IEEE/ACM
2022 INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN
ICCAD
41st Edition

Aging-Aware Training for Printed Neuromorphic Circuits

Haibin Zhao[1], Michael Hefenbrock[1], Michael Beigl[1], Mehdi B. Tahoori[2]

[1]Chair for Pervasive Computing Systems
Karlsruhe Institute of Technology
Karlsruhe, Germany

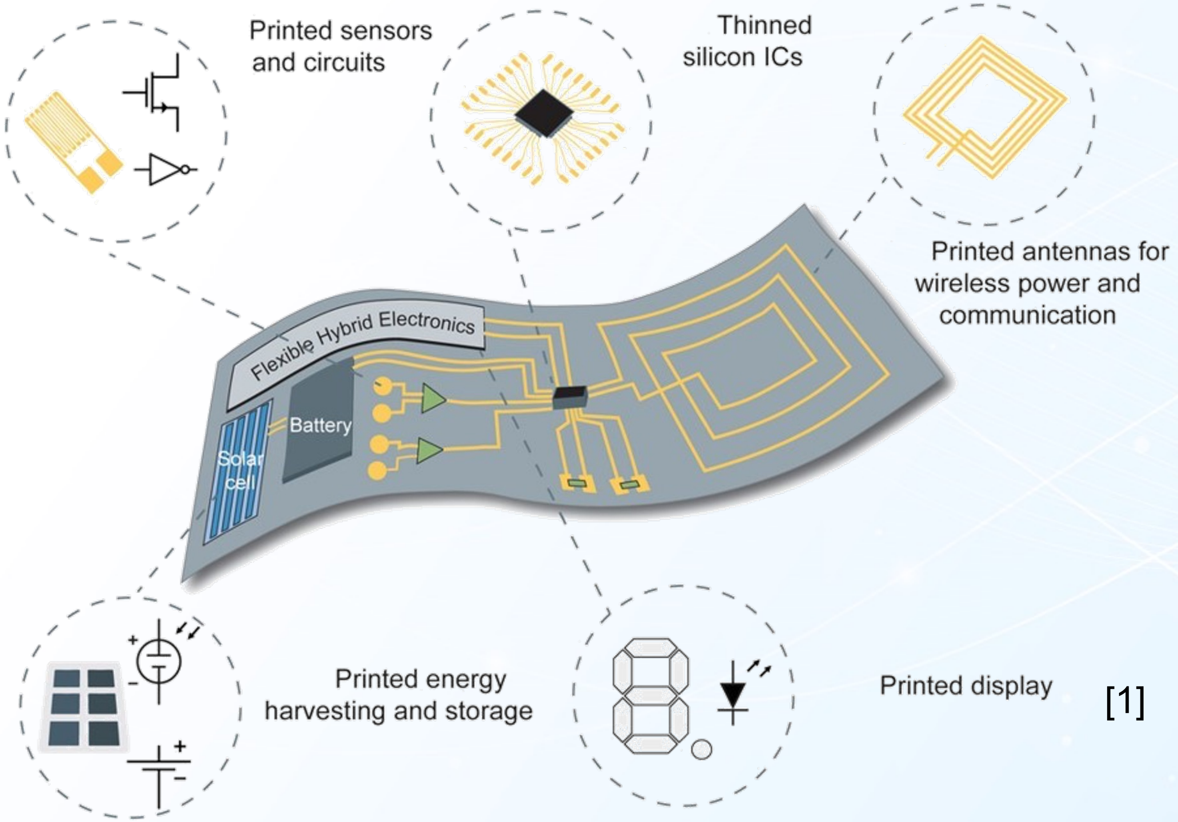
[2]Chair of Dependable Nano Computing
Karlsruhe Institute of Technology
Karlsruhe, Germany

Outline

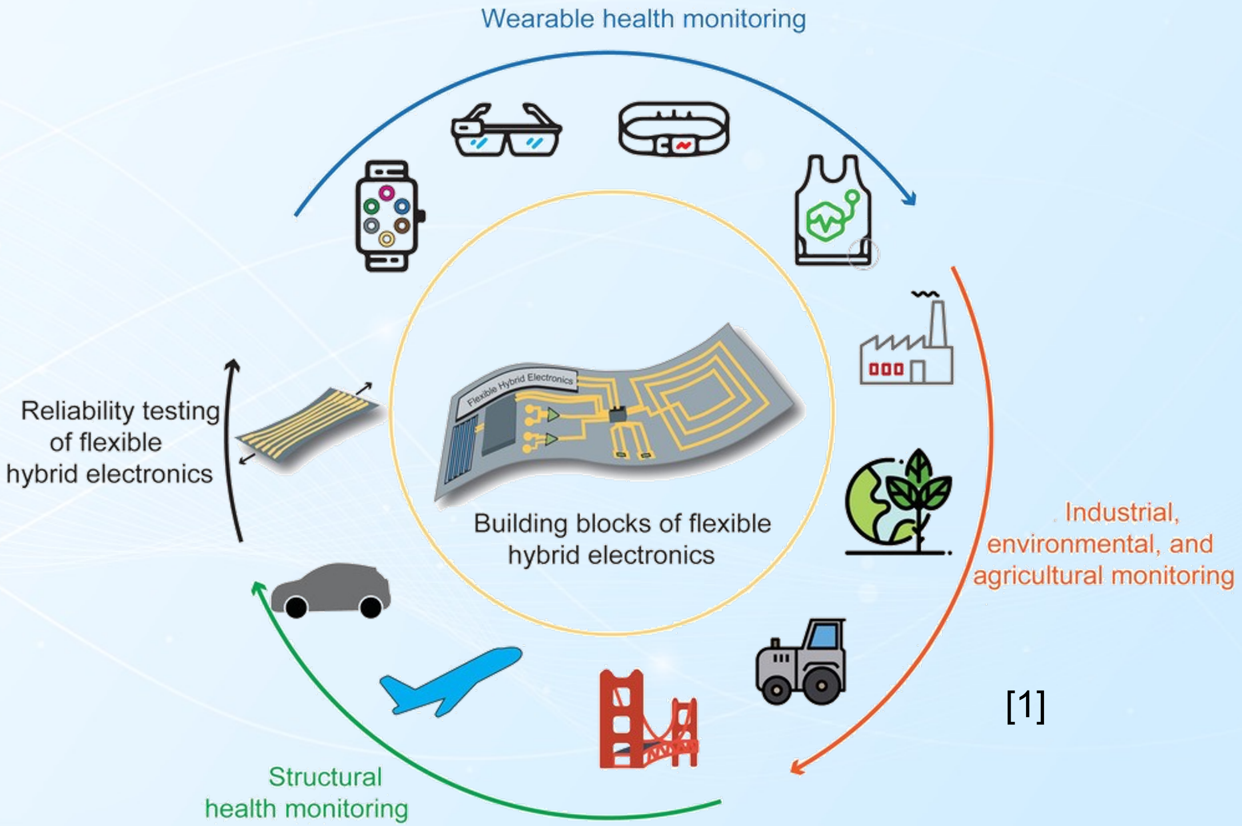
- Printed Electronics
- Printed Neural Networks
- Aging-aware Training
- Results
- Conclusions

Printed Electronics - Overview

➢ *Producing electronic devices by printing them on flexible or rigid substrates*



[1]



[1]

[1] Khan, Y., Thielens, A., Muin, S., Ting, J., Baumbauer, C., & Arias, A. C. “A new frontier of printed electronics: flexible hybrid electronics”. *Advanced Materials,* 2020

Printed Electronics vs Silicon-based Technologies

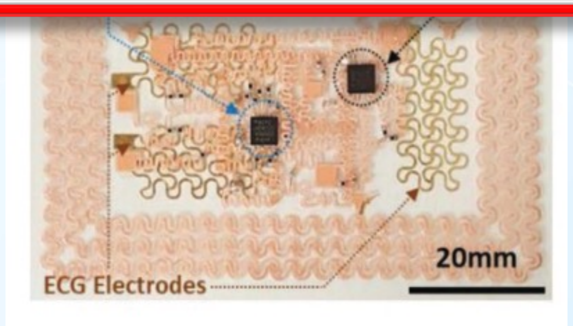
- Conventional Electronics
  - + High performance
  - + Low area
  - – High per-unit manufacturing cost

[Apple]

Complementary Technologies

- Target different application domains
- Offer different advantages

- Printed Electron
  - + Low fabricati
  - + Low Non-Rec
    Engineering (NRE) cost
  - + Flexibility
  - – Large feature sizes
  - – High power consumption



[Khan2020]

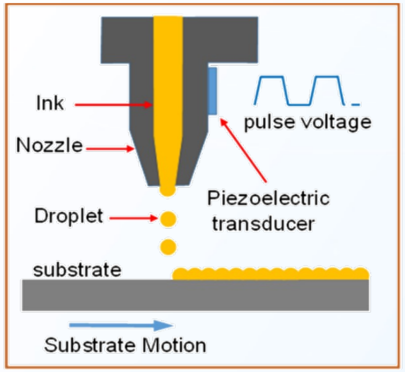
[Bonna2018]

[Bonna2018]

