differentiable spiking function

- Use surrogate gradients



Visualization of a spiking Leaky Integrate-And-Fire (LIF) neuron

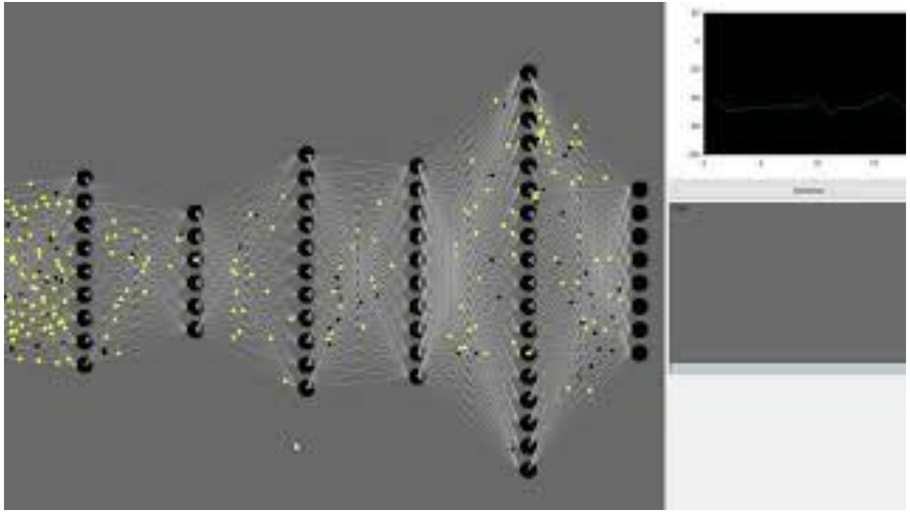
Spiking Neural Networks

Bio-inspired neuron dynamics

- Membrane potential
- Sparse, binary spikes

Non-differentiable spiking function

- Use surrogate gradients



Visualization of a spiking neural network from https://www.youtube.com/watch?v=oG0PTP3ogCA

