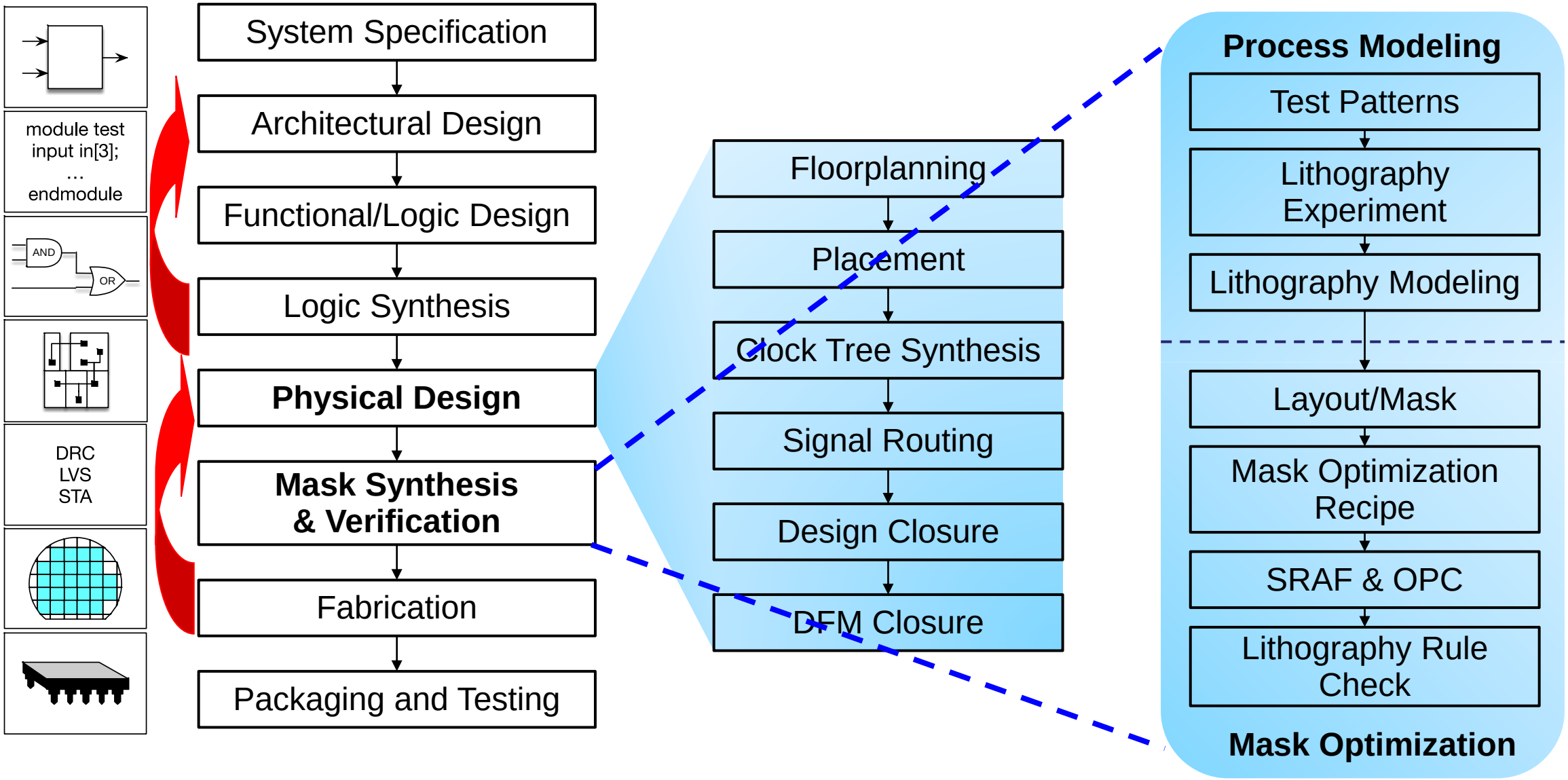
EDPS 2023 Dinner Keynote

AI for Chip Design & EDA: *Everything, Everywhere, All at Once (?)!*

David Z. Pan, Silicon Labs Endowed Chair
The University of Texas at Austin
dpan@ece.utexas.edu

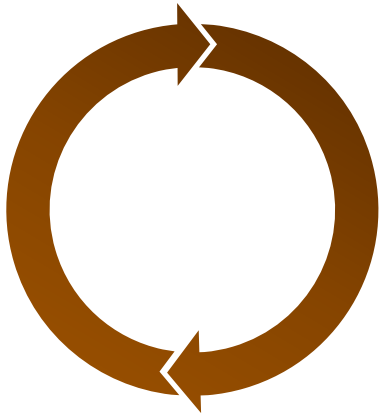
IC Design/Manufacturing Flow



What Can AI Do?

- AI/ML can serve as hammers, bridges, and optimizers

- Prediction
- Generation
- Acceleration
- Optimization
- ……

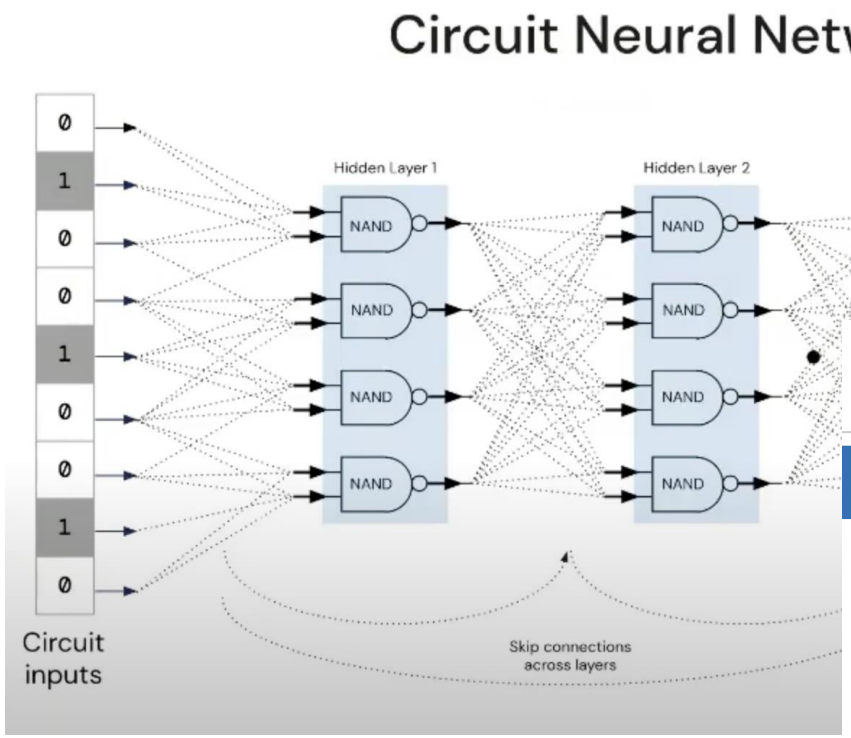


- AI/ML algorithms: supervised learning, unsupervised, transfer learning, reinforcement learning, generative AI, transformers, …

Everything, Everywhere, All at Once!

- **Everything:** power, performance, area (PPA), cost, turn-around time, reliability, yield, + new things like security …
- **Everywhere:** All levels of design abstractions, testing, verification, DFM, mask synthesis / ILT, … (e.g., DAC'23)
- **All at once:** not quite there yet, but …
  - Currently, parameter / flow tweaking
  - Generative AI, from specification to layout (?)

Logic Synthesis: DeepMind IWLS'23 1st Place



THE WALL STREET JOURNAL.

SUBSCRIBE SIGN IN

CIO JOURNAL

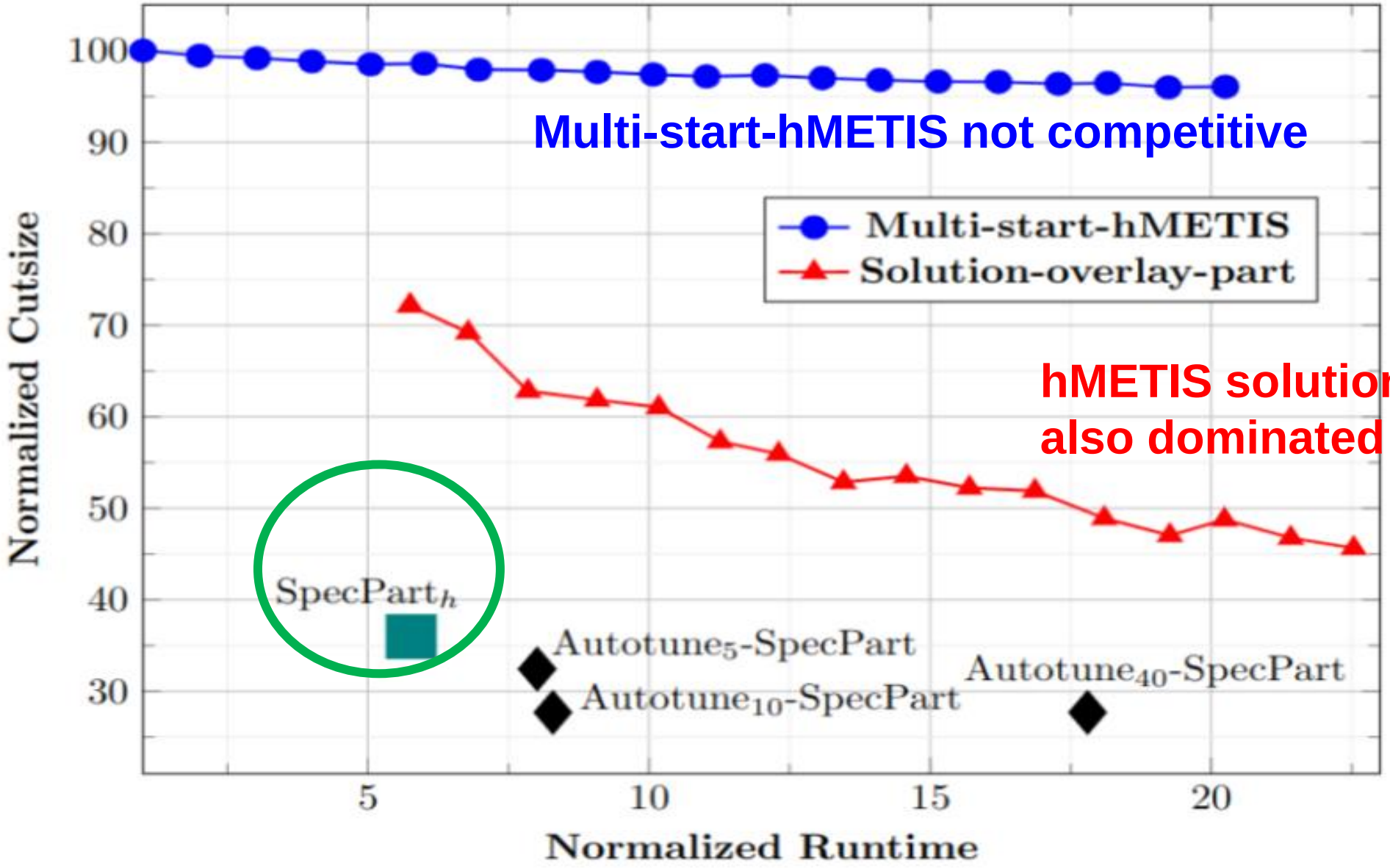
In Race for AI Chips, Google DeepMind Uses AI to Design Specialized Semiconductors

The Alphabet research lab said it wants to make specialized chip design faster, less reliant on solely human engineers

By Belle Lin Follow

July 20, 2023 12:49 pm ET

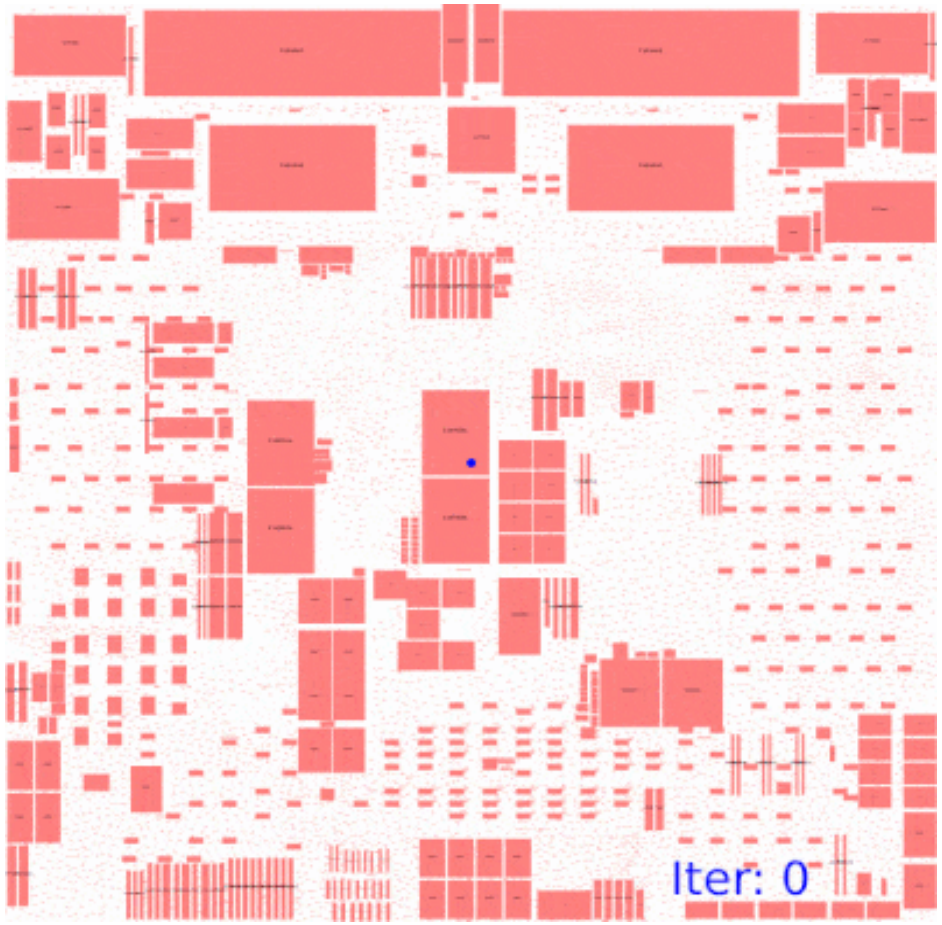
SpecPart: *Supervised* Spectral Partitioning



- From UCSD
- ICCAD'22 BPA

VLSI Placement

- A classical NP-hard problem!
- Modern huge designs: many billions of transistors, 10M+ cells, and hundreds of macros
- Plays a **central** role in modern IC design closure
  - Largely determines interconnects
  - Interconnect-centric/limited designs



DREAMPlace: Deep Learning Toolkit-Enabled GPU Acceleration for Modern VLSI Placement [Lin … Pan, DAC'19 Best Paper Award; IEEE TCAD 2021 Donald O. Pederson Best Paper Award]

Source code release: https://github.com/limbo018/DREAMPlace
Widely used by industry (Google, Nvidia, Intel, …) and academia

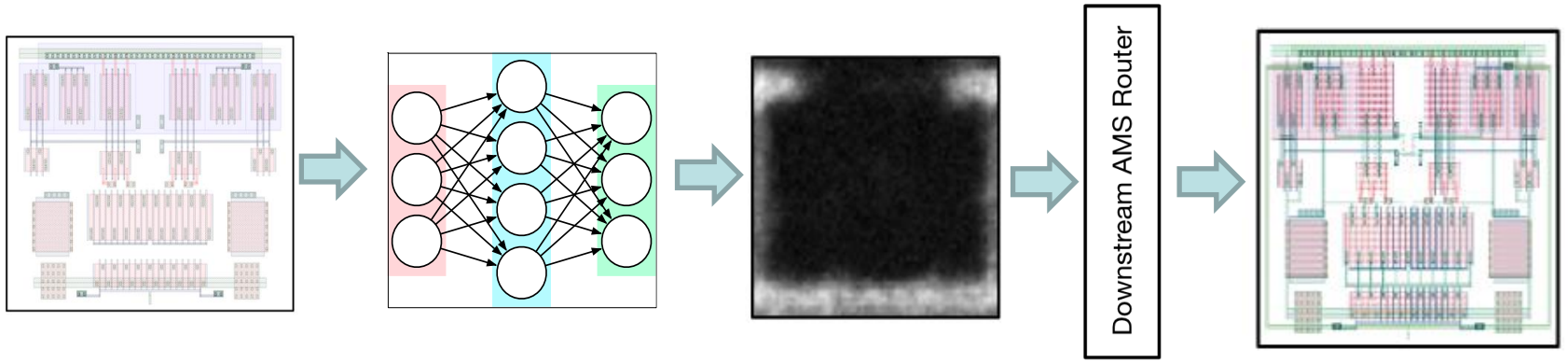
DREAMPlace Strategies and Applications

- We propose a **novel analogy** by casting the nonlinear placement optimization into a neural network training problem
- Greatly leverage deep learning **hardware** (GPU) and open-source **software** toolkits (e.g., PyTorch)
- Enable ultra-high parallelism and acceleration while getting state-of-the-art results
- Obtained state-of-the-art quality, **yet 30-40x faster**
- Better design space exploration
- DREAMPlace for macro placement, FPGA (DREAMPlaceFPGA)

  …

ML-Guided Routing [Zhu+, ICCAD'19]

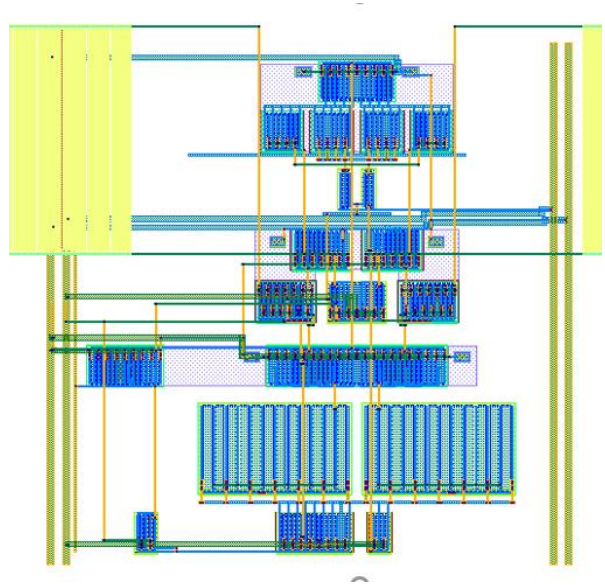
- Learn from manual layouts ➔ Generative AI, GeniusRoute
- Encode designer expertise into neural networks
- Incorporate ML into automatic routing
- Part of DARPA MAGICAL project (2018-2023)



Unrouted Placement

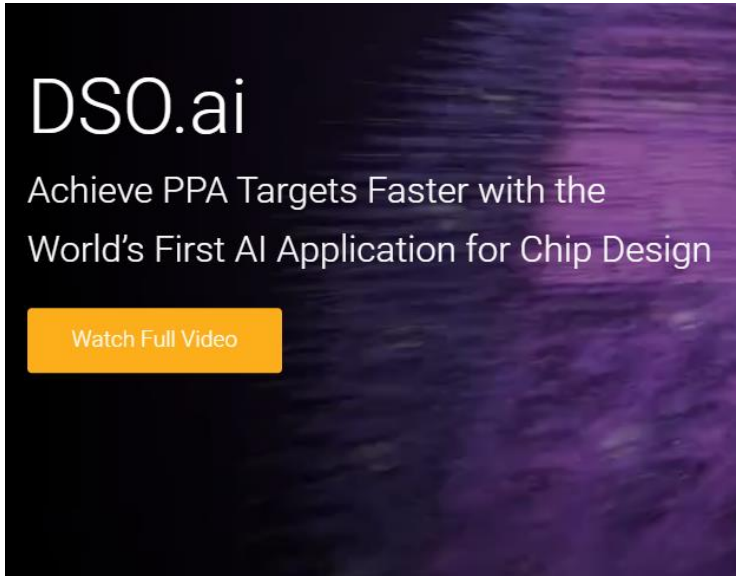
Generated Routing Region

Routed Layout

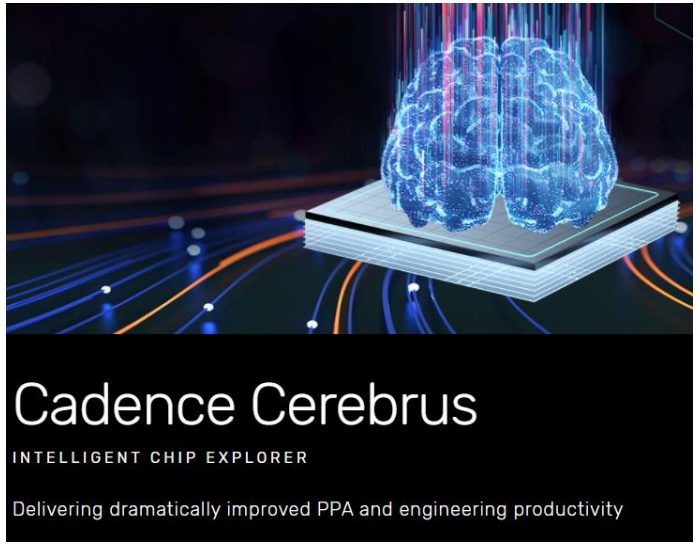


RL for EDA

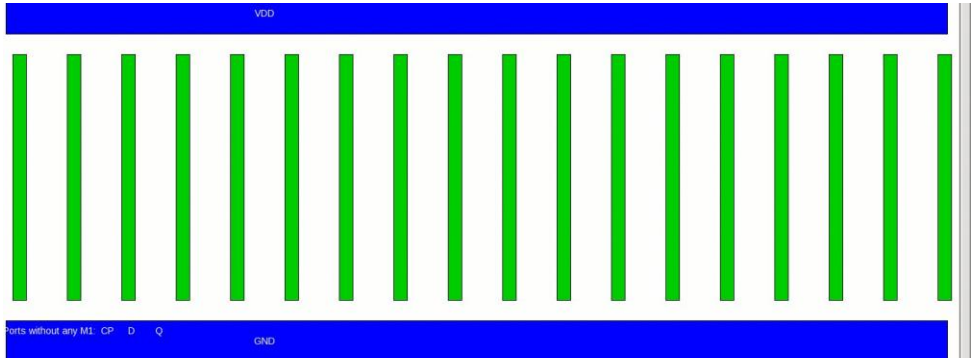
- Synopsys DSO.ai



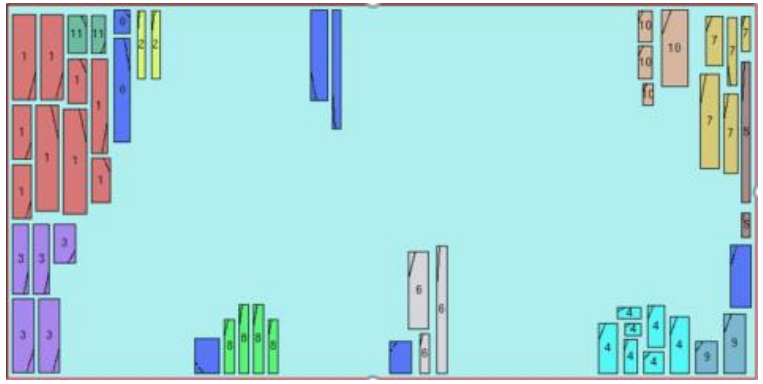
- Cadence Cerebrus



- Nvidia NVCell



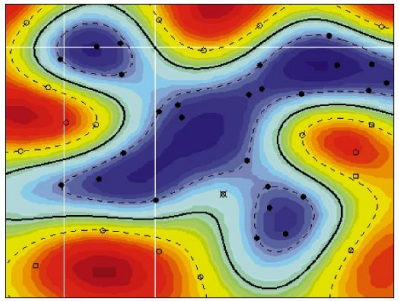
- Google RL macro placement



Lithography Hotspot Detection

Question 1: Without going through detailed litho-simulations, can we directly predict lithography hotspot to avoid poor yield?

- Our work [Ding+, ICICDT **2009** Best Paper] is **among the first** to use machine learning (SVM) for litho-hotspot detection
  - Very active research topic in the last 12+ years
  - **Inspired ICCAD 2012 CAD Contest**, run by Mentor Graphics
  - Meta-classification combining ML and PM [Ding+, ASPDAC'12 BPA]
  - Deep neural network [Yang+, DAC'17]
  - Big data vs. small data: transfer learning, active learning, semi-supervised learning [Lin+, ISPD'18], [Chen+, ASPDAC'19] …
  - Litho-GPA: confidence estimation [Ye+, DATE 2019]
  - ……

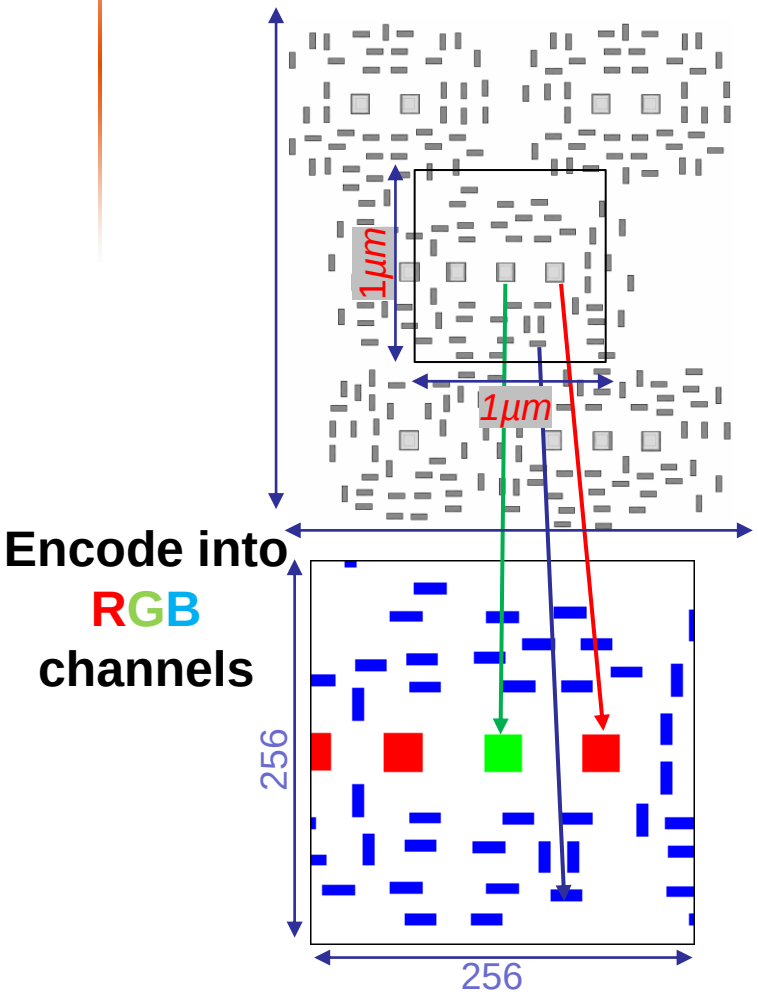


LithoGAN: End-to-End Lithography Modeling with Generative Adversarial Networks [Ye+, DAC'19 Best Paper Finalist]

Question 2 (much harder): Without going through litho-simulations, can we directly get printed images?

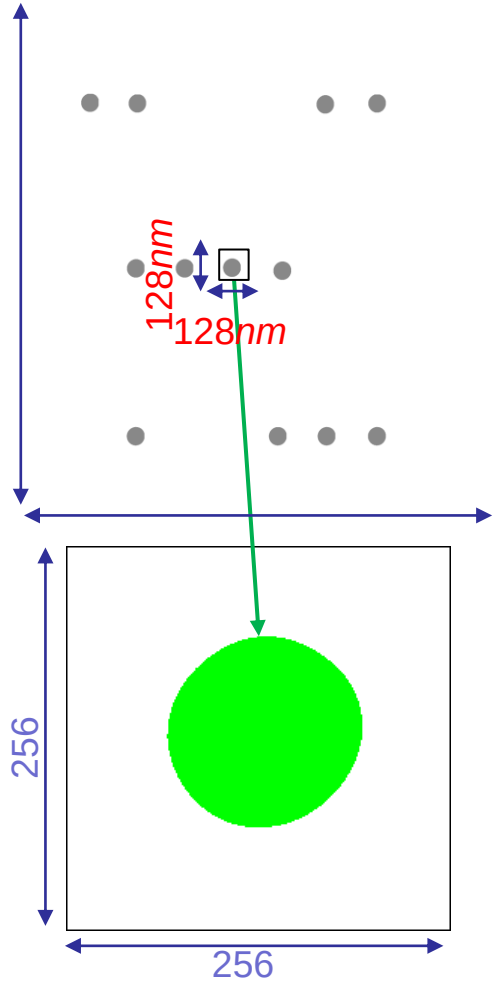
Image Translation for Litho Modeling

[Ye+, DAC'19]



Expensive Litho Simulation

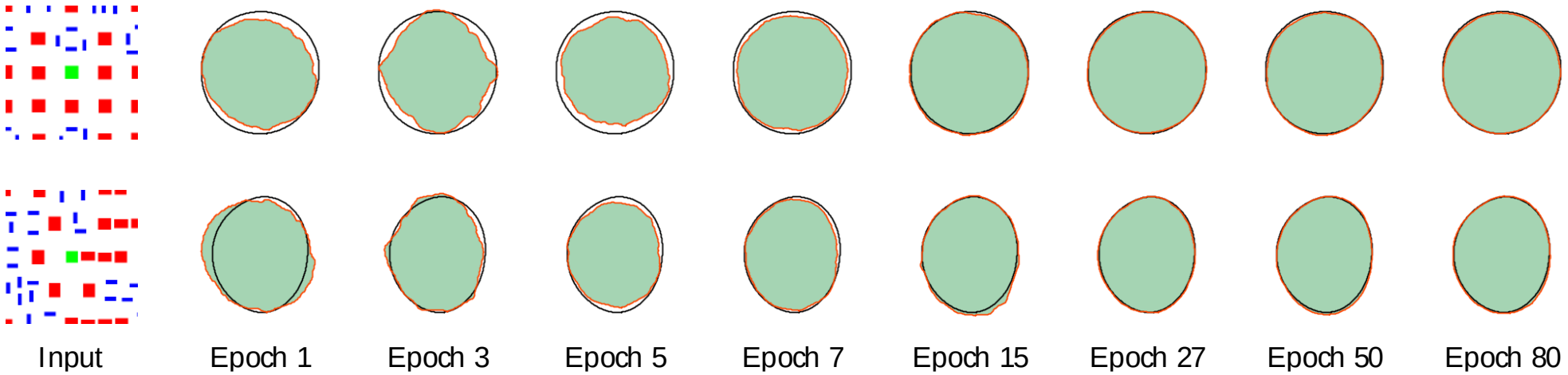
This is now a (modified) **image translation** task



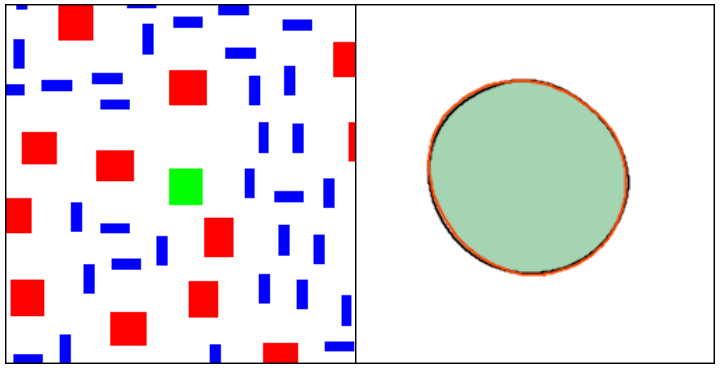
- Different elements encoded on different image channels
- Resist pattern zoomed in for high-resolution/accuracy

LithoGAN Results

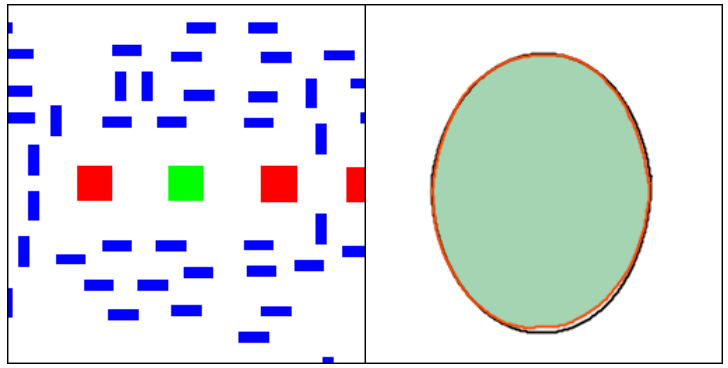
[Ye+, DAC'19]



Model advancement progress



Input LithoGAN output

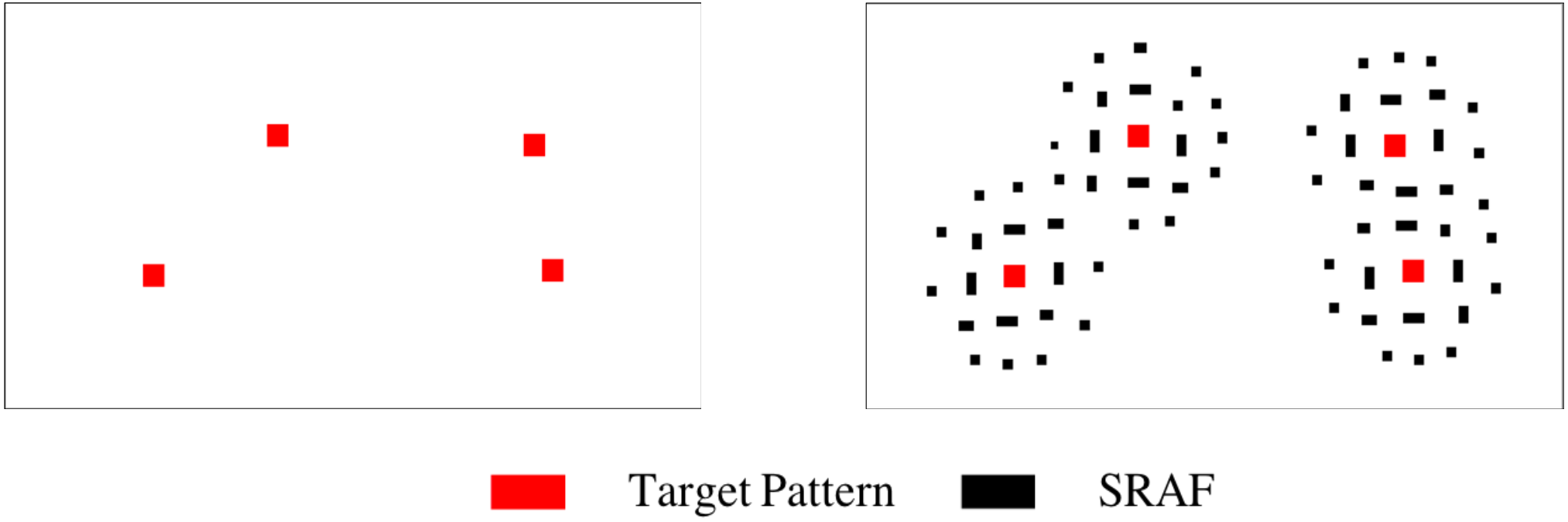


Input LithoGAN output

LithoGAN is **1800x** faster than rigorous simulations, with acceptable error (in consultation with industry)

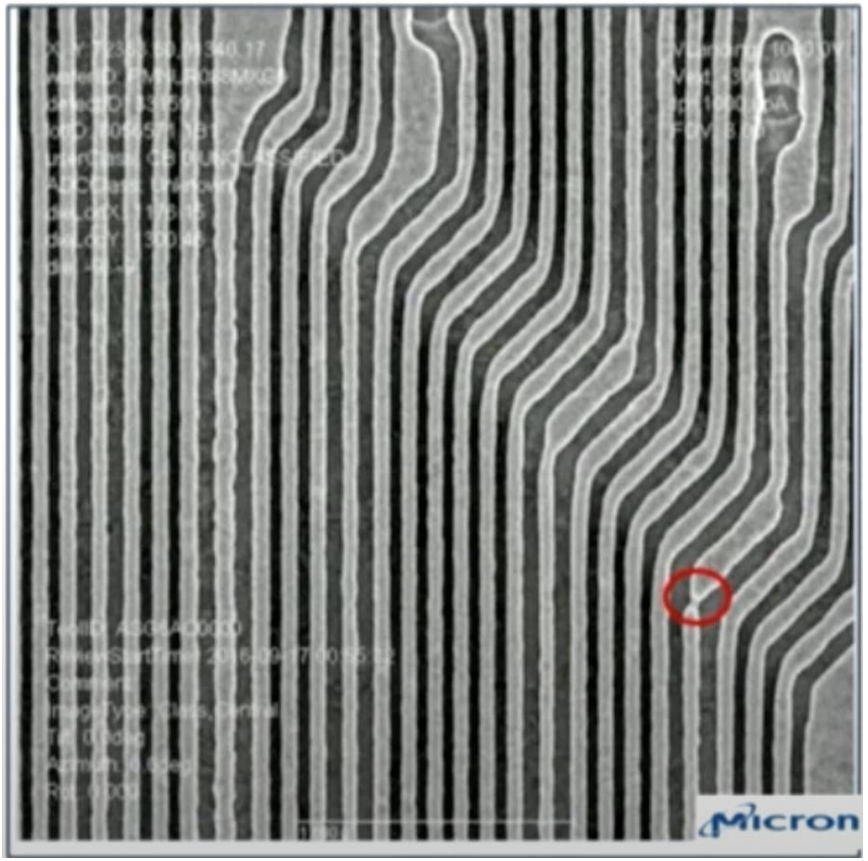
Mask Synthesis: GAN-SRAF [Alawieh+, DAC'19]

- Sub-Resolution Assist Feature (SRAF) generation using Conditional Generative Adversarial Networks (CGAN)
- **144x** faster than model-based SRAF insertion with similar QoR



Layout & Lithography: DTCO

- Designer's intent or dream (free form!): wire sizing/shaping, any-direction jogs, …
- Power, Performance, Area
- ……
- Curvilinear design when in demand
- Can lithography accommodate?



Layout → Lithography

- Actually, real printed image on wafer is never truly rectangular, like that in mask!
- Industry trend: curvilinear mask
- Curvilinear ILT

