

Algorithm-Driven Design and Optimization of Printed Analog Neuromorphic Circuits

Exposé of Haibin Zhao, M.Sc.
supervised by

Prof. Mehdi B. Tahoori and Prof. Michael Beigl

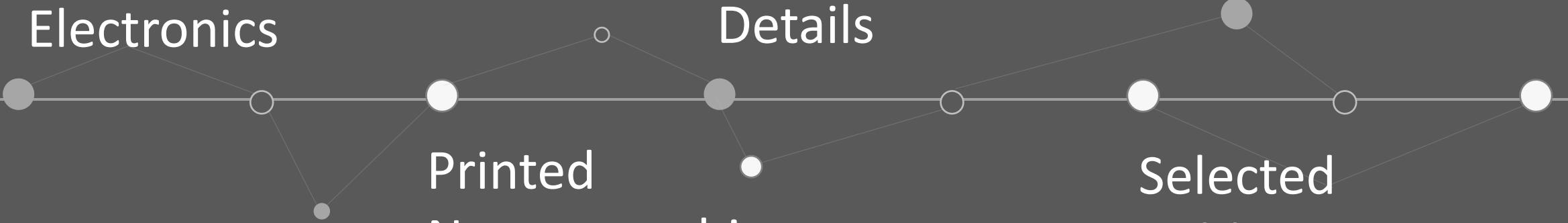
CONTENT

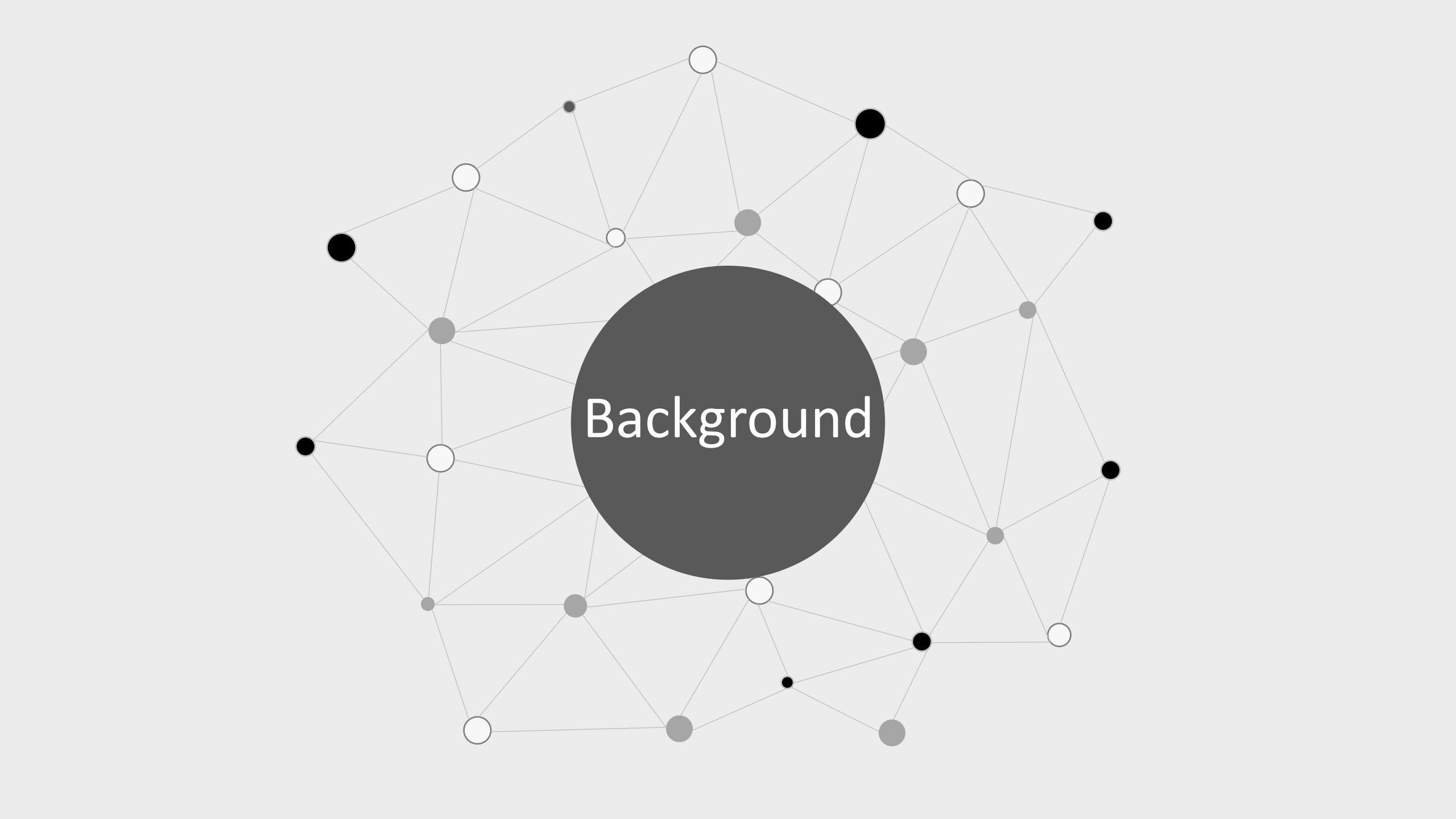
Printed
Electronics

Research
Details

Selected
Publications

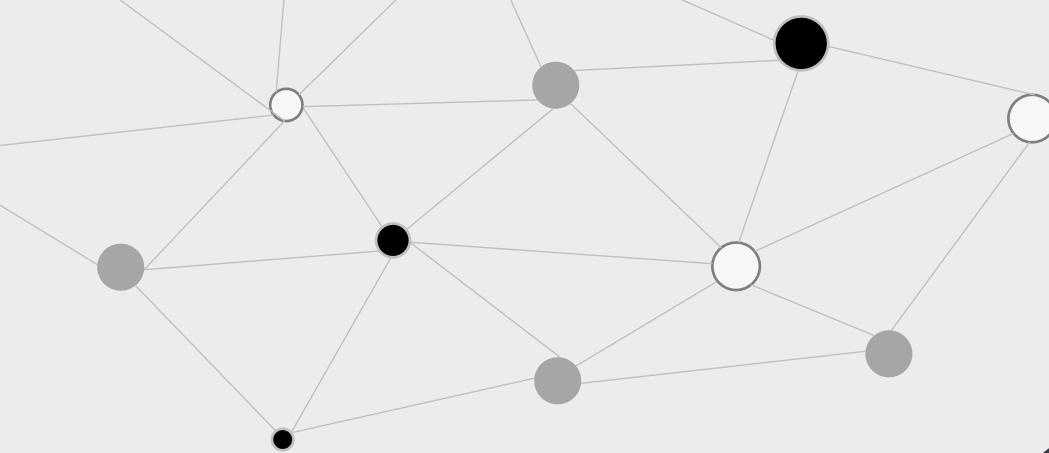
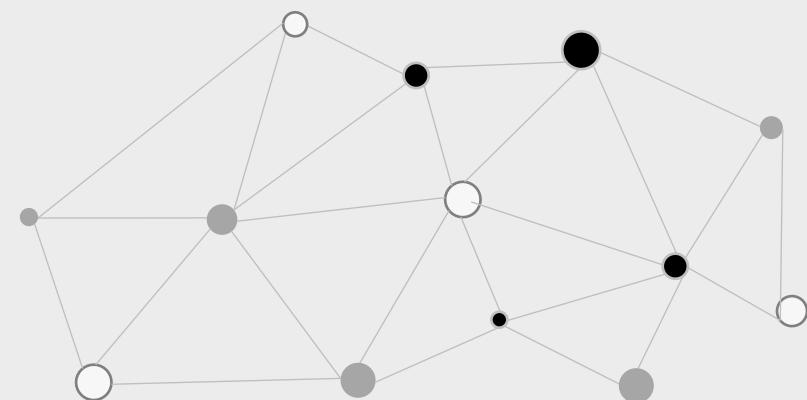
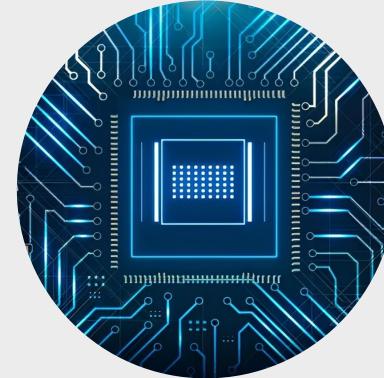
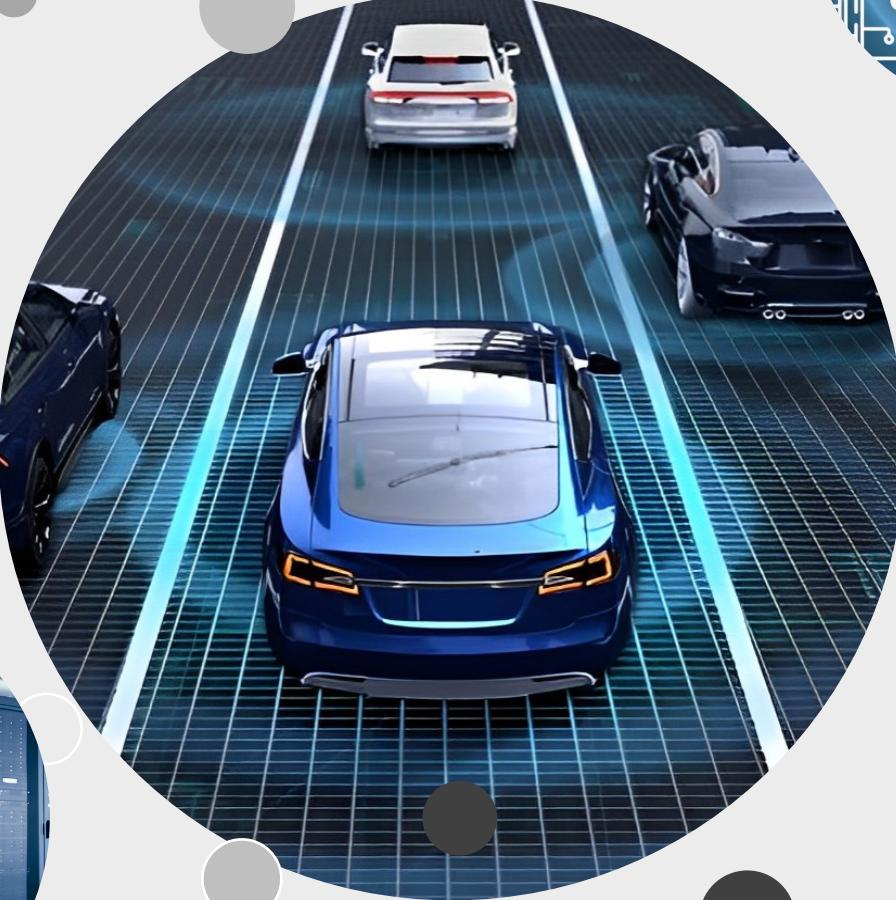
Printed
Neuromorphic
Circuits





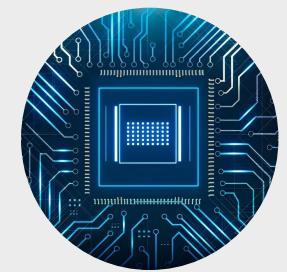
Background

high performance
resource intensive
general purpose

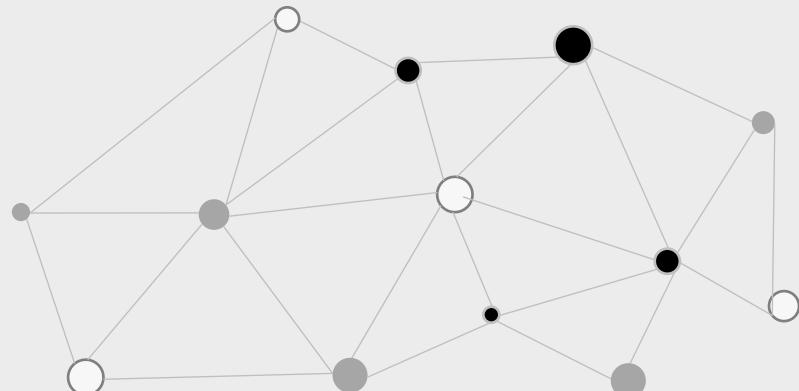
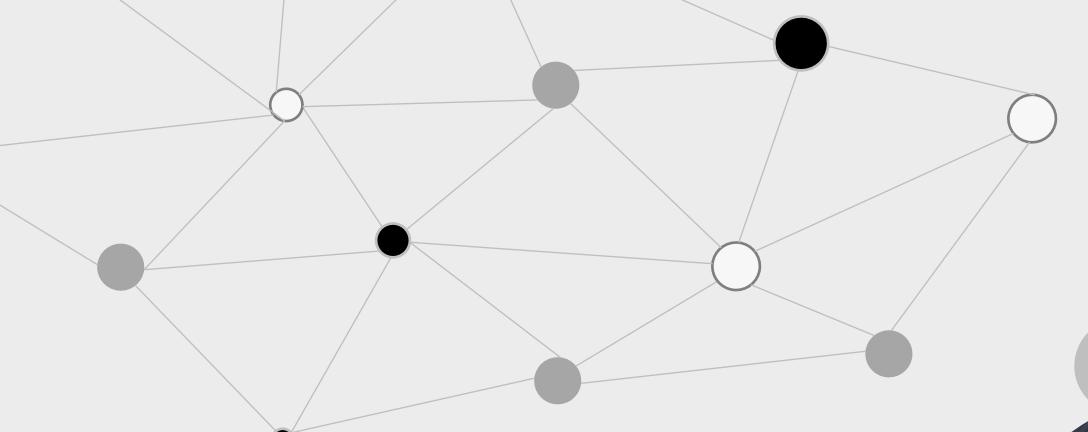


**ultra-low cost
bio-compatible
degradable
stretchable**

...



**high performance
resource intensive
general purpose**



Printed Electronics

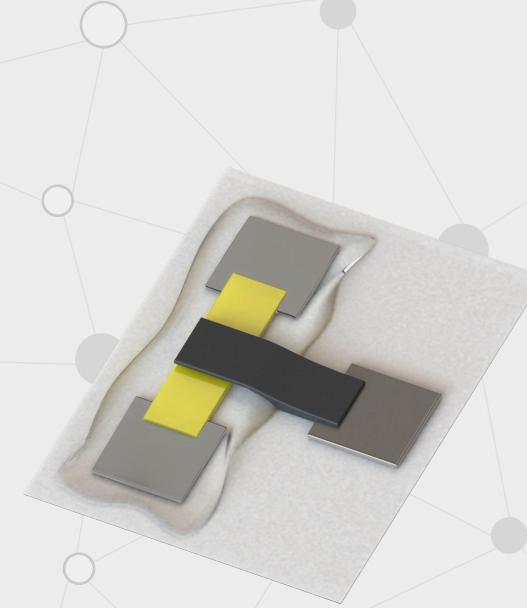
enabler of the next-generation electronics



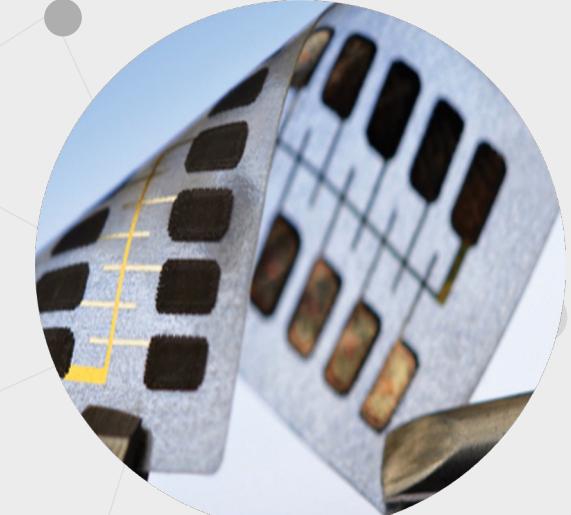
Gravure Printing



Inkjet Printing



Printed Transistor



Printed Circuit

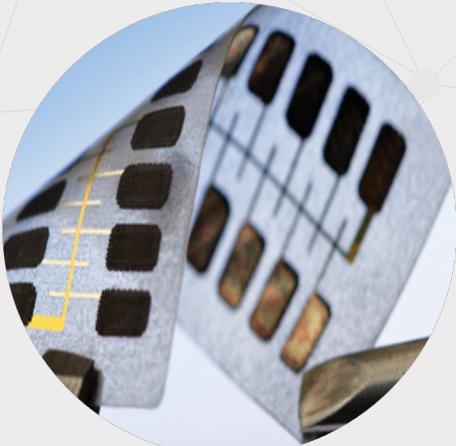
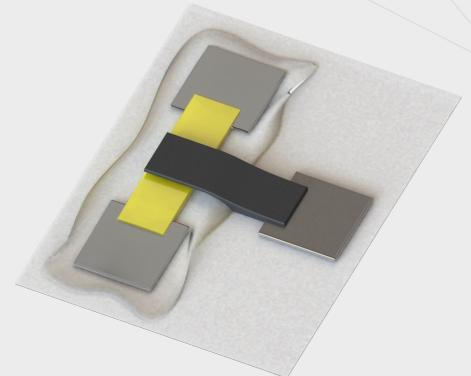
Maskless, fully additive process
avoid photo-lithograph based manufacturing
ensures ultra-low fabrication cost

Various printing technologies
enables both high- and low- throughput

Abundant functional inks
allows for printed semiconductors
allows for non-toxicity, degradability, ...
Diverse substrate materials
provides flexibility, porosity, ...

Printed Electronics

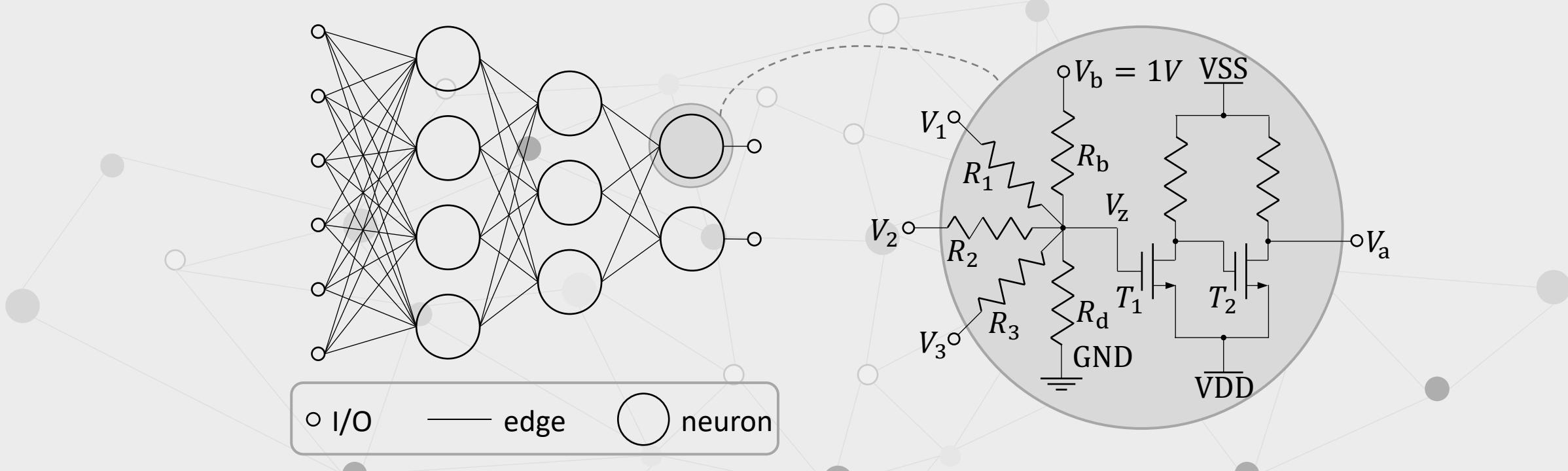
enabler of the next-generation electronics



- Maskless, fully additive process ensures ultra-low fabrication cost
- Various printing technologies enables both high- and low- throughput
- Abundant functional inks allows for printed semiconductors
- allows for non-toxicity, degradability, ...
- Diverse substrate materials provides flexibility, porosity, ...
- Large feature size limits the device counts in the printed circuits
- High variation impact the circuit output from designed values

Printed Analog Neuromorphic Circuit

hardware implementation of artificial neural networks



Artificial neuron:

weighted-sum

$$z = w_1x_1 + w_2x_2 + w_3x_3 + b$$

nonlinear activation

$$a = \tanh(z)$$

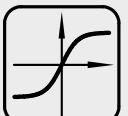
Neuromorphic circuit:

resistor crossbar

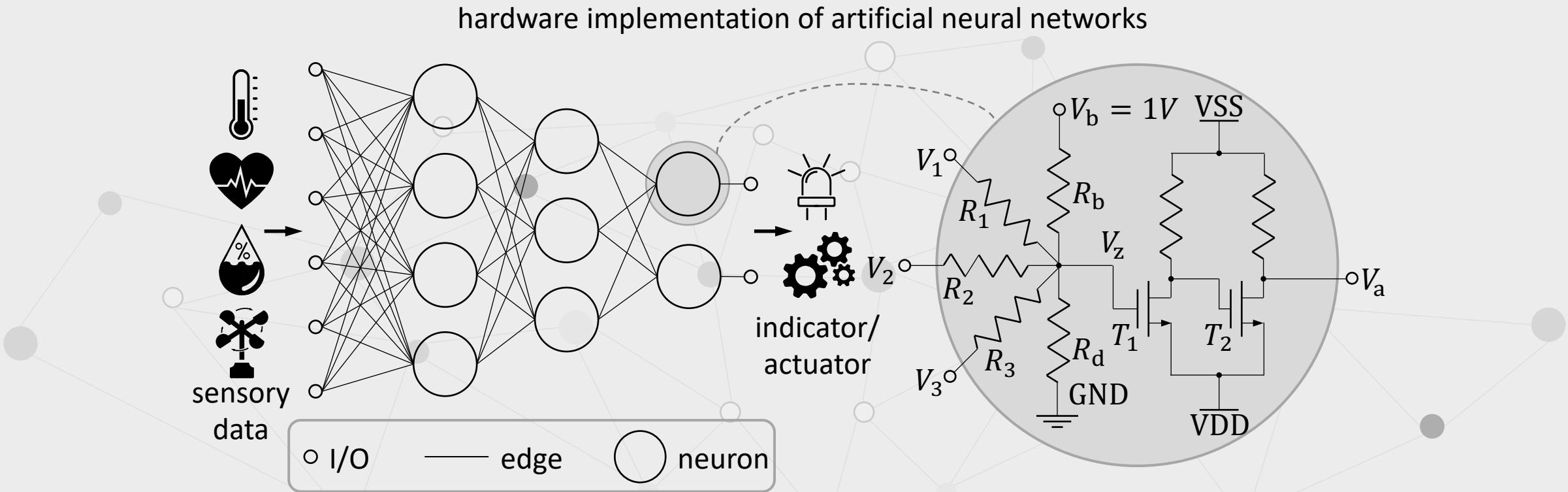
$$V_z = \frac{g_1}{G}V_1 + \frac{g_2}{G}V_2 + \frac{g_3}{G}V_3 + \frac{g_b}{G} \text{ with } g_1 = \frac{1}{R_i}, G = \sum_i g_i$$

inverter-based nonlinear circuit

$V_a = \text{ptanh}(V_z)$, a tanh-like function measured from experiment



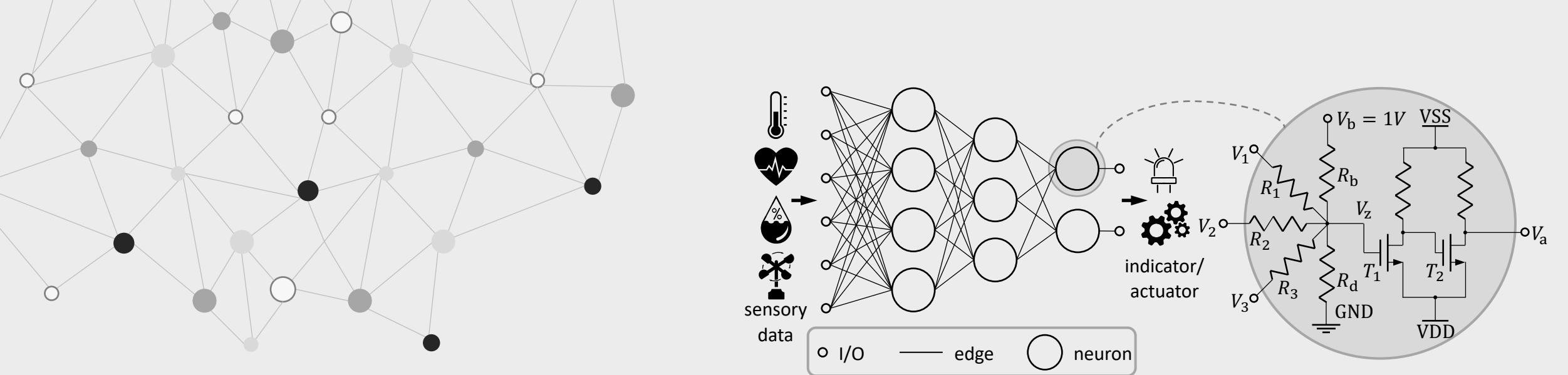
Printed Analog Neuromorphic Circuit



Analog vs. digital implementation

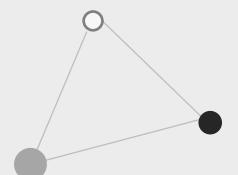
Components	# Transistors	
	4-bit digital	analog
Input ADC	185	—
Weighted-sum	265	≤ 4
Activation	10	2

→ respect the high feature size
of printed electronics



Algorithm-Driven Design and Optimization of Printed Analog Neuromorphic Circuits

The improvement of the circuits is a multifaceted undertaking that spans multiple disciplines, including material science, electrical and electronic engineering. This dissertation primarily focuses on optimizing the circuit at the algorithmic level.



1

Modeling of pNCs

2

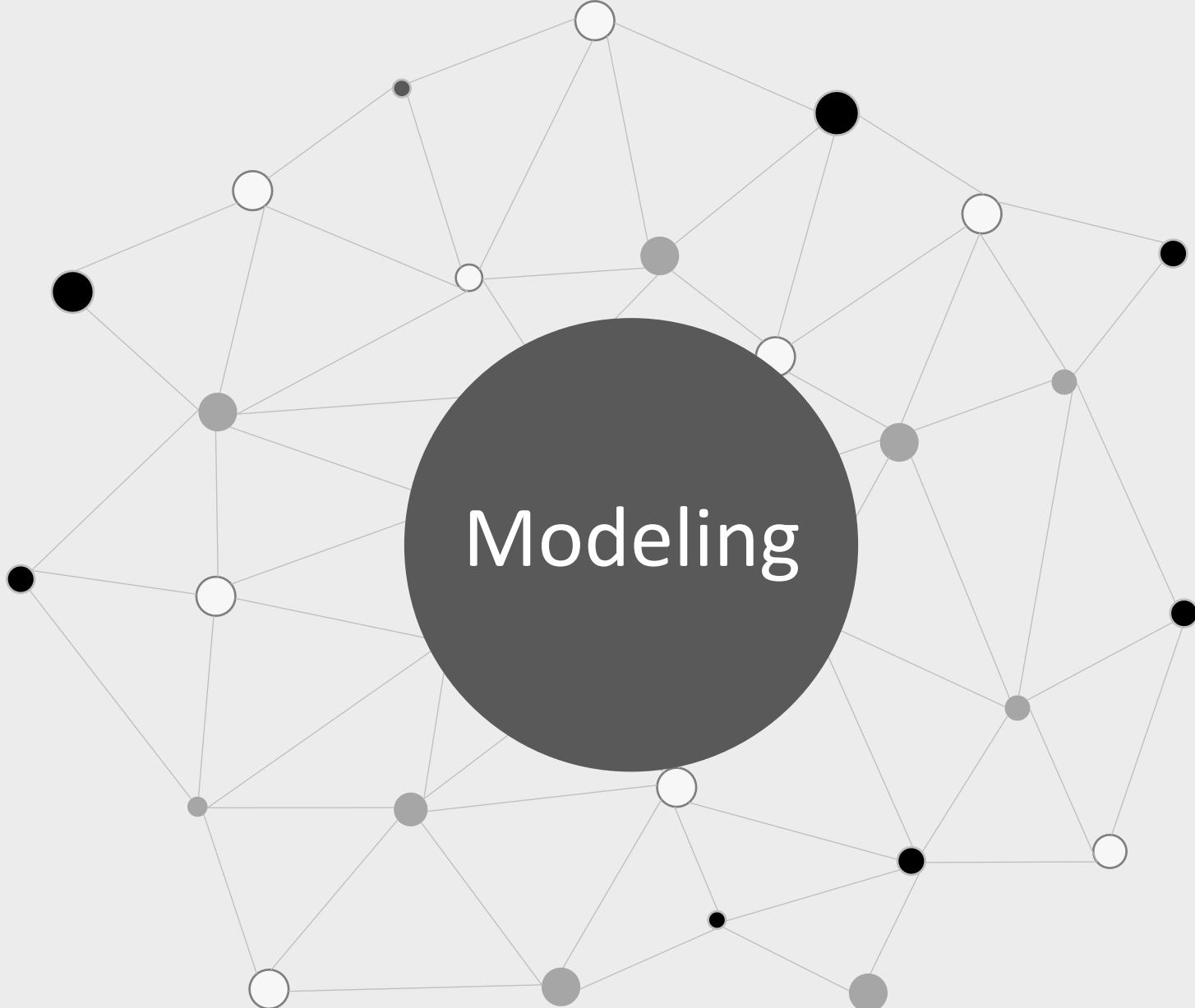
Reliability design of the pNCs

3

Practicality design of the pNCs

4

Extensions of the pNCs

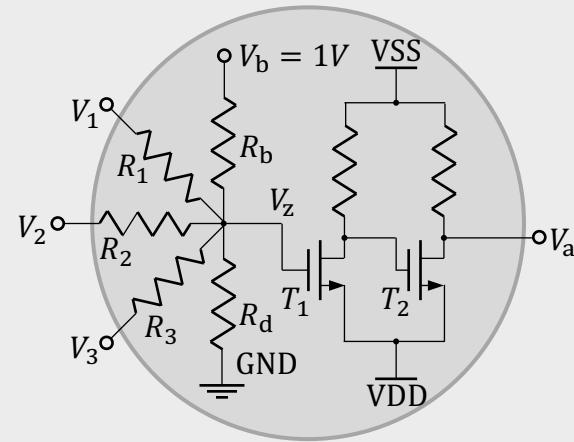


Reliability

Practicality

Extension

Modeling of the Printed Analog Neuromorphic Circuit



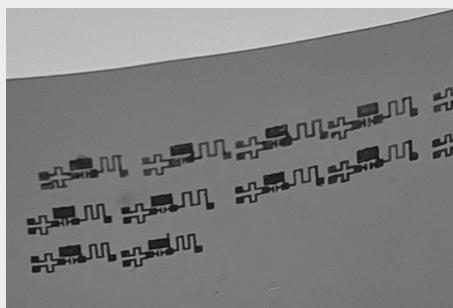
$$V_z = \frac{g_1}{G} V_1 + \frac{g_2}{G} V_2 + \frac{g_3}{G} V_3 + \frac{g_b}{G}$$

$$V_a = \text{ptanh}(V_z)$$

↓ training

optimal weight (i.e., conductance)

↓ printing

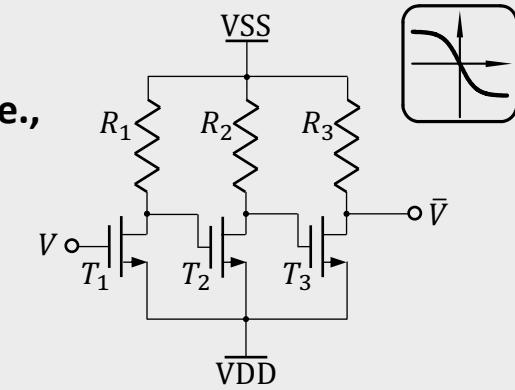


- Unidirectional mapping from pNCs to ANNs
 $w_i \leftarrow g_i/G$, however $w_i \not\rightarrow g_i/G$
Directly model the conductance $\{g_i\}$ as learnable parameters instead of weights $\{w_i\}$
- Physical constraint: negative weight
 $w_i \leftarrow g_i/G$, however $g_i \geq 0$ thus $w_i \geq 0$
Negation circuit, transferring the input into negative one, i.e.,
 $(-w_i) \cdot x_i \rightarrow w_i \cdot (-x_i) \rightarrow w_i \cdot \text{neg}(x_i)$
Resulting term in weighted-sum operation

$$\begin{cases} w_i \cdot \text{neg}(x_i), & \text{if negation circuit exists} \\ w_i \cdot x_i, & \text{otherwise} \end{cases}$$

Existence of negation circuit?
Surrogate conductance θ_i , encoding g_i by $|\theta_i|$ and the negation circuit by $\text{sign}(\theta_i)$
 $w_i \cdot (x_i \cdot \mathbb{I}_{\{\theta_i \geq 0\}} + \text{neg}(x_i) \cdot \mathbb{I}_{\{\theta_i < 0\}})$
- Manufacturability: limited technological fabrication range
 $|\theta_i| = g_i \in [g_{\min}, g_{\max}] \cup \{0\}$
Ensure constraints through projection

$$\tilde{\theta}_i = \begin{cases} \text{sign}(\theta_i) \cdot g_{\max}, & |\theta_i| \geq g_{\max} \\ 0, & |\theta_i| \leq g_{\min} \\ \theta_i, & \text{otherwise} \end{cases}$$



Modeling of the Printed Analog Neuromorphic Circuit

Training objective of printed neuromorphic circuit

$$\min_{\theta} \mathcal{L}\{\mathbf{y}, \hat{\mathbf{y}}(\mathbf{x}, \tilde{\theta})\}$$

(\mathbf{x}, \mathbf{y}) : target dataset, $\hat{\mathbf{y}}$: output of pNC, θ : collection of learnable surrogate conductance

$$\text{elemental operation: } w_i \cdot \left(x_i \cdot \mathbb{I}_{\{\tilde{\theta}_i \geq 0\}} + \text{neg}(x_i) \cdot \mathbb{I}_{\{\tilde{\theta}_i < 0\}} \right)$$

Machine Learning based Modeling

Highly efficient training

Vectorization for multiple neurons in each layer

$$\text{ptanh}\left(\mathbf{X} \cdot (\mathbf{W} \odot \mathbb{I}_{\{\tilde{\theta} \geq 0\}}) + \text{neg}(\mathbf{X}) \cdot (\mathbf{W} \odot \mathbb{I}_{\{\tilde{\theta} < 0\}})\right)$$

Gradient-based optimization through backpropagation

Problem with non-differentiable operations

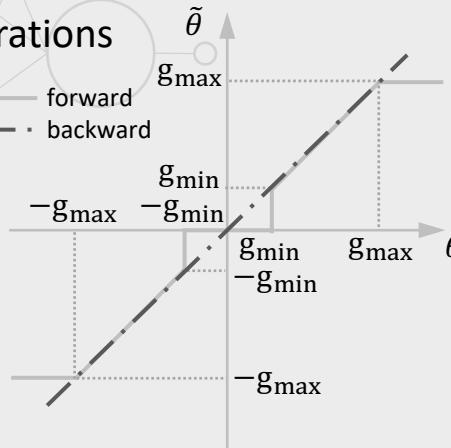
Projection from θ to $\tilde{\theta}$

Heuristic gradient estimator

Circuit topology

Neural Architecture Search

Neural Network Pruning



Evolutionary Algorithm based Modeling

Less efficient training

Non-layer-based structure, inference neuron-by-neuron

Evolving through random mutation and crossover

Only suitable for small-scale problems

Justified by simple target applications of PE

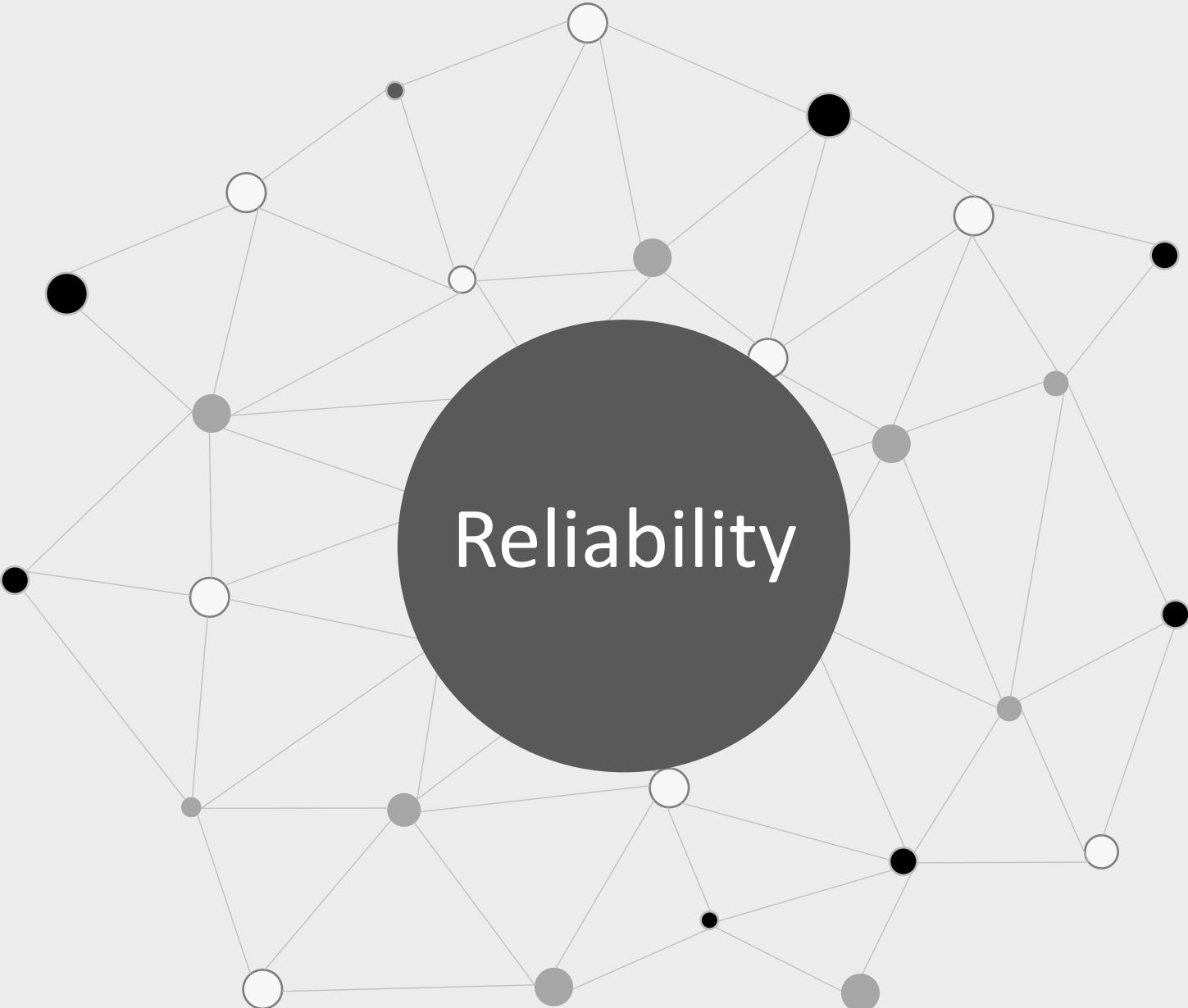
Feasible for non-differentiable operations

Directly project from θ to $\tilde{\theta}$

Topology optimization through skillful encoding

Allows using accuracy as training objective

Reliability



Modeling

Practicality

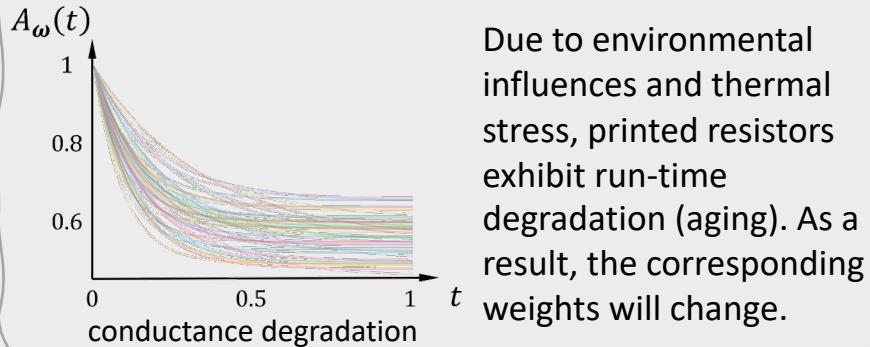
Extension

Reliability of the Printed Analog Neuromorphic Circuit

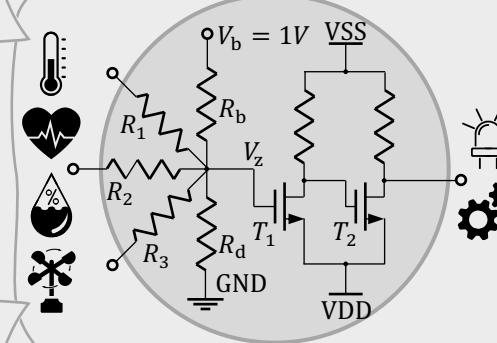
Sensing uncertainty

- intrinsic fabrication errors of the sensor
- coupling between sensor and the system being measured
- changes in measurement conditions
- imperfections in the calibration

Aging of printed resistors



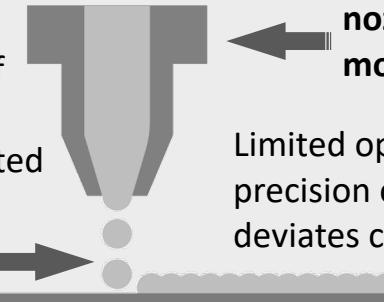
H Zhao et al. Aging-Aware Training for Printed Neuromorphic Circuits. ICCAD, 2022.



Variation of printed devices

Dispersion of the droplets impacts printed components

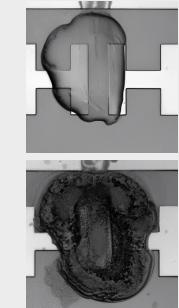
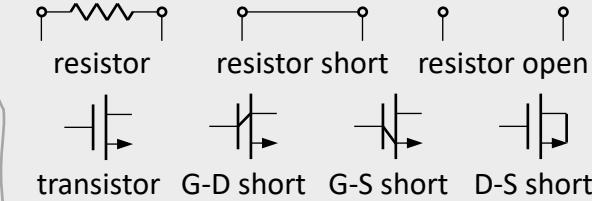
droplet



Limited operational precision of the printer deviates component values

H Zhao et al. Highly-Bespoke Robust Printed Neuromorphic Circuit. DATE, 2023.
H Zhao et al. Highly-dependable printed neuromorphic circuits. IOP FPE, 2023.

Device fault

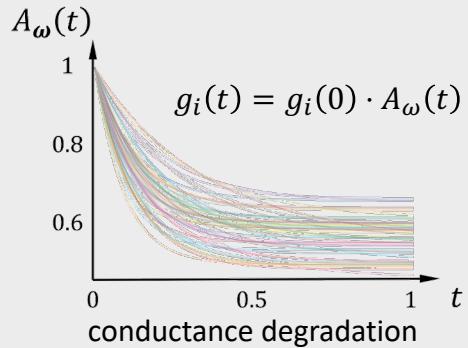


Due to catastrophes during manufacturing and usage, e.g., high voltage breakdown, devices can incur faults. They tend to be low-probability but with high impacts.

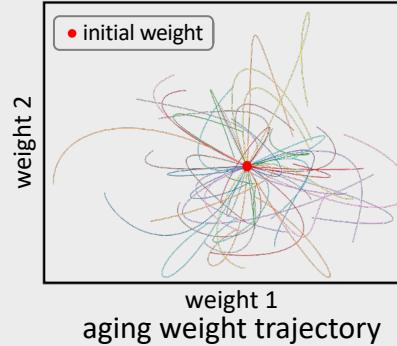
(submitted) P Pal et al. Fault Detection and Tolerance in Printed Neural Networks. ETS, 2024.

Reliability of the Printed Analog Neuromorphic Circuit

Resistor aging

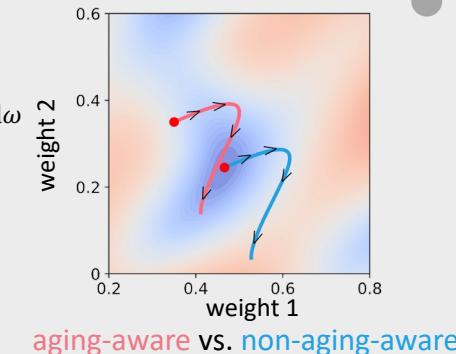


$$w_i(t) = \frac{g_i(t)}{g_1(t)+g_2(t)}$$



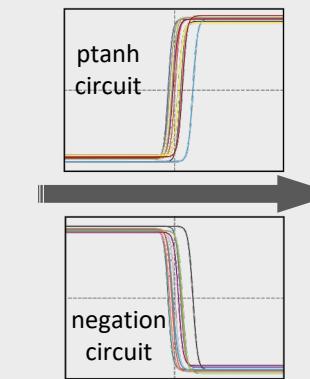
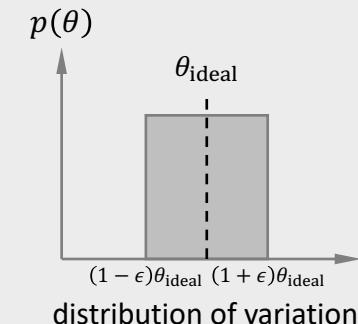
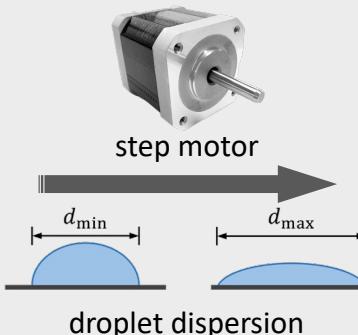
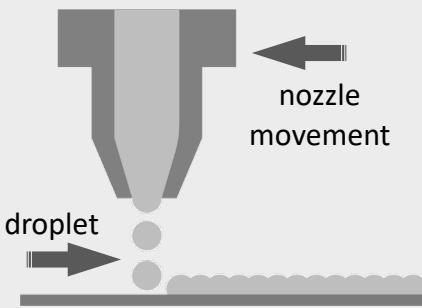
$$\min_{\theta_0} \int_{\omega}^1 \int_0^1 \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}, \tilde{\theta}_0, A_\omega(t)), \mathbf{y}) dt p(\omega) d\omega$$

aging aware training



aging-aware vs. non-aging-aware

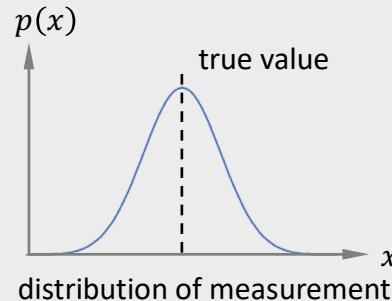
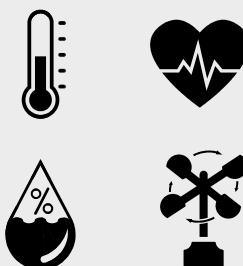
Printing variation



Objective function of variation-aware training

$$\min_{\theta} \int_{\varepsilon} \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}, \varepsilon \cdot \tilde{\theta}), \mathbf{y}) p(\varepsilon) d\varepsilon$$

Sensing uncertainty



Objective function of sensing uncertainty-aware training

$$\min_{\theta} \int_v \mathcal{L}(\hat{\mathbf{y}}(v \cdot \mathbf{x}, \tilde{\theta}), \mathbf{y}) p(v) dv$$