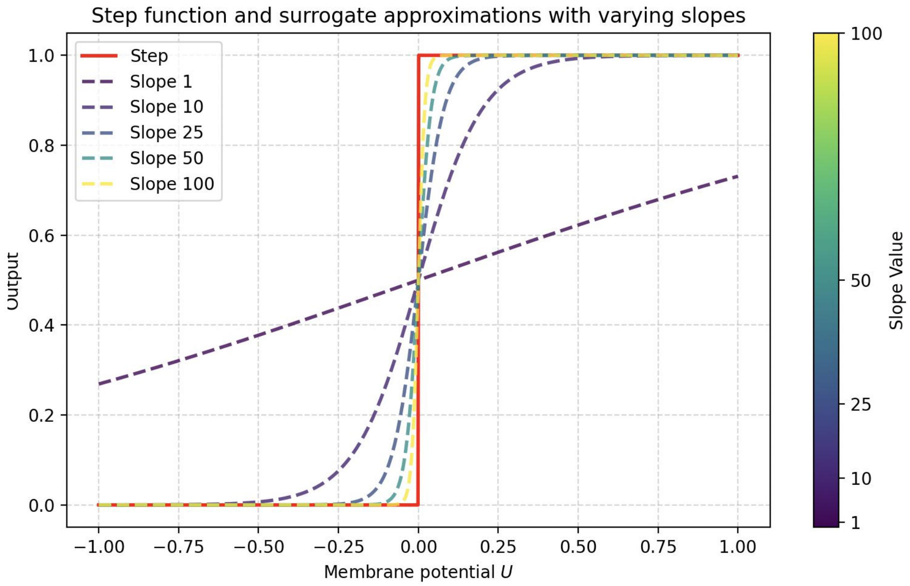
Spiking Neural Networks

Bio-inspired neuron dynamics

- Membrane potential
- Sparse, binary spikes

Non-differentiable spiking function

- Use surrogate gradients



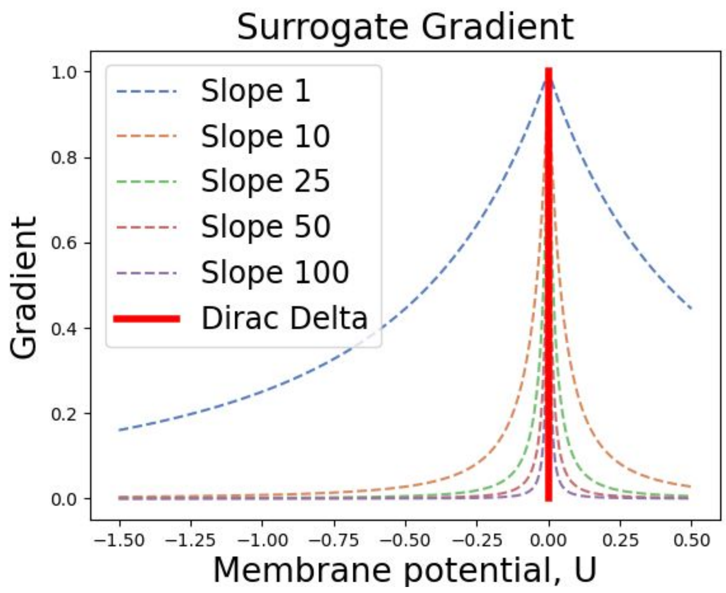
Spiking Neural Networks

Bio-inspired neuron dynamics

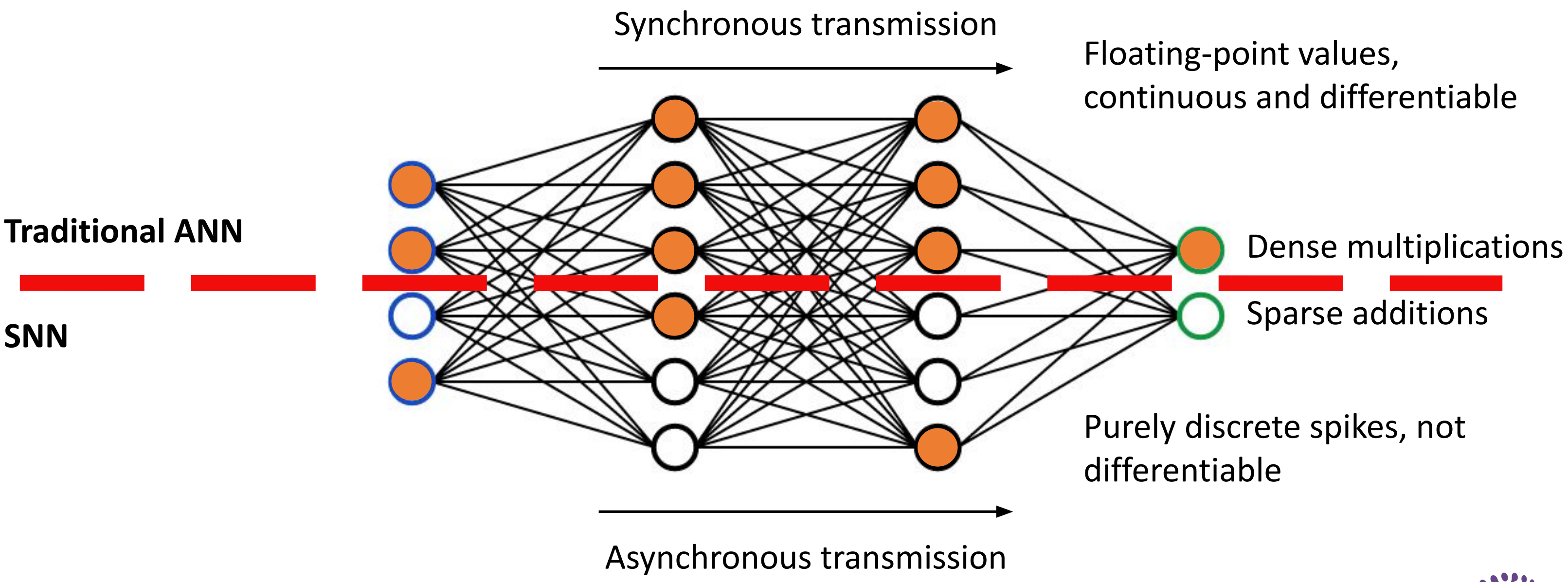
- Membrane potential
- Sparse, binary spikes

Non-differentiable spiking function

- Use surrogate gradients



Differences between ANNs and SNNs



State of the art SNNs

- Training still harder due to more complex neural dynamics
  - Stateful
  - Non-differentiable spiking function
    - Dead and/or saturated neurons
    - Commonly use surrogate gradients
- Accuracy slightly less, but faster and more energy efficient

NEURAL INFORMATION PROCESSING SYSTEMS

Neuromorphic AI for autonomous robots

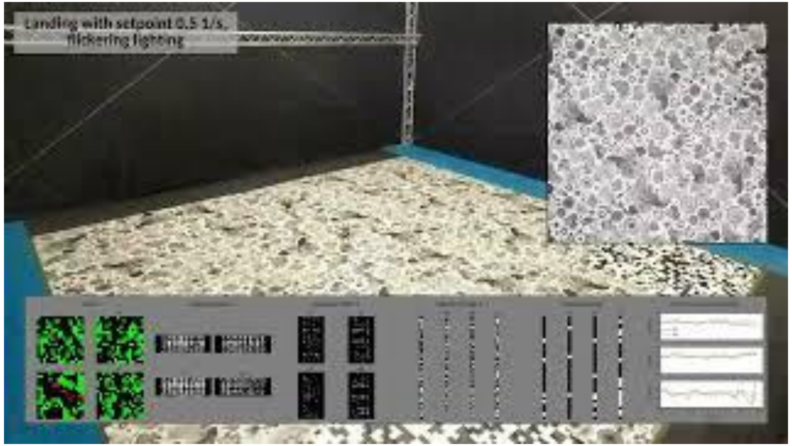
To fully exploit neuromorphic AI the **entire autonomy stack** should run **on a single neuromorphic chip**.

This implies performing all tasks from low-level attitude control to high-level navigation with spiking neural networks.

Neuromorphics at the MAVLab

Towards end-to-end control with SNN:

- Perception network:
  - Optic flow from event camera
  - SNN trained self-supervised
- Control network:
  - SNN trained through evolution
- Merge perception and control SNN
- Sends Attitude commands

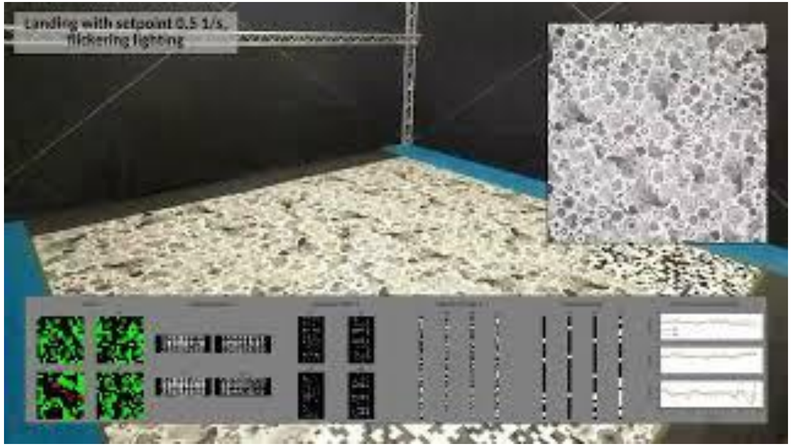


F. Paredes-Vallés et al., Fully neuromorphic vision and control for autonomous drone flight. Science Robotics

Neuromorphics at the MAVLab

Towards end-to-end control with SNN:

- Perception network:
  - Optic flow from event camera
  - SNN trained self-supervised
- Control network:
  - SNN trained through evolution
- Merge perception and control SNN
- Sends Attitude commands



F. Paredes-Vallés et al., Fully neuromorphic vision and control for autonomous drone flight. Science Robotics

Hardware deployment:

Neuromorphic: 0.021mJ/inf, max 411inf/s

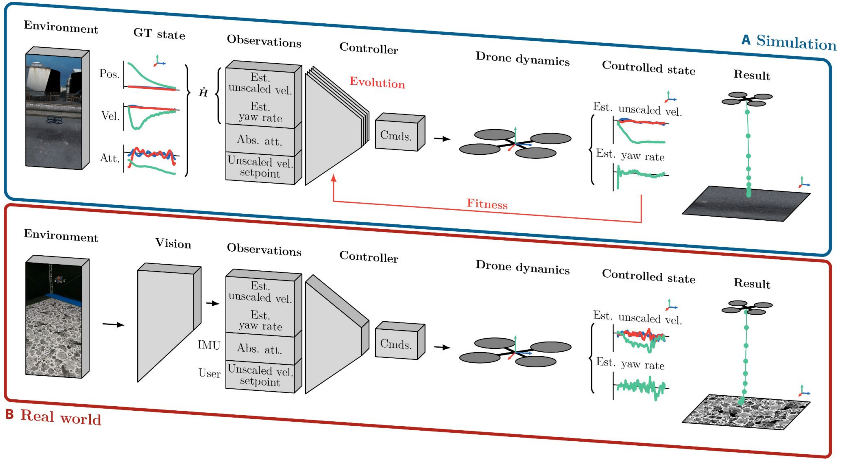
Jetson Nano: 75mJ/inf, max 27inf/s

Neuromorphics at the MAVLab

NEURAL INFORMATION PROCESSING SYSTEMS

Fully neuromorphic vision and control for autonomous drone flight

- Optic flow for ego-motion estimation
- Single layer SNN for control
- Trained through evolution
- No need for external sensors!



Need to merge two nets together

F. Paredes-Vallés et al. ,Fully neuromorphic vision and control for autonomous drone flight.Sci. Robot.9, eadi0591(2024). DOI:10.1126/scirobotics.adi0591

TU Delft

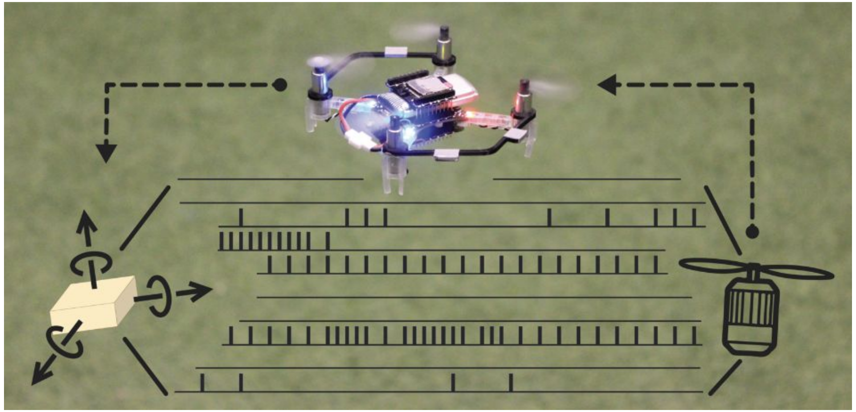
NEURAL INFORMATION PROCESSING SYSTEMS

Neuromorphics at the MAVLab

NEURAL INFORMATION PROCESSING SYSTEMS

Neuromorphic Attitude Estimation and Control

- State-estimation from IMU
- Attitude control
- Trained using Supervised Learning
- No need for external sensors!



Stroobants, S., De Wagter, C., & De Croon, G. C. (2025). Neuromorphic Attitude Estimation and Control. IEEE Robotics and Automation Letters.

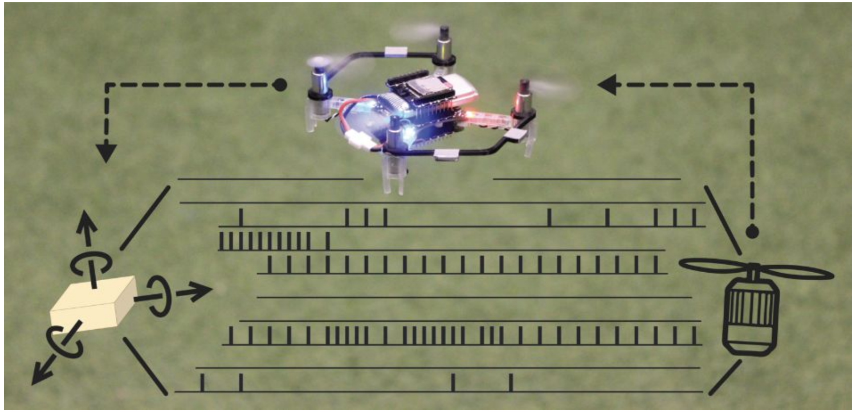
Neuromorphics at the MAVLab

NEURAL INFORMATION PROCESSING SYSTEMS

Neuromorphic Attitude Estimation and Control

- State-estimation from IMU
- Attitude control
- Trained using Supervised Learning
- No need for external sensors!

But: requires gathering dataset
And still merge nets together



Stroobants, S., De Wagter, C., & De Croon, G. C. (2025). Neuromorphic Attitude Estimation and Control. IEEE Robotics and Automation Letters.

NEURAL INFORMATION PROCESSING SYSTEMS

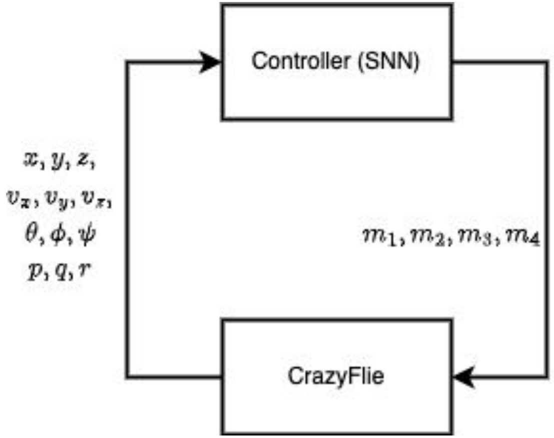
Towards end-to-end RL

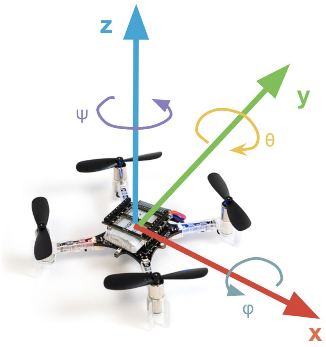
(Our work)

SNNs for Drone Control

Task description:

- Position control of Crazyflie
- Output motor commands
- No action or observation history (use temporal information)





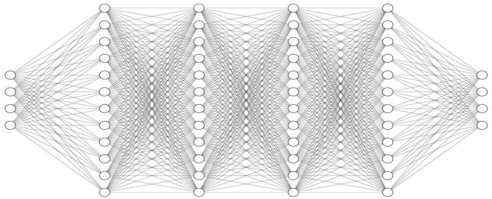
TU Delft

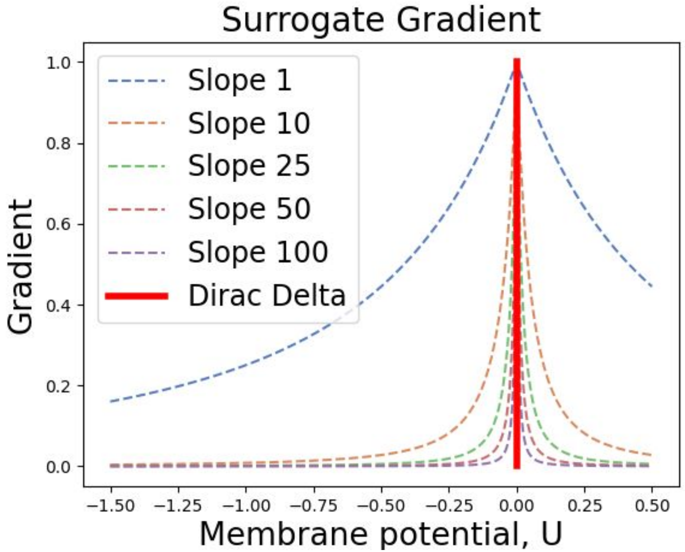
HARVARD UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Surrogate Gradient Analysis

Dead neurons: neurons that never spike
Saturated neurons: neurons that always spike

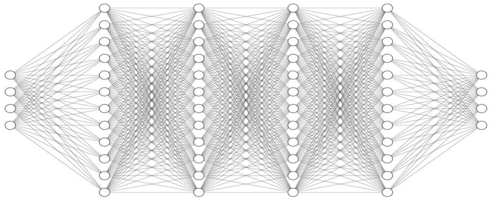


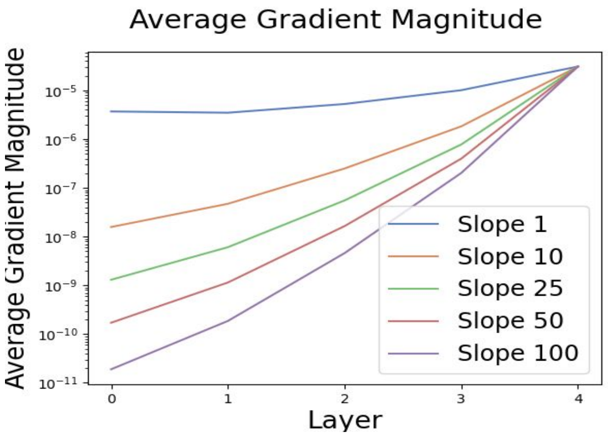


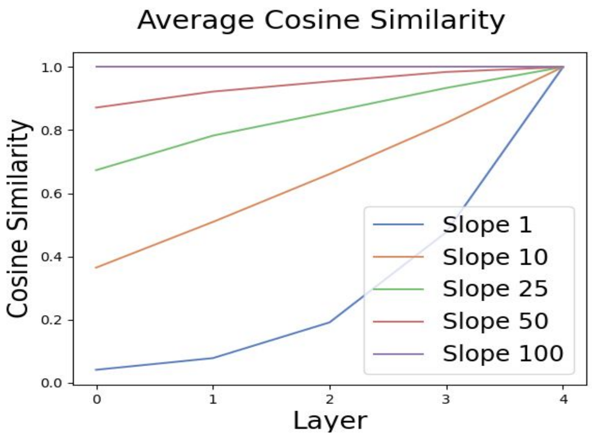
TU Delft
HARVARD UNIVERSITY
NEURAL INFORMATION PROCESSING SYSTEMS

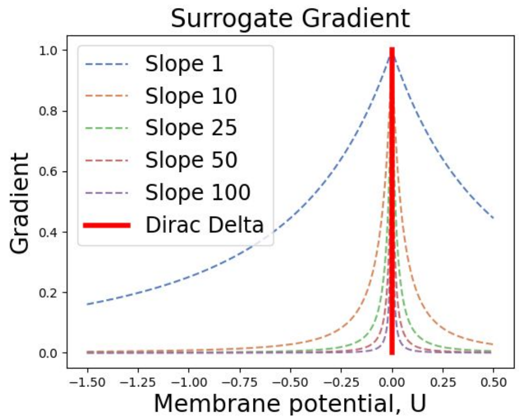
Surrogate Gradient Analysis

Looking at gradient magnitude and cosine similarity of deeper networks:



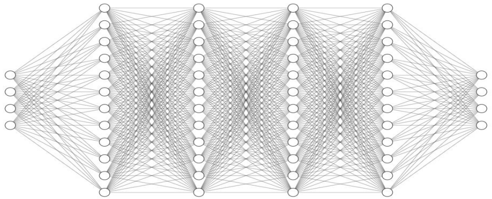


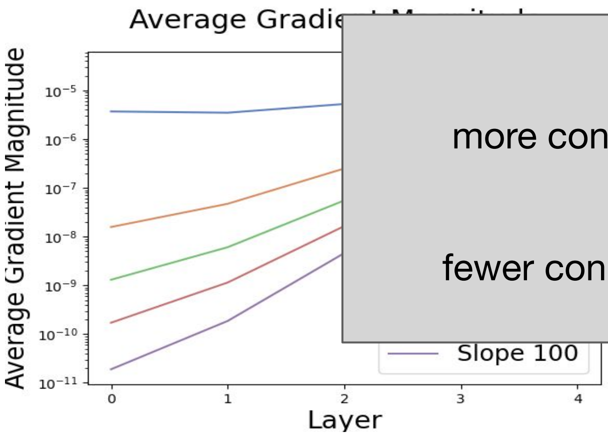


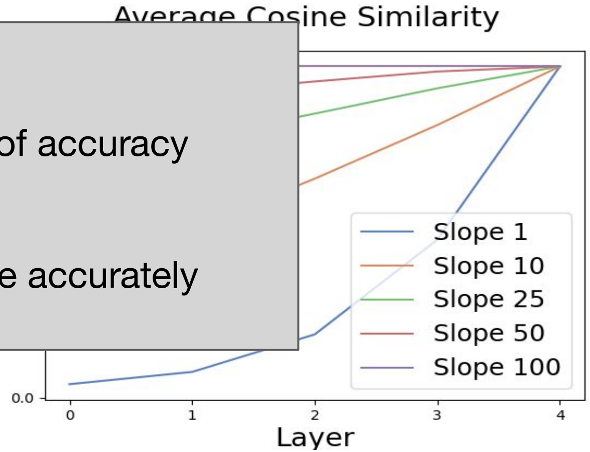


Surrogate Gradient Analysis

Looking at gradient magnitude and cosine similarity of deeper networks:







<u>*Shallow slope:*</u>
more connections are updated at cost of accuracy

<u>*Steep slope:*</u>
fewer connections are updated but more accurately

Adaptive Surrogate Gradients (this work)

When performance is bad → prioritize large gradient update
When performance reaches limit → prioritize accurate gradient update

$$k_t = \frac{1}{10}\sum_{i=0}^{9}\left[0.5r_{t-i} + 0.5r'_{t-i}\right]$$

Adaptive Surrogate Gradients

When performance is bad → prioritize large gradient update
When performance reaches limit → prioritize accurate gradient update

Directly proportional to performance

$$k_t = \frac{1}{10} \sum_{i=0}^{9} \left[0.5 r_{t-i} + 0.5 r'_{t-i} \right]$$

Adaptive Surrogate Gradients

When performance is bad → prioritize large gradient update
When performance reaches limit → prioritize accurate gradient update

Proportional to change performance

- positive change in performance: more accurate gradient
- negative change in performance: noisier gradient (exploration)

$$k_t = \frac{1}{10} \sum_{i=0}^{9} [0.5 r_{t-i} + 0.5 r'_{t-i}]$$

Adaptive Surrogate Gradients

When performance is bad → prioritize large gradient update
When performance reaches limit → prioritize accurate gradient update

Average over 10 timesteps

$$k_t = \frac{1}{10}\sum_{i=0}^{9}[0.5r_{t-i} + 0.5r'_{t-i}]$$

TU Delft

HARVARD UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Adaptive Surrogate Gradients

When performance is bad → prioritize large gradient update
When performance reaches limit → prioritize accurate gradient update

We demonstrate this on a complex
end-to-end control task

$$k_t = \frac{1}{10} \sum_{i=0}^{9} \left[0.5 r_{t-i} + 0.5 r'_{t-i}\right]$$

Neuromorphics for Robotics Challenges

- Surrogate Gradients are used widely
  - Limited understanding in optimization characteristics

Neuromorphics for Robotics Challenges

- Surrogate Gradients are used widely
    - Limited understanding in optimization characteristics
- Lack of data in robotics
    - Using RL with SNN challenging

NEURAL INFORMATION PROCESSING SYSTEMS

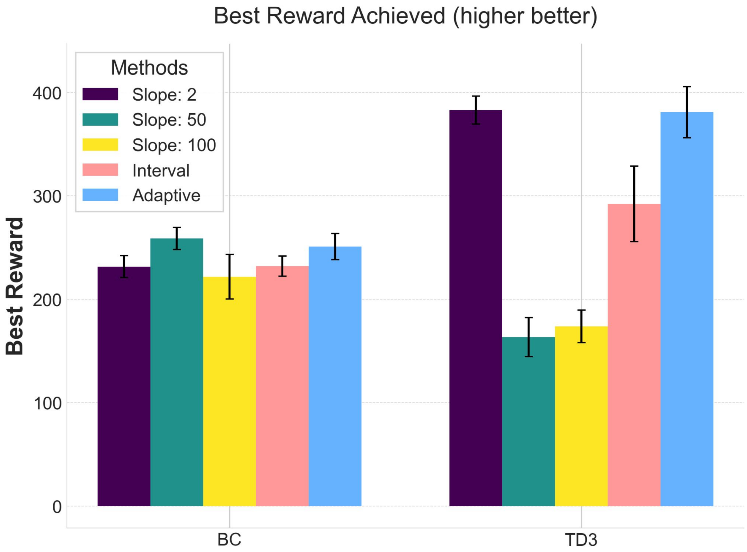
RL for SNNs

- Surrogate Gradients are used widely
  - Limited understanding in optimization characteristics
- Lack of data in robotics
  - Using RL with SNN challenging
    - Imperfect behavior → short rollouts
      - Short rollouts too short to learn temporal information

Effect on training?

<u>Supervised Learning versus Online RL</u>

- Supervised
  - Effect minimal
- Online RL
  - Shallow slope better
  - Adaptive keeps slope in right regime



Train performance of supervised learning (BC) and online RL (TD3), trained on single transitions

HARVARD UNIVERSITY

TU Delft

HARVARD UNIVERSITY

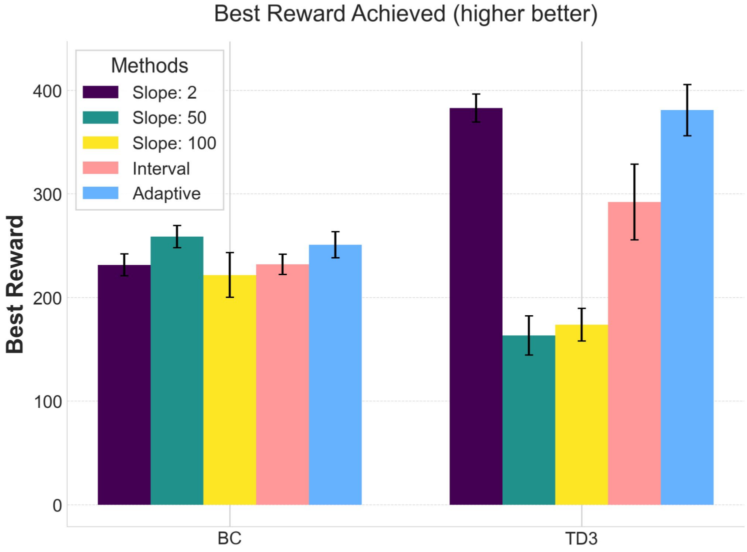
NEURAL INFORMATION PROCESSING SYSTEMS

Effect on training?

<u>Supervised Learning versus Online RL</u>

- Supervised
  - Effect minimal
- Online RL
  - Shallow slope better
  - Adaptive keeps slope in right regime

Adaptive Scheduling can eliminate hyperparameter sweeps, and tune the slope to its optimal value



Train performance of supervised learning (BC) and online RL (TD3), trained on single transitions

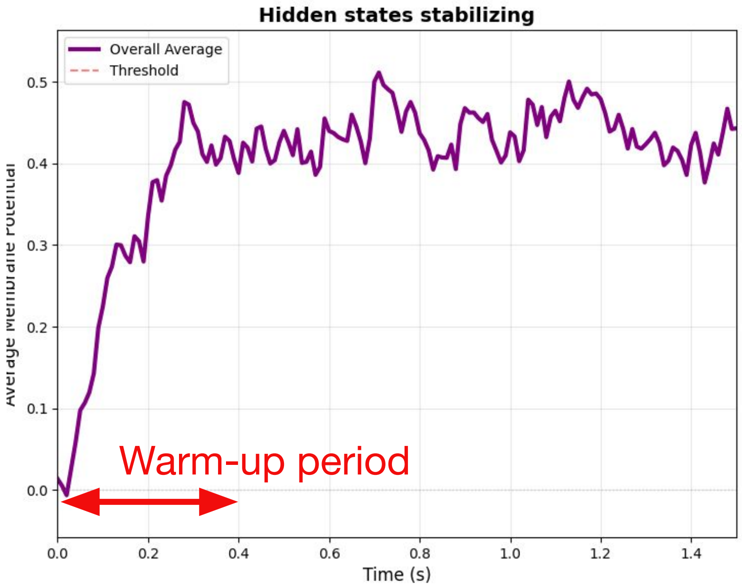
TU Delft

HARVARD UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Remaining Issue: sequence lengths

- Spiking networks need warm-up period
- Bad policy in drone control → crashes!



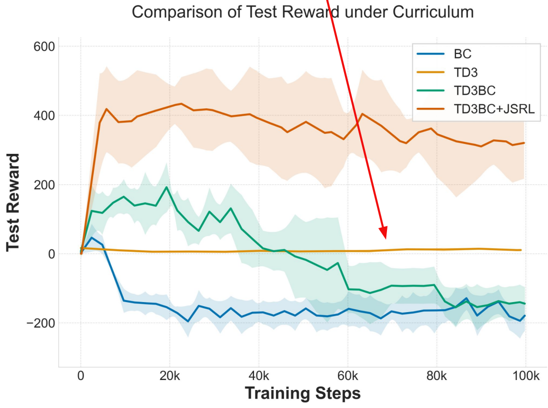
TU Delft

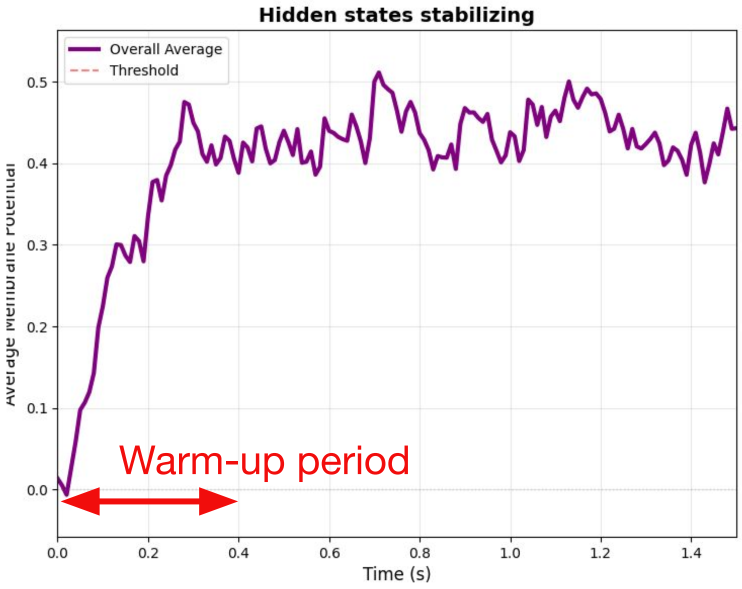
HARVARD UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Remaining Issue: sequence lengths

- Spiking networks need warm-up period
- Bad policy in drone control → crashes!
- Online RL fails to learn controller that bridges this period





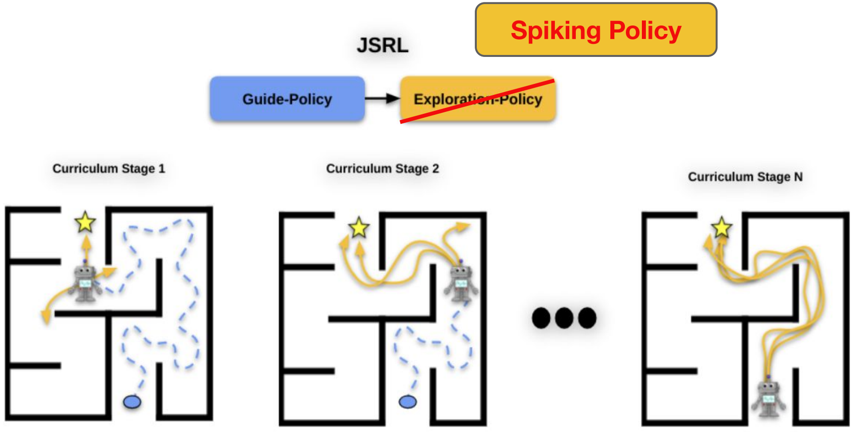
TU Delft

HARVARD UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Proposed: TD3BC+JSRL?

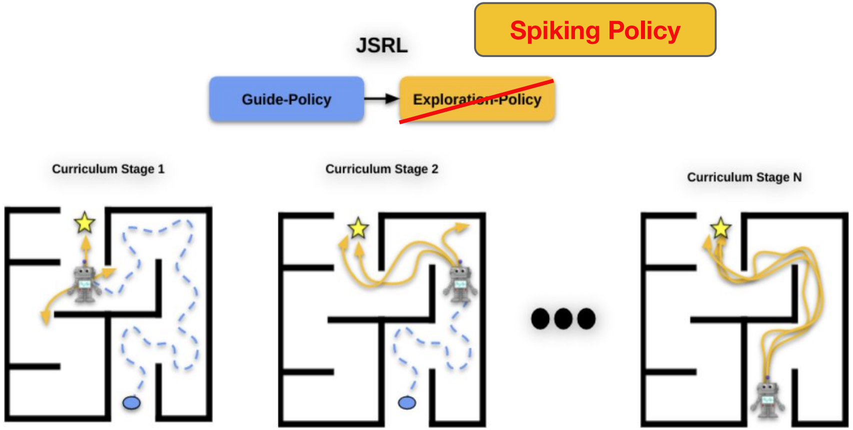
- JSRL: bridge warm-up period
  - Use privileged, non-spiking guiding policy



Uchendu, I., Xiao, T., Lu, Y., Zhu, B., Yan, M., Simon, J., ... & Hausman, K. (2023, July). Jump-start reinforcement learning. In International Conference on Machine Learning (pp. 34556-34583). PMLR.

Proposed: TD3BC+JSRL?

- JSRL: bridge warm-up period
    - Reduce number of jump-start steps as spiking policy improves
- TD3: leverage rewards
- BC: leverage privileged, non-spiking guiding policy
    - Decay BC term as spiking policy improves



Uchendu, I., Xiao, T., Lu, Y., Zhu, B., Yan, M., Simon, J., ... & Hausman, K. (2023, July). Jump-start reinforcement learning. In International Conference on Machine Learning (pp. 34556-34583). PMLR.

TU Delft — HARVARD UNIVERSITY — NEURAL INFORMATION PROCESSING SYSTEMS

Guide Policy for Warm-Up Period

Guide policy to bridge warm-up period?

- Replay buffer fills with guide policy interactions
- Use imitation to leverage

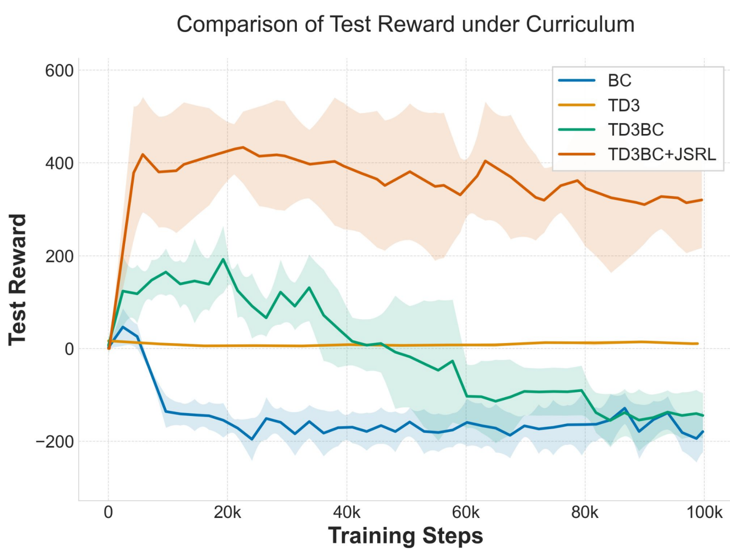
→ TD3BC with decaying BC factor

TD3BC+JSRL

- Privileged, non-spiking guiding policy
  - Bridge warm-up period
  - BC term to leverage guiding demonstrations
  - TD3 term to leverage rewards
  - Decay BC term as spiking policy improves

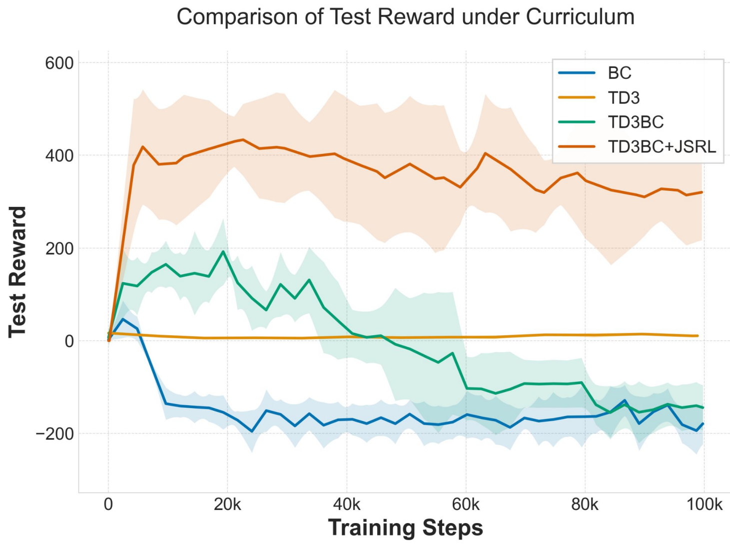
TD3BC+JSRL: Results

- Curriculum reward
  - Stricter penalties on:
    - Position
    - Velocity
    - Orientation

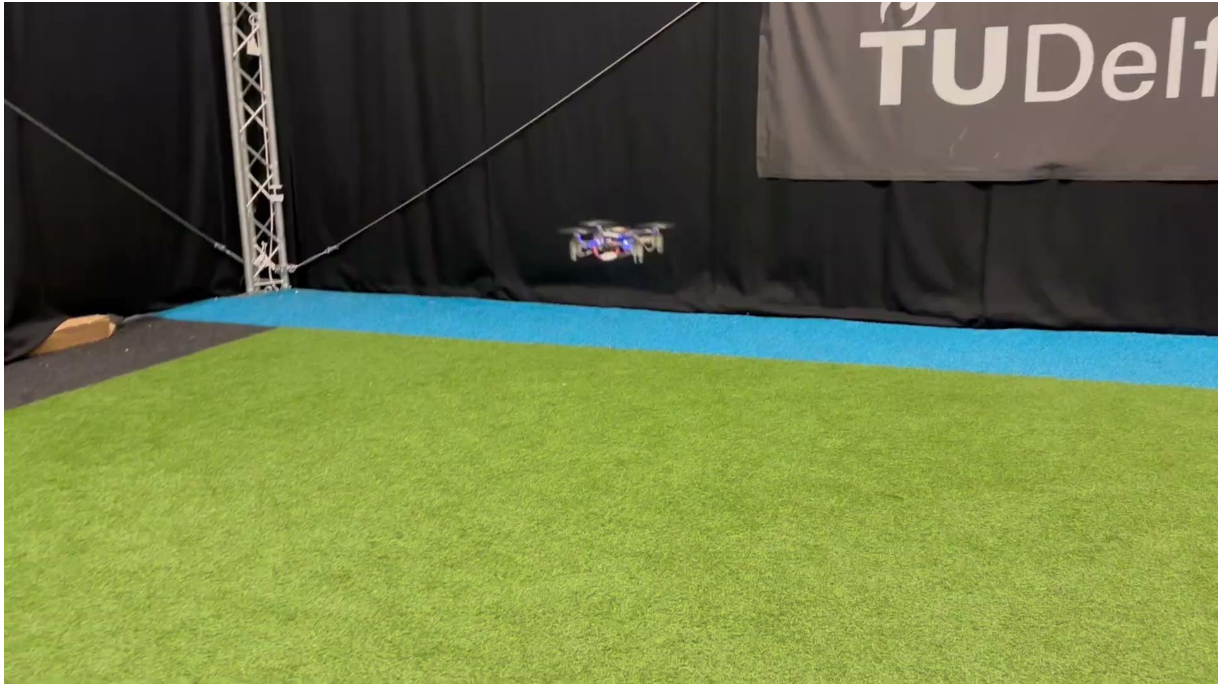


TD3BC+JSRL: Results

- Curriculum reward
- TD3 → not bridging warm-up period
- BC → not improving
- TD3BC → improves until reward function diverges too much
- TD3BC+JSRL → outperforms all methods



Bridging the reality gap



Bridging the reality gap

- ANN
  - 146 inputs (action history of 32 timesteps)
  - 2 hidden layers [64,64]
  - Multiplications
- SNN (ours)
  - 18 inputs
  - 2 hidden layers [256,128]
  - Mostly Additions

Bridging the reality gap

- Position Error: hover task
- Trajectory Error: trajectory following task

	ANN action history [14]	ANN no action history [14]	SNN no action history [Ours]
Position Error [m]	0.1	0.25	0.04
Trajectory Error [m]	0.21	NA	0.24

Bridging the reality gap

- Position Error: hover task
- Trajectory Error: trajectory following task

	ANN action history [14]	ANN no action history [14]	SNN no action history [Ours]
Position Error [m]	0.1	0.25	0.04
Trajectory Error [m]	0.21	NA	0.24

TU Delft

HARVARD UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Conclusion

- First Step Toward End-to-End Neuromorphic Control using RL
- Current Performance
    - First end-to-end deployed network without action or observation history

Conclusion

- First Step Toward End-to-End Neuromorphic Control using RL
- Current Performance
  - First end-to-end deployed network without action or observation history
  - Oscillatory

Conclusion

- First Step Toward End-to-End Neuromorphic Control using RL
- Current Performance
  - First end-to-end deployed network without action or observation history
  - Oscillatory
  - Expected to outperform ANN (energy efficiency and low latency)

Conclusion

- First Step Toward End-to-End Neuromorphic Control using RL
- Current Performance
  - First end-to-end deployed network without action or observation history
  - Oscillatory
  - Expected to outperform ANN (energy efficiency and low latency)
- Broader Impact
  - Other resource-constrained devices
    - Edge: smartwatches, video monitoring
    - Robotics: humanoids, AVs

Adaptive Surrogate Gradients for Sequential Reinforcement Learning in Spiking Neural Networks

Poster: #2308

Contact:
korneel.vandenberghe@hotmail.be

Paper:

Code:

TU Delft

HARVARD UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS