Intelligent Circuit Design and Implementation with ML in EDA

Zhiyao Xie

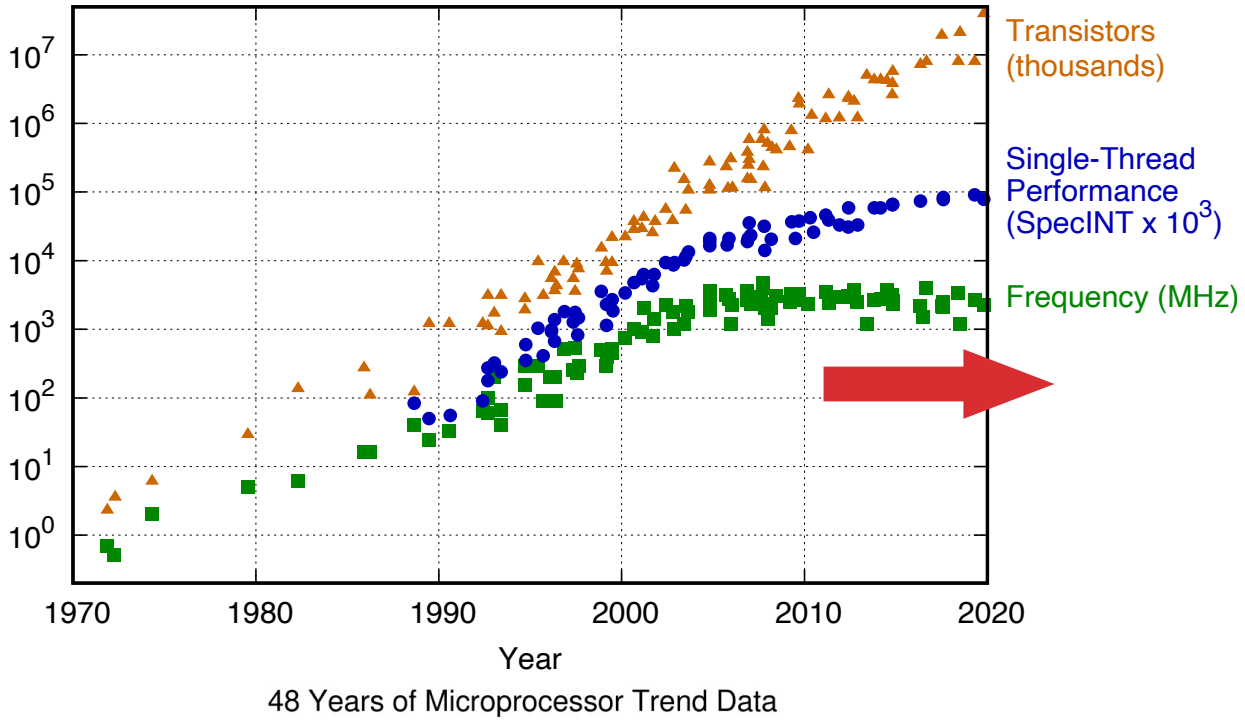
Duke University

Advisors: Yiran Chen, Hai (Helen) Li

Chip Design Challenges

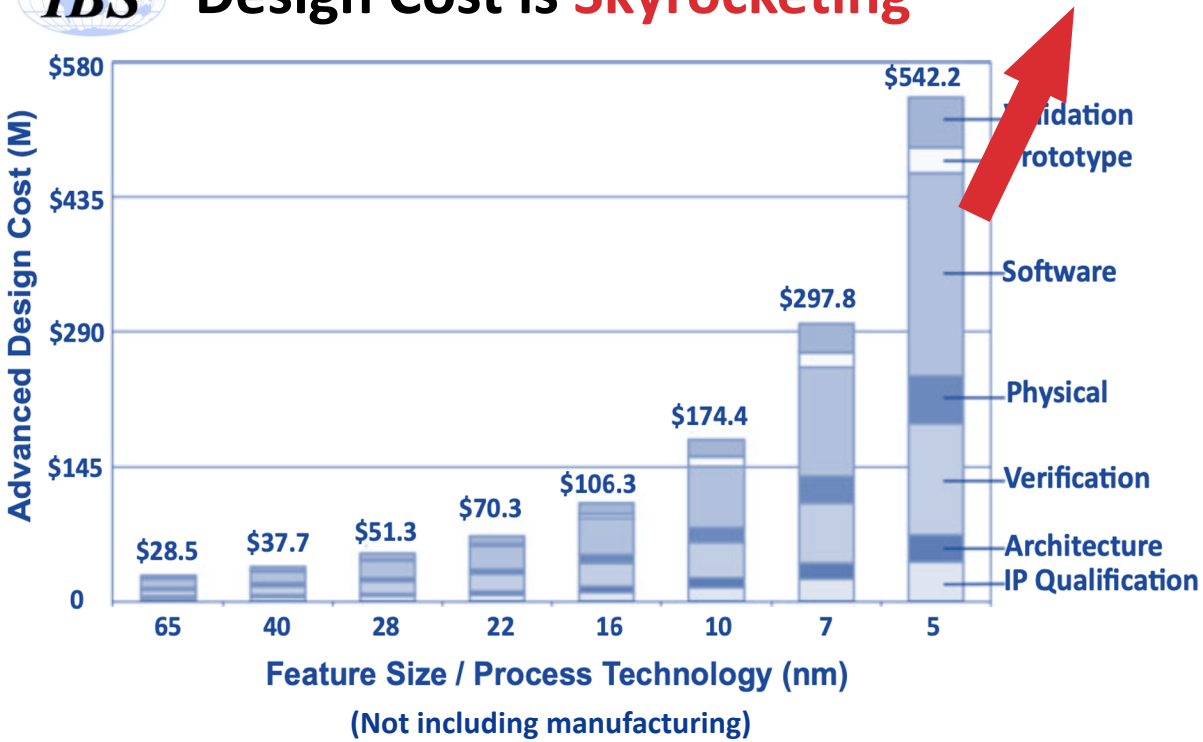
Diminishing performance gain and increasing design cost

Per-Core Performance Gain is Diminishing



Partially collected by M. Horowitz et al. Plotted by Karl Rupp, 2020

IBS Design Cost is Skyrocketing



International Business Strategies, 2020

Duke UNIVERSITY

My Related Works

PPA

Power

Power & Power Delivery Challenges

[ICCAD'20], [ASPDAC'20],

[MICRO'21] (Best Paper Award)

Performance

Timing & Interconnect Challenges

[ICCAD'20], [ASPDAC'21],

[TCAD'21] (under review)

Area

Routability Challenges

[ICCAD'18], [DATE'18], [ICCAD'21]

Overall Flow Tuning

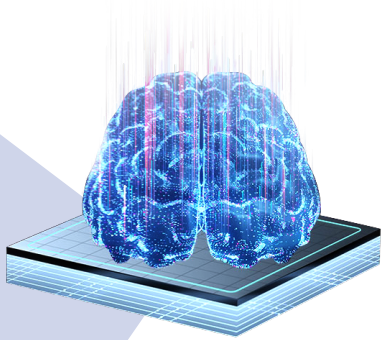
[ASPDAC'20]

Covered in this talk

Case Study 1:

Routability Challenges

My Roadmap Towards Intelligent Chip Design



True intelligence in chip design

ML for Chip Design

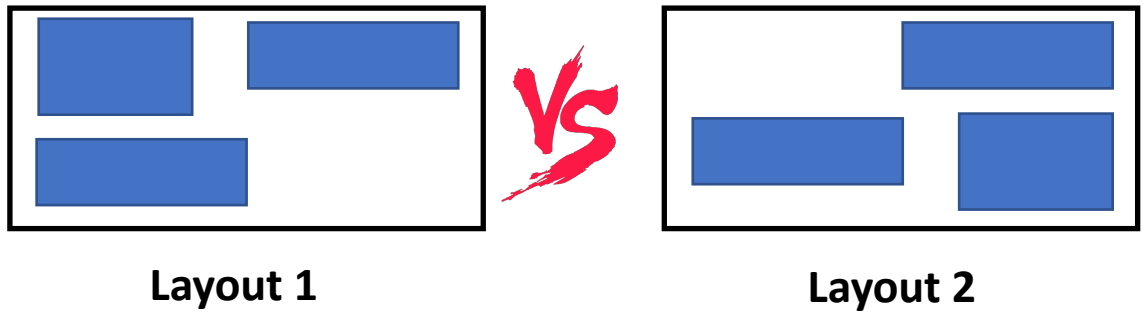
Well-studied in recent years

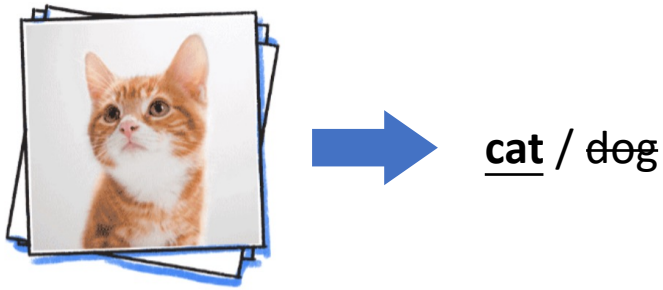
Traditional Chip Design

Source: Image adapted from Cadence Cerebrus.

First Deep Learning Method for Routability Prediction

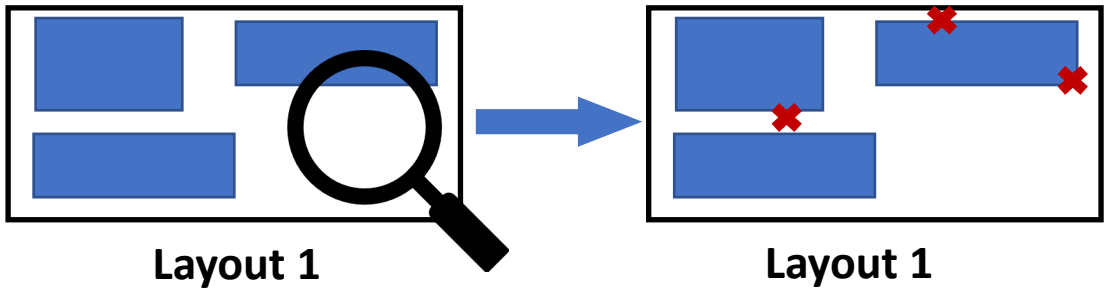
- Task 1: which one will result in less DRV count?

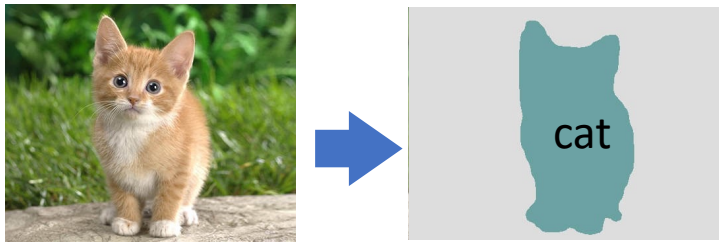




Customized CNN methods

- Task 2: where are DRC violations?



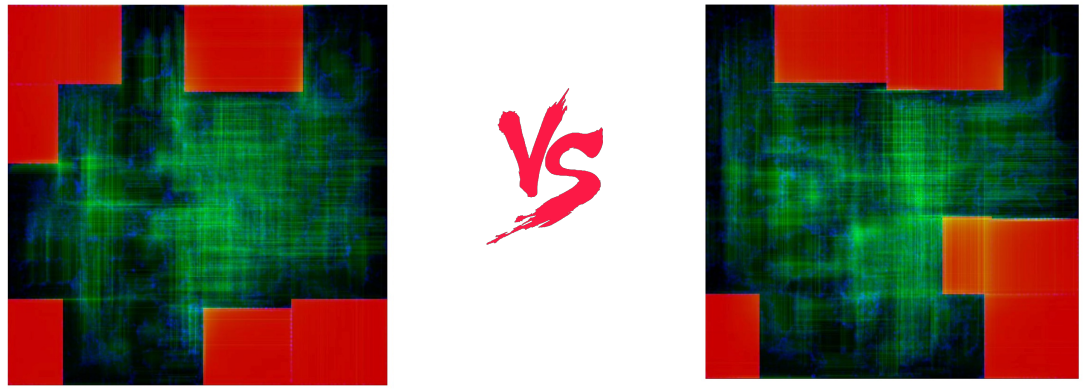


Customized FCN methods

RouteNet [Xie, et al., ICCAD'18]

First Deep Learning Method for Routability Prediction

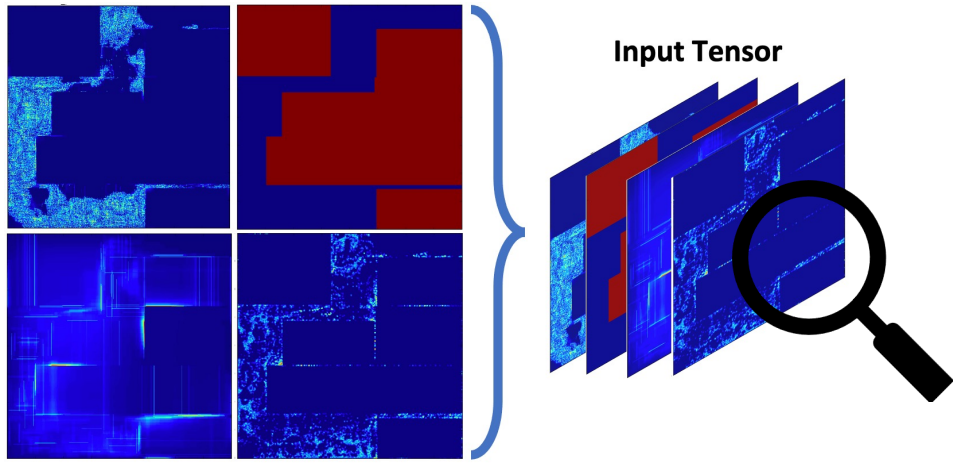
- Task 1: which one will result in less DRV count?



- Requires global routing:
  ~~Hours~~ * Number of Layouts

In seconds, with close accuracy

- Task 2: where are DRC violations?



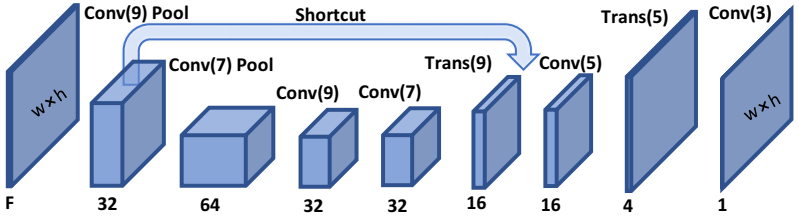
- Requires detailed routing
  ~~More hours~~ * Iterations

In seconds, outperform previous works

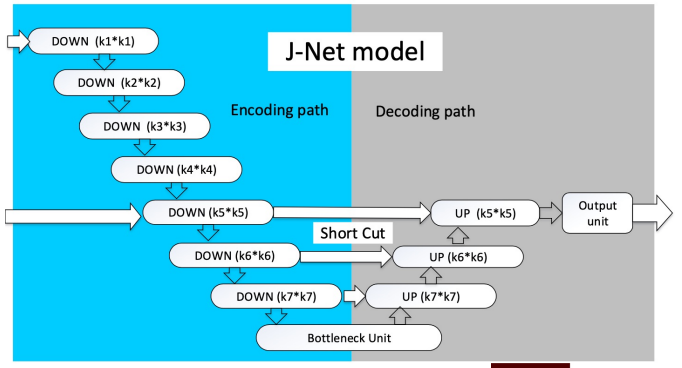
RouteNet [Xie, et al., ICCAD'18]

Duke UNIVERSITY

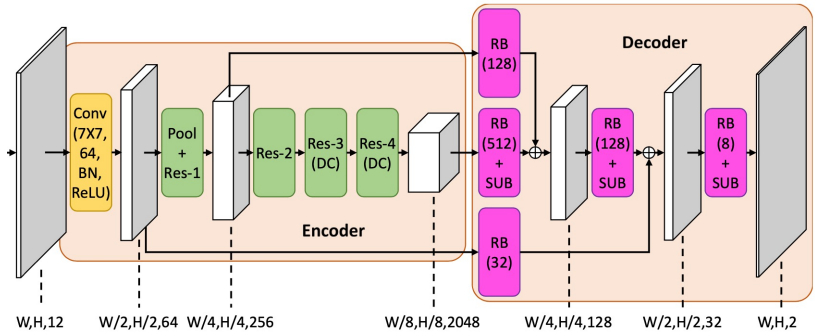
Many Excellent Deep Learning Methods



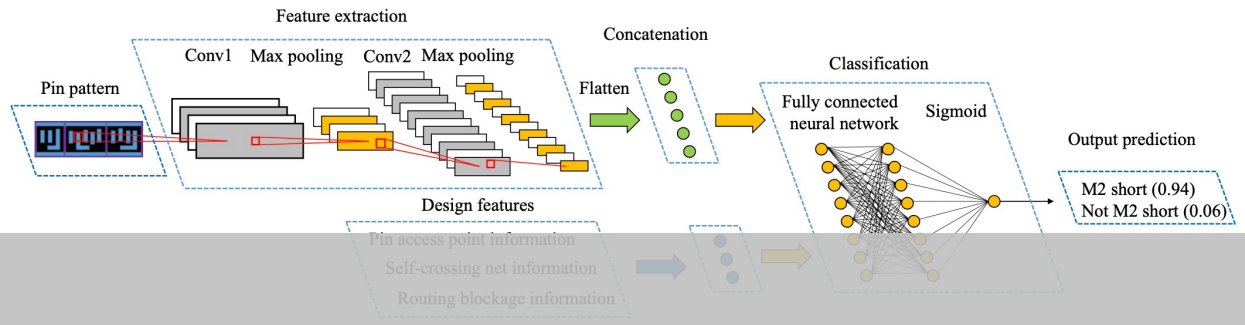
RouteNet [Xie, et al., ICCAD'18] Duke UNIVERSITY



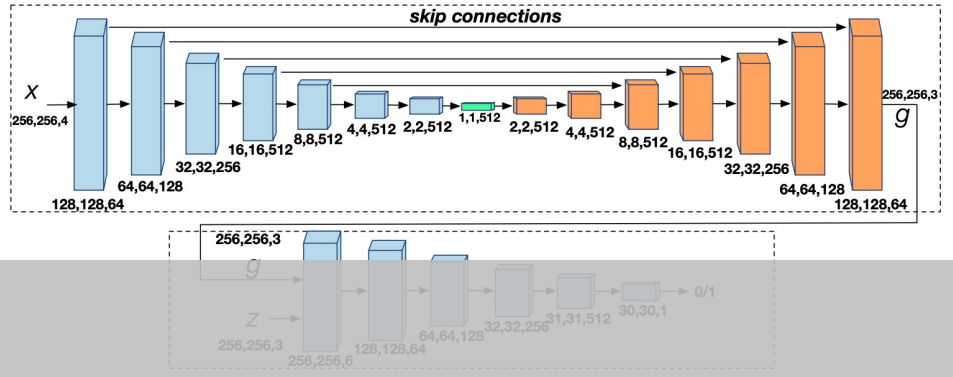
J-Net [Liang, et al., ISPD'20]



PROS [Chen, et al., ICCAD'20]

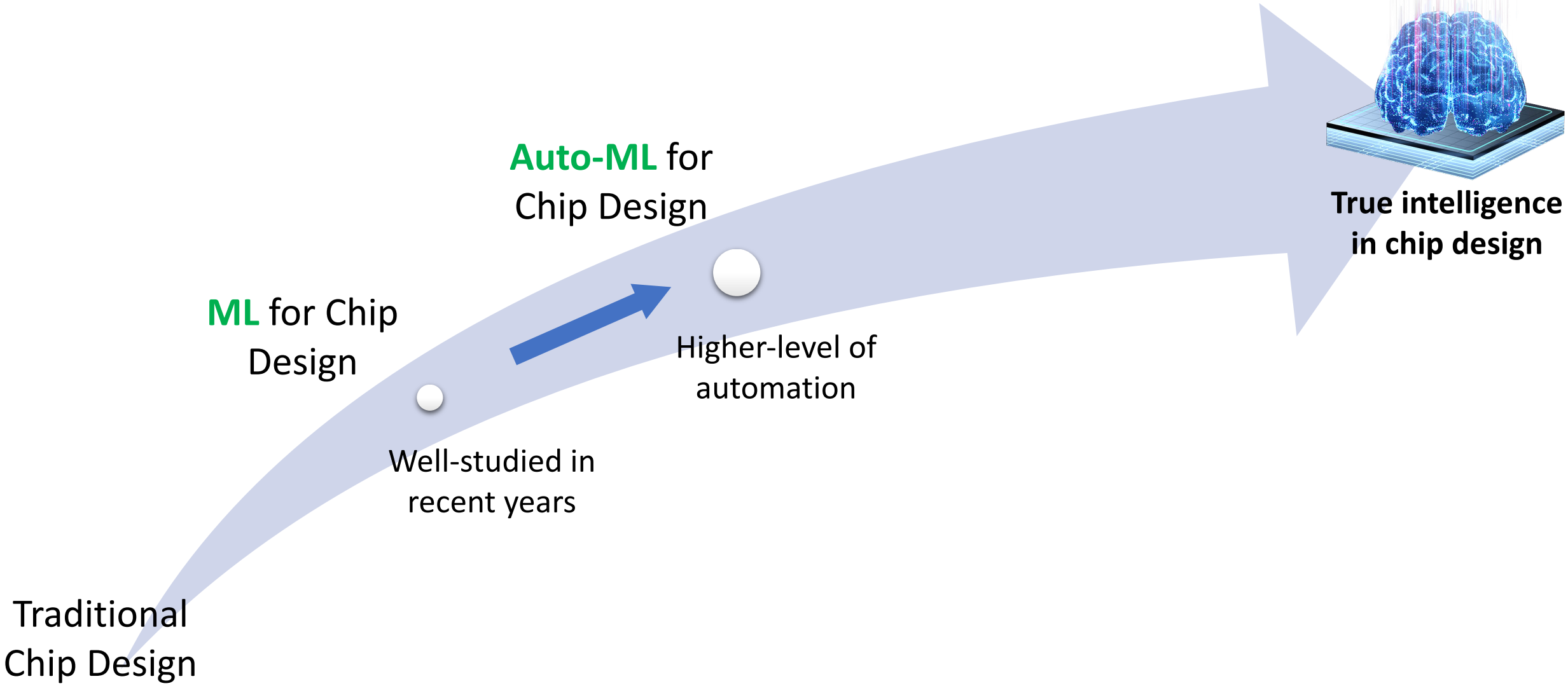


[Yu, et al., DAC'19]



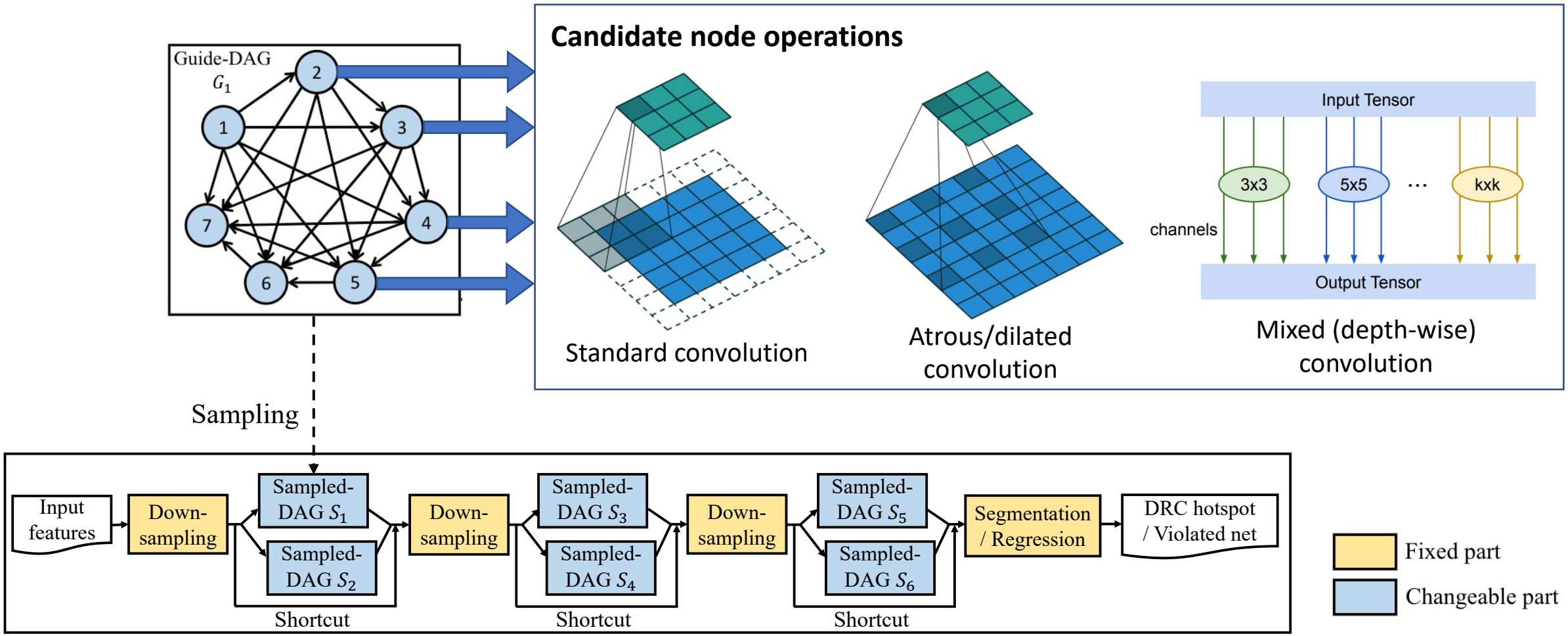
Tremendous Engineering Efforts Required!

