IEEE/ACM 2023 INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN ICCAD 42nd Edition

Power-Aware Training for Energy-Efficient Printed Neuromorphic Circuits

Haibin Zhao, Priyanjana Pal, Michael Hefenbrock, Michael Beigl, Mehdi B. Tahoori

Karlsruhe Institute of Technology (KIT)

Karlsruhe, Germany

Outline

- Printed Electronics
- Printed Neuromorphic Circuit
- Power-Aware Training
- Experiment
- Conclusions

The Cliché vs Computational Deserts

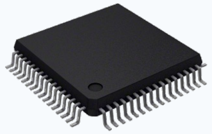
high performance
resource-intensive
general purpose

simple tasks
resource-limited
tailored functionality
disposable,
flexible,
degradable…

Cost Wall

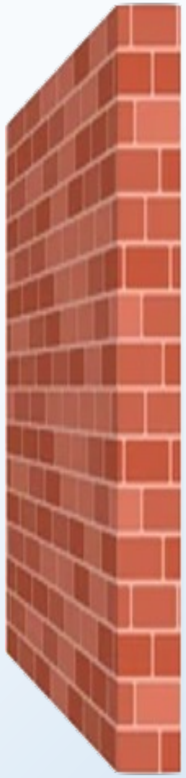
Electronics

Java Card: €0.20



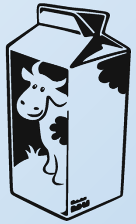
Microcontroller: €0.35

RFID Tag: €0.15



Applications

Milk Carton: €0.01

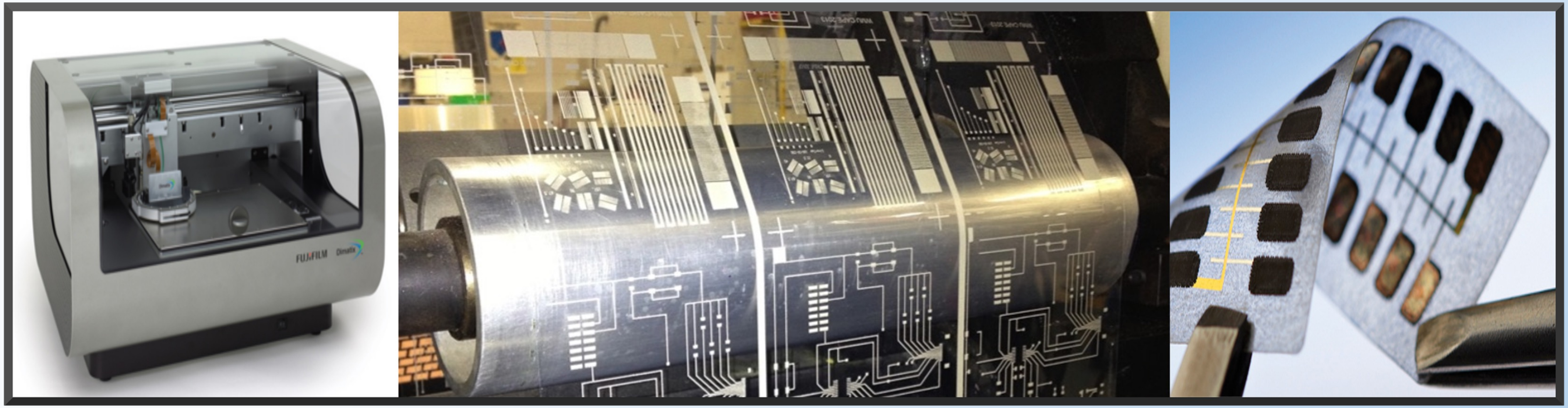


Adhesive Bandage: €0.02

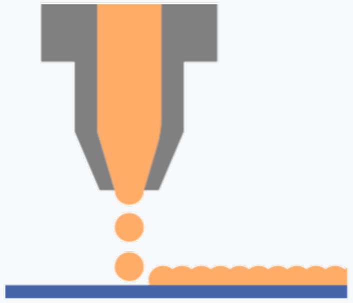
Packaging Label: €0.02

Additive Printed Technologies

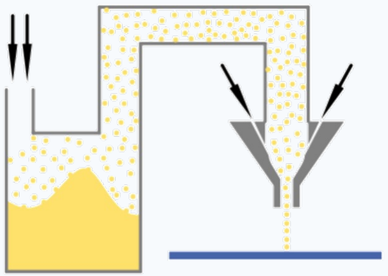
- Maskless, fully additive processes
- Flexible, stretchable, and porous substrates
- Non-toxic, bio-compatible inks and substrates



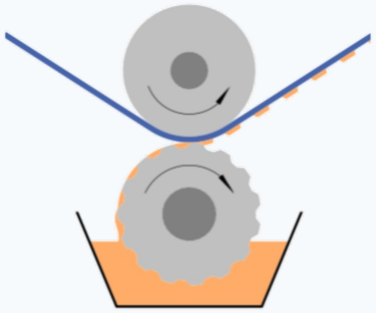
Printed Electronics Technology



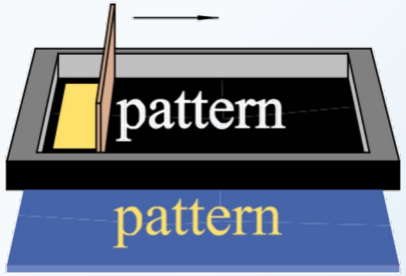
Inkjet printing



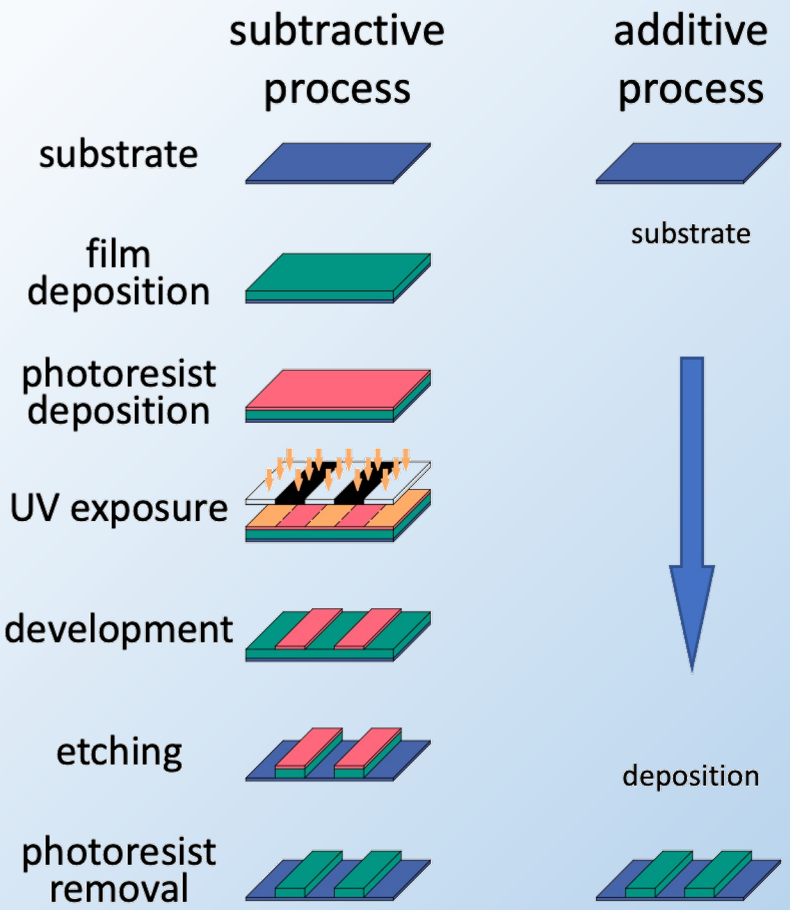
Aerosol jet printing



Gravure printing

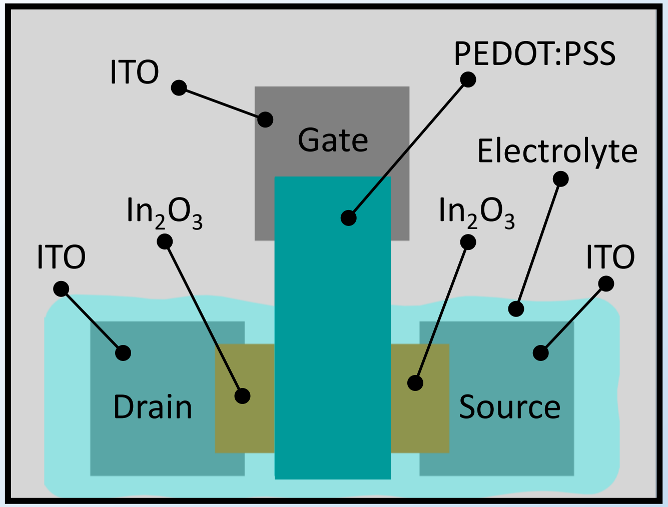


Screen printing

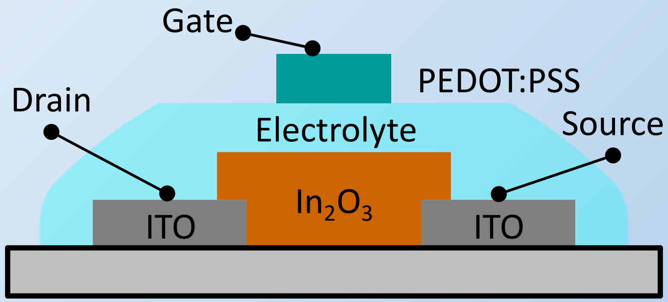


Electrolyte-Gated Transistor (EGT)

- Thin-film transistor
  - Signal routing: indium tin oxide (ITO)
  - Semiconductor: indium oxide (In_2O_3)
  - Gate insulator: composite solid polymer electrolyte
  - Top gate: PEDOT:PSS
- Voltage levels: $\leq$ 1.5 V, $\approx$ 100 µA – 1 mA
- Frequency range: 100 Hz – 1 kHz

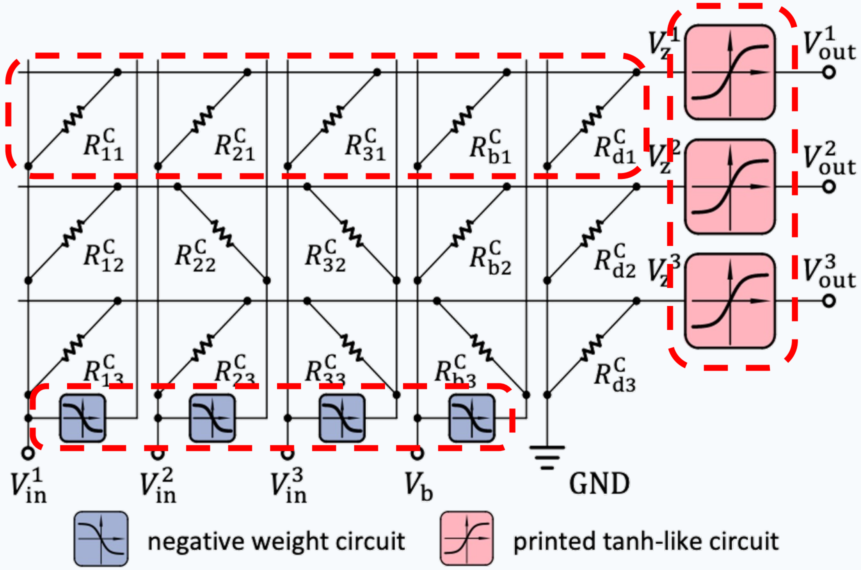


(a) top view of nEGT

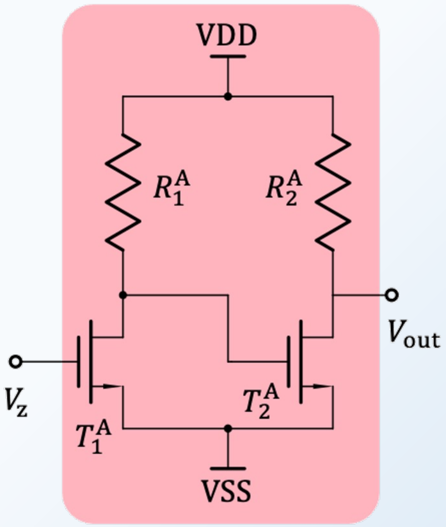


(b) front view of nEGT

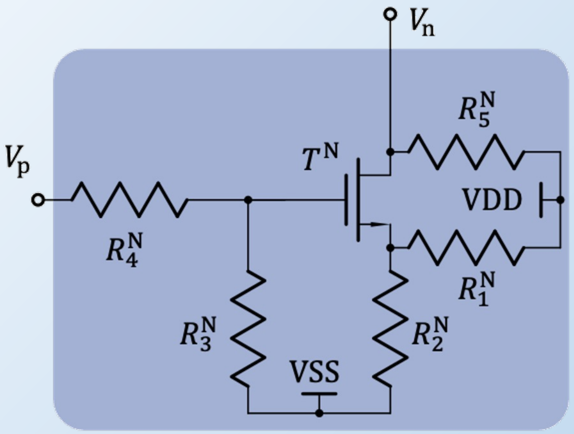
Printed Neuromorphic Circuit



(a) Exemplary printed neuron



(b) Printed tanh-like circuit



(c) Negative weight circuit

$$V_z^1 = \frac{g_{11}}{G_1} V_{in}^1 + \frac{g_{21}}{G_1} V_{in}^2 + \frac{g_{31}}{G_1} V_{in}^3 + \frac{g_{b1}}{G_1} V_b$$

where $g_{ij} = \frac{1}{R^C_{ij}}$, G_i is the sum of g_{ij}, $V_b \equiv 1V$.

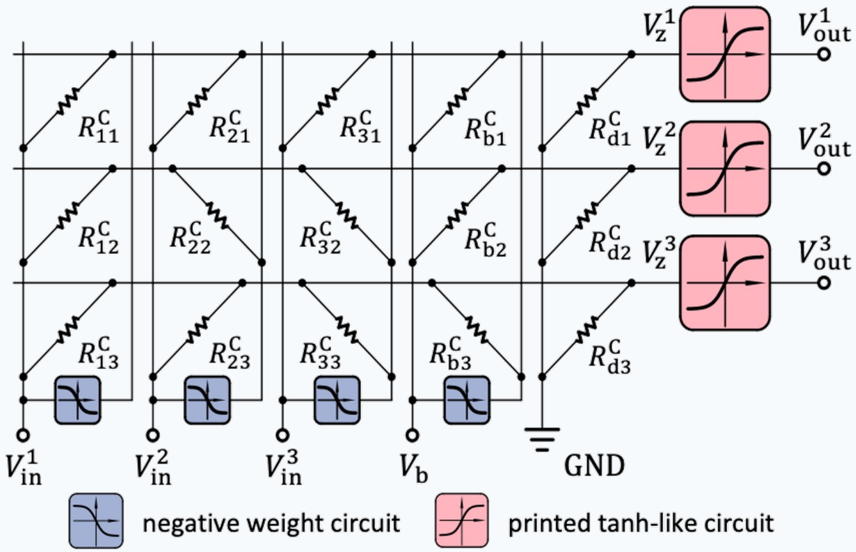
$$\begin{aligned} V_{out} &= \mathrm{ptanh}(V_{in}) \\ &= \eta_1^A + \eta_2^A \cdot \tanh\left(\left(V_{in} - \eta_3^A\right) \cdot \eta_4^A\right) \end{aligned}$$

where η_i^A is auxiliary parameter determined by physical quantities $\boldsymbol{q}^A = \left[R_1^A, R_2^A, T_1^A, T_2^A\right]$

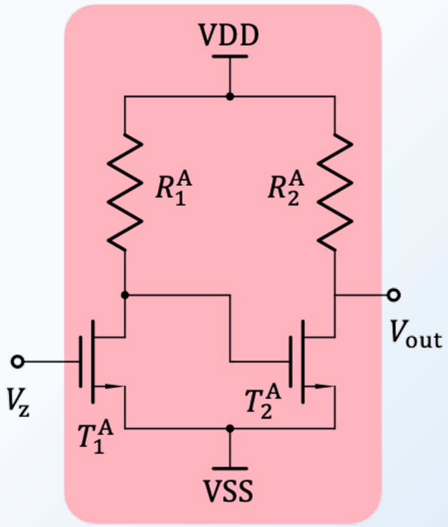
$$\begin{aligned} V_n &= \mathrm{neg}(V_p) \\ &= -\left(\eta_1^N + \eta_2^N \cdot \tanh\left(\left(V_p - \eta_3^N\right) \cdot \eta_4^N\right)\right) \end{aligned}$$

where η_i^N is auxiliary parameter determined by physical quantities $\boldsymbol{q}^N = \left[R_1^N, \ldots, T^N\right]$

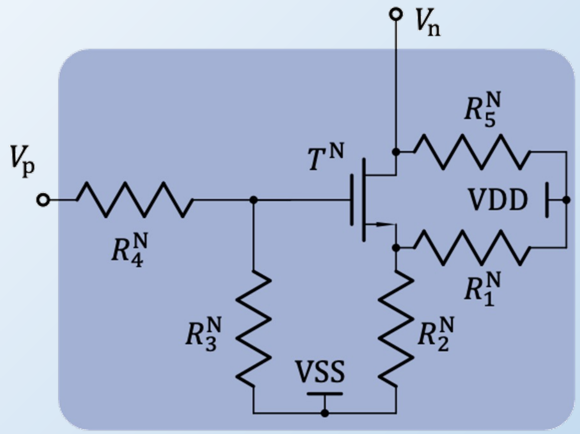
ML-based Circuit Optimization



(a) Exemplary printed neuron



(b) Printed tanh-like circuit



(c) Negative weight circuit

$$V_{out} = \mathrm{ptanh}_{\boldsymbol{q}^{A}}\left(\boldsymbol{V}_{in} \cdot \left(\boldsymbol{W} \odot \mathbb{1}_{\{\boldsymbol{\Theta} \geq \boldsymbol{0}\}}\right) + \mathrm{neg}_{\boldsymbol{q}^{N}}(\boldsymbol{V}_{in}) \cdot \left(\boldsymbol{W} \odot \mathbb{1}_{\{\boldsymbol{\Theta} < \boldsymbol{0}\}}\right)\right)$$

with $\boldsymbol{W} = |\boldsymbol{\Theta}| \cdot \mathrm{diag}(\boldsymbol{\Theta}^{\top} \cdot \boldsymbol{1})^{-1}$

$\boldsymbol{\Theta}$ learnable surrogate conductance, $|\boldsymbol{\Theta}|$ printable conductance, $\mathrm{sign}(\boldsymbol{\Theta})$ requirement of negative weight circuit

ML-based Circuit Optimization

$$\boldsymbol{V}_{\text{out}} = \text{ptanh}_{\boldsymbol{q}^{\text{A}}}\Big(\boldsymbol{V}_{\text{in}} \cdot \big(\boldsymbol{W} \odot \mathbb{I}_{\{\boldsymbol{\Theta}\geq\boldsymbol{0}\}}\big) + \text{neg}_{\boldsymbol{q}^{\text{N}}}(\boldsymbol{V}_{\text{in}}) \cdot \big(\boldsymbol{W} \odot \mathbb{I}_{\{\boldsymbol{\Theta}<\boldsymbol{0}\}}\big)\Big)$$

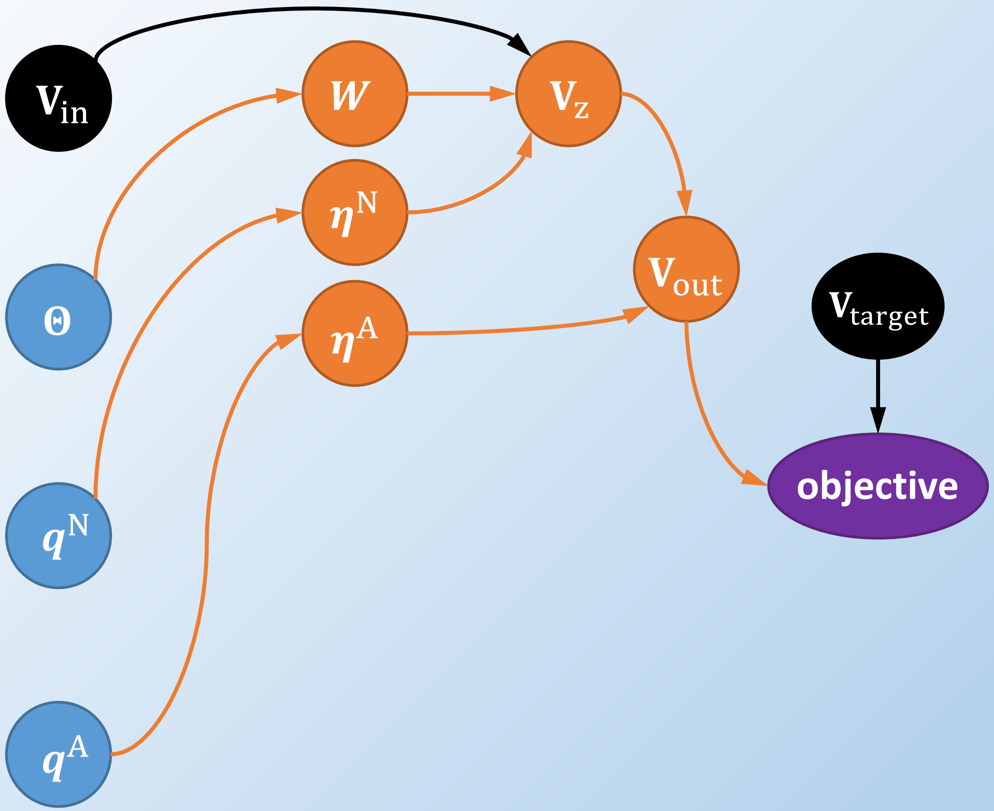
$\boldsymbol{\Theta}$ learnable surrogate conductance,

$|\boldsymbol{\Theta}|$ printable conductance,

$\text{sign}(\boldsymbol{\Theta})$ requirement of negative weight circuit

η_i^{A} in $\text{ptanh}(\cdot)$ can be determined by physical quantities $\boldsymbol{q}^{\text{A}}$ through a surrogate nonlinear circuit model [1]

η_i^{N} in $\text{neg}(\cdot)$ can be determined by physical quantities $\boldsymbol{q}^{\text{N}}$ through a surrogate nonlinear circuit model [1]

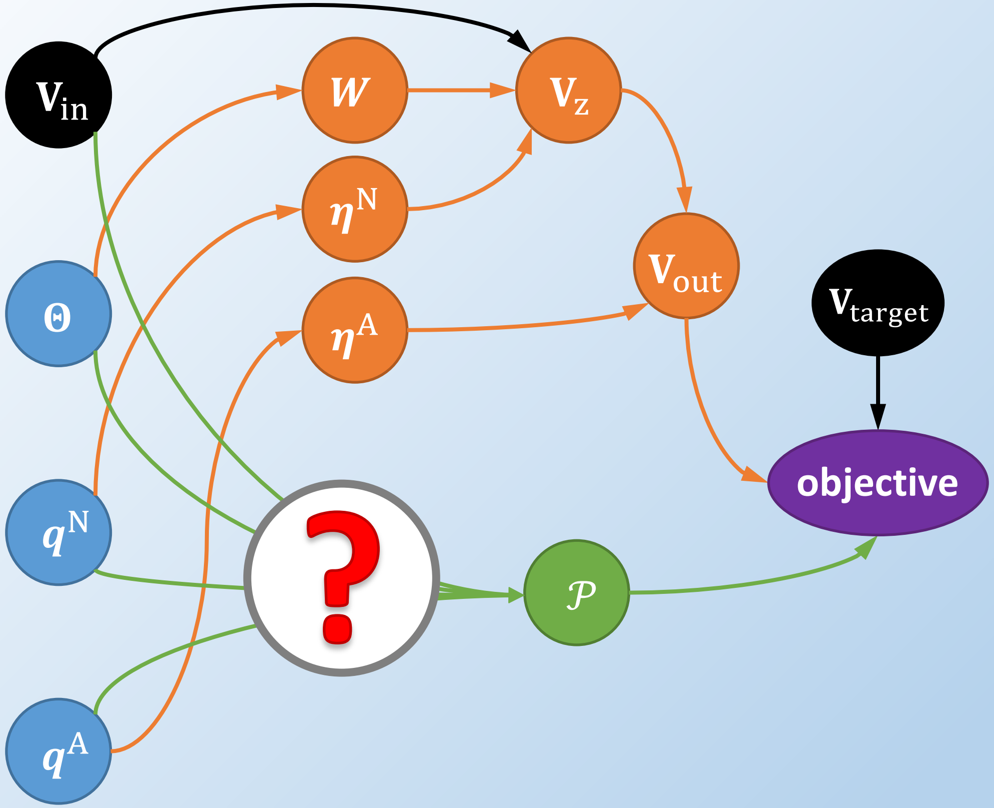


[1] Zhao, et al. "Highly-bespoke robust printed neuromorphic circuits." *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2023.

Power-Aware Training

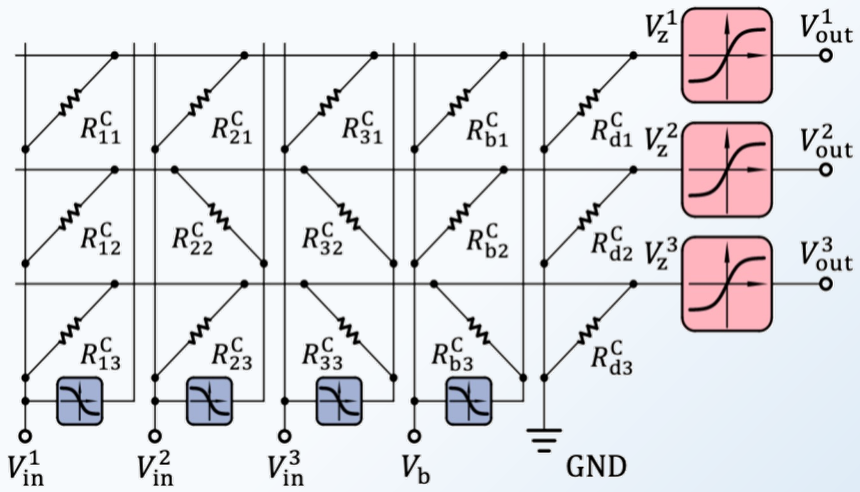
- Target applications of pNCs

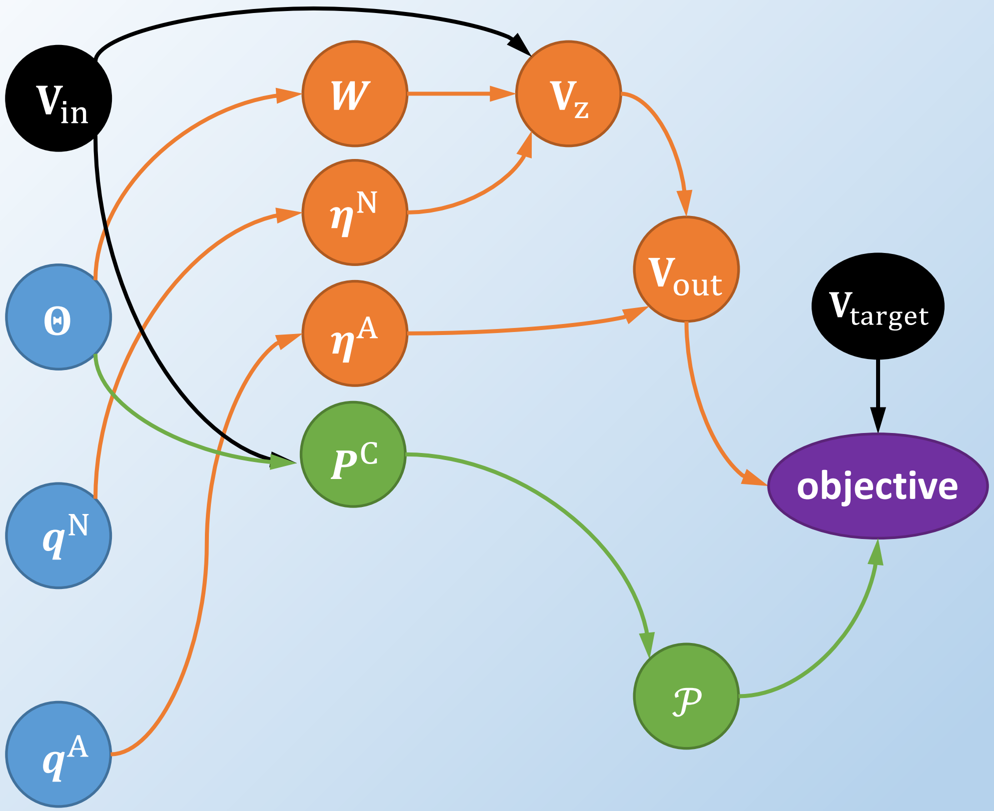
  - Printed batteries with limited power
  - Supplied by energy harvesters
- Low-power design
  - Develop differentiable power model of pNCs
  - Integrate power model into training objective