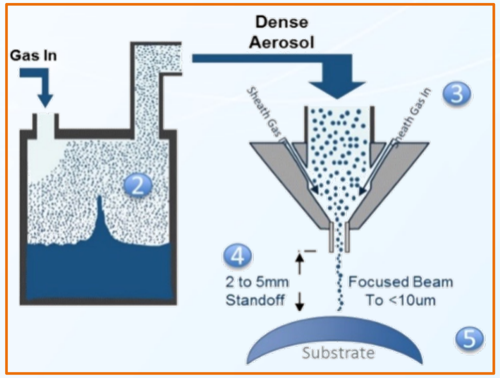
Printed Electronics - Manufacturing

- Simple manufacturing process
- Additive processes for reduced fabrication costs

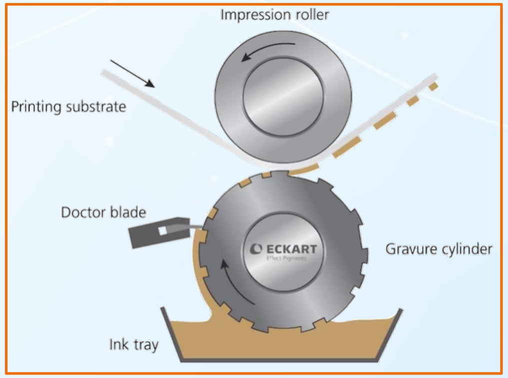
	Conventional	Printed
Switching times	Short	Long
Integration density	High	Low
Area usage	Small	Large
Substrate type	Rigid	Flexible/Rigid
Fabrication complexity	Sophisticated	Simple
Fabrication cost	High	Low



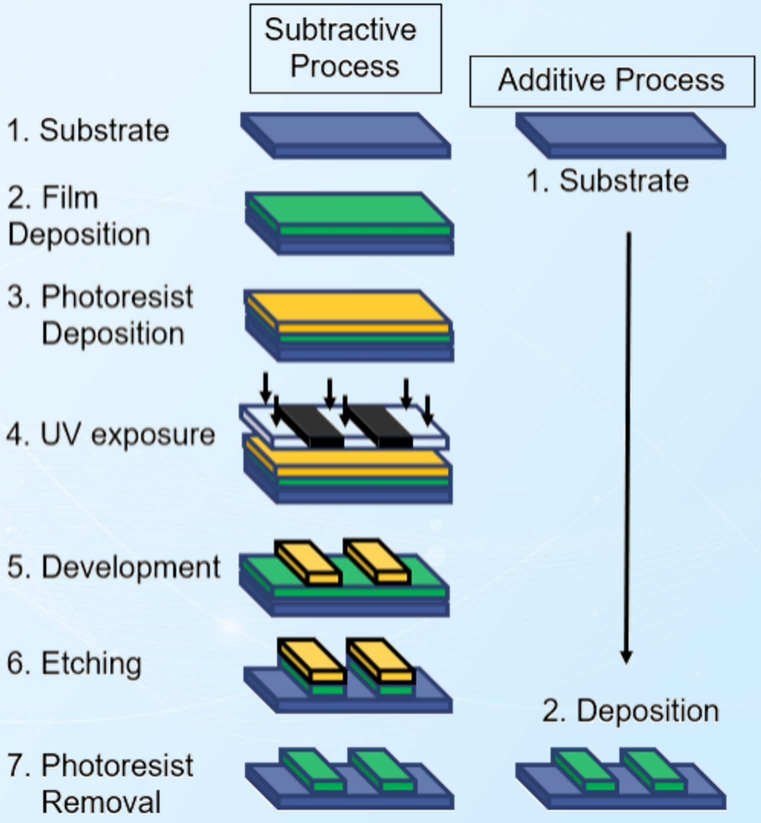
Inkjet Printing



Aerosol Jet Printing

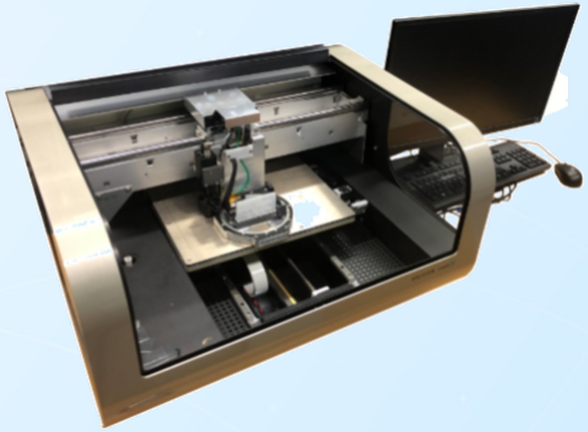


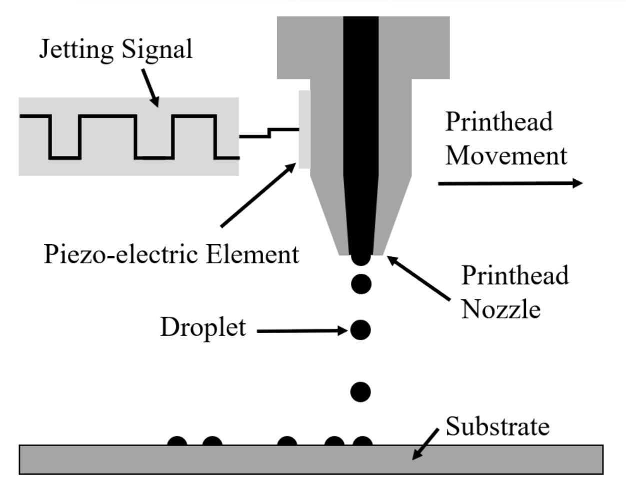
Gravure Printing

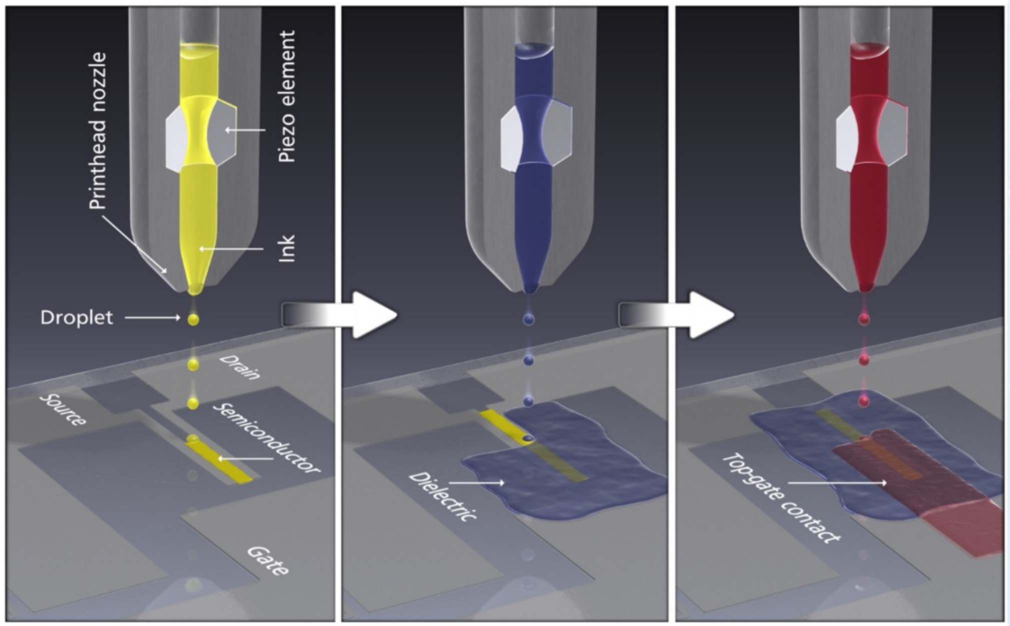


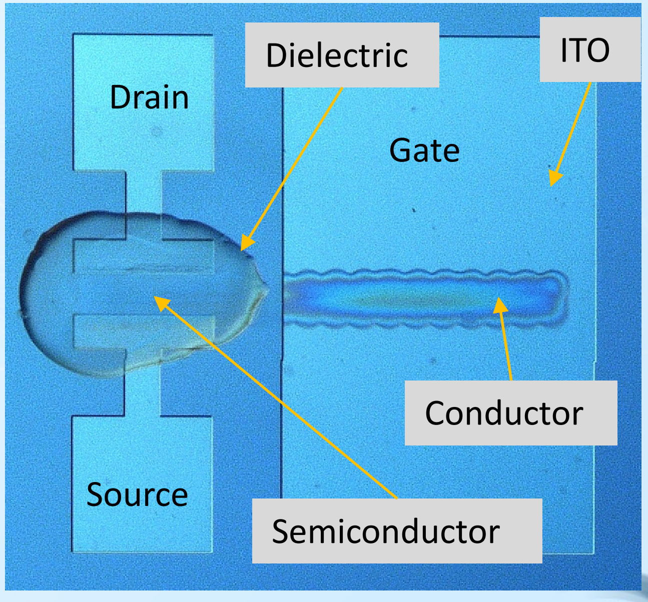
Inkjet-Printed Circuits

- Mask-less, non-contact and point-of-use manufacturing process
  - Low-cost fabrication
  - Printing on wide range of (flexible) substrates
  - Digital printing (enables highly customized designs)
  - Low voltage operation with possible with electrolyte-based oxide transistors