Reliability of the Printed Analog Neuromorphic Circuit

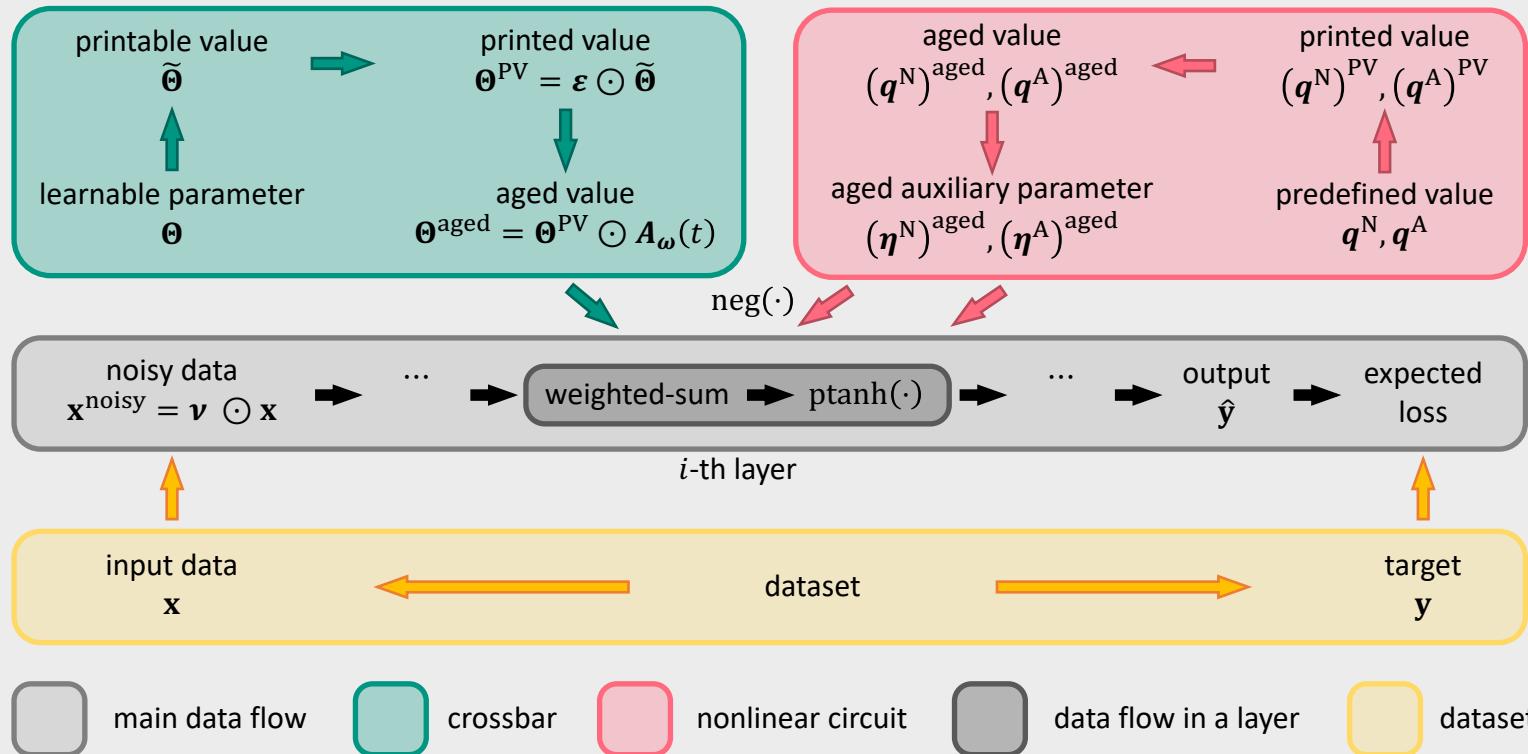


Figure: Data flow and computing graph in the dependability-aware training for printed neuromorphic circuits. The gray part indicates the main data flow, the green box contains the data processing for weighted-sum resistor crossbar, the red box indicates the processing of nonlinear circuits, while the bottom yellow box refers to the target dataset.

Conclusion

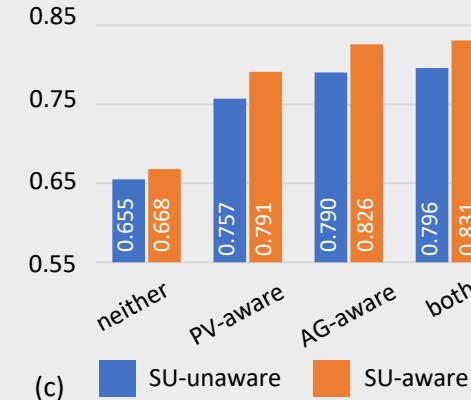
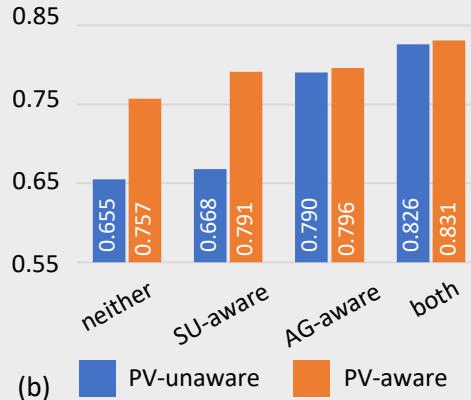
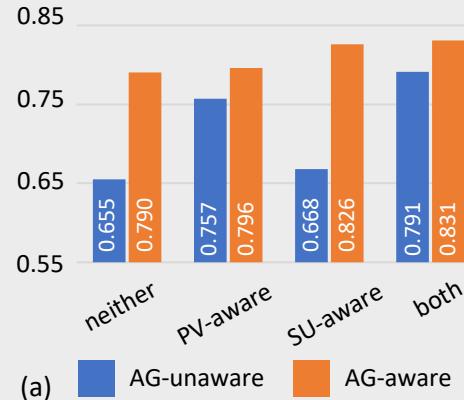
By considering aging, printing variation and sensing uncertainty into the training process

27% improvement in **accuracy**

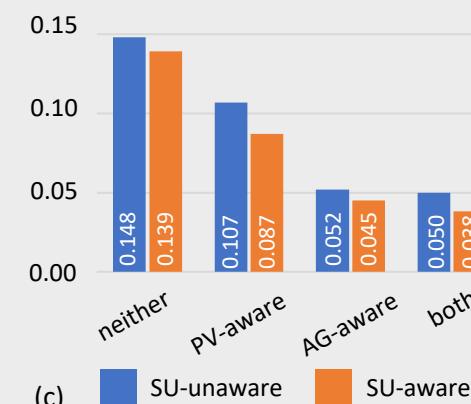
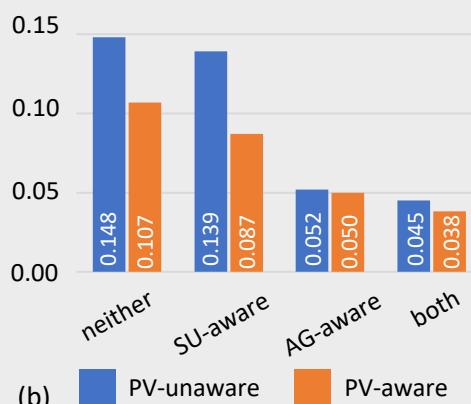
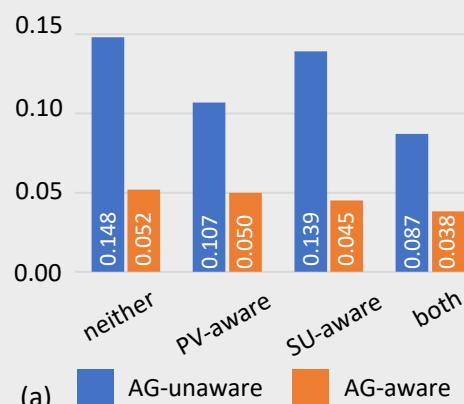
74% improvement in **robustness**
can be observed.

Reliability of the Printed Analog Neuromorphic Circuit

Mean accuracy



Robustness (standard deviation)



Conclusion

Aging-aware training contributes **most**

Variation-aware training effects **fewer**

Sensing uncertainty-aware training provides **lowest** improvement

Aging-aware and Variation-aware training **overlap**

Sensing uncertainty-aware training offers **consistent improvement**



Modeling

Reliability

Extension

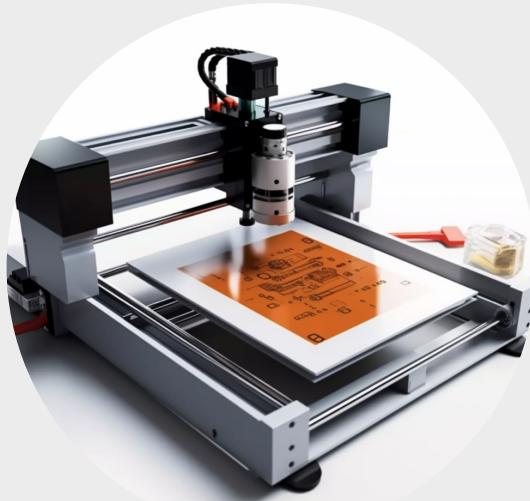
Practicality of the Printed Analog Neuromorphic Circuit

Ultra-Low Cost Design



Gravure Printing

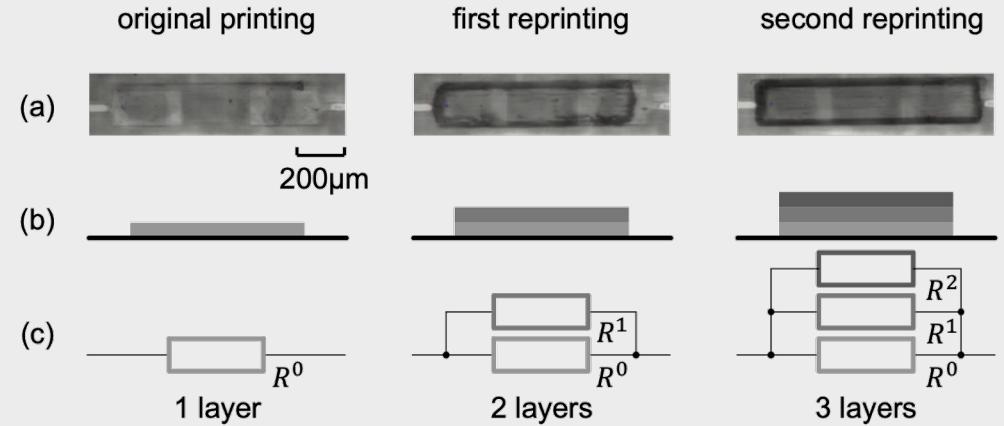
high volume printing
+ low cost per circuit
– low fabrication flexibility



Inkjet Printing

low volume printing
– high cost per circuit
+ high customization

Resistor reprinting



Resulting conductance
 $g = g_0 + g_1 + g_2 + \dots$

Bridging the gap between different printing technologies and combining their advantages

Practicality of the Printed Analog Neuromorphic Circuit

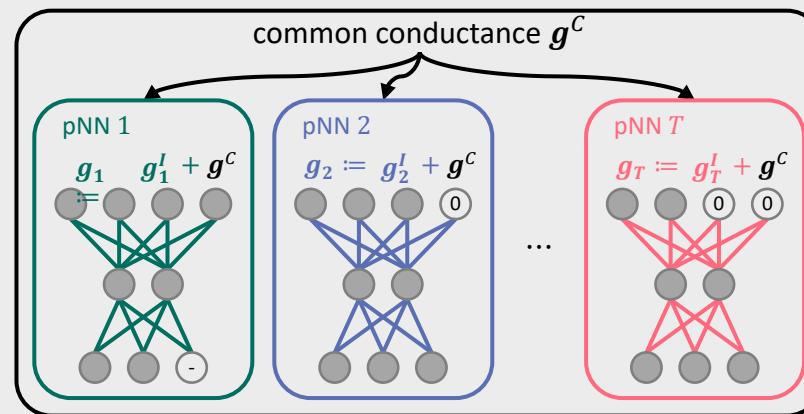
Ultra-Low Cost Design

Question: How to print large number of neuromorphic circuits for **different tasks** with only **small quantities** for each?

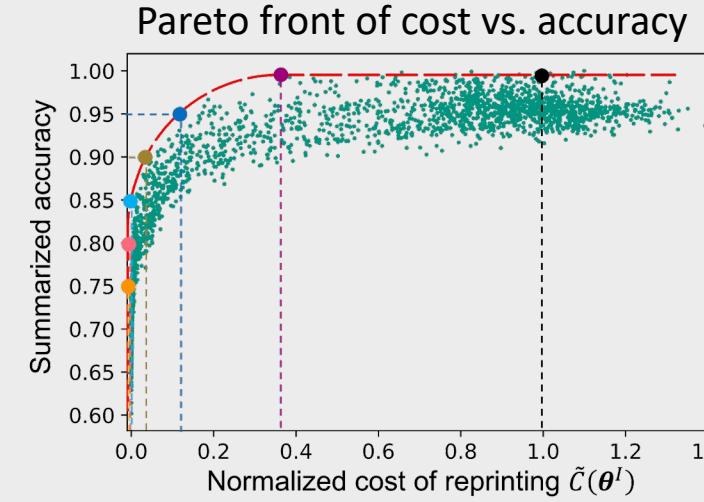
High volume printing
+ low cost
– same circuit for different tasks

Low volume printing
+ bespoke circuit for each task
– high printing cost

Super printed neural network



Split conductance into **common** and **individual** parts
Common part by high volume process -> **low cost**
Individual part by low volume process -> **point of use**