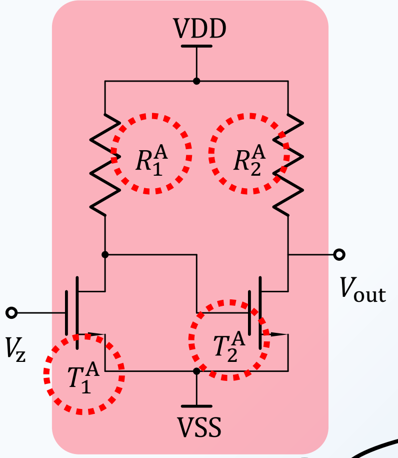
Power-Aware Training

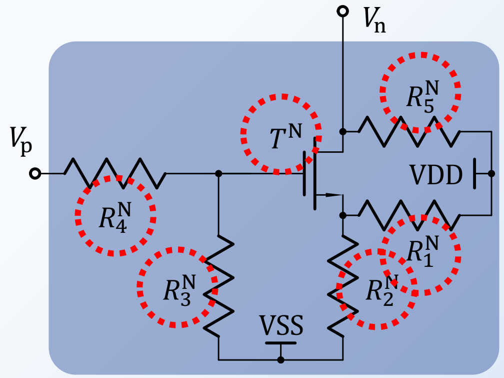
- Power Model
  - Power of a single resistor
$$P = (V_{\text{in}} - V_{\text{out}})^2 \cdot g$$
  - Power of the crossbar
$$P^{\text{C}} = \left(\left(\boldsymbol{V}_{\text{in}} \odot \mathbb{I}_{\{\boldsymbol{\Theta} \geq \boldsymbol{0}\}} + \boldsymbol{V}_{\text{in}} \odot \mathbb{I}_{\{\boldsymbol{\Theta} < \boldsymbol{0}\}}\right) - \boldsymbol{V}_{\text{z}}\right)^2 \odot \boldsymbol{\Theta}$$





Power-Aware Training

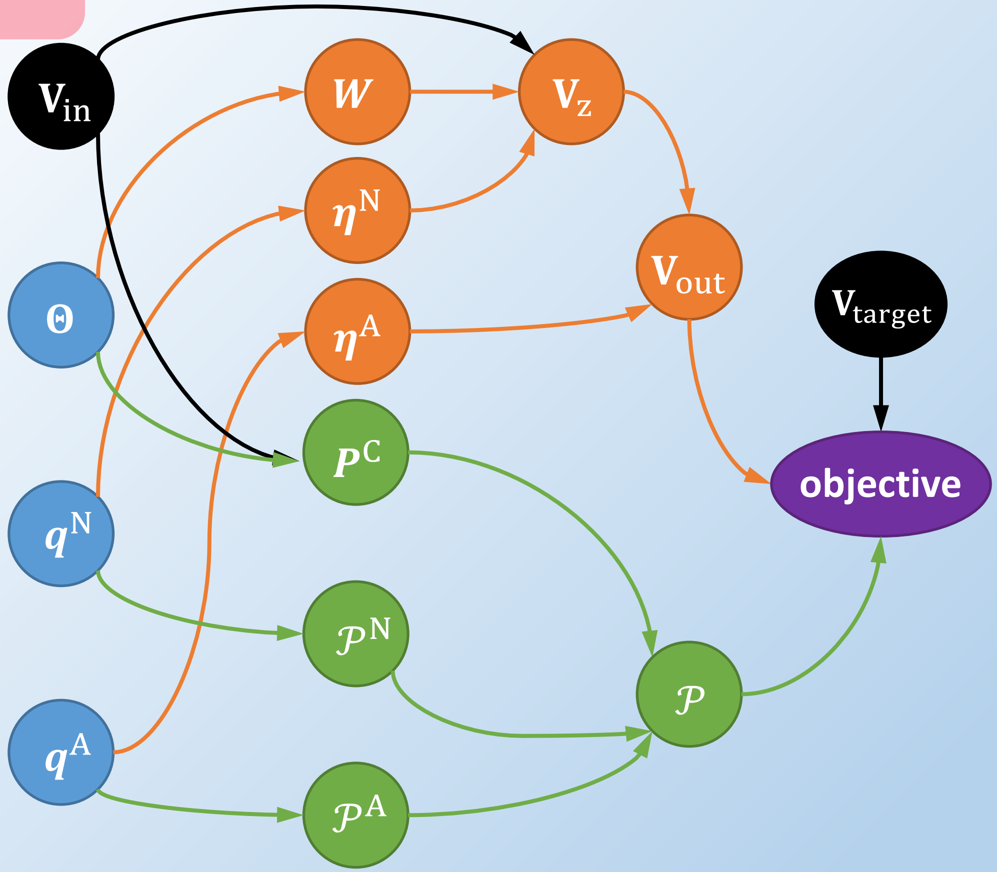




- Power Model
  - Power of the crossbar

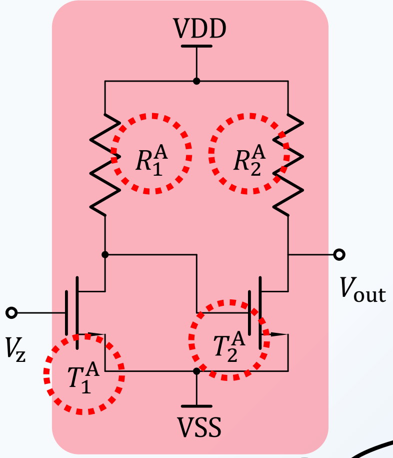
$$P^{\mathrm{C}} = \left(\left(\boldsymbol{V}_{\mathrm{in}} \odot \mathbb{1}_{\{\boldsymbol{\Theta}\geq\boldsymbol{0}\}} + \boldsymbol{V}_{\mathrm{in}} \odot \mathbb{1}_{\{\boldsymbol{\Theta}<\boldsymbol{0}\}}\right) - \boldsymbol{V}_{\mathrm{z}}\right)^2 \odot \boldsymbol{\Theta}$$

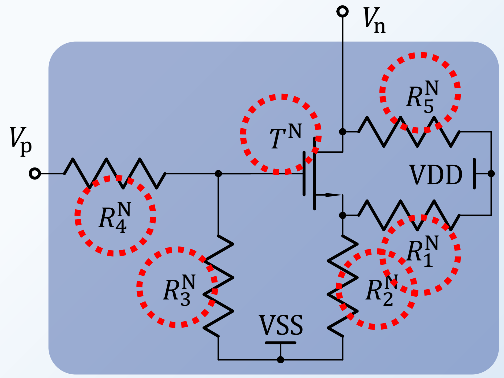
  - Power of the nonlinear circuits
    - Estimating power consumptions P^{A} and P^{N} from physical quantities $\boldsymbol{q}^{\mathrm{A}}$ and $\boldsymbol{q}^{\mathrm{N}}$ is complicated
    - Employ NN-based models as *surrogate power consumption models, denoted by* $\mathrm{SP}(\cdot)$
      - Sampled 10 000 $\boldsymbol{q}_i^{\mathrm{A}}$ and $\boldsymbol{q}_i^{\mathrm{N}}$ with Quasi-Monte Carlo
      - Simulated their power P_i^{A} and P_i^{N} based on SPICE with pPDK [2]
      - Trained NN to estimate P_i by $\boldsymbol{q}_i$, i.e., $\mathrm{SP}(\cdot)\colon \boldsymbol{q}_i \mapsto P_i$



[2] Rasheed, et al. "Variability modeling for printed inorganic electrolyte-gated transistors and circuits." *IEEE transactions on electron devices* 66.1 (2018): 146-152.