My Roadmap Towards Intelligent Chip Design



Duke UNIVERSITY

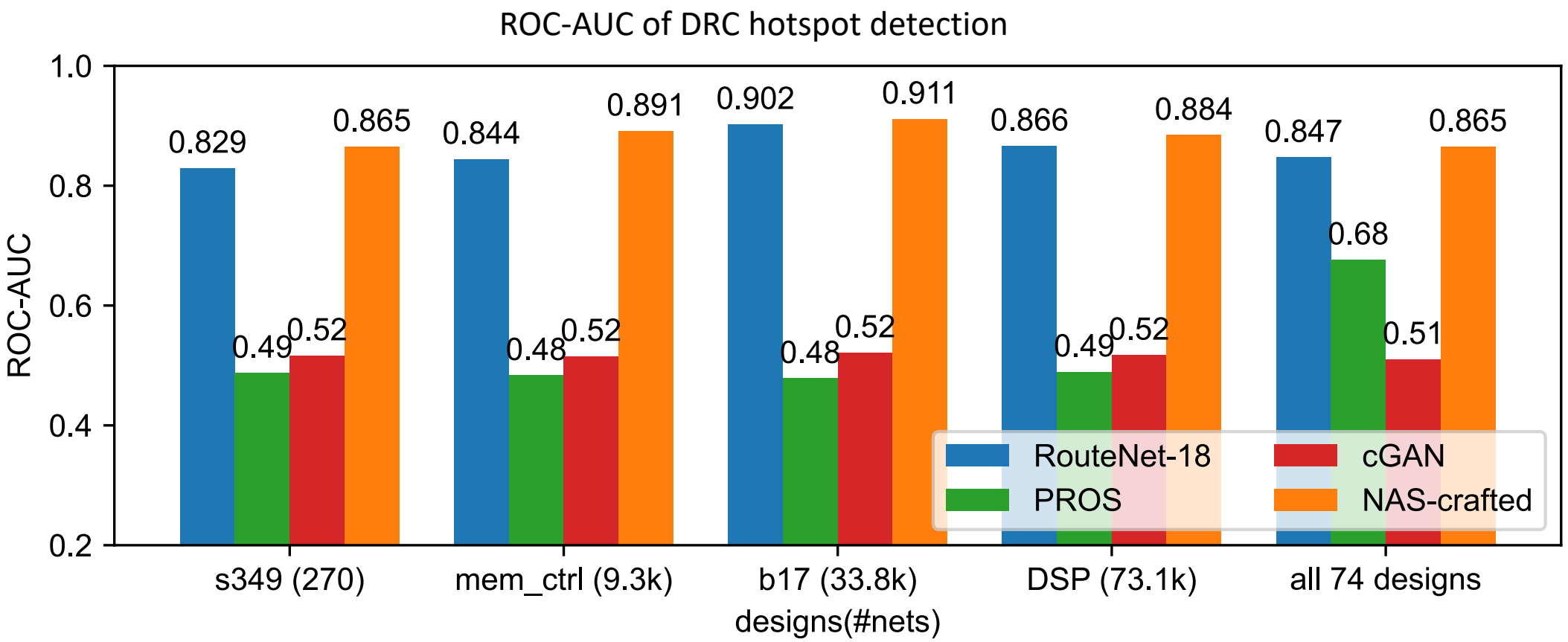
Automatic Estimator Development – Search Space



[Chang, et al., ICCAD'21]

DRC Hotspot Detection Results

- This model **outperforms** RouteNet [ICCAD'18], PROS [ICCAD'20], and cGAN [DAC'19]
- Developed without human in **one day**



Duke UNIVERSITY

Case Study 2:

Power & Power Delivery Challenges

My Roadmap Towards Intelligent Chip Design

Traditional Chip Design

ML for Chip Design

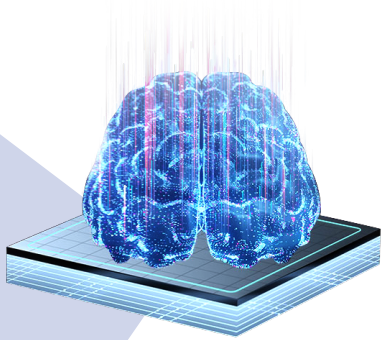
Well-studied in recent years

Auto-ML for Chip Design

Higher-level of automation

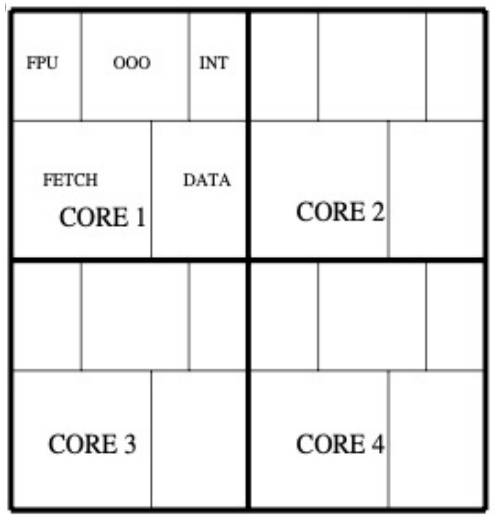
Unified ML for Both Design & Runtime

Benefit the whole chip life cycle

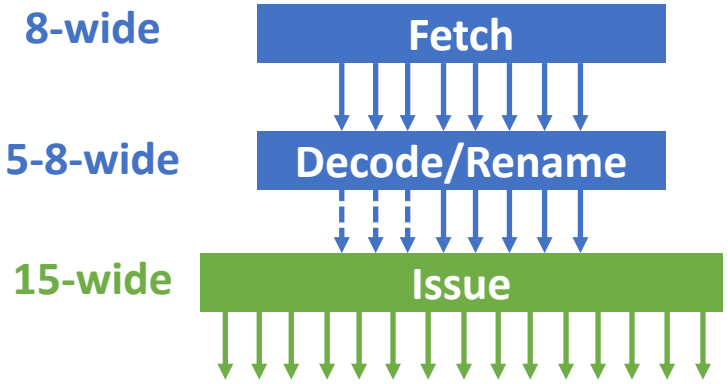


True intelligence in chip design

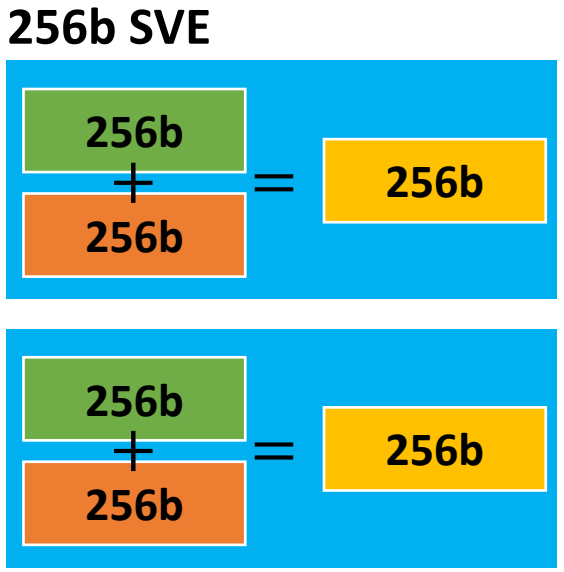
Challenge 1 – Design-time Power Introspection



Many-core CPU with more transistors



Wider issue



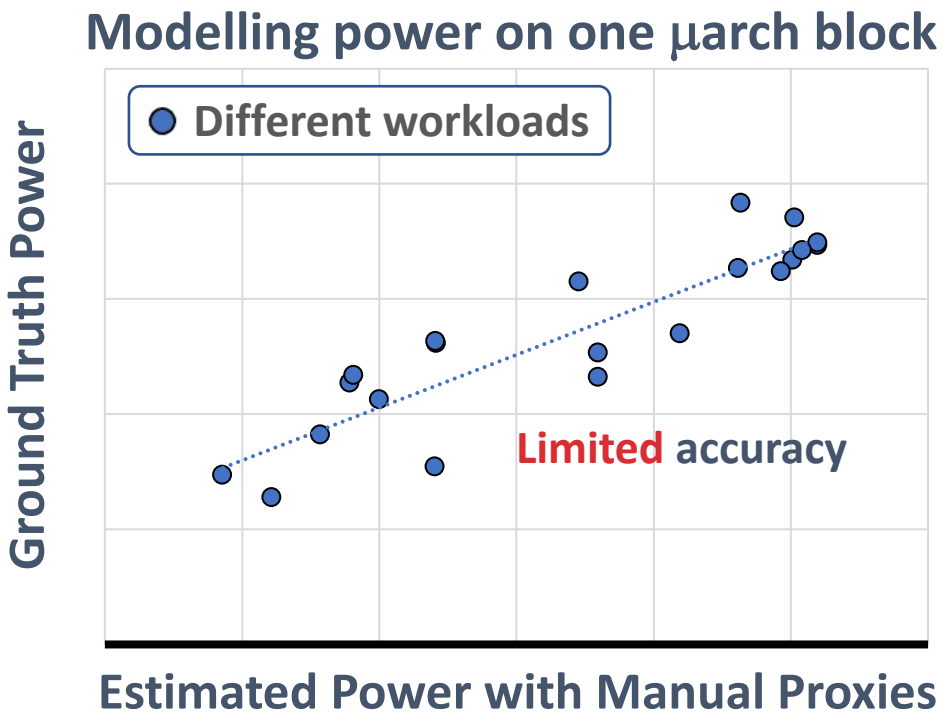
More vectored execution

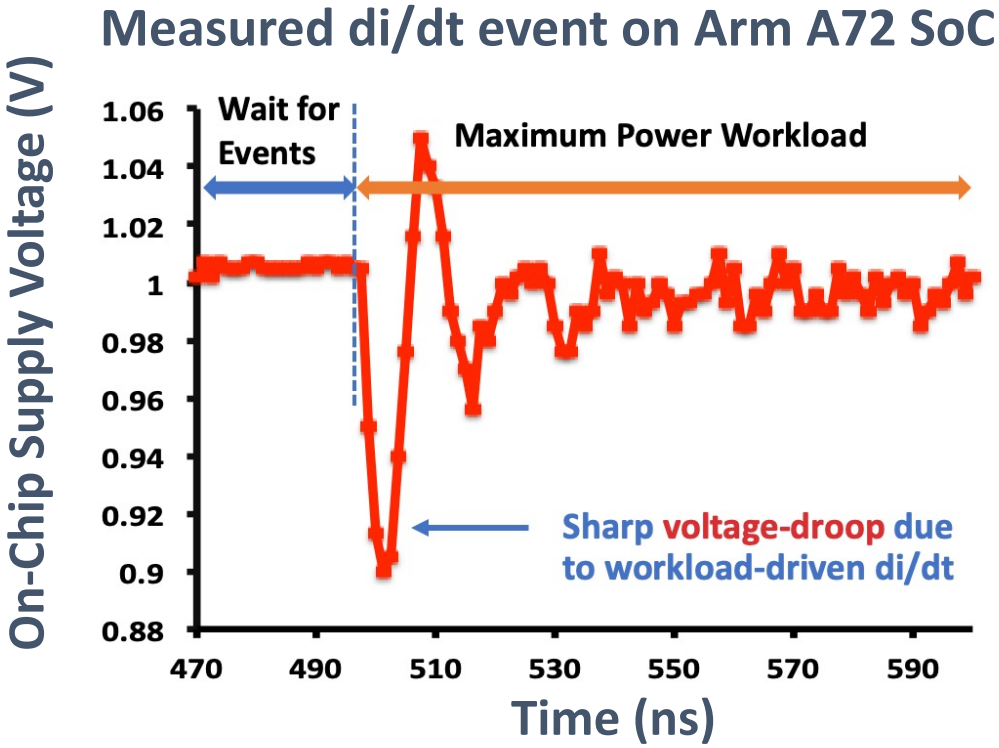
- Delivering generational performance gains **adversely impacts** CPU power
- Power-delivery resources **not keeping pace** with CPU power demands
- **Increasing power-sensitivity drives the need for design-time introspection**

APOLLO [**Xie**, et al., MICRO'21] (**Best Paper Award**)

Source: Arm Neoverse V1, 2021

Challenge 2 – Run-time Power Introspection





- **Peak-Power mitigation requires accurate power estimation to drive throttling**
  - Manually inferring proxies is very difficult in complex modern CPUs
- **Abrupt changes in CPU current-demand (di/dt event) leading to deep voltage-droop**