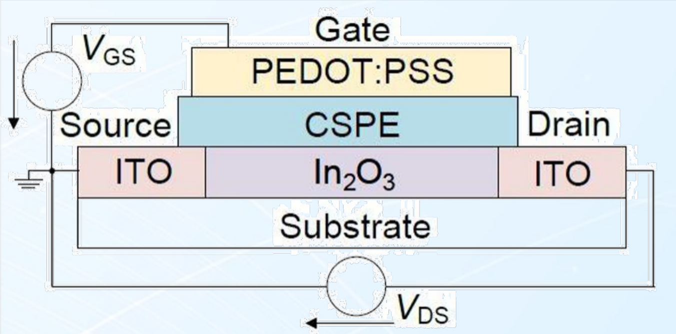
Electrolyte-gated transistors (EGTs)

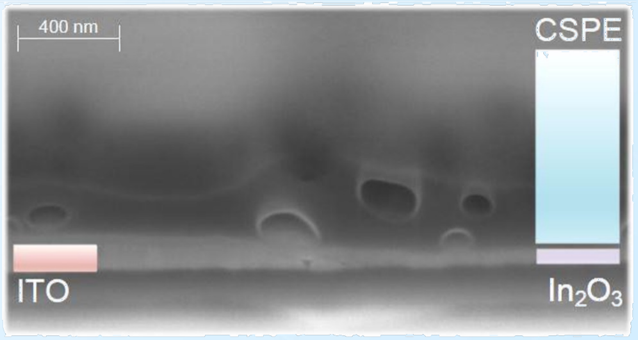
Thin-film transistor

- Signal routing: Indium tin oxide (ITO)
- Semiconductor: indium oxide (In_2O_3)
- Gate insulator: Composite solid polymer electrolyte (CSPE)
- Top gate: PEDOT:PSS

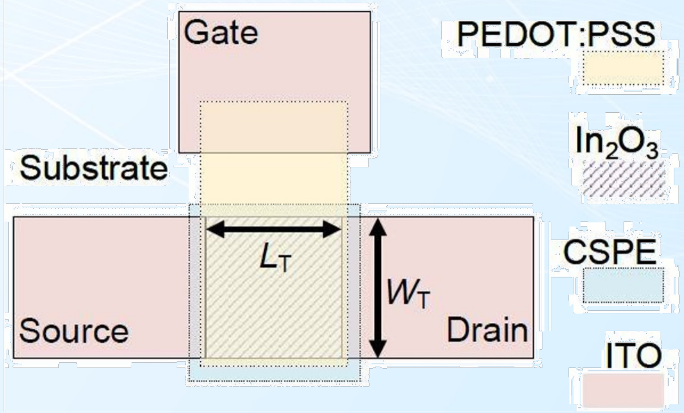
- Voltage levels: $\lesssim 1.5$ V, $\approx 100\ \mu A - 1\ mA$
- Frequency range: 100 Hz – 1 kHz

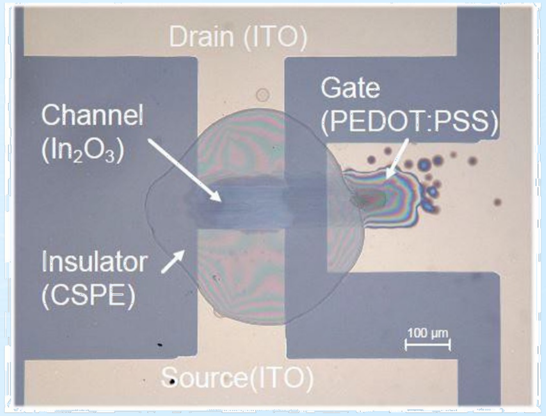
Cross section





Top view



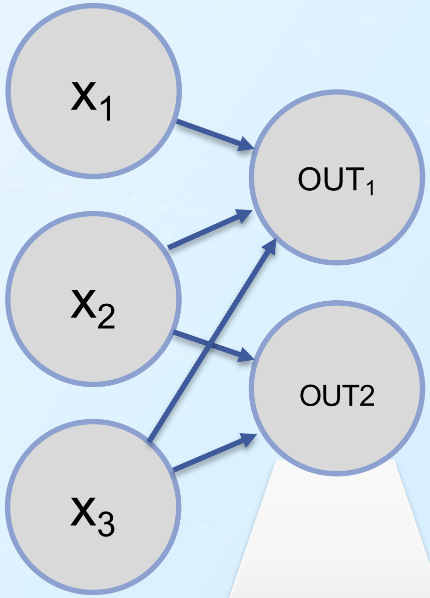


Outline

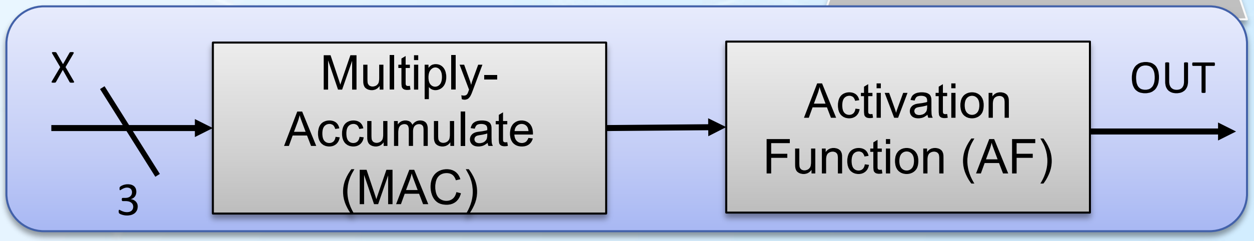
- Printed Electronics
- **Printed Neural Networks**
- Aging-aware Training
- Results
- Conclusions

Printed Artificial Neural Network (ANN)

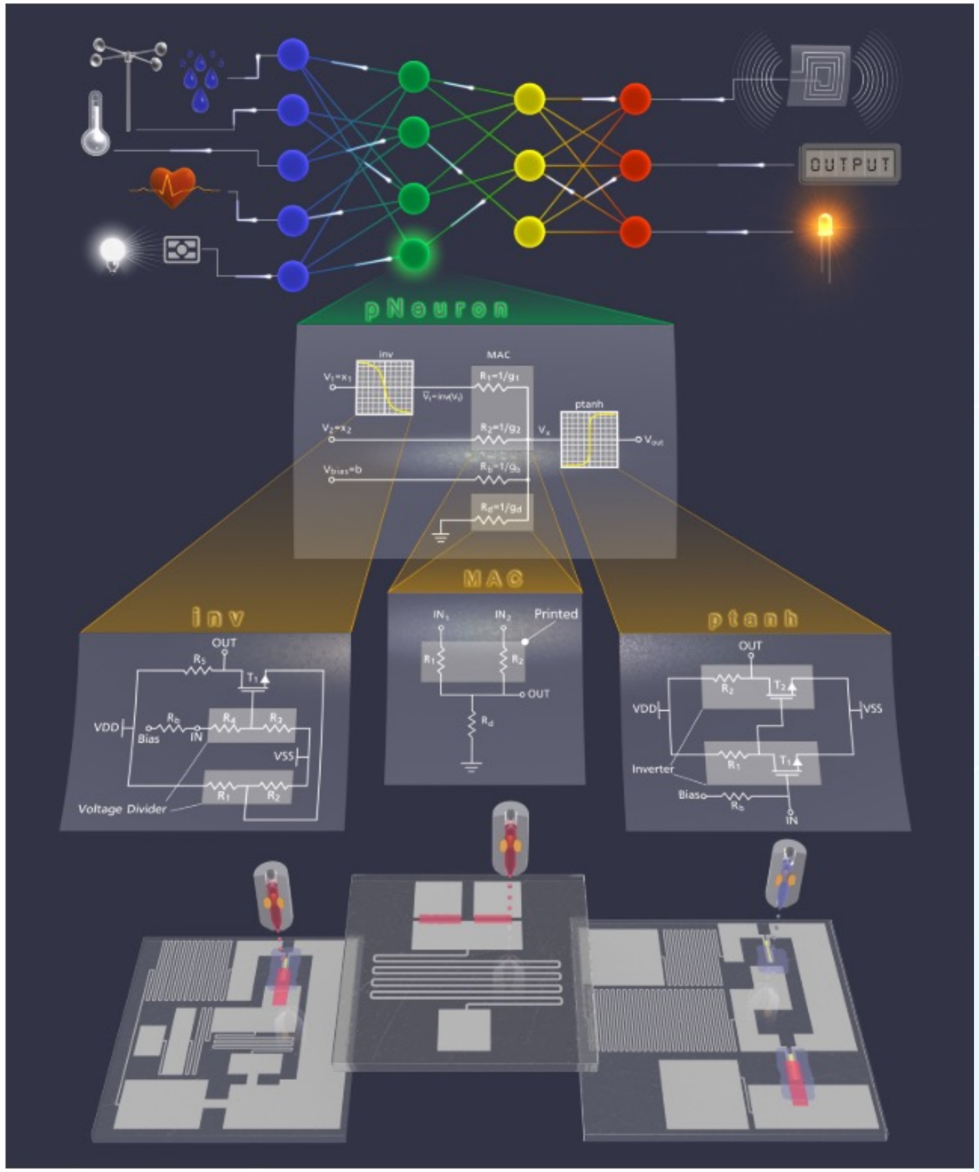
- Conventional designs for machine learning (ML) are infeasible
- Analog (neuromorphic) NN design was developed

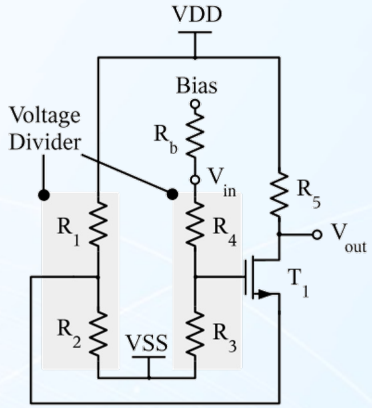


Components	Number of transistors	
	4-bit Digital ANN	**Proposed Analog ANN**
Input Converter	185	**-**
MAC	265	$\leq$ **4**
AF	10	**2**

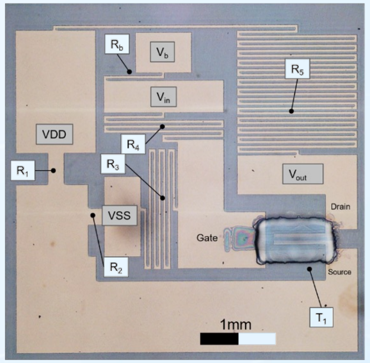


Printed Artificial Neural Network (ANN)

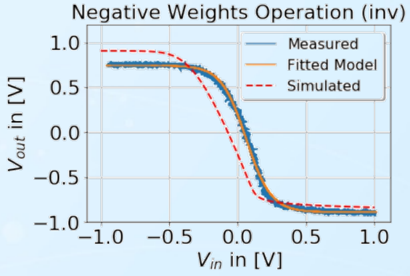




(a) Schematic

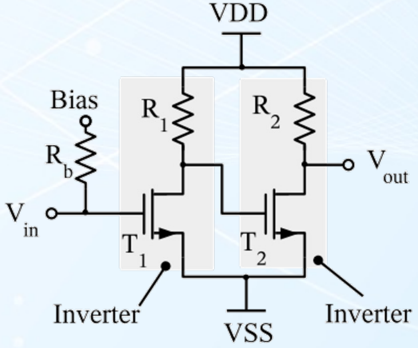


(b) Microscopic Photo

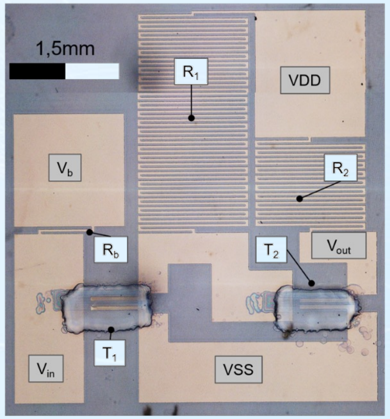


(c) Measured Waveforms

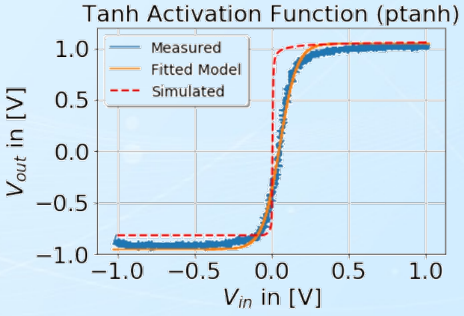
$$\mathrm{inv}(x) = -(a + b \cdot \tanh((x - c) \cdot d)),$$



(a) Schematic



(b) Microscopic Photo



(c) Measured Waveforms

$$\mathrm{ptanh}(x) = a + b \cdot \tanh((x - c) \cdot d)$$

Neuromorphic Circuit

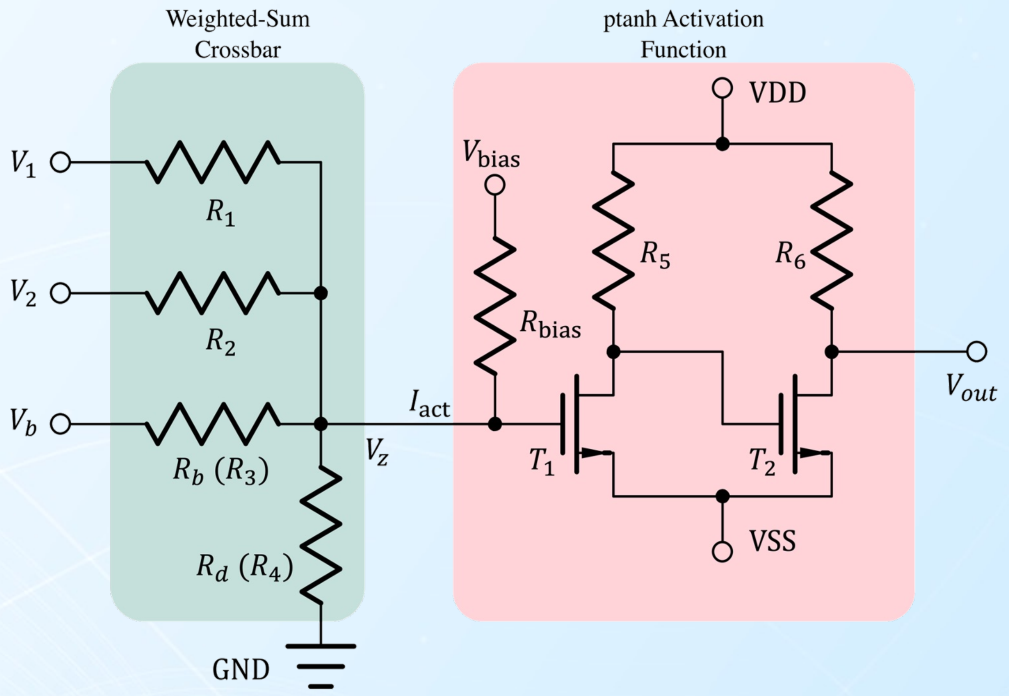
- Crossbar

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

Express resistance R_i by conductance $1/g_i$,
let G be the sum of g_i,
set $V_b \equiv 1V$.

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

V_z is the weighted-sum of input voltages V_i



Circuit schematic of a neuron.

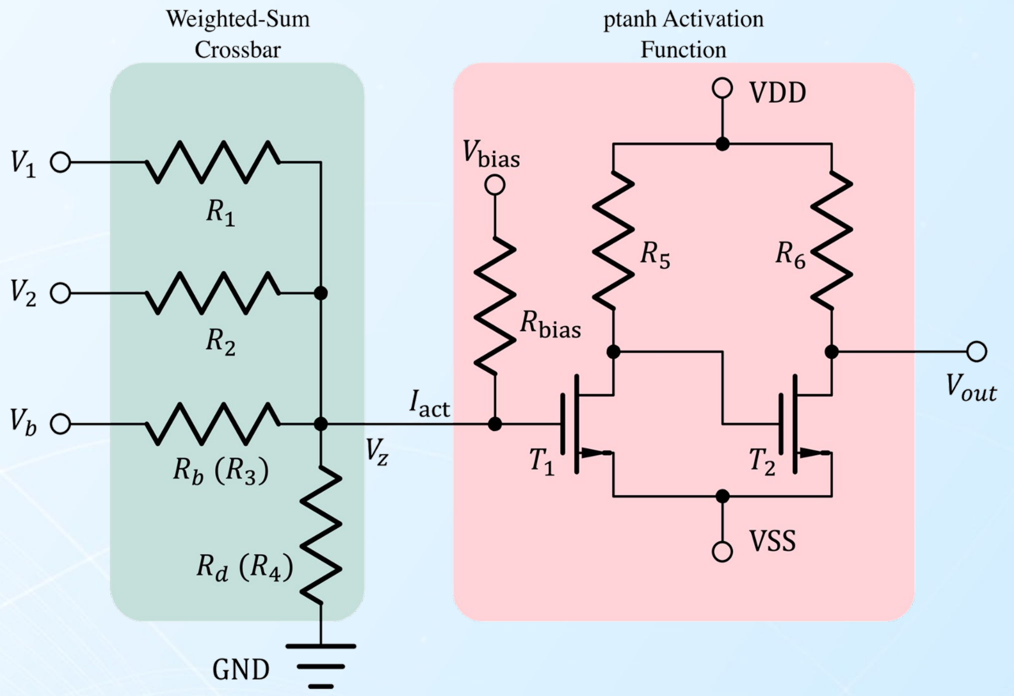
Neuromorphic Circuit

- Crossbar

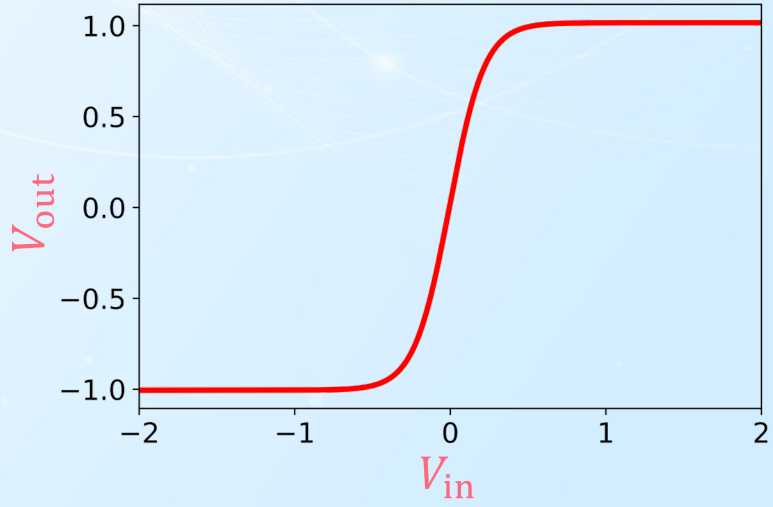
$$V_z = \frac{g_1}{G} V_1 + \frac{g_2}{G} V_2 + \frac{g_b}{G}$$

- printed activation function

$$\begin{aligned} V_{\text{out}} &= \text{ptanh}(V_{\text{in}}) \\ &= \eta_1 + \eta_2 \cdot \tanh\big((x - \eta_3) \cdot \eta_4\big) \end{aligned}$$



Circuit schematic of a neuron.