Power-Aware Training

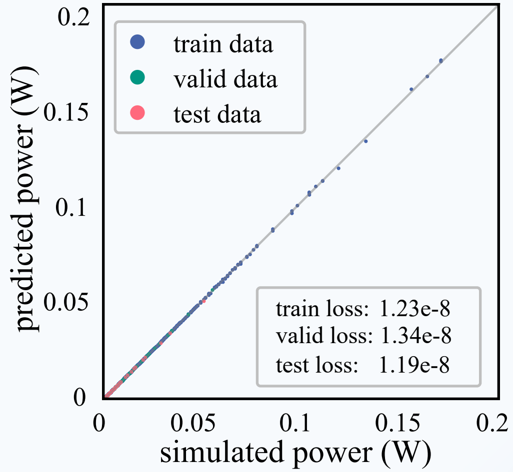




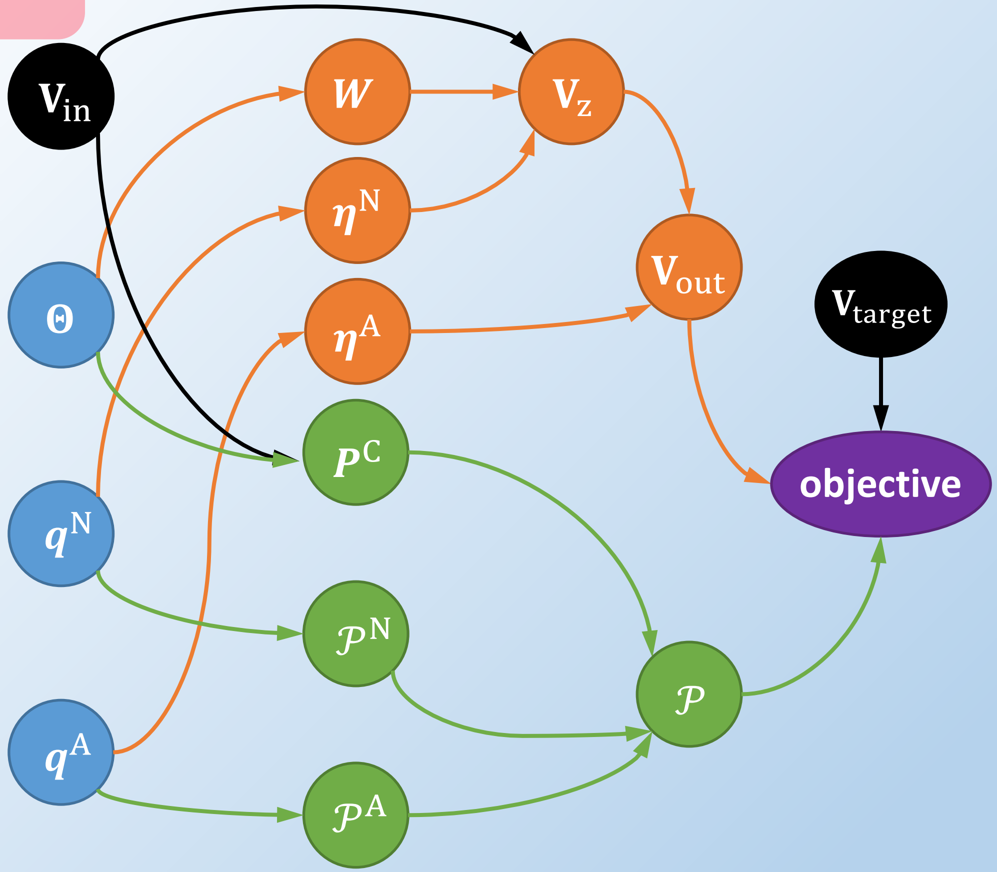
- Power Model
  - Power of the crossbar

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

  - Power of the nonlinear circuits
    - Employ NN-based models as *surrogate power consumption models, denoted by* $\mathrm{SP}(\cdot)$



$$P^{A} = \mathrm{SP}^{A}(\boldsymbol{q}^{A})$$
$$P^{N} = \mathrm{SP}^{N}(\boldsymbol{q}^{N})$$



Power-Aware Training

- Power Model
  - Power of the crossbar

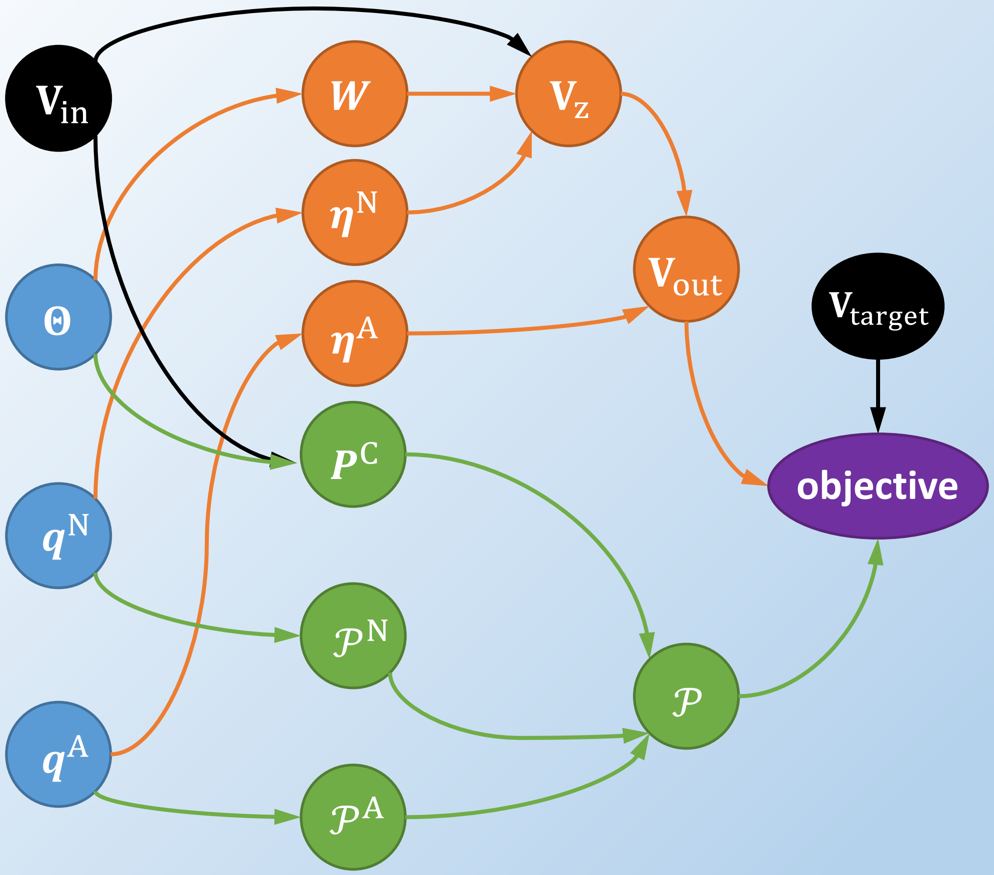
$$P^{\mathrm{C}} = \left(\left(\boldsymbol{V}_{\mathrm{in}} \odot \mathbb{1}_{\{\boldsymbol{\Theta}\geq\boldsymbol{0}\}} + \boldsymbol{V}_{\mathrm{in}} \odot \mathbb{1}_{\{\boldsymbol{\Theta}<\boldsymbol{0}\}}\right) - \boldsymbol{V}_{\mathrm{z}}\right)^{2} \odot \boldsymbol{\Theta}$$

  - Power of the nonlinear circuits
    - Employ NN-based models as *surrogate power consumption models, denoted by* $\mathrm{SP}(\cdot)$