Duke UNIVERSITY

Power-Performance Trade-offs

Generational gains in both IPC and FMAX

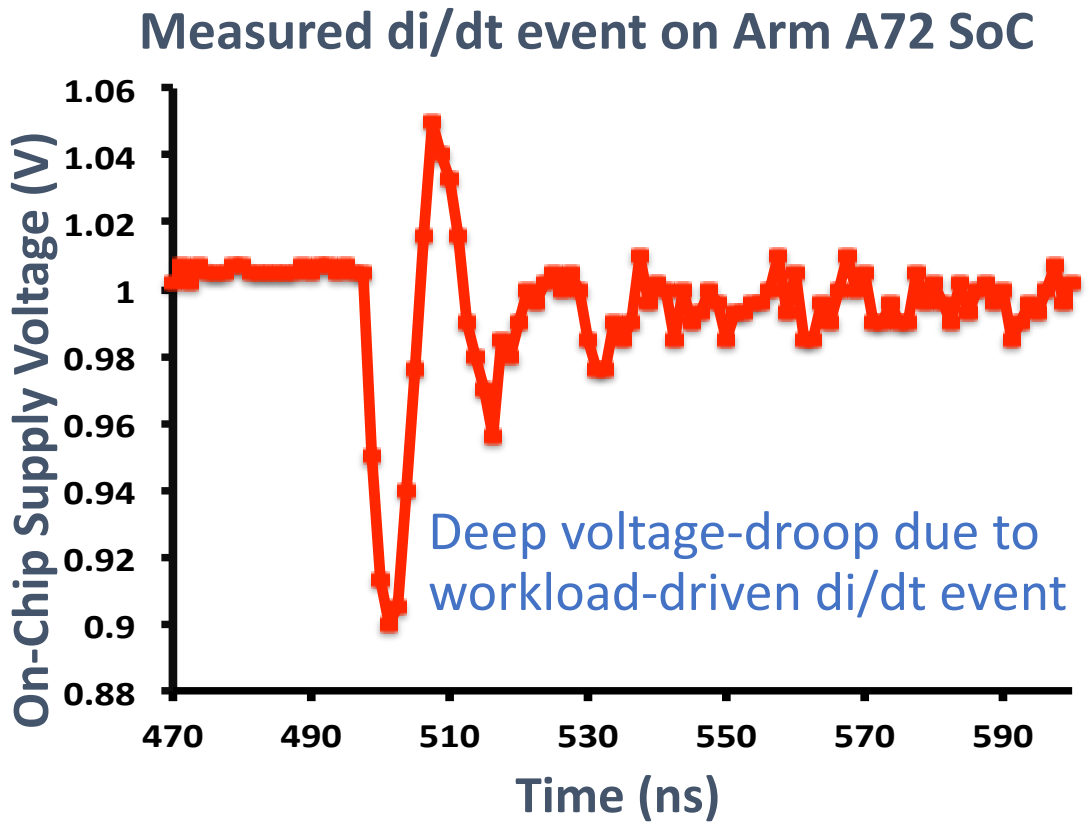
- Wide issues queues, vector-execution

Power consumption is adversely impacted

- Diminishing returns with scaling
- Increased transistor integration

Power-delivery resources not keeping pace

- Resistive-interconnects in scaled nodes
- Package-technology unable to sustain di/dt demands



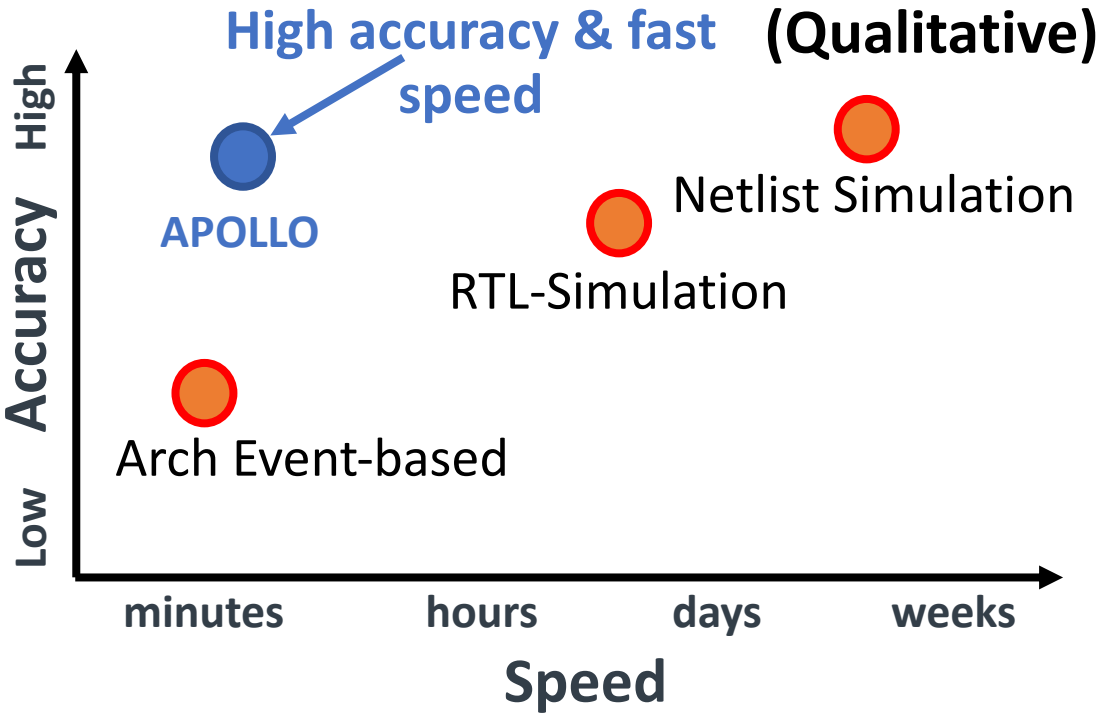
Increasing power-sensitivity drives the need for power-introspection at design and runtime

Duke UNIVERSITY

Challenges from Both Design-time and Runtime

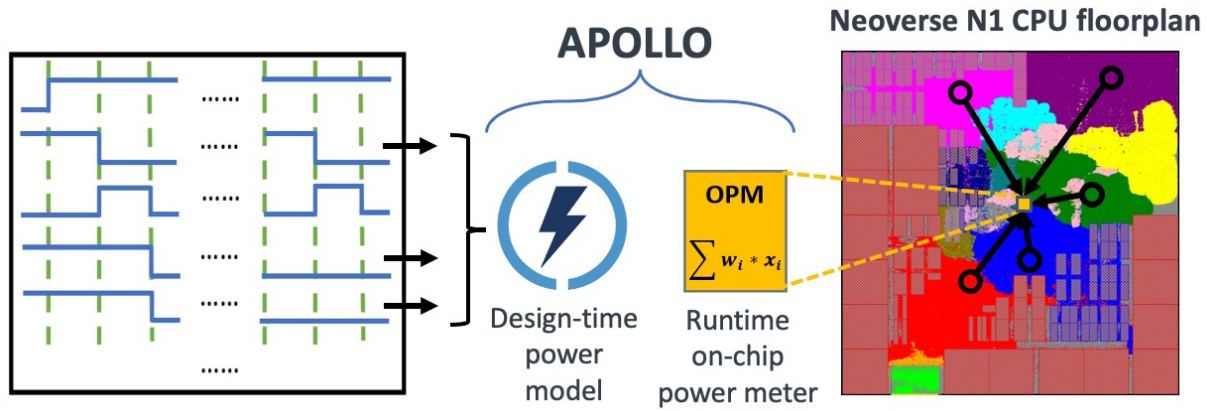
A unified solution for both scenarios

Design-time Power Simulation Trade-offs (Qualitative)



APOLLO: A Unified Power Modeling Framework

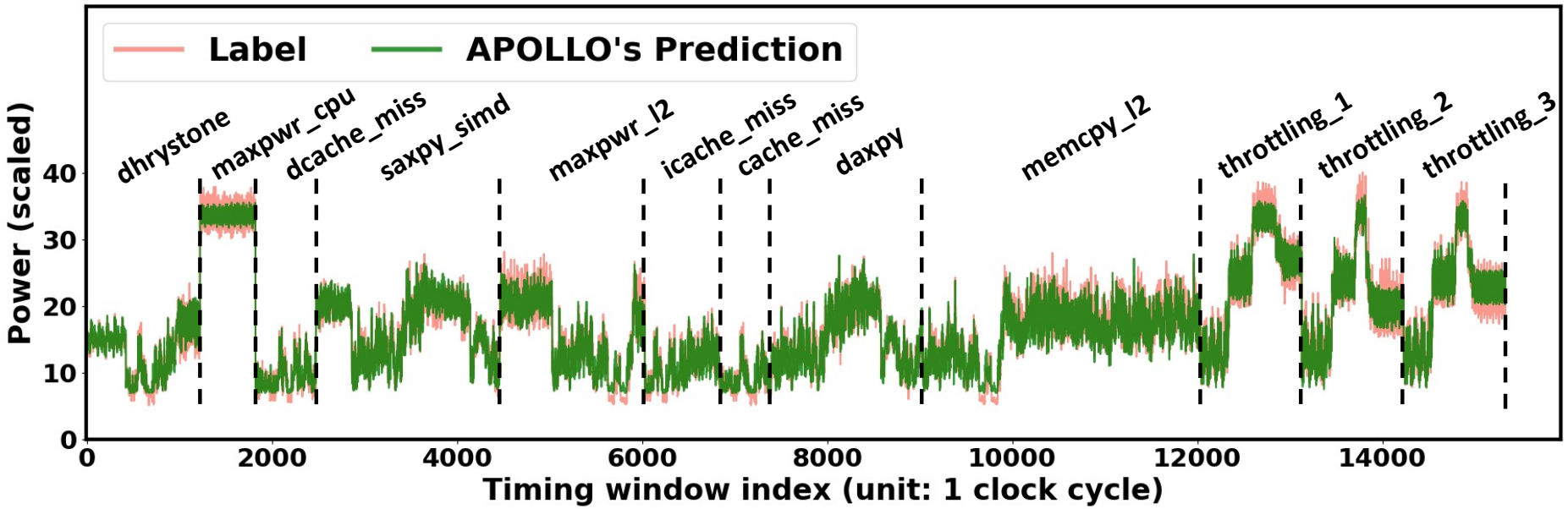
- **Fast**, yet **accurate** design-time simulation
- **Low-cost**, yet **accurate** runtime monitoring
- Design-agnostic **automated** development



Duke UNIVERSITY

Prediction Accuracy as Design-Time Power Model

Per-cycle prediction from APOLLO with $\boldsymbol{Q}$=159 proxies



- MAE = 7.19%
- R^2 = 0.953

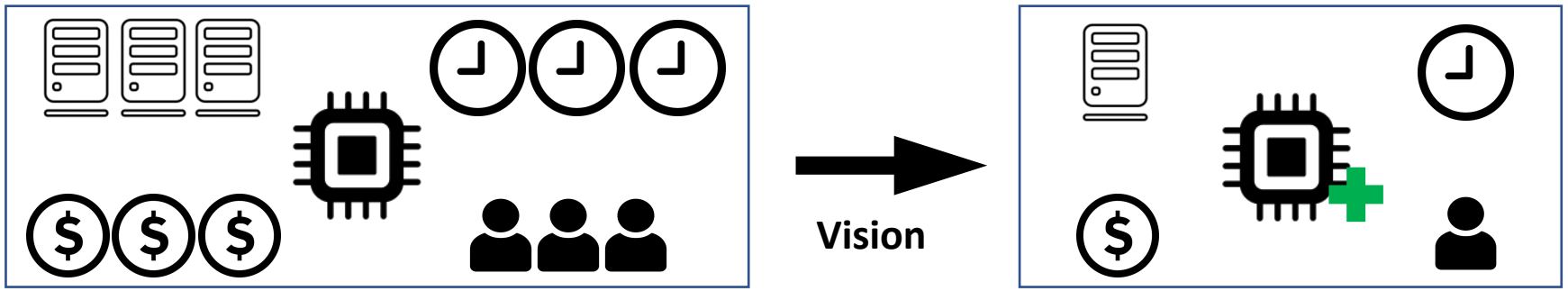
Neoverse N1 (infra)
Deployed in AWS Graviton

Cortex A77 (mobile)
Deployed In Snapdragon 865

Duke UNIVERSITY

Summary and Takeaway

- Problem: Increasing Challenges in Chip Design
  - Cost, time-to-market, reliance on designers, diminishing performance return, ……
- **ML** in chip design
  - Less simulation time, faster feedback, less designer effort
- **AutoML** in chip design
  - Reduces months of model development to hours, no developers
- **Unified ML** in both design & runtime
  - Benefit the entire chip life cycle



Thanks! Questions?

If you have further questions, please contact me:

zhiyao.xie@duke.edu

Duke