An Energy-efficient Classifier via Boosted Spin Channel Networks

Ameya D. Patil[1], Sasikanth Manipatruni[2], Dmitri Nikonov[2], Ian A. Young[2], and Naresh R. Shanbhag[1]

[1]University of Illinois at Urbana-Champaign
[2]Intel Corporation

ECE ILLINOIS

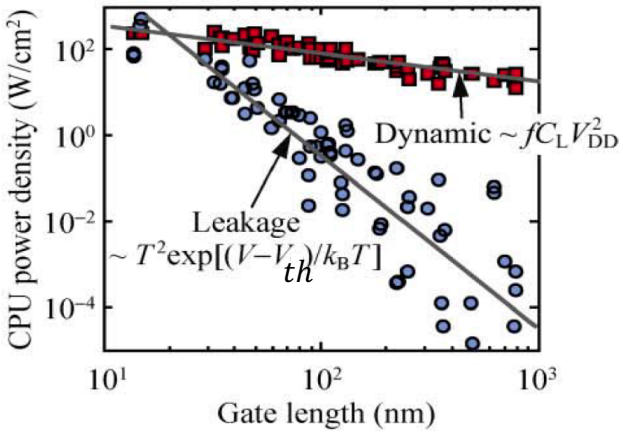
ILLINOIS

Energy-efficient Computing Challenge

Emerging Cognition at the Edge

computationally intensive
severe energy constraints

CMOS Device Scaling

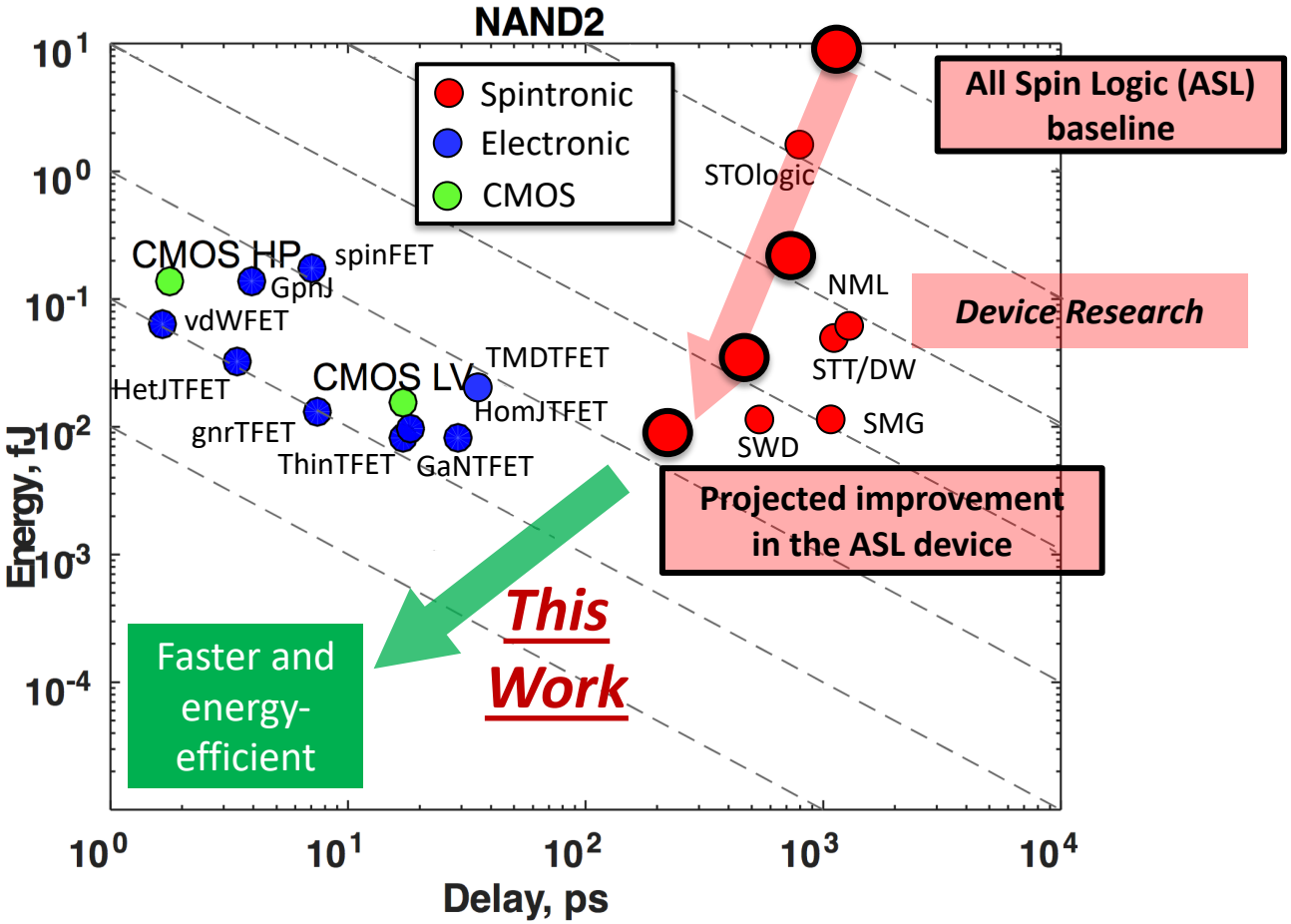


[Pop-NanoResearch-2010]

diminishing returns

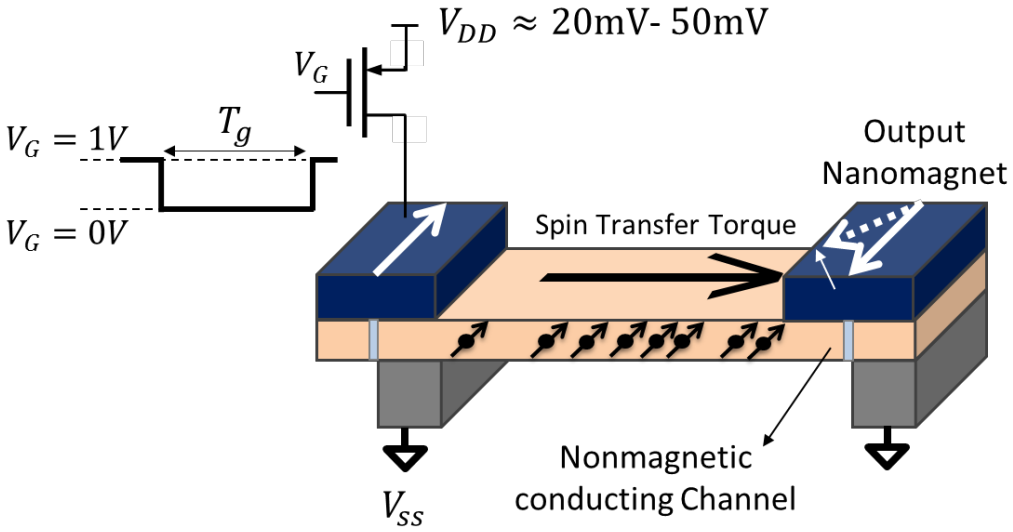
need to explore computing based on beyond-CMOS devices

Beyond-CMOS Space



[Nikonov-JXCDC-2015] [Manipatruni-PRApplied-2016]

Devices under exploration are more energy efficient than CMOS *but slower*



[Behin-Aein-NatureNano-2010]

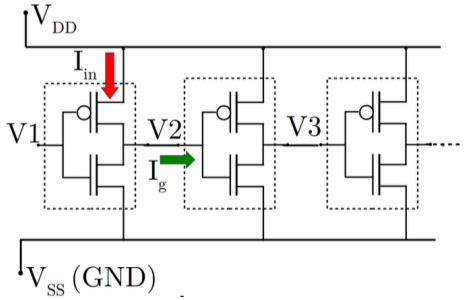
All Spin Logic (ASL)

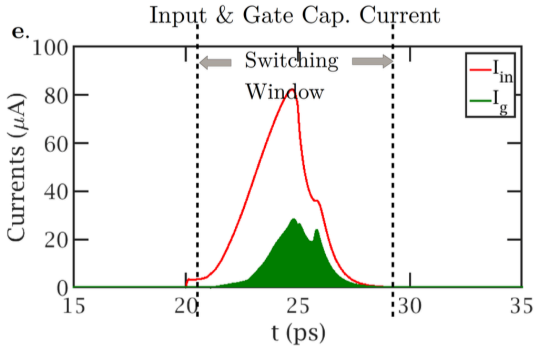
- **Advantages**: [Kim-IEEEProc-2015]
  - ultra-low voltage ($\approx 20\text{mV}$), non-volatility, higher logical density
- **Not competitive to CMOS** [Pajouhi-TCAD-2015]
  - large current (10-100 μA) and delay (few ns) to switch the nanomagnet
  - exponential decay of spin current
  - large static power consumption

CMOS vs. Spin-torque Switching

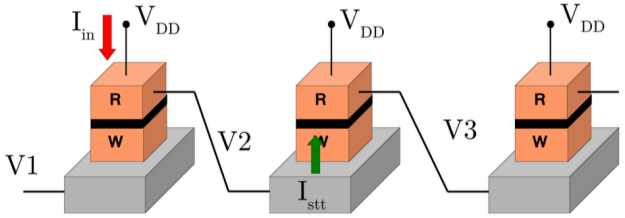
[Ganguly-JXCDC-2017]

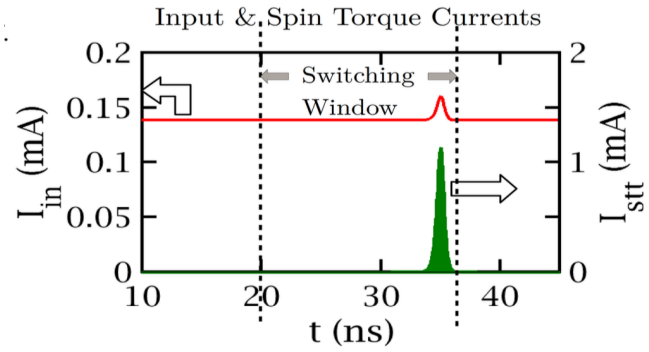
CMOS





Spin-torque



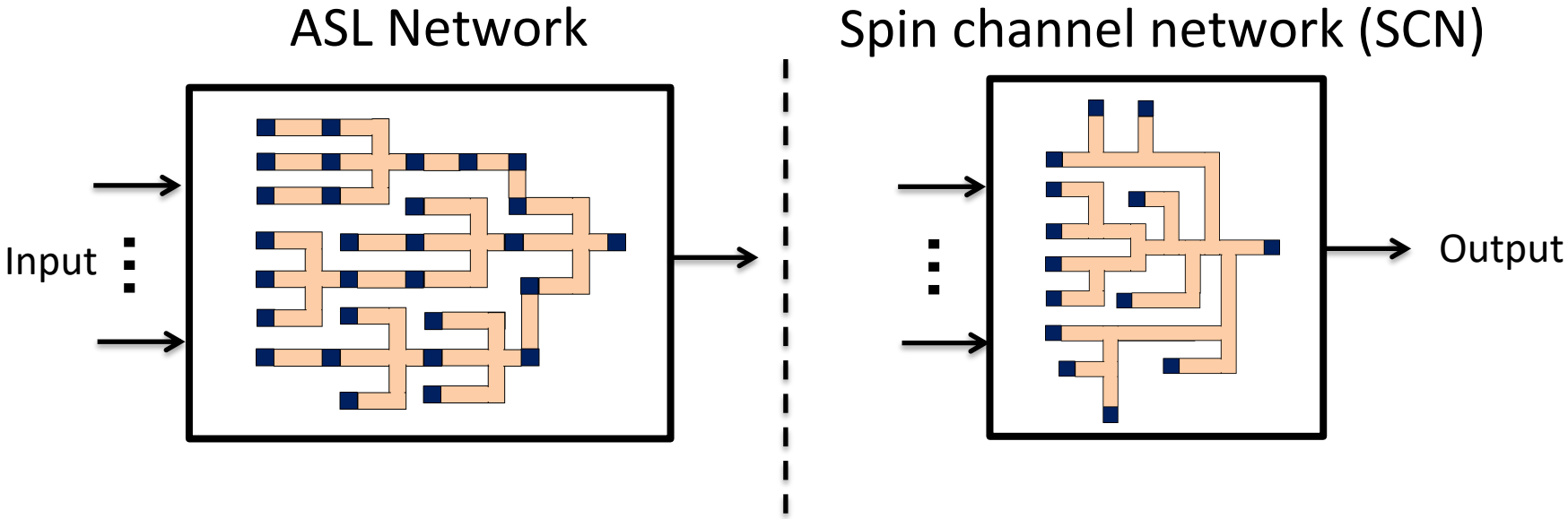


$220e^-$ ⟵ minimum charge needed to switch ⟶ $10^5 - 10^6 e^-$

spin-torque devices need 3-4 orders-of-magnitude higher charge for switching compared to CMOS

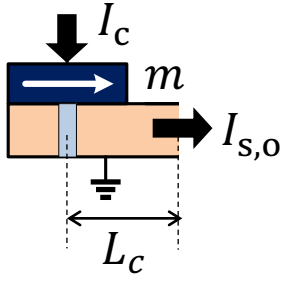
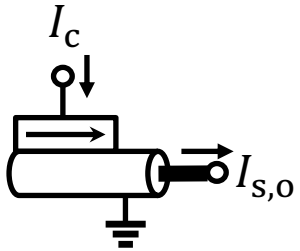
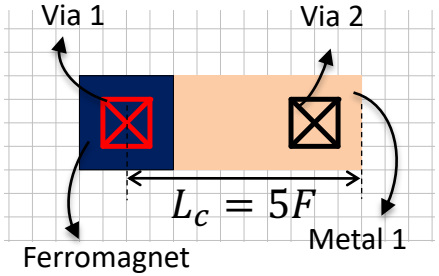
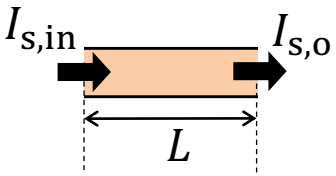
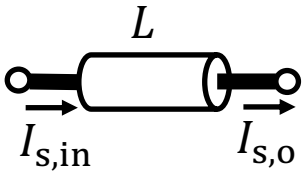
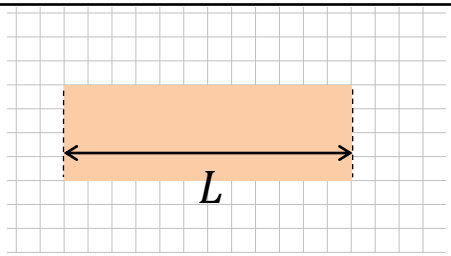
Fundamental Question

Do we really need switching at every ASL gate?



- how does one realize arbitrary computation while accumulating analog spin currents from multiple nanomagnets?
- will such approach scale with the input vector dimensionality (complexity) of the inference kernel?

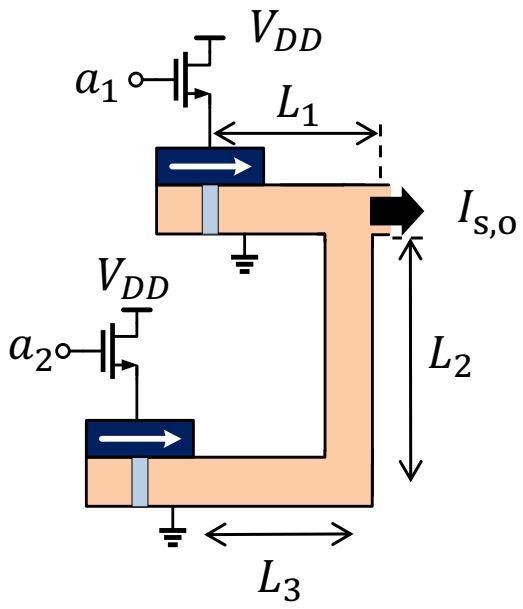
ASL Building Blocks

	Conceptual diagram	Transfer function	Symbol	Layout
Channelized nano-magnet		$I_{s,o} = \beta_m I_c m$		
Spin channel		$I_{s,o} = I_{s,in} e^{-\frac{L}{\lambda}}$		

Spin channel networks can be systematically composed using these ASL building blocks

Proposed Idea: Spin Channel Networks (SCNs)

Conceptual diagram



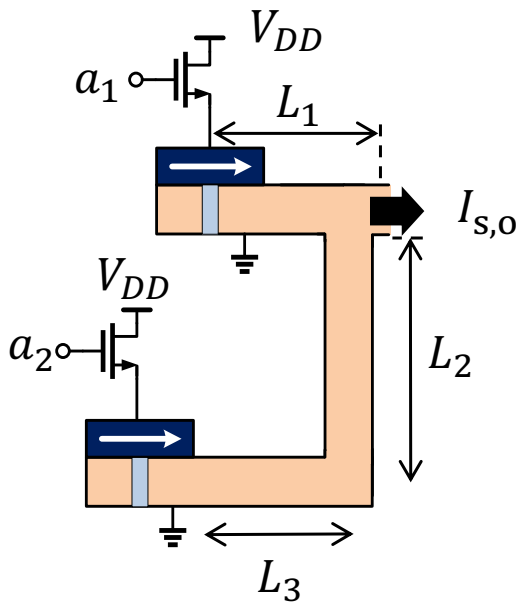
- **weighted** spin current **summing** via tailored spin channel lengths and wired-OR summing
- exploits exponential decay of spin current for computation

$$I_{s,o} \propto I_c \left(a_1 e^{-\frac{L_1}{\lambda}} + a_2 e^{-\frac{L_2+L_3}{\lambda}} \right)$$

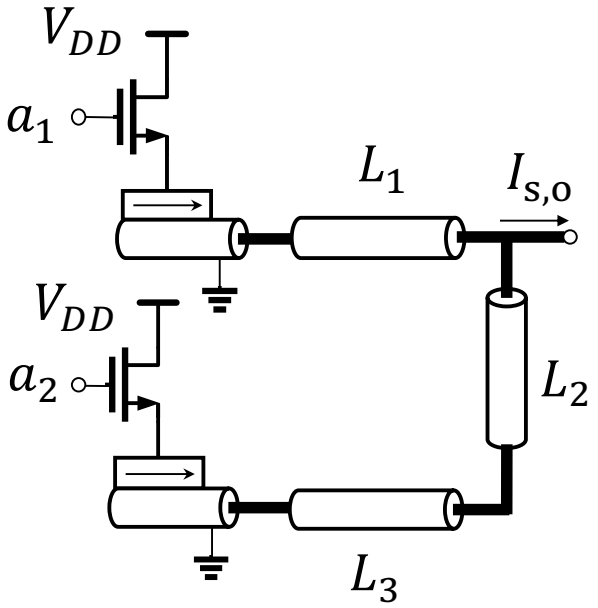
input bits $a_1, a_2 \in \{0,1\}$
λ: device constant
I_c: ON current of NMOS

Proposed Idea: Spin Channel Networks (SCNs)

Conceptual diagram



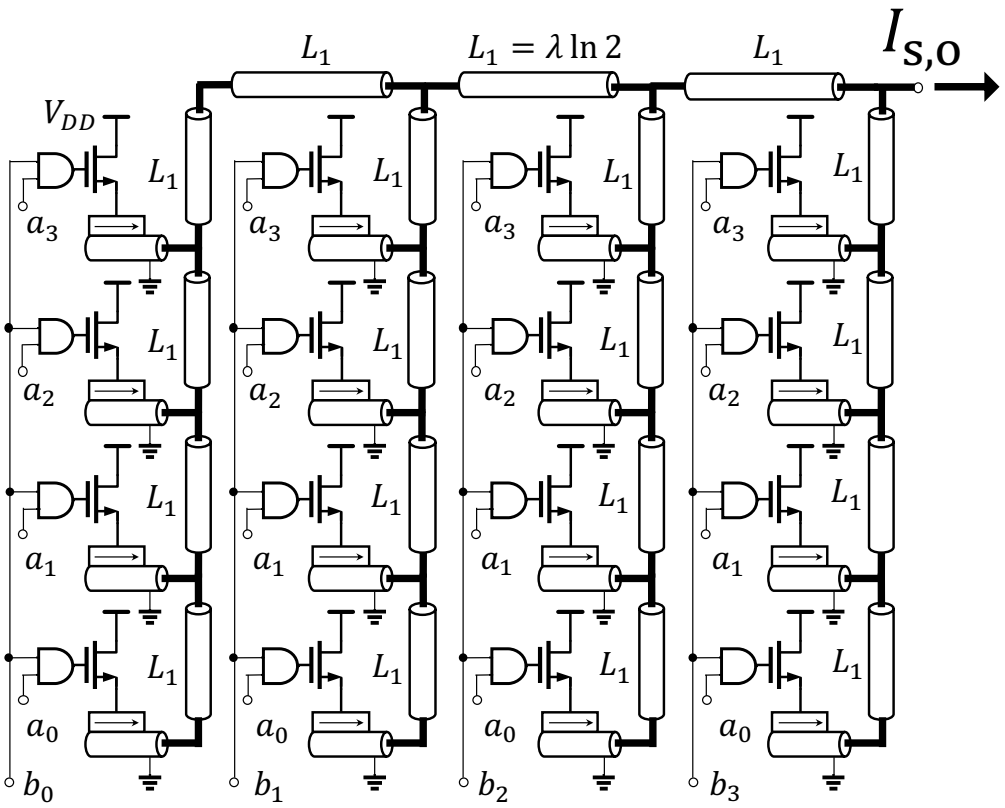
Schematic



$$I_{s,o} \propto I_c \left(a_1 e^{-\frac{L_1}{\lambda}} + a_2 e^{-\frac{L_2+L_3}{\lambda}} \right)$$

input bits $a_1, a_2 \in \{0,1\}$
λ: device constant
I_c: ON current of NMOS

4×4 SCN Multiplier



$$I_{s,o} \propto I_c \sum_{j=0}^{3} 2^j b_j \left(\sum_{i=0}^{3} 2^i a_i \right)$$

$a_0 \ldots, a_3 \in \{0,1\}$

$b_0 \ldots, b_3 \in \{0,1\}$

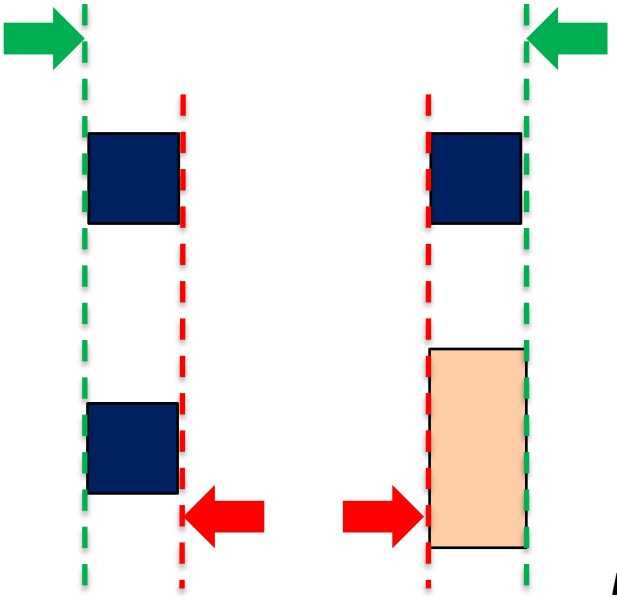
I_c: ON current of NMOS

- digital inputs; zero static power consumption
- output spin current can be combined with outputs of other channel networks

4×4 SCN Multiplier Layout

Higher energy-efficiency

(with smaller layout area)



Layout λ rules

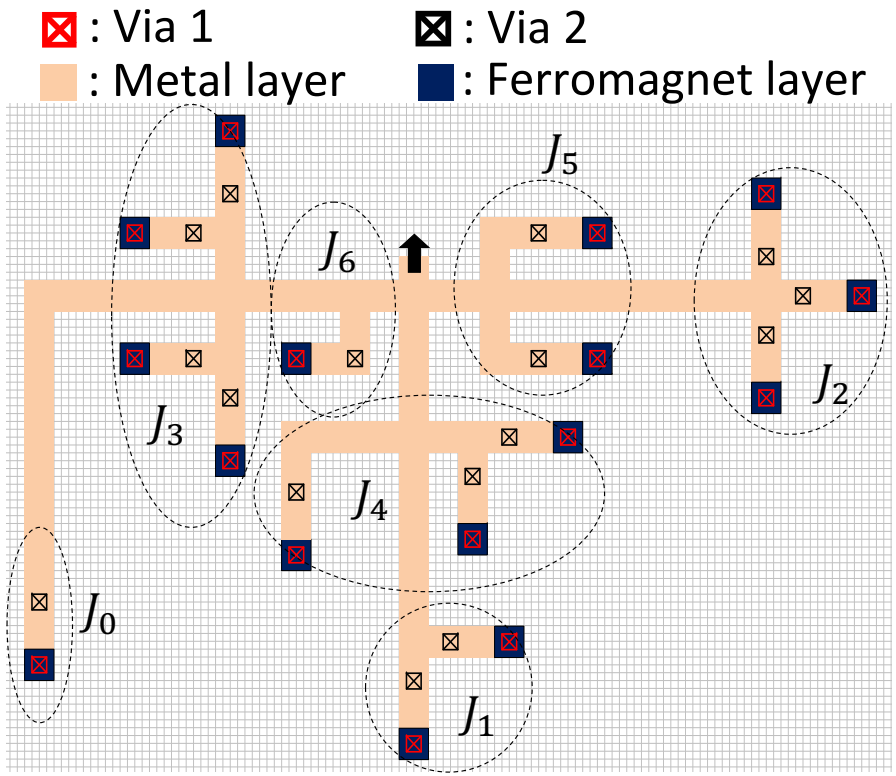
(dictate minimum spacing requirements)

[nikonov-JXCDC-2015]

Proposed layout methodology

[A. D. Patil et al, IEEE JxCDC, 2019]

***Star-of-ladders* network**

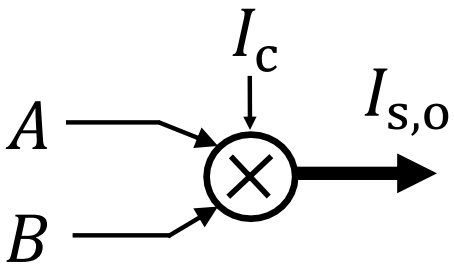


cluster J_i spin current

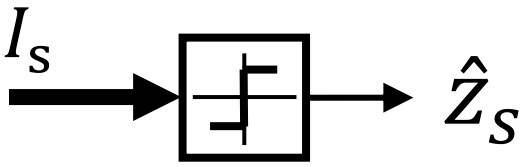
$\propto \sum_{p+q=J_i} a_p b_q$

SCN Components

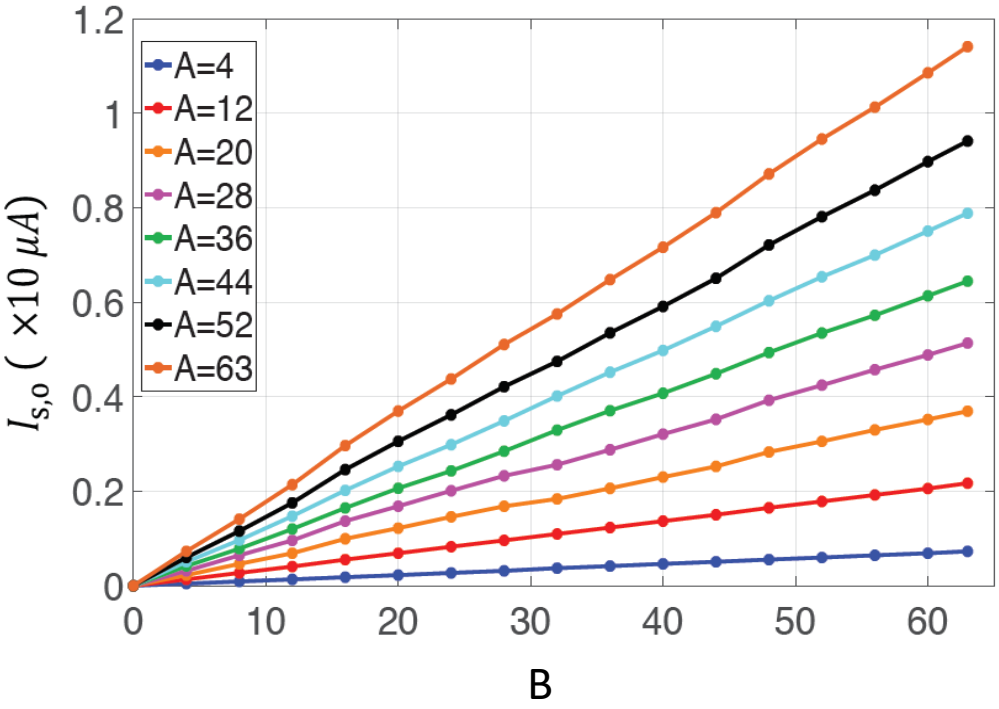
SCN multiplier



nanomagnetic slicer



6×6 SCN multiplier



(via SPICE models [Naeemi, Intel])

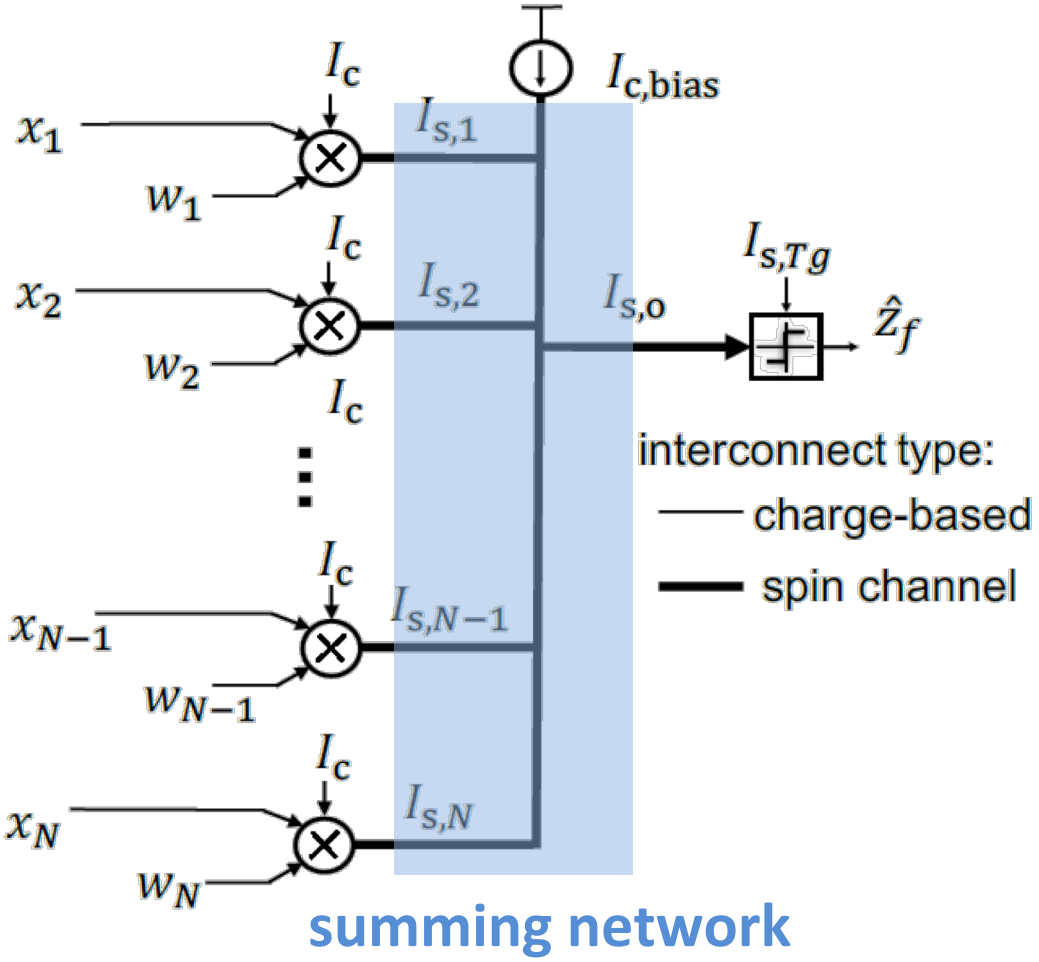
[Bonhomme-TED -2014]

SCN-based classifiers are composed using these components

Support Vector Machine (SVM) Classifier

$\boldsymbol{x}$: input feature
$\boldsymbol{w}$: weight vector
N: vector dimension
$\hat{z}_f$: final decision

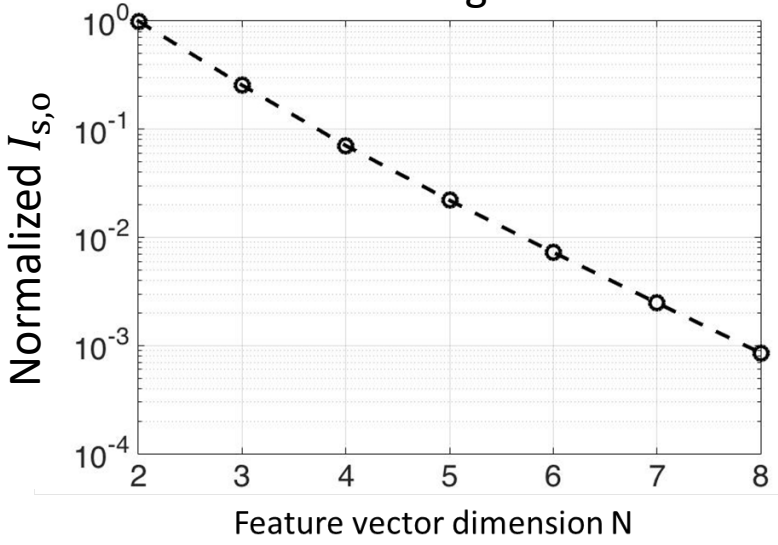
$$\hat{z}_f = 1$$
$$\boldsymbol{w^T x} + b \gtrless 0$$
$$\hat{z}_f = -1$$

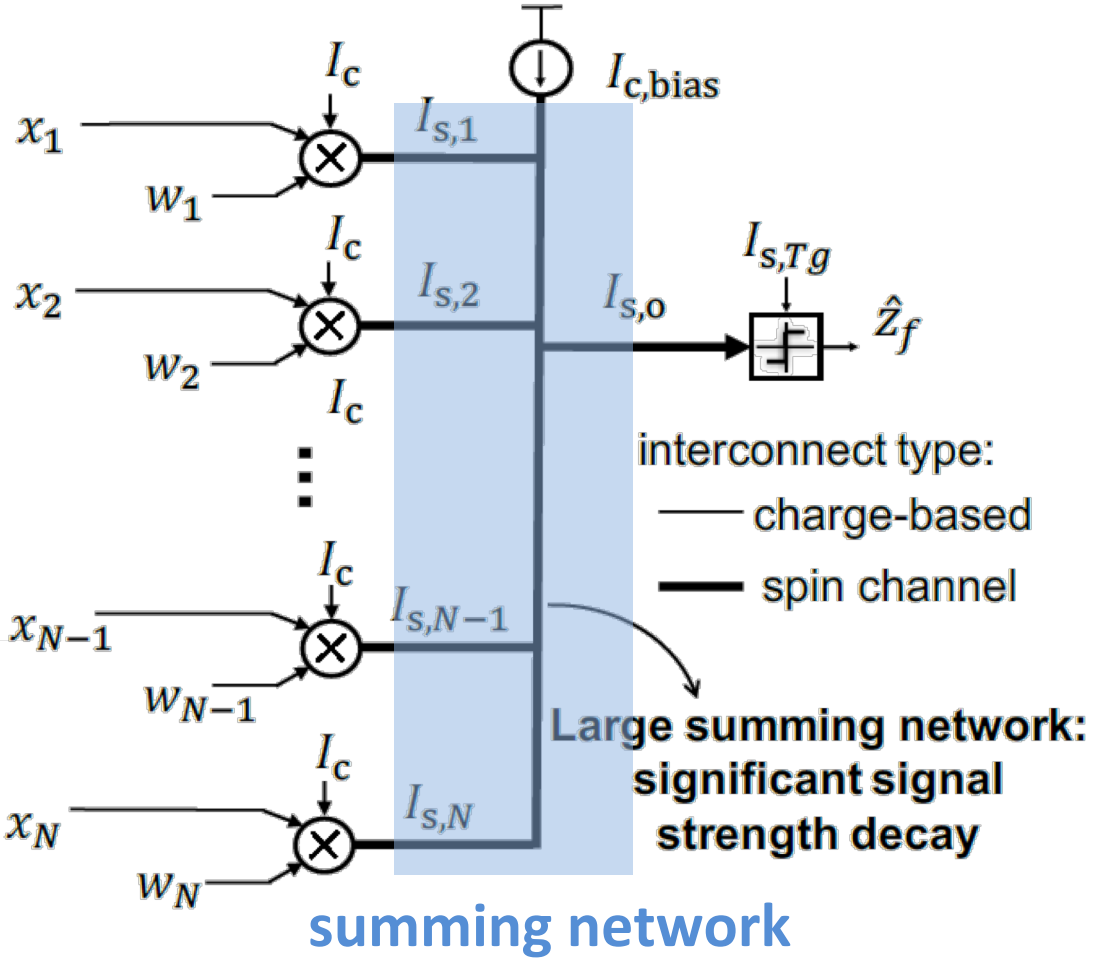


Requires a large summing network to accumulate spin currents

Signal Decay over the Summing Network

Exponentially decaying $I_{s,o}$ with increasing N





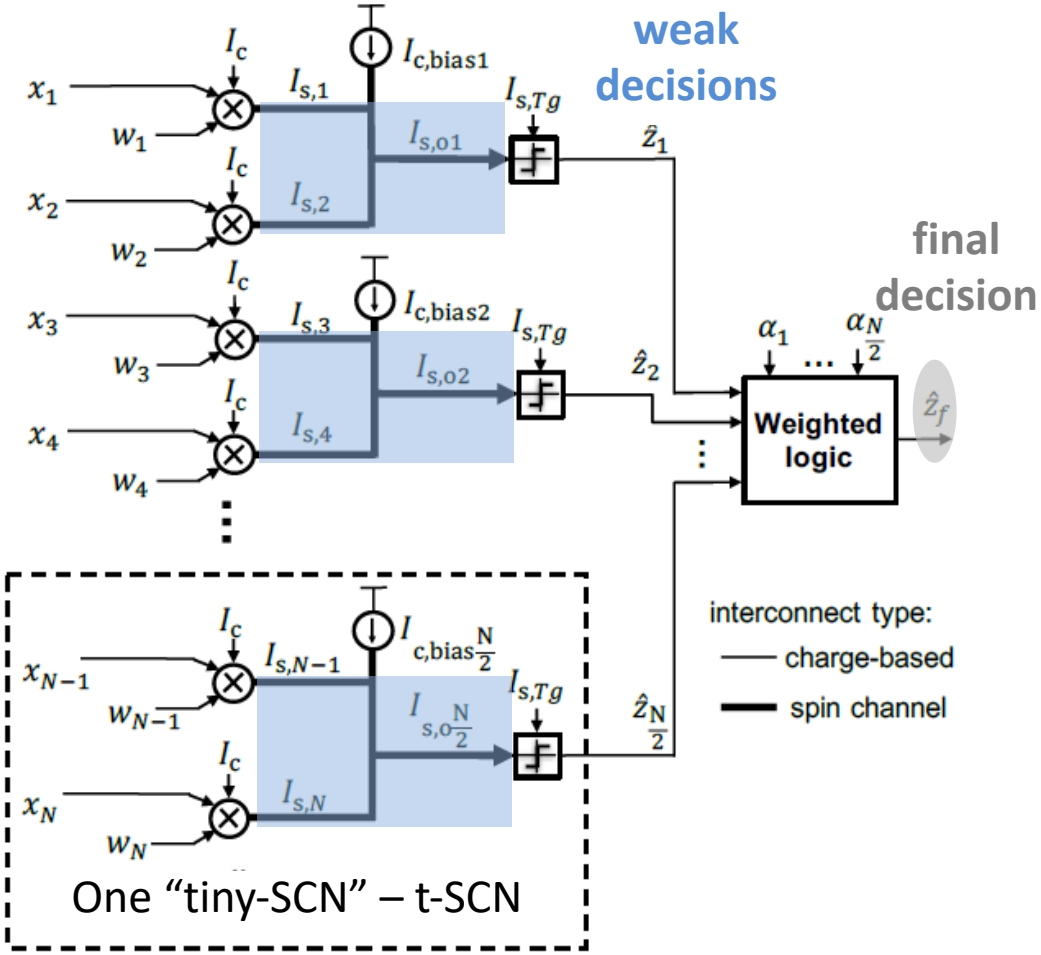
Energy-efficiency ***scalability***

Boosted SVM Classifier:

- split given classification into multiple small ($N = 2$) classifiers
- combine weak decisions to obtain a strong decision

  [Freund & Schapire-ICML-1996]

- $N/2$ smaller summing networks



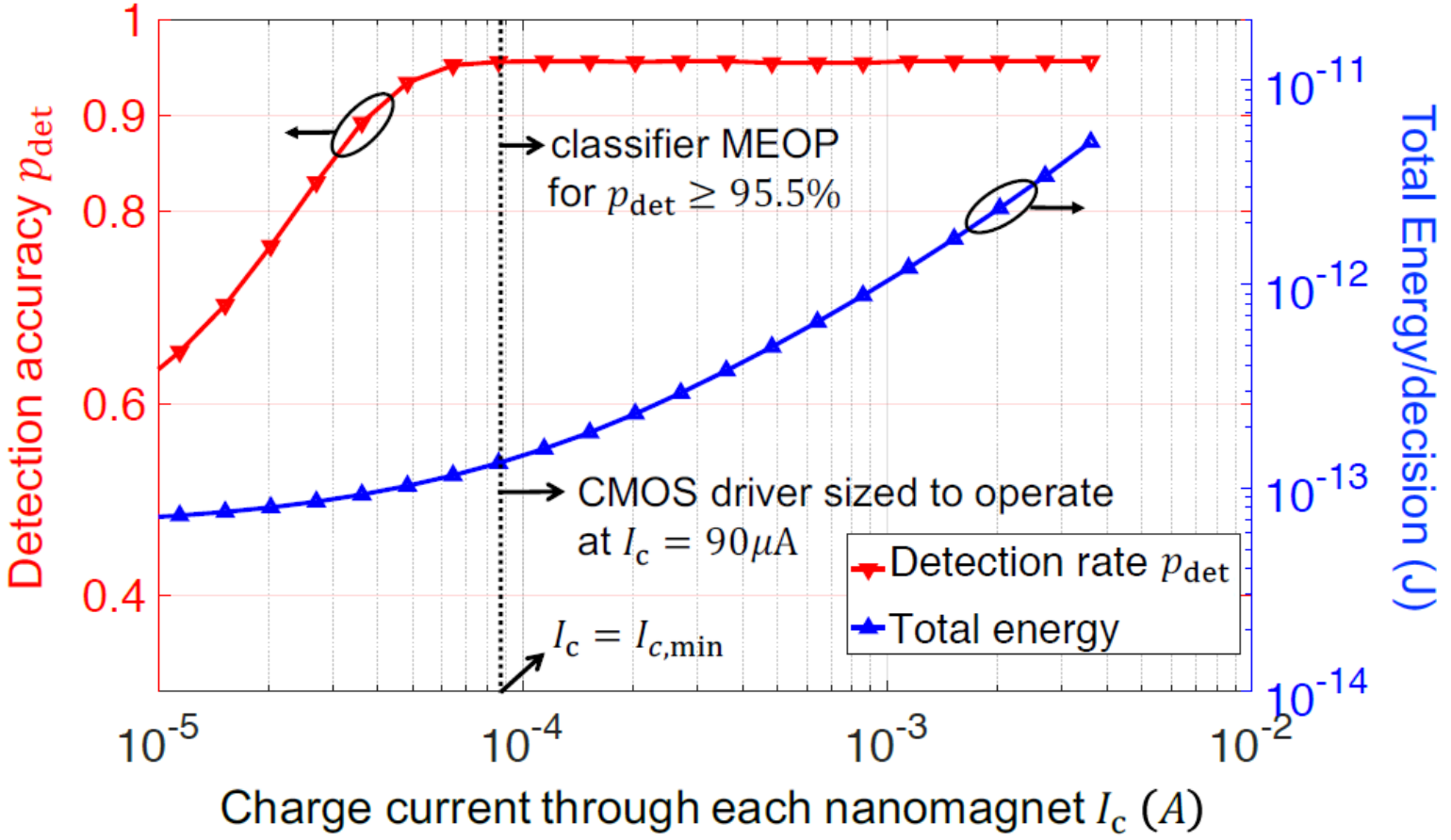
by employing systems theory along with circuit design