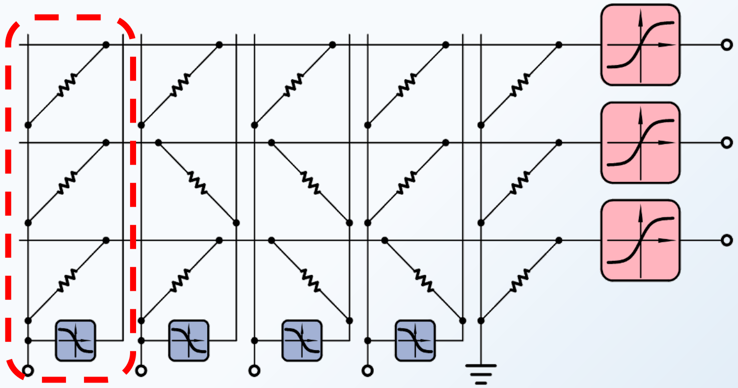
$$P^{\mathrm{A}} = \mathrm{SP}^{\mathrm{A}}\left(\boldsymbol{q}^{\mathrm{A}}\right)$$
$$P^{\mathrm{N}} = \mathrm{SP}^{\mathrm{N}}\left(\boldsymbol{q}^{\mathrm{N}}\right)$$

    - Power overhead

$$\mathcal{P}^{\mathrm{A}} = N^{\mathrm{A}} \cdot \mathrm{SP}^{\mathrm{A}}\left(\boldsymbol{q}^{\mathrm{A}}\right)$$
$$\mathcal{P}^{\mathrm{N}} = N^{\mathrm{N}} \cdot \mathrm{SP}^{\mathrm{N}}\left(\boldsymbol{q}^{\mathrm{N}}\right)$$



Power-Aware Training



- Power Model
  - Power of the crossbar

$$P^{\mathrm{C}} = \left(\left(\boldsymbol{V}_{\mathrm{in}} \odot \mathbb{1}_{\{\boldsymbol{\Theta}\geq\boldsymbol{0}\}} + \boldsymbol{V}_{\mathrm{in}} \odot \mathbb{1}_{\{\boldsymbol{\Theta}<\boldsymbol{0}\}}\right) - \boldsymbol{V}_{\mathrm{z}}\right)^2 \odot \boldsymbol{\Theta}$$

  - Power of the nonlinear circuits
    - Employ NN-based models as *surrogate power consumption models, denoted by* $\mathrm{SP}(\cdot)$
    - Power overhead

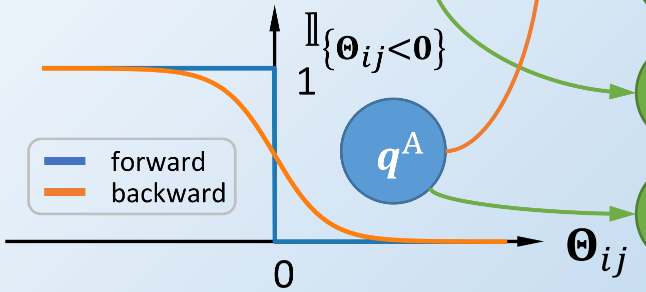
$$\mathcal{P}^{\mathrm{A}} = N^{\mathrm{A}} \cdot \mathrm{SP}^{\mathrm{A}}(\boldsymbol{q}^{\mathrm{A}})$$

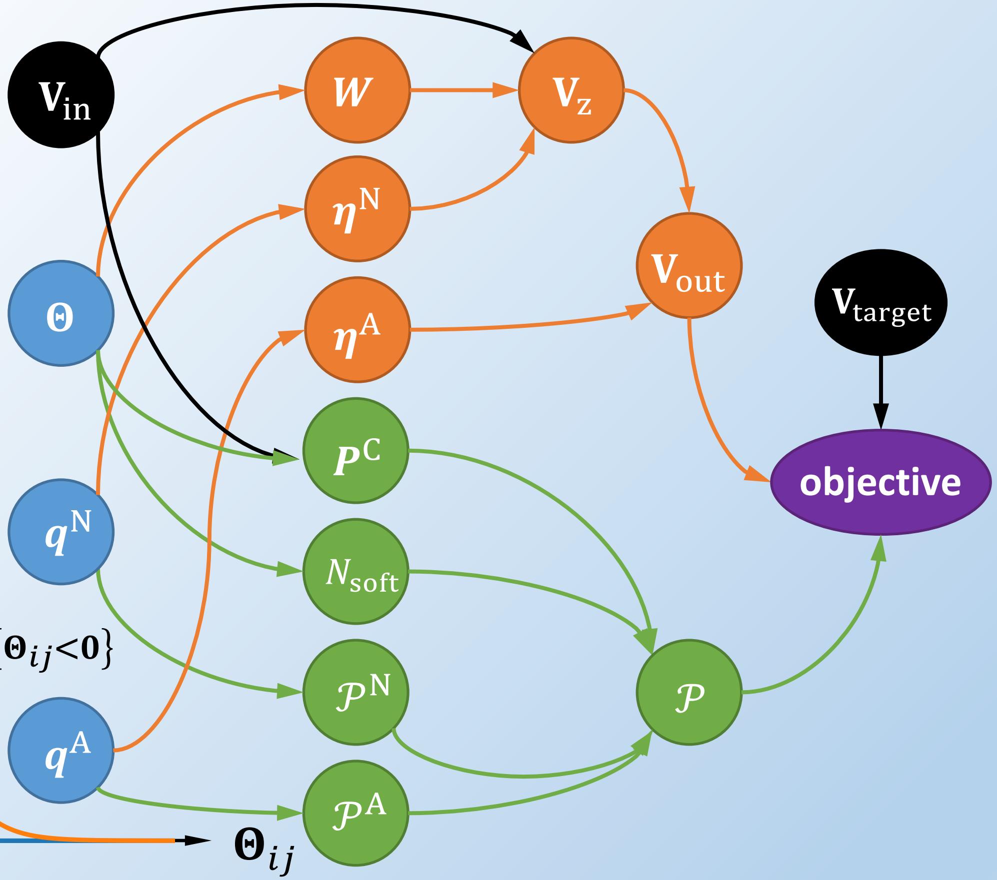
$$\mathcal{P}^{\mathrm{N}} = N^{\mathrm{N}} \cdot \mathrm{SP}^{\mathrm{N}}(\boldsymbol{q}^{\mathrm{N}})$$

    - Soft-Count

$$N^{\mathrm{N}} = \sum \text{column max}\left\{\mathbb{1}_{\{\boldsymbol{\Theta}<\boldsymbol{0}\}}\right\}$$