Neuromorphic Circuit

- Crossbar

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

- printed activation function

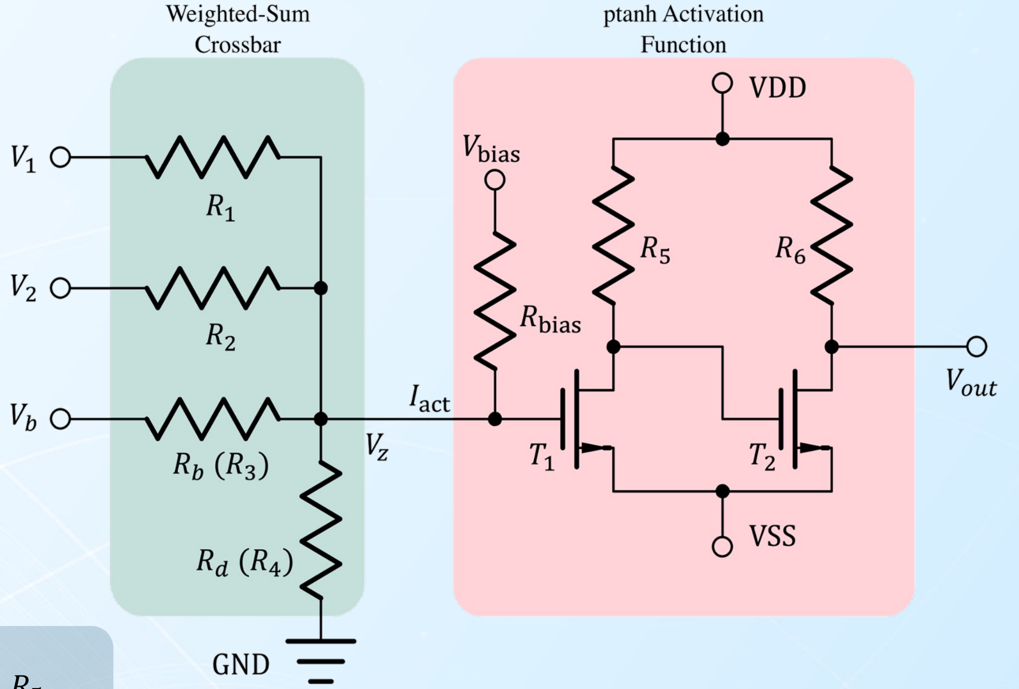
$$V_{\mathrm{out}} = \mathrm{ptanh}(V_{\mathrm{in}})$$
$$= \eta_1 + \eta_2 \cdot \tanh\big((x - \eta_3) \cdot \eta_4\big)$$

- negative weight circuit

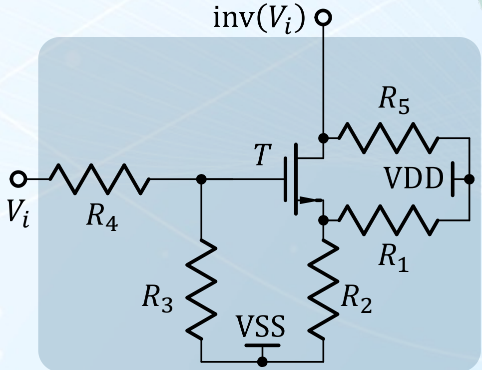
$$V_{\mathrm{out}} = \mathrm{inv}(V_{\mathrm{in}})$$
$$= -\Big(\eta_1 + \eta_2 \cdot \tanh\big((x - \eta_3) \cdot \eta_4\big)\Big)$$

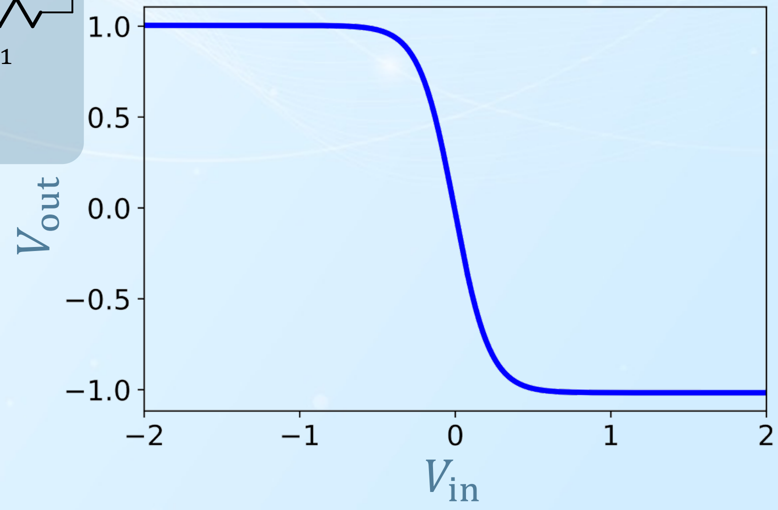
If negative weight is required:

$$\left(-\frac{g_1}{G}\right) V_1 \leftarrow \frac{g_1}{G} \mathrm{inv}(V_1)$$



Circuit schematic of a neuron.





Neuromorphic Circuit

- Crossbar

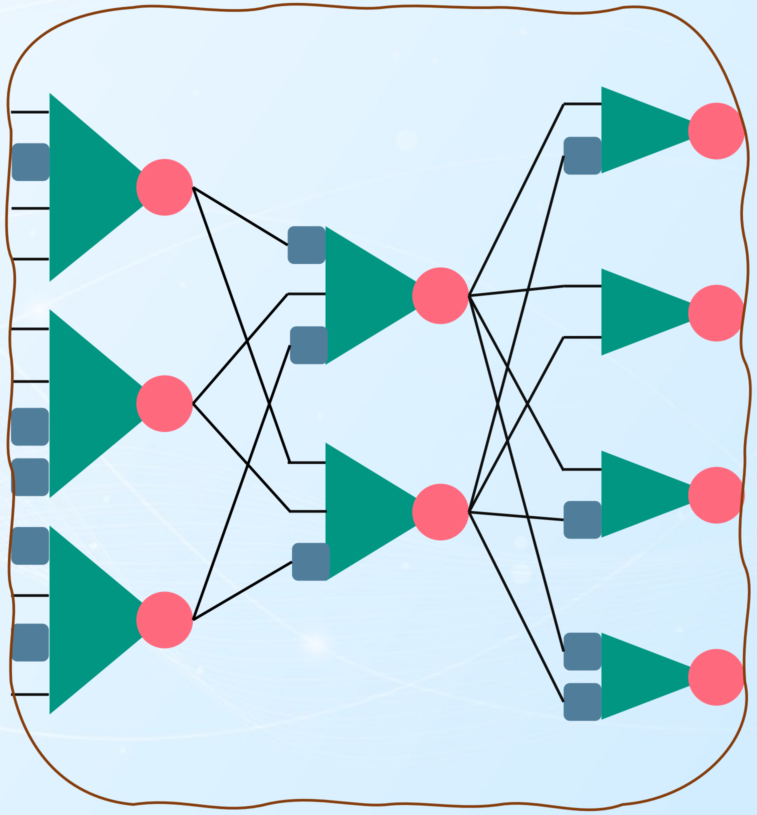
$$V_z = \frac{g_1}{G} V_1 + \frac{g_2}{G} V_2 + \frac{g_b}{G}$$

- printed activation function

$$\begin{aligned} V_{\text{out}} &= \text{ptanh}(V_{\text{in}}) \\ &= \eta_1 + \eta_2 \cdot \tanh\big((x - \eta_3) \cdot \eta_4\big) \end{aligned}$$

- negative weight circuit

$$\begin{aligned} V_{\text{out}} &= \text{inv}(V_{\text{in}}) \\ &= -\Big(\eta_1 + \eta_2 \cdot \tanh\big((x - \eta_3) \cdot \eta_4\big)\Big) \end{aligned}$$



Example of a neuromorphic circuit with 3-2-4 neurons

Printed Neural Network (pNN)

- A machine learning based model of printed neuromorphic circuits
  - Models the circuit input and output
    - learnable parameter: surrogate conductance $\boldsymbol{\theta} \in \mathbb{R}$

$$\theta_i = \mathrm{sign}(\theta_i) \cdot |\theta_i| = \mathrm{sign}(\theta_i) \cdot g_i$$

weights: $w_i = \frac{g_i}{\sum_j g_j},\ g_i = |\theta_i|$

inputs: $\begin{cases} V_{\mathrm{in}}, & \mathrm{sign}(\theta_i) \geq 0 \\ \mathrm{inv}(V_{\mathrm{in}}), & \mathrm{sign}(\theta_i) < 0 \end{cases}$

weighted-sum: $V_{\mathrm{z}} = \sum_{\theta_i \geq 0} w_i \cdot V_{\mathrm{in}} + \sum_{\theta_i < 0} w_i \cdot \mathrm{inv}(V_{\mathrm{in}})$

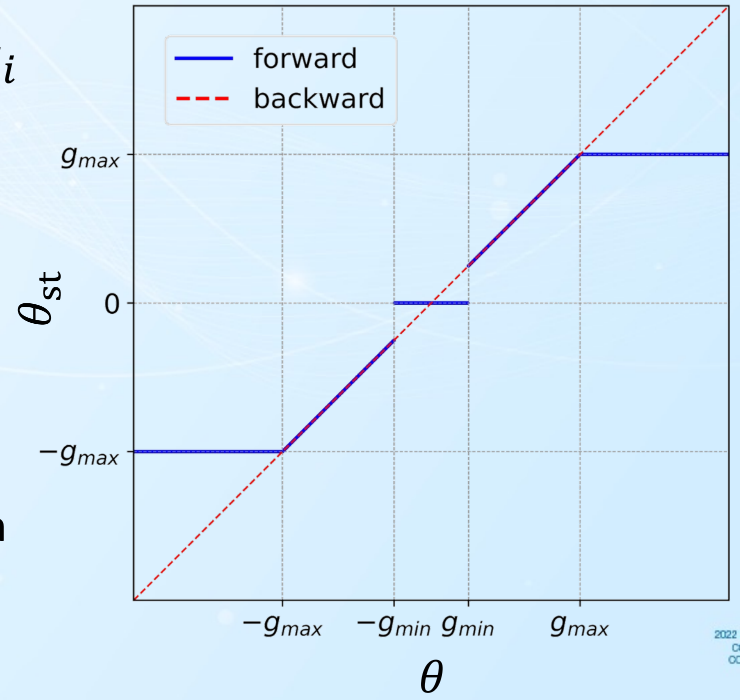
    - training pNN is to find the best design parameters in neuromorphic circuits

Printed Neural Network (pNN)

- A machine learning based model of printed neuromorphic circuits
  - Models the circuit input and output
    - learnable parameter: surrogate conductance $\boldsymbol{\theta} \in \mathbb{R}$
    - training pNN is to find the best design parameters in neuromorphic circuits

$$\theta_i = \mathrm{sign}(\theta_i) \cdot |\theta_i| = \mathrm{sign}(\theta_i) \cdot g_i$$

  - Considers hardware constraints
    - Printing technology $g_i = |\theta_i| \in [\mathrm{g_{min}}, \mathrm{g_{max}}]$
      - Straight-through estimator:
        - forward: project $\boldsymbol{\theta}$ in feasible range ($\boldsymbol{\theta}_{\mathrm{st}}$)
        - drawback: zero gradient in some area
        - solution: manual gradient to guide the optimization



Printed Neural Network (pNN)

- A machine learning based model of printed neuromorphic circuits
  - Models the circuit input and output
    - learnable parameter: surrogate conductance $\boldsymbol{\theta} \in \mathbb{R}$
    - training pNN is to find the best design parameters in neuromorphic circuits

$$\theta_i = \text{sign}(\theta_i) \cdot |\theta_i| = \text{sign}(\theta_i) \cdot g_i$$

  - Considers hardware constraints
    - Printing technology $g_i = |\theta_i| \in [\mathrm{g}_{\min}, \mathrm{g}_{\max}]$
    - Printing error (variation-aware training)
      - model the deterministic $\boldsymbol{\theta}$ as stochastic variable $\boldsymbol{\varepsilon} \cdot \boldsymbol{\theta}$, with each $\boldsymbol{\varepsilon}_{\boldsymbol{i},\boldsymbol{j}} \sim \mathcal{U}\{1-\delta, 1+\delta\}$
      - objective

$$\underset{\boldsymbol{\theta}}{\text{minimize}}\, L \mapsto \underset{\boldsymbol{\theta}}{\text{minimize}}\, \mathbb{E}_{\boldsymbol{\varepsilon}}\{L\}$$

Printed Neural Network (pNN)

- A machine learning based model of printed neuromorphic circuits
  - Models the circuit input and output
    - learnable parameter: surrogate conductance $\boldsymbol{\theta} \in \mathbb{R}$
    - training pNN is to find the best design parameters in neuromorphic circuits

$$\theta_i = \mathrm{sign}(\theta_i) \cdot |\theta_i| = \mathrm{sign}(\theta_i) \cdot g_i$$

  - Considers hardware constraints
    - Printing technology $g_i = |\theta_i| \in [\mathrm{g_{min}}, \mathrm{g_{max}}]$
    - Printing error (variation-aware training)
    - **Aging of printed resistors**

Outline

- Printed Electronics
- Printed Neural Networks
- **Aging-aware Training**
- Results
- Conclusions

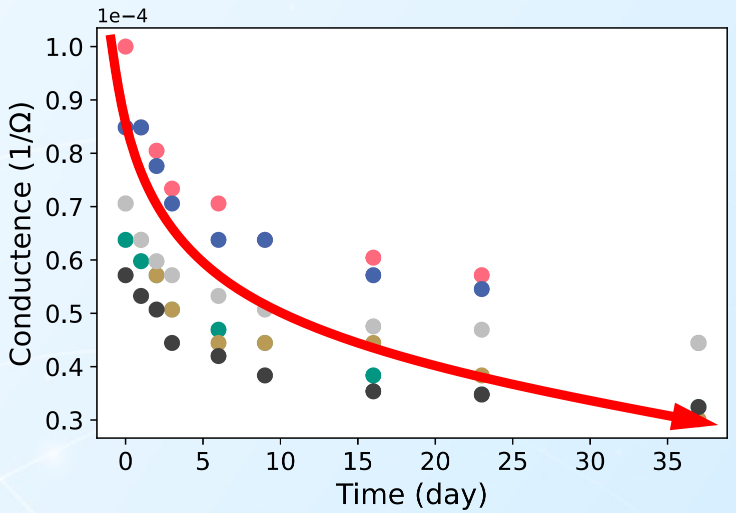
Aging in Printed Electronics

- Aging effect
  - degradation of printed components
  - changes of component values

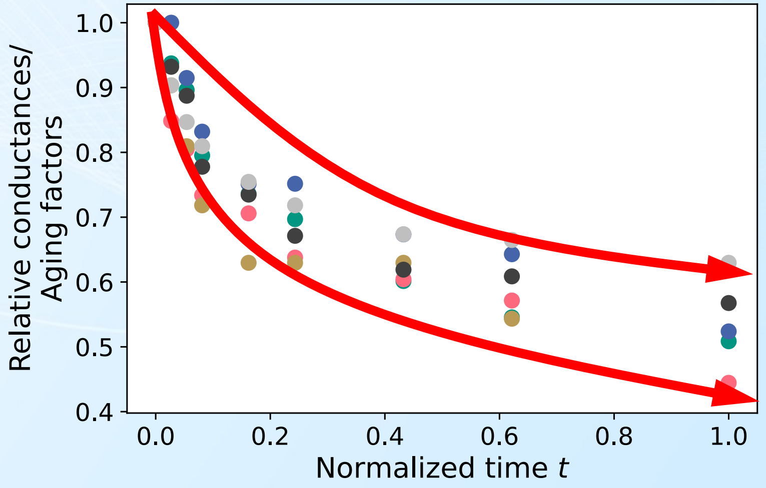
- Pipeline to model aging behavior

Measurements of printed components over time → Choose functional form $A_{\boldsymbol{\omega}}(t)$ → Extract aging model coefficients $\boldsymbol{\omega}$ → Model coefficient distributions

Aging of Printed Resistors



- Data collection
  - 6 printed resistors
  - 37-day measurement



- 2-stage aging behavior (similar to work [1,2])
  - fast degradation phase
  - gradual phase

- Aging model

$$\theta(t) = \theta_0 \cdot A_\omega(t)$$

where $A_\omega(t) = \omega_1 \cdot e^{-\omega_2 \cdot t} - \omega_1 + 1$

[1] Rzasa, et al. Thick film resistors on dielectrics as temperature detectors. *Active and passive electronic components* 12.2 (1986): 137-147.
[2] Hamasha, et al. 2012. Stability of ITO thin film on flexible substrate under thermal aging and thermal cycling conditions. *Journal of Display Technology* 8, 7 (2012), 385–390

Aging of Printed Resistors

- Methodology
  - Aging model

$$\theta(t) = \theta_0 \cdot A_\omega(t)$$

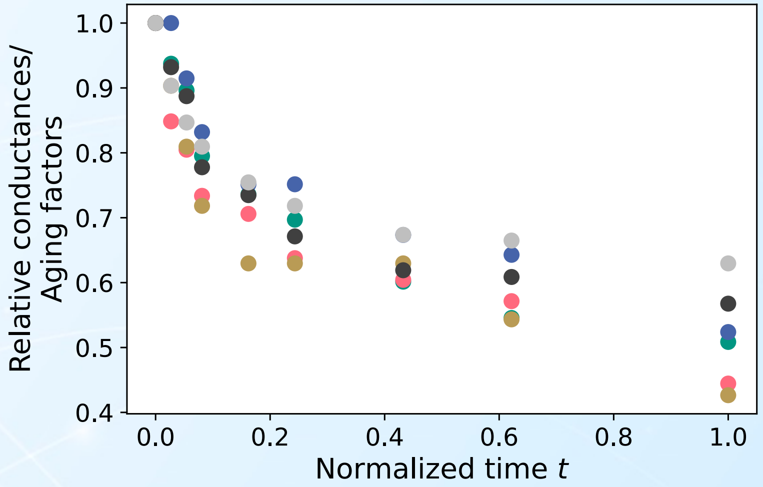
$$\text{where } A_\omega(t) = \omega_1 \cdot e^{-\omega_2 \cdot t} - \omega_1 + 1$$

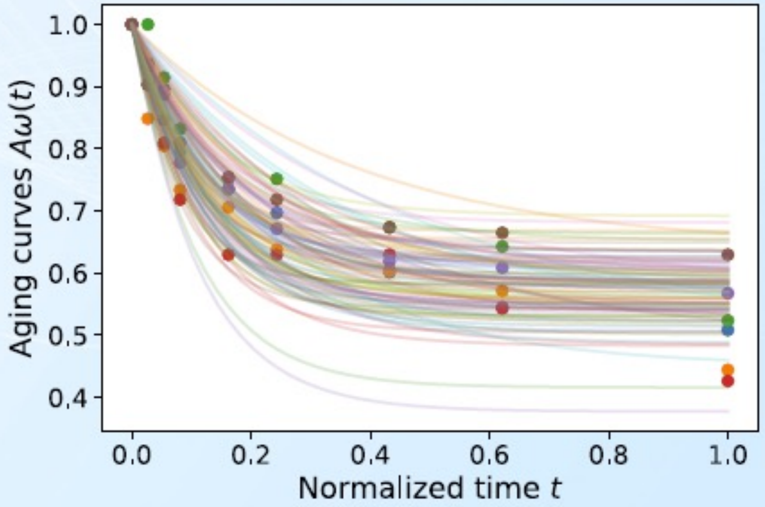
  - Stochastic aging model

$$\theta(t) = \theta_0 \cdot A_\omega(t)$$

$$\text{where } A_\omega(t) = \omega_1 \cdot e^{-\omega_2 \cdot t} - \omega_1 + 1$$

with $\omega \sim p(\omega)$, the log-normal distribution fitted from printed resistors



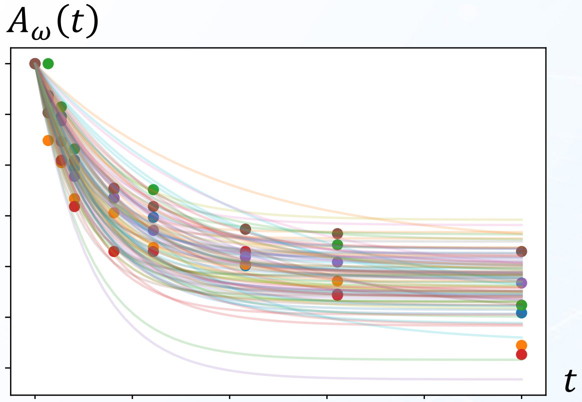


Objective of Aging-Aware Training

model		**Objective**
nominal (non aging-aware)	θ	$\min_{\theta} \mathcal{L}(\hat{y}(x, \theta), y)$
with aging-aware	$\theta(t)$	$\min_{\theta(t)} \int_0^1 \mathcal{L}(\hat{y}(x, \theta(t)), y) \mathrm{d}t$
with aging model	$\theta(t) = \theta_0 \cdot A_\omega(t)$	$\min_{\theta_0} \int_0^1 \mathcal{L}(\hat{y}(x, \theta_0, A_\omega(t)), y) \mathrm{d}t$
with stochastic aging model	$\theta(t) = \theta_0 \cdot A_\omega(t)$ $\omega \sim p(\omega)$	$\min_{\theta_0} \int_\omega \int_0^1 \mathcal{L}(\hat{y}(x, \theta_0, A_\omega(t)), y) \mathrm{d}t\, p(\omega) \mathrm{d}\omega$

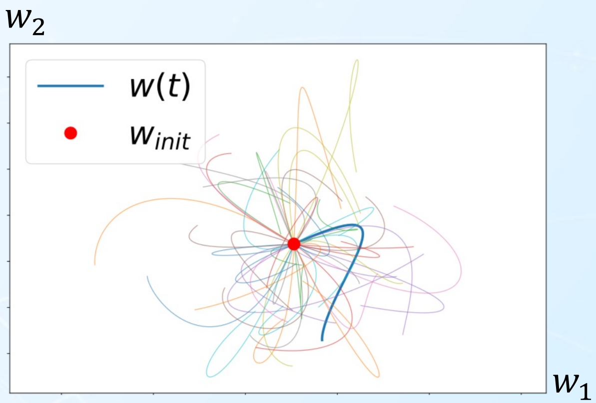
Monte-Carlo approximation $\min_{\theta_0} \frac{1}{N^t} \frac{1}{N^\omega} \sum_{t'} \sum_{\omega'} \mathcal{L}(\hat{y}(x, \theta_0, A_{\omega'}(t')), y)$

Visualization of Aging-Aware Training

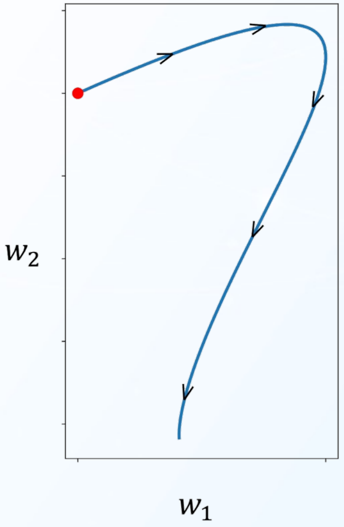


stochastic aging degradations

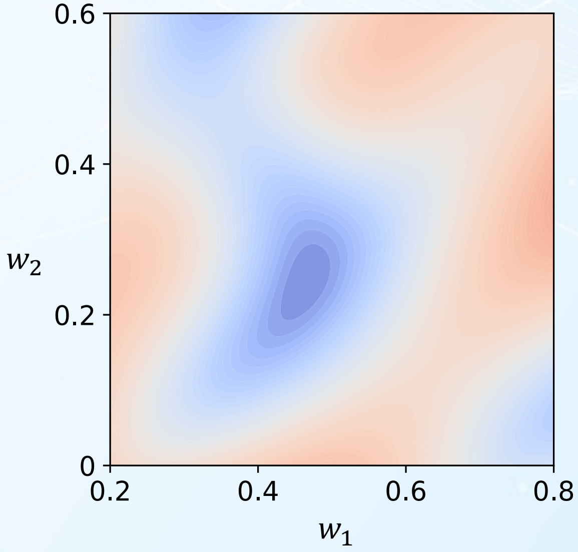
$$w_i = \frac{\theta_i}{\sum_j \theta_j}$$



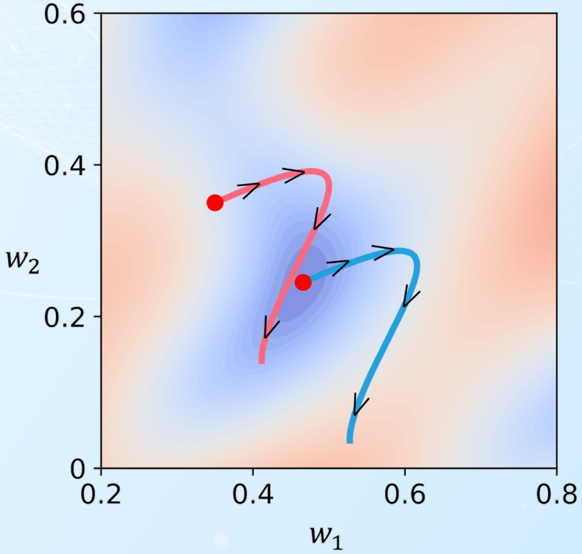
aging trajectories



exemplary trajectory



exemplary loss contour



aging-aware vs. non-aging-aware

Outline

- Printed Electronics
- Printed Neural Networks
- Aging-aware Training
- **Results**
- Conclusions