61.4% ↓ printing cost without accuracy loss

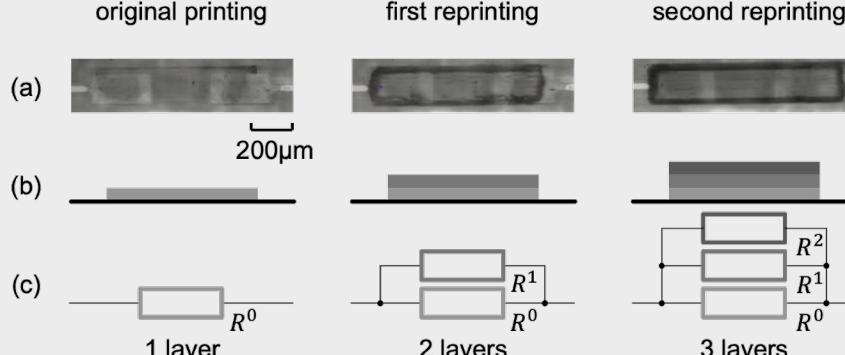
Enabling other trade-offs
between accuracy and printing cost



Gravure Printing

Inkjet Printing

Resistor reprinting



Resulting conductance

$$g = g_0 + g_1 + g_2 + \dots$$

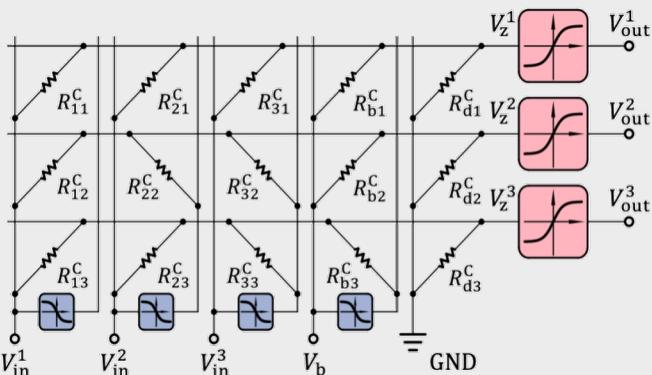
Practicality of the Printed Analog Neuromorphic Circuit

Low-Power Circuit Design

Target applications of printed electronics



Printed batteries with limited power
Supplied by energy harvesters



improved architecture of a printed layer
each input is only inverted once

Power model

Power of the crossbar

$$P^C = \left((V_{in} \odot \mathbb{I}_{\{\Theta \geq 0\}} + V_{in} \odot \mathbb{I}_{\{\Theta < 0\}}) - V_z \right)^2 \odot \Theta$$

Power of nonlinear circuits

Analytical solution is impossible

NN-based surrogate power model

prepare data through SPICE simulation
train NN to map q to \mathcal{P}

Power consumption

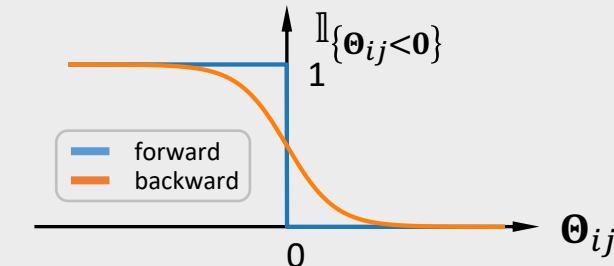
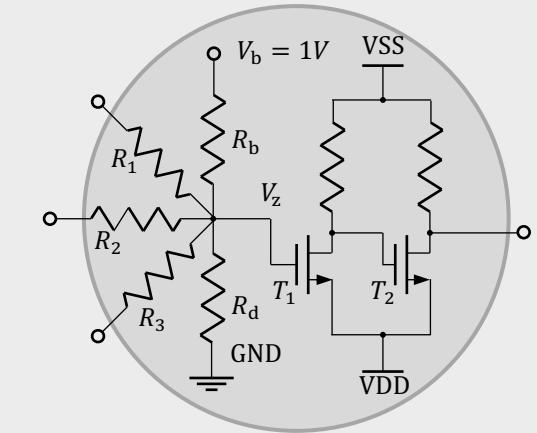
$$\mathcal{P} = P^C + N^A \cdot \mathcal{P}^A + N^N \cdot \mathcal{P}^N$$

Power can be reduced not only through P^C , \mathcal{P}^A , and \mathcal{P}^N , but also through reducing device count N . However, N is **non-differentiable**, e.g.,

$$N^N = \sum \text{column max } \{\mathbb{I}_{\{\Theta < 0\}}\}$$

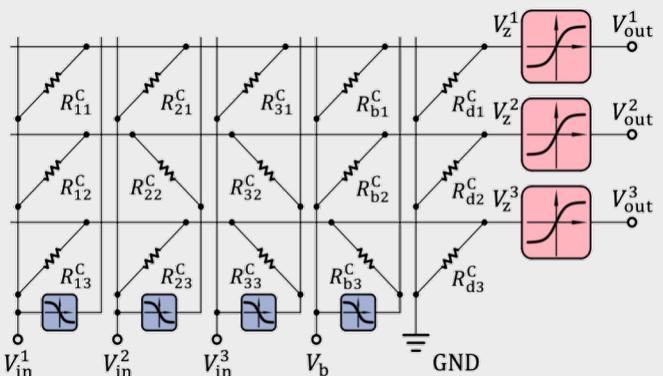
Soft-count

$$\frac{\partial N_{\text{soft}}^N}{\partial \Theta} = \frac{\partial \sum \text{column max } \{1 - \text{sigmoid}(\Theta)\}}{\partial \Theta}$$



Heuristic gradient for soft-count

Practicality of the Printed Analog Neuromorphic Circuit



improved architecture of a printed layer
each input is only inverted once

Power consumption

$$\mathcal{P} = P^C + N^A \cdot \mathcal{P}^A + N^N \cdot \mathcal{P}^N$$

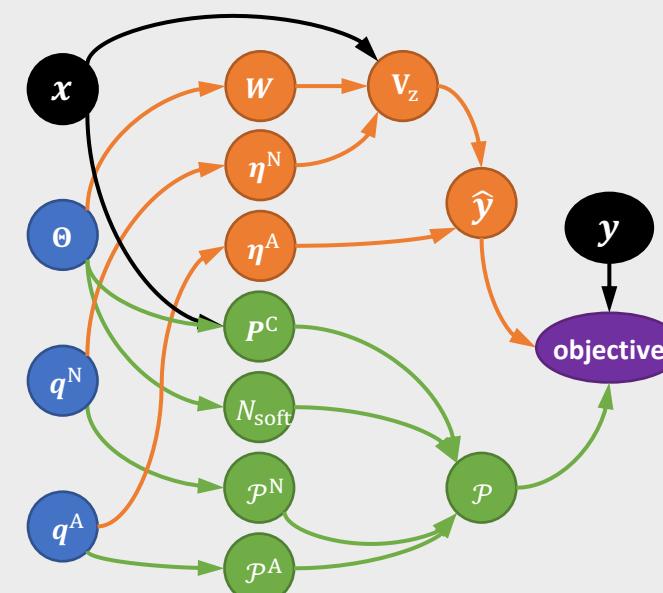
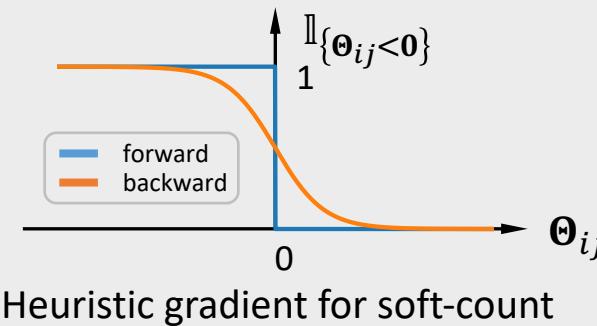
Power can be reduced not only through P^C , \mathcal{P}^A , and \mathcal{P}^N , but also through reducing device count N . However, N is **non-differentiable**, e.g.,

$$N^N = \sum \text{column max} \{\mathbb{I}_{\{\Theta < 0\}}\}$$

Soft-count

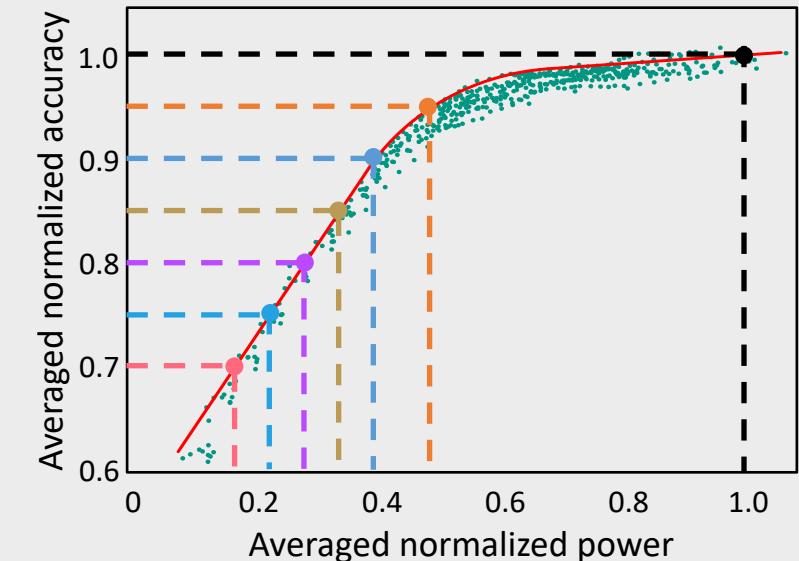
$$\frac{\partial N_{\text{soft}}^N}{\partial \Theta} = \frac{\partial \sum \text{column max} \{1 - \text{sigmoid}(\Theta)\}}{\partial \Theta}$$

Low-Power Circuit Design



Computing graph of the power-aware training

Pareto front of power vs. accuracy



Conclusion

50% power ↓ with only 5% accuracy ↓

Other trade-offs between power and accuracy can be selected on demand

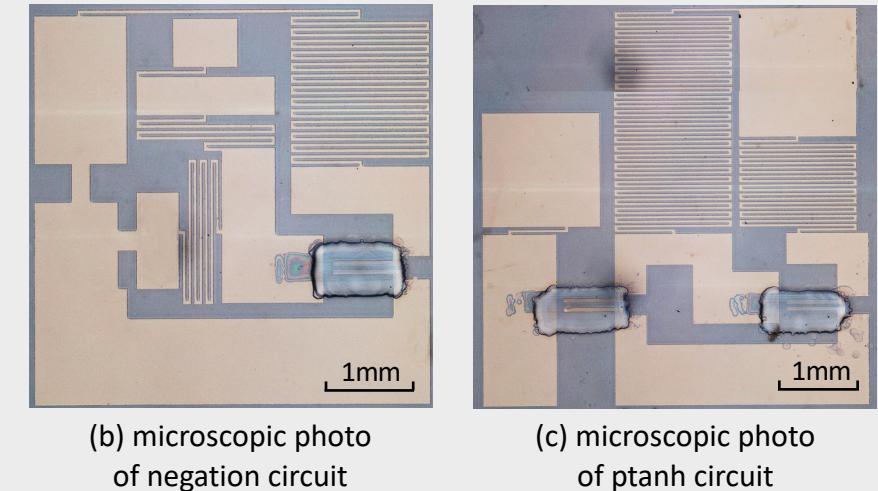
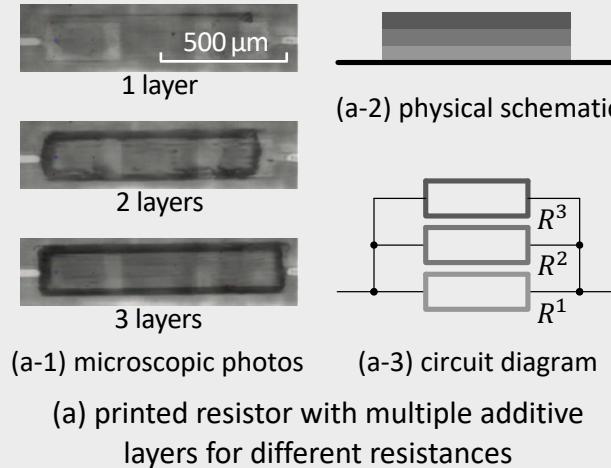
Practicality of the Printed Analog Neuromorphic Circuit