Simulation Set-up

- Implementations to be compared
  - Boolean ASL network w/ and wo/ boosting
  - Boolean CMOS network w/ and wo/ boosting
  - Proposed boosted SCNs
- Evaluation metrics:
  - energy/decision, decision throughput
  - accuracy of the classifier (p_{det}: probability of correct classification)
- We choose identical decision throughput (of $\frac{1}{4\mathrm{ns}}$) for all implementations

Boosted SCN Accuracy vs Energy Tradeoff

(at a fixed decision delay)

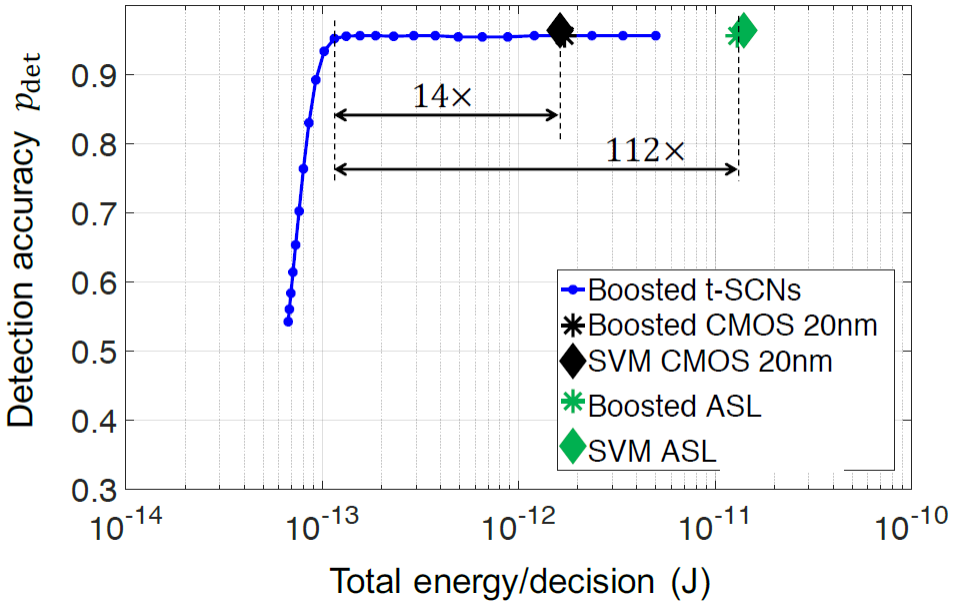


There exists a minimum energy operating point (MEOP) for given accuracy

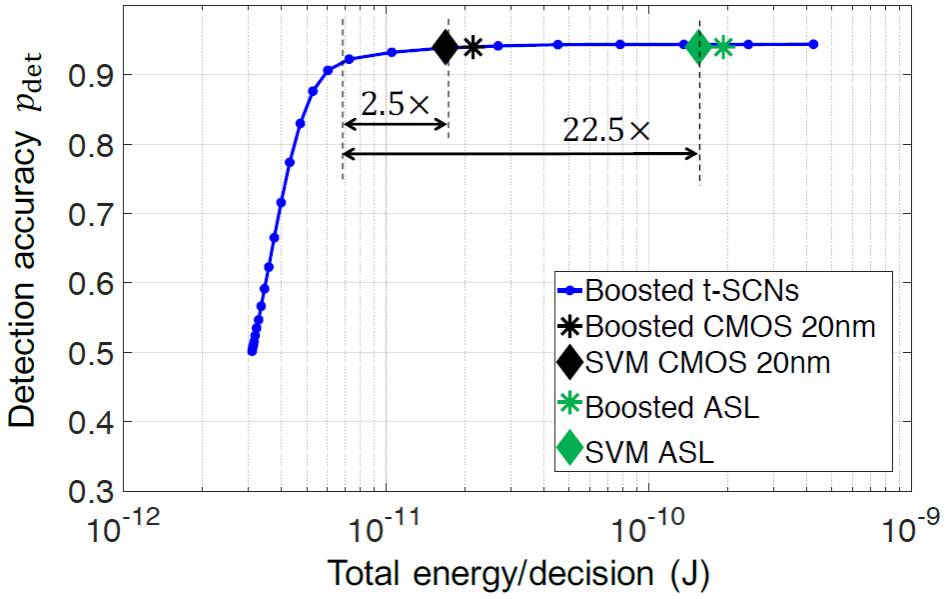
Final Simulation Results

(at a fixed decision delay)

Breast Cancer dataset
$(N = 10)$
UCI repository



Face Detection dataset
$(N = 100)$
MIT CBCL



Boosted SCNs achieve 2.5x-to-14x higher energy-efficiency compared to CMOS implementations

See [A. D. Patil et al, IEEE JxCDC, 2019] for more details

Conclusion and Future Work

- Boosted spin channel networks achieve up to 112x and 14x lower energy compared conventional spin and CMOS implementations, respectively

- Employing systems theory along with circuit innovations opens up fundamentally new design spaces

- Next steps:
  - Evaluate the impact of process variations on the decision accuracy
  - Extend the proposed approach to efficiently implement multi-layer deep neural networks

Thank You!

This work was supported in part by C-BRIC, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) Program sponsored by DARPA, and in part by SONIC, one of the six SRC STARnet Centers sponsored by MARCO and DARPA.