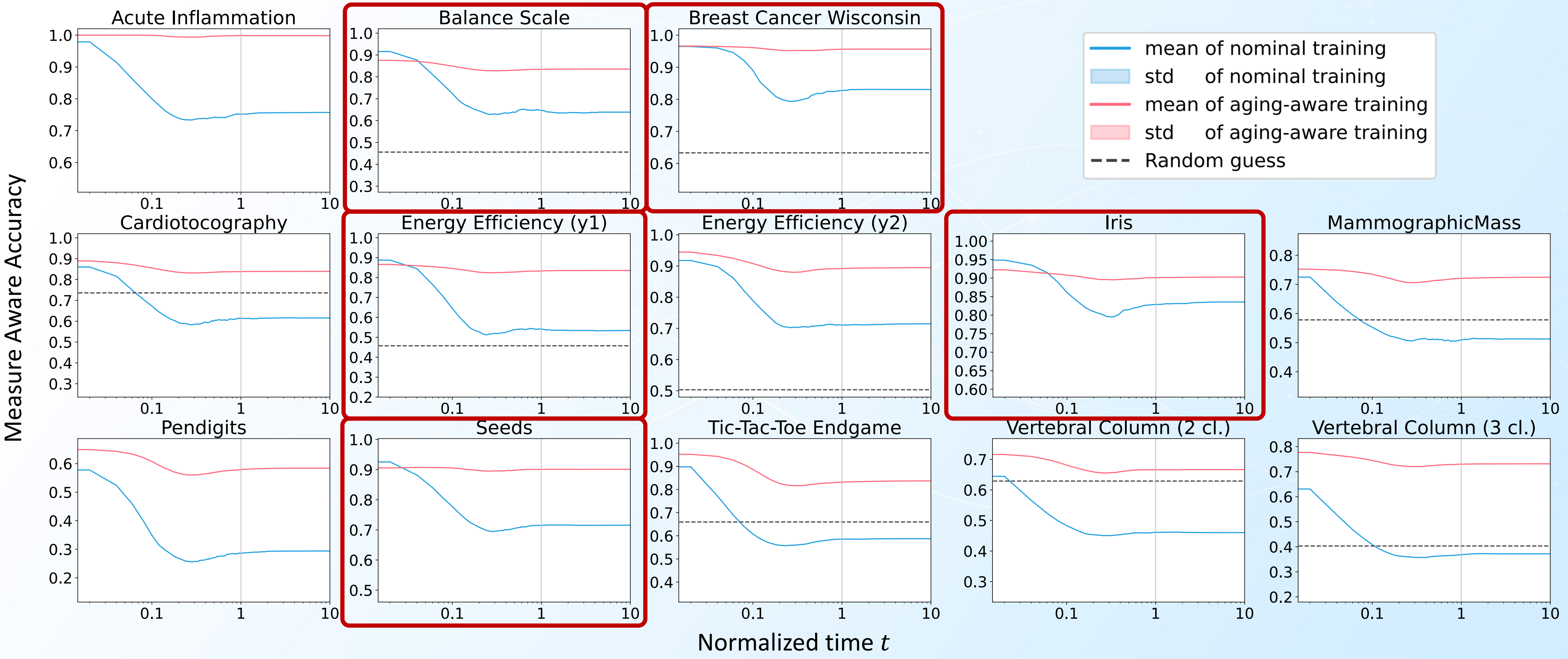
Experiment

- 13 benchmark datasets
- nominal training vs. aging-aware training
- 10 runs with different random seeds {0,1,…,9}
- training time in [0,1], but evaluation time is extrapolated to [0,10]
- evaluation metric
  - accuracy
  - measuring-aware accuracy
- baseline: random guess

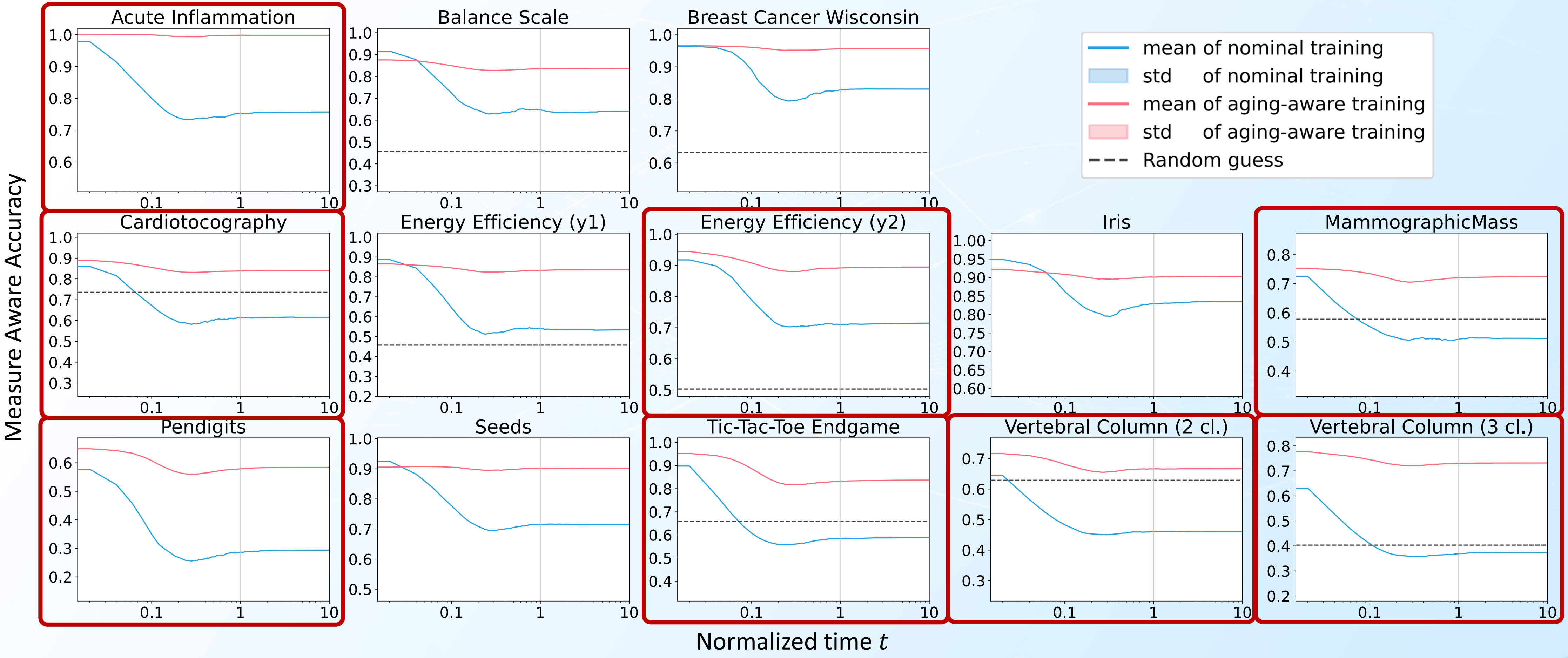
Result

Dataset	Topology	Classic accuracy		Measuring-aware accuracy		Baseline
		nominal	aging-aware	nominal	aging-aware	
Acute Inflammation [21]	6-3-2	0.857 ± 0.146	0.999 ± 0.006	0.757 ± 0.167	0.998 ± 0.012	0.417
Balance Scale [49]	4-3-3	0.832 ± 0.087	0.891 ± 0.031	0.641 ± 0.324	0.836 ± 0.038	0.456
Breast Cancer Wisconsin [40]	9-3-2	0.941 ± 0.088	0.966 ± 0.005	0.831 ± 0.204	0.956 ± 0.024	0.633
Cardiotocography [4]	21-3-3	0.832 ± 0.095	0.884 ± 0.019	0.616 ± 0.309	0.839 ± 0.037	0.736
Energy Efficiency [52] (y_1)	8-3-3	0.789 ± 0.154	0.871 ± 0.015	0.537 ± 0.285	0.835 ± 0.042	0.458
Energy Efficiency (y_2)	8-3-3	0.873 ± 0.072	0.925 ± 0.027	0.715 ± 0.157	0.894 ± 0.047	0.503
Iris [2]	4-3-3	0.927 ± 0.063	0.918 ± 0.022	0.834 ± 0.135	0.902 ± 0.033	0.300
Mammographic Mass [23]	5-3-2	0.618 ± 0.146	0.767 ± 0.012	0.514 ± 0.151	0.724 ± 0.042	0.578
Pendigits [1]	16-3-10	0.563 ± 0.112	0.746 ± 0.024	0.294 ± 0.131	0.583 ± 0.057	0.099
Seeds [14]	7-3-3	0.845 ± 0.116	0.936 ± 0.020	0.717 ± 0.168	0.901 ± 0.036	0.333
Tic-Tac-Toe Endgame [41]	9-3-2	0.671 ± 0.226	0.875 ± 0.086	0.588 ± 0.214	0.837 ± 0.106	0.660
Vertebral Column [5] (2 cl.)	6-3-2	0.628 ± 0.109	0.737 ± 0.032	0.462 ± 0.181	0.667 ± 0.053	0.629
Vertebral Column (3 cl.)	6-3-3	0.586 ± 0.144	0.778 ± 0.030	0.373 ± 0.151	0.731 ± 0.047	0.403
Average	-	0.766 ± 0.120	0.869 ± 0.025	0.606 ± 0.198	0.823 ± 0.044	0.477

Result


Acute Inflammation
Balance Scale
Breast Cancer Wisconsin
Cardiotocography
Energy Efficiency (y1)
Energy Efficiency (y2)
Iris
MammographicMass
Pendigits
Seeds
Tic-Tac-Toe Endgame
Vertebral Column (2 cl.)
Vertebral Column (3 cl.)
Measure Aware Accuracy
Normalized time t
mean of nominal training
std of nominal training
mean of aging-aware training
std of aging-aware training
Random guess

Result



Acute Inflammation
Balance Scale
Breast Cancer Wisconsin
Cardiotocography
Energy Efficiency (y1)
Energy Efficiency (y2)
Iris
MammographicMass
Pendigits
Seeds
Tic-Tac-Toe Endgame
Vertebral Column (2 cl.)
Vertebral Column (3 cl.)
mean of nominal training
std of nominal training
mean of aging-aware training
std of aging-aware training
Random guess
Measure Aware Accuracy
Normalized time t

Outline

- Printed Electronics
- Printed Neural Networks
- Aging-aware training
- Results
- **Conclusions**

Conclusions

- Printed electronics provides complementary advantages
  - Compared to traditional silicon-based VLSI technologies
- Low device count, large feature sizes, large latencies
  - Constraints for sub-cent printed circuits
- Printed neural networks (pNNs)
  - Mixed-signal computing to reduce device count
- Aging of printed resistors
  - Can severely impact the accuracy of pNNs over time
- Aging-aware training
  - Can ensure acceptable accuracy over device lifetime

Thank you for your attention

European Research Council

IEEE/ACM 2022 INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN ICCAD 41st Edition