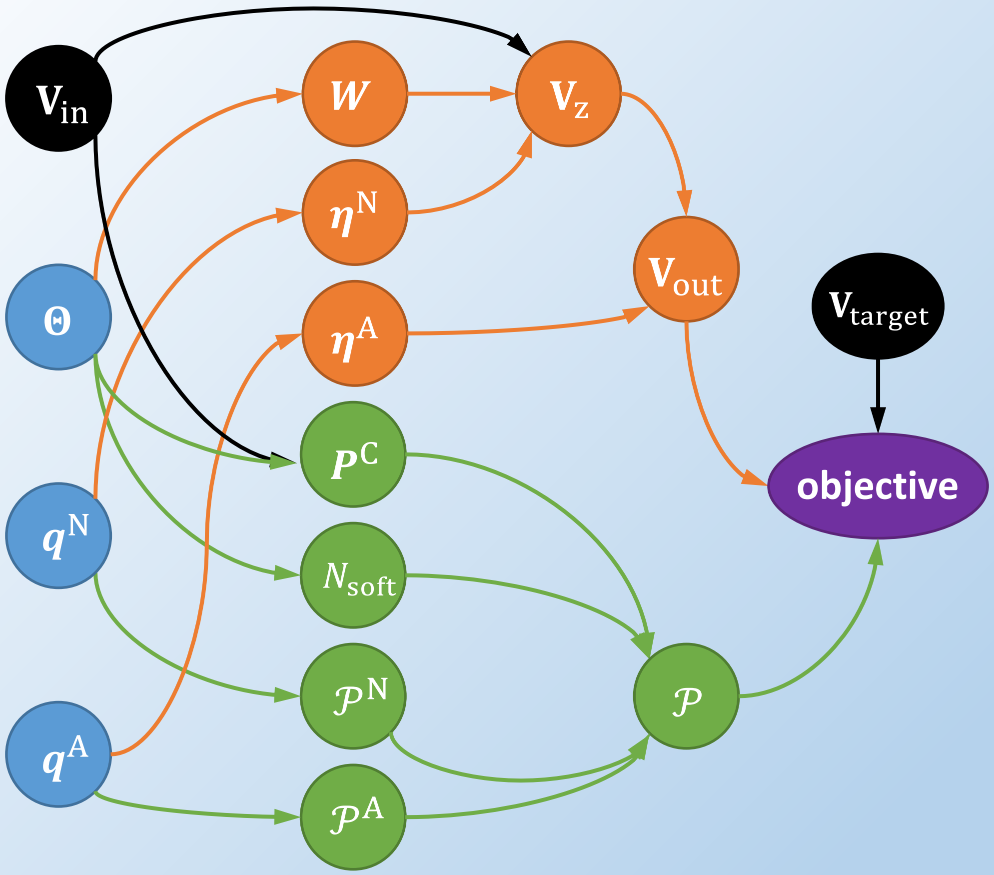
$$\frac{\partial N^{\mathrm{N}}_{\mathrm{soft}}}{\partial \Theta} = \frac{\partial \sum \text{column max}\{1-\text{sigmoid}(\boldsymbol{\Theta})\}}{\partial \Theta}$$





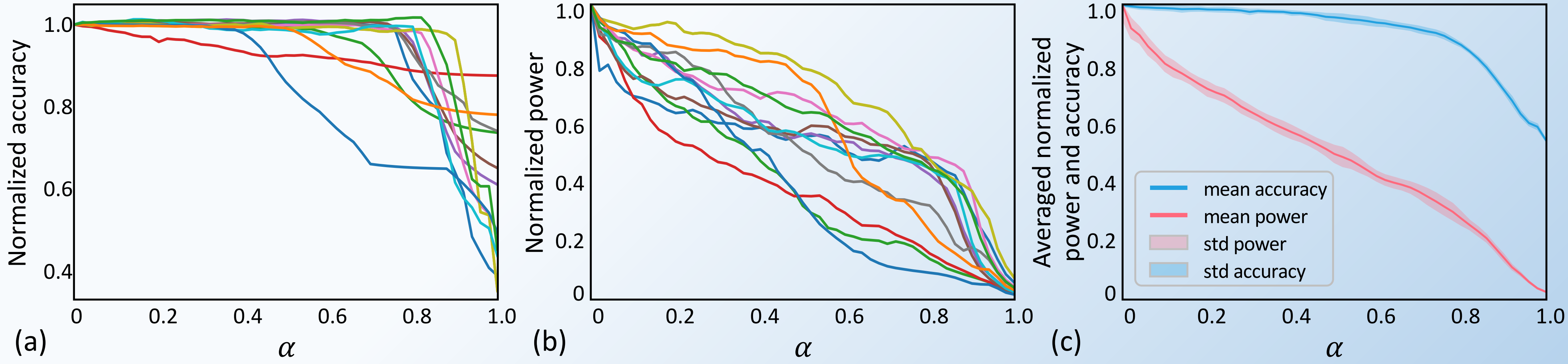
Experiment

- Training objective
$$\mathcal{L} = (1-\alpha)\cdot \mathrm{CELoss} + \alpha \cdot \mathcal{P}$$
  - CELoss: cross-entropy for classification accuracy
  - $\mathcal{P}$: overall power consumption of the pNC
- 13 benchmark datasets
- 100 different α values in $[0,1]$
- Baseline: $\alpha = 0$ (power-**un**aware training)
- Evaluation metric
  - Accuracy
    - Normalized accuracy (by baseline)
    - Averaged normalized accuracy (across all datasets)
  - Power
    - normalized power (by baseline)
    - Averaged normalized power (across all datasets)



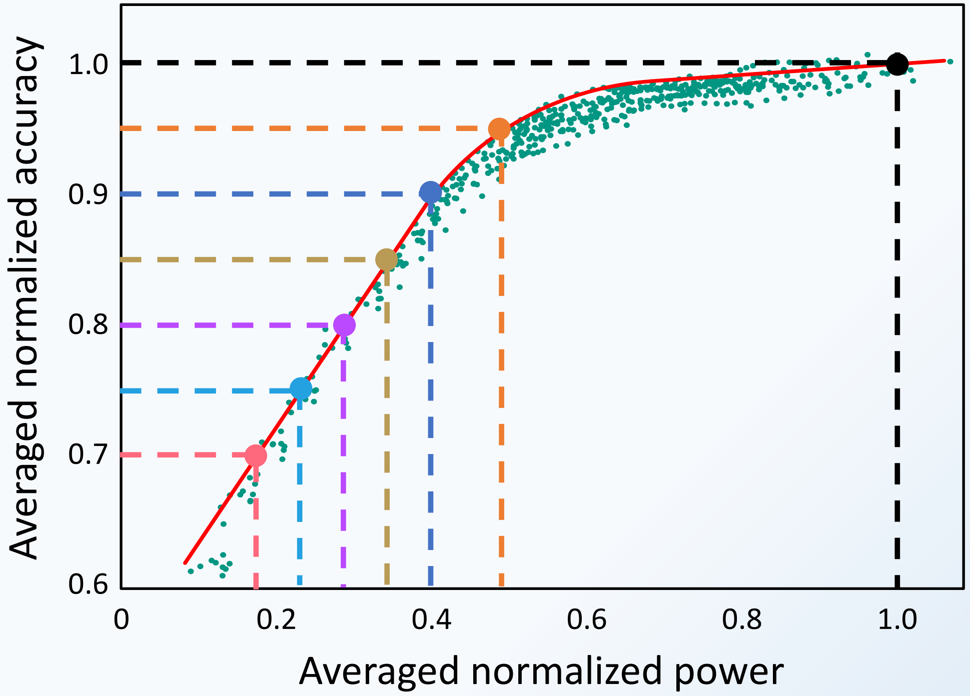
Result

- Normalized Accuracy vs. α
- Normalized Power vs. α
- Averaged normalized accuracy and power vs. α



Result

• Pareto analysis



Accuracy	Power	
100%	100%	1×
95%	50%	2×
90%	40%	2.5×
85%	34%	3×
80%	28%	3.6×
75%	23%	4.4×
70%	18%	5.5×

Conclusion

- Printed electronics provides complementary advantages, i.e.,
  - flexibility, bio-degradability, high customization, ultra-low cost, …
- Printed neuromorphic circuits (pNCs)
  - implement effective computational functionalities as in ANNs
  - through interconnection of simple-structured circuit primitives
  - favored for circuit design, optimization, and manufacturing
- Power-aware training for pNCs
  - derive analytical power model for the resistor crossbar
  - develop *surrogate power models* to precisely estimate the power of nonlinear circuits
  - propose soft-count to enable the reduce the number of required negative weight circuits
  - achieve 2× power-saving of the whole pNC with only 5% loss in classification accuracy

IEEE/ACM

2023 INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN

ICCAD

42nd Edition

Thank you for your attention

Power-Aware Training for Energy-Efficient Printed Neuromorphic Circuits

Haibin Zhao, Priyanjana Pal, Michael Hefenbrock
Michael Beigl, Mehdi B. Tahoori

Karlsruhe Institute of Technology (KIT)