Compactness shows significance when targeting to area-limited scenarios, e.g.,



smart packaging for small fruits
smart bandage on fingers

Compact Circuit Design



Area assessment through the summed area of each component

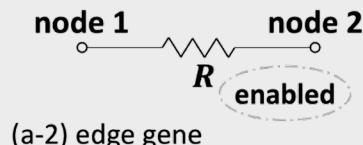
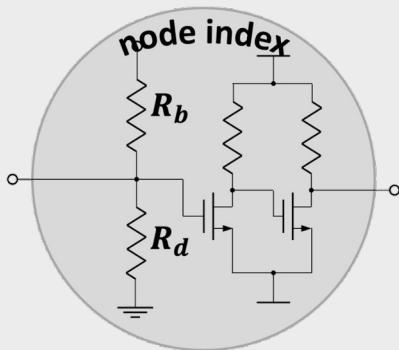
$$A = N^R \cdot A^R + N^A \cdot A^A + N^N \cdot A^N$$

soft-count

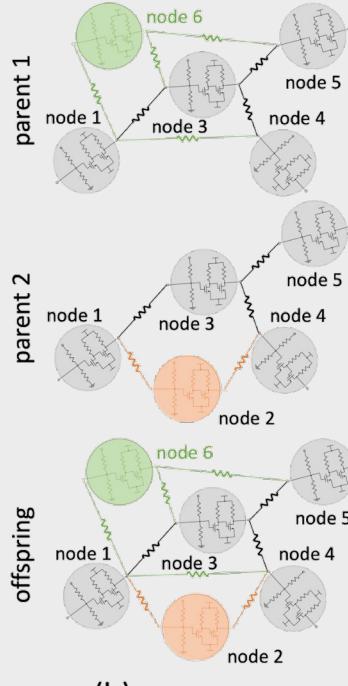
$$A = N_{\text{soft}}^R \cdot A^R + N_{\text{soft}}^A \cdot A^A + N_{\text{soft}}^N \cdot A^N$$

Analogous to **neural network pruning**

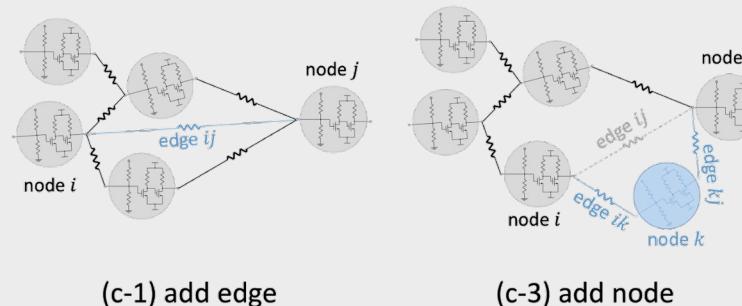
Practicality of the Printed Analog Neuromorphic Circuit



(a) node gene and edge gene

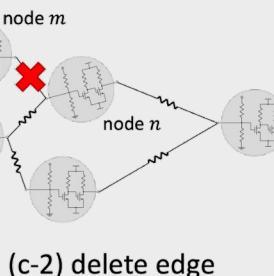


Compact Circuit Design

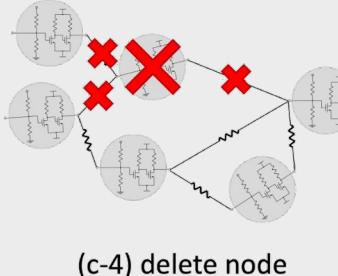


(c-1) add edge

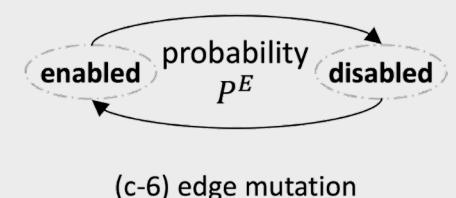
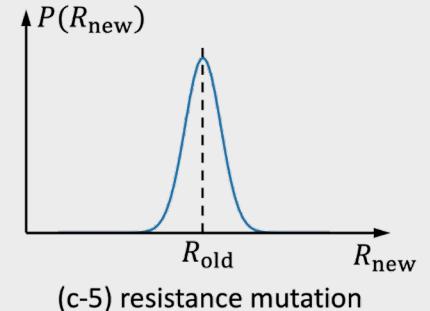
(c-3) add node



(c-2) delete edge



(c) topology and parameter mutation

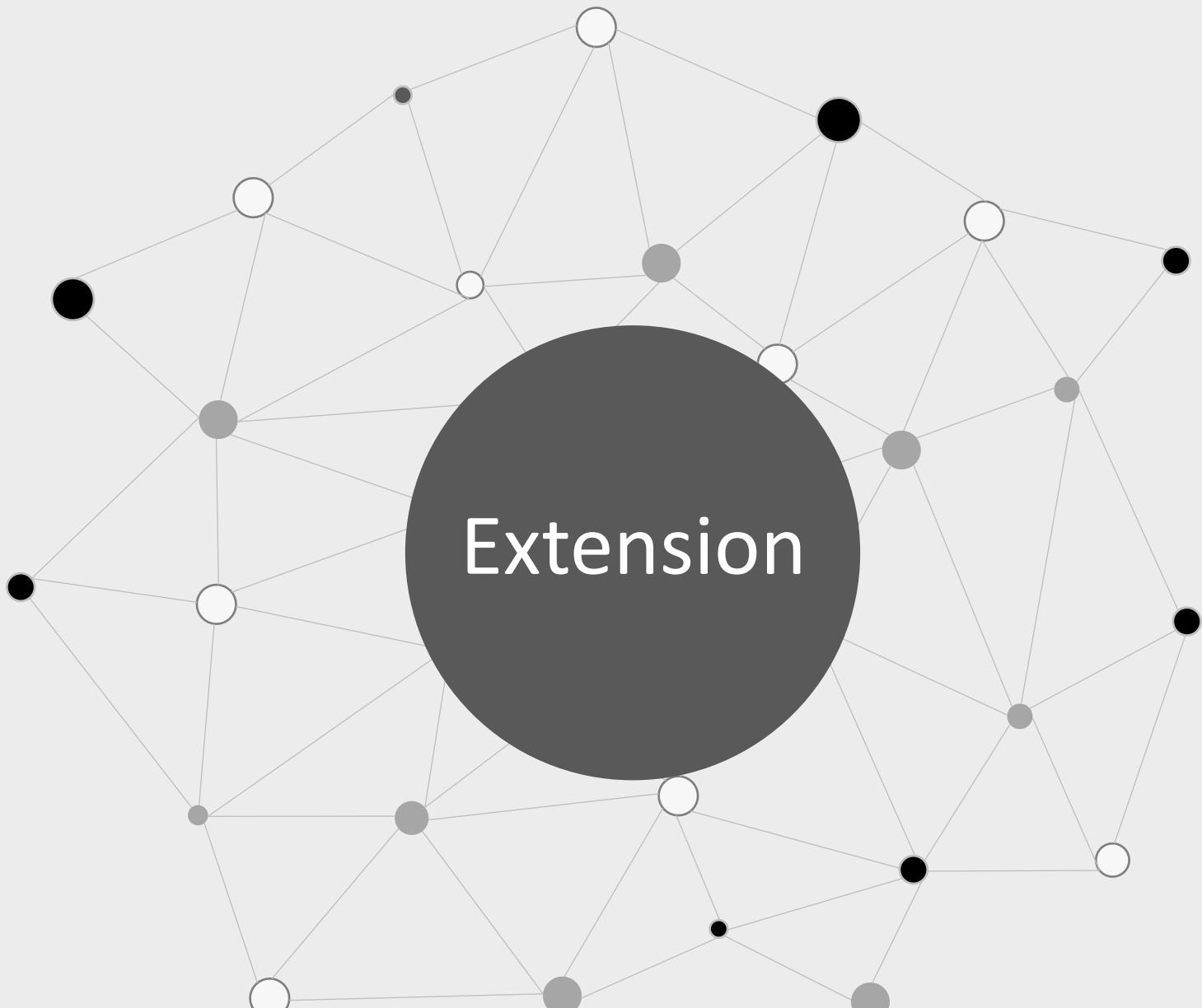


(c-6) edge mutation

Figure: evolutionary architecture search for compact printed analog neuromorphic circuit.

Accuracy		100%	95%	90%	85%
Area	Gradient method	89%	55%	42%	35%
	Evolutionary method	32%	19%	17%	15%

3 times more efficient than
gradient-based method



Modeling

Reliability

Practicality

Extension of the Printed Analog Neuromorphic Circuit

Printed Recurrent Neural Network

Target applications of printed analog neuromorphic circuits



In some cases, classification result depends only on atemporal (non-temporal) data, e.g., meat quality detection.

In other cases, e.g., wound healing detection, due to the difference among individuals, the temporal change of the data is more informative than the specific data.

Lacking of component with time dependencies, existing pNCs are unable to handle temporal data.

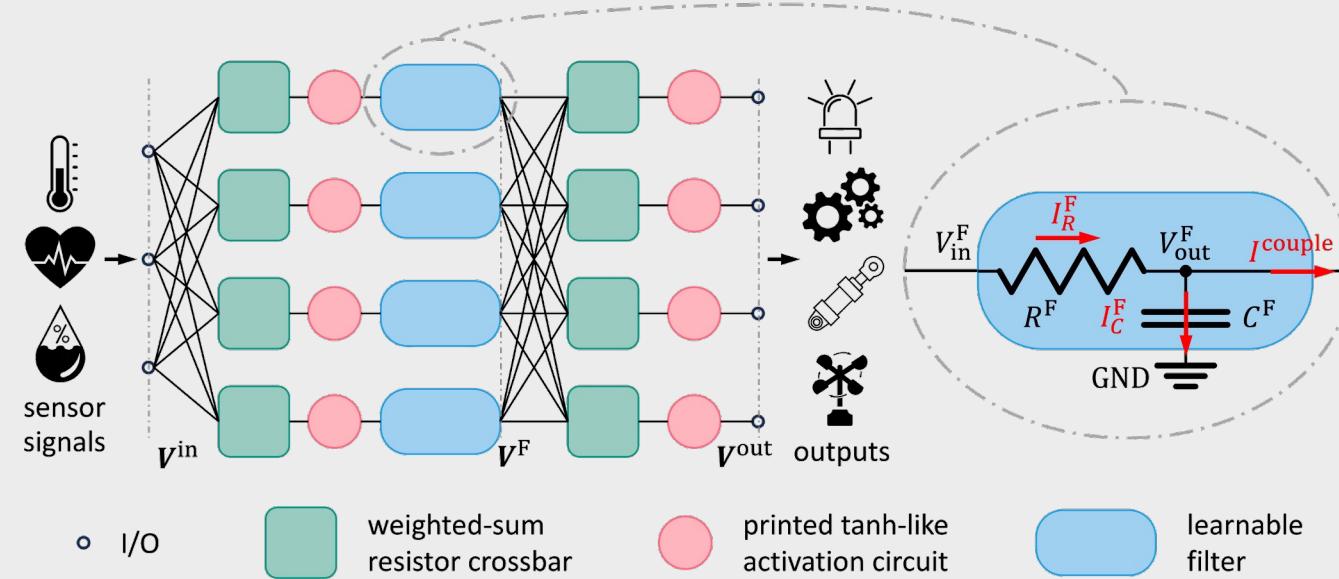


Figure: printed temporal processing block (pTPB).

$$V_k^F = \beta' \odot V_{k-1}^F + (1 - \beta') \odot \text{ptanh}(W_1 V_k^{\text{in}} + b_1)$$
$$V_k^{\text{out}} = \text{ptanh}(W_2 V_k^F + b_2)$$

with signal decay in β' each filter

$$\beta' = \frac{\mu R^F C^F}{\mu R^F C^F + \Delta t}$$

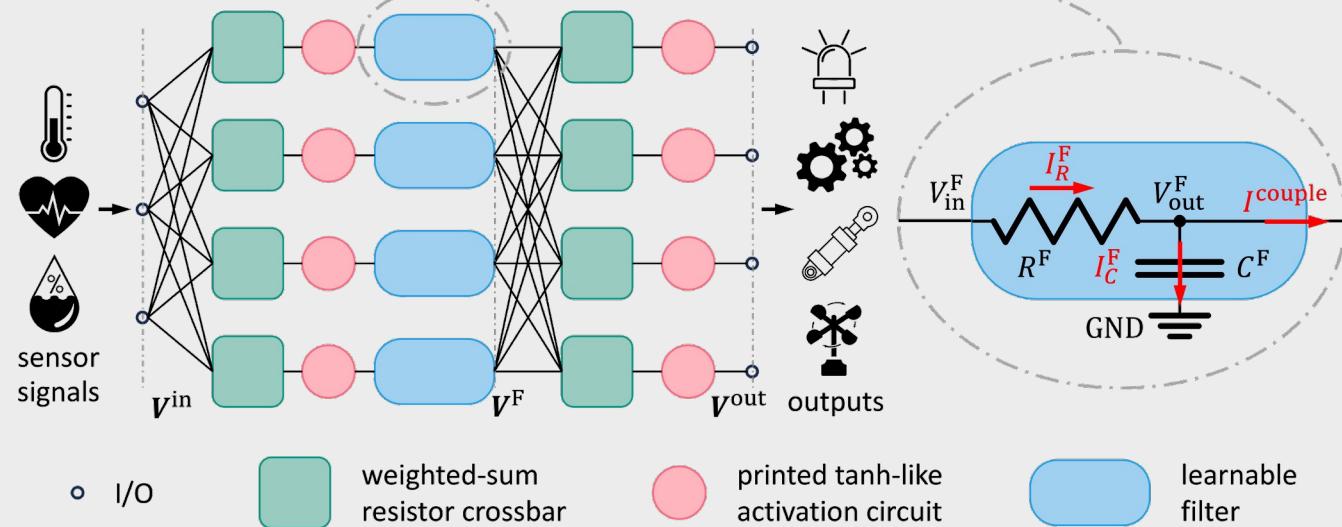
R^F and C^F : learnable parameters

Δt : temporal discretization

μ : coupling between filter and crossbar

Extension of the Printed Analog Neuromorphic Circuit

Printed Recurrent Neural Network



$$V_k^F = \beta' \odot V_{k-1}^F + (1 - \beta') \odot \text{ptanh}(W_1 V_k^{\text{in}} + b_1)$$

$$V_k^{\text{out}} = \text{ptanh}(W_2 V_k^F + b_2)$$

with signal decay in β' each filter

$$\beta' = \frac{\mu R^F C^F}{\mu R^F C^F + \Delta t}$$

R^F and C^F : learnable parameters

Δt : temporal discretization

μ : coupling between filter and crossbar

Expected classification result w.r.t. time

$$\min_{R^F, C^F, \theta} \underbrace{\int_t L\{\hat{y}_t(\beta', x_t), y\} dt}_{\mathcal{L}\{\beta', \theta, x_t, y\}}$$

Overcome dependencies on coupling factor μ and initial voltage caused by preceding signals

$$\min_{R^F, C^F, \theta} \int_{V_0^F} \int_{\mu} \mathcal{L}\{\beta', \theta, x_t, y, \mu, V_0^F\} d\mu dV_0^F$$

Result on 15 benchmark time-series datasets

Model	RG	pNC	RNN	pTPNC
Accuracy	43.7%	50.1%	77.9%	76.4%

Enables temporal data processing

Comparable to hardware-agnostic Elman RNN

Extension of the Printed Analog Neuromorphic Circuit

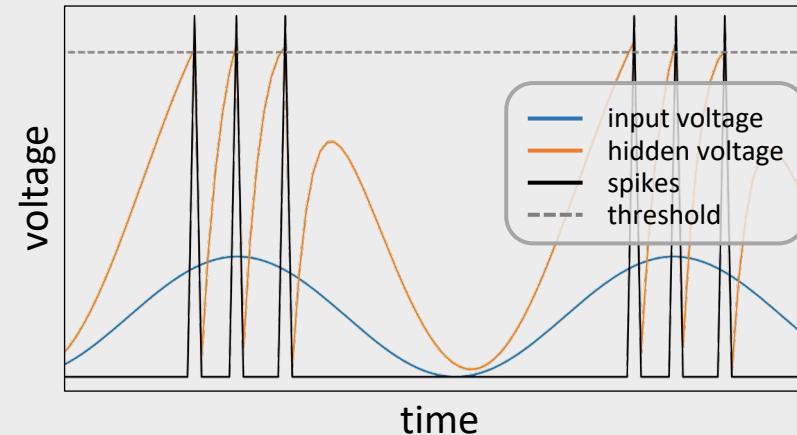
Printed Spiking Neural Network



biological nervous system

event-based data processing
energy-efficient
robust against perturbation

leaky-integration-fire (LIF) neuron



decayed integration -> threshold
-> spikes -> voltage reset

simple in-silico implementation

$$V_k = \beta V_{k-1} + (1 - \beta)x_k$$

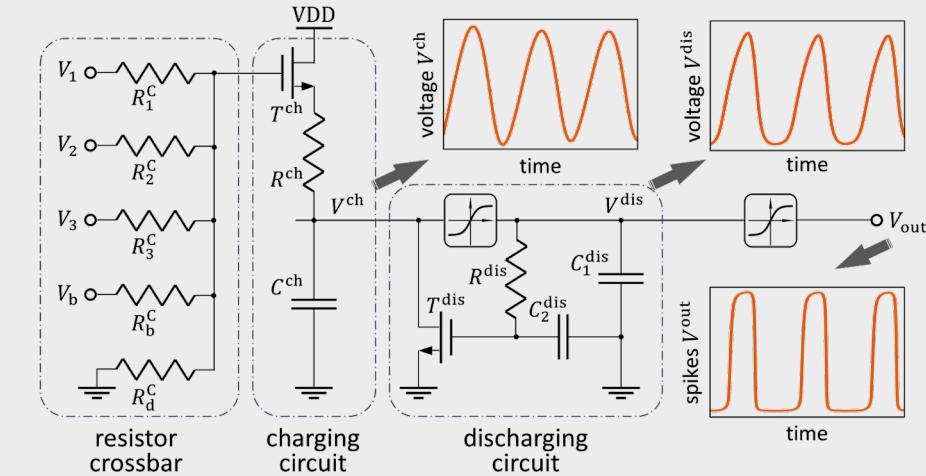
$$y_k = \mathbb{I}_{\{V_k \geq \text{threshold}\}}$$

$$V_k = V_k - \text{threshold} \cdot \mathbb{I}_{\{V_k \geq \text{threshold}\}}$$

requires ideal comparator

requires oscillator

printed spiking neuromorphic circuit



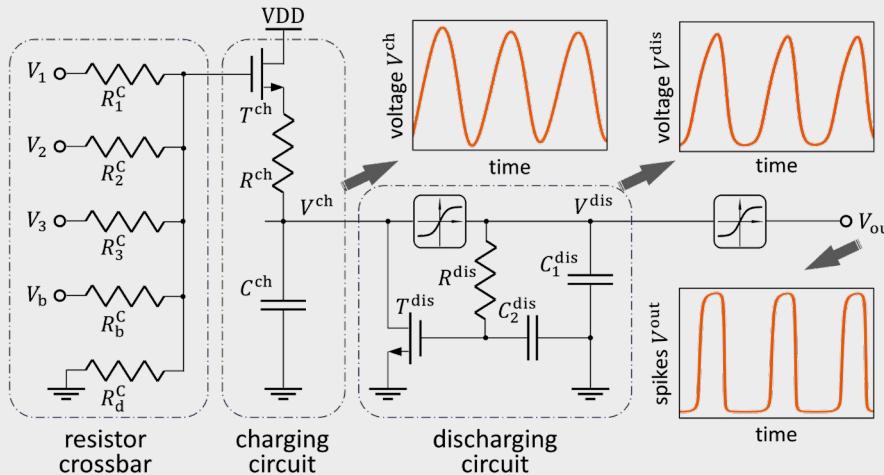
Emulating spikes through
charging & discharging

Complicated system behavior

Extension of the Printed Analog Neuromorphic Circuit

Printed Spiking Neural Network

printed spiking neuromorphic circuit



Emulating spikes through
charging & discharging

Complicated system behavior

Transformer-based surrogate spike-generator model

Transformer is a well-known machine learning model to process sequential data. It is widely used in many emerging applications such as ChatGPT.

1. Prepare various input sequence V_t^{in}
2. Conduct SPICE simulation to obtain output V_t^{out}
3. Train Transformer that can predict output sequence V_t^{out} from the input sequence V_t^{in} , denoted by
$$\hat{V}_t^{\text{out}} = \text{SG}(V_t^{\text{in}})$$
4. Output of the printed spiking neuron
$$\hat{y}_t = \text{SG}(\mathbf{W}_2 \cdot \text{SG}(\mathbf{W}_1 \cdot x_t + \mathbf{b}_1) + \mathbf{b}_2) \dots$$
5. Training objective

$$\min_{\theta} \int_t L\{\hat{y}_t(\theta, x_t), y\} dt$$

3.86 times power saving with similar classification accuracy compared to pNCs

Selected Publications

H Zhao et al. Aging-Aware Training for Printed Neuromorphic Circuits. In Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2022. (Top EDA conference)

H Zhao et al. Highly-dependable printed neuromorphic circuits based on additive manufacturing. In Flexible and Printed Electronics 8.2 (2023), p. 025018. (Impact factor 3.8)

H Zhao et al. Power-Aware Training for Energy-Efficient Printed Neuromorphic Circuits, In Proceedings of the 42nd IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2023. (Top EDA conference)

(submitted) **H Zhao et al.** Neural Evolutionary Architecture Search for Compact Printed Analog Neuromorphic Circuits. In Proceedings of Design Automation Conference (DAC), 2024. (Top EDA conference)

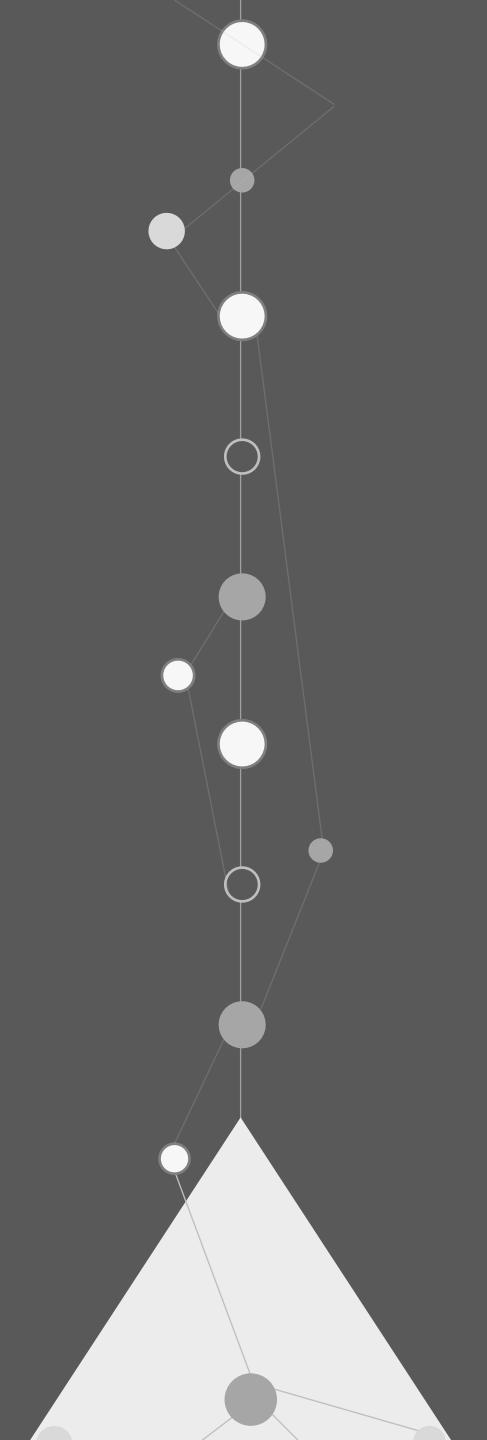
H Zhao et al. Split Additive Manufacturing for Printed Neuromorphic Circuits. In Proceedings of Design, Automation & Test in Europe Conference & Exhibition (DATE), 2023. (Top EDA conference)

H Zhao et al. Highly-Bespoke Robust Printed Neuromorphic Circuits. In Proceedings of Design, Automation & Test in Europe Conference & Exhibition (DATE), 2023. (Top EDA conference)

H Zhao et al. Towards Temporal Information Processing — Printed Neuromorphic Circuits with Learnable Filters. In I Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures (NanoArch), 2023.

P Pal, H Zhao et al. Analog Printed Spiking Neuromorphic Circuits, In Proceedings of Design, Automation & Test in Europe Conference & Exhibition (DATE), 2024. (Top EDA conference)

Bibliography

- 
- [1] E Shirzaei *et al.* A Stretchable Wireless Wearable Bioelectronic System for Multiplexed Monitoring and Combination Treatment of Infected Chronic Wounds. *Science Advances*, 9(12):7388, 2023.
 - [2] C Zhang *et al.* Time-Temperature Indicator for Perishable Products Based on Kinetically Programmable Ag Overgrowth on Au Nanorods. *ACS nano*, 7(5):4561–4568, 2013.
 - [3] D D Weller *et al.* Realization and training of an inverter-based printed neuromorphic computing system. In *Scientific reports* 11.1 (2021): 9554.
 - [4] M Hefenbrock *et al.* In-situ tuning of printed neural networks for variation tolerance. In *Proceedings of 2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2022.
 - [5] M Hefenbrock. Modelling and Training Printed Neuromorphic Circuits. Dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), 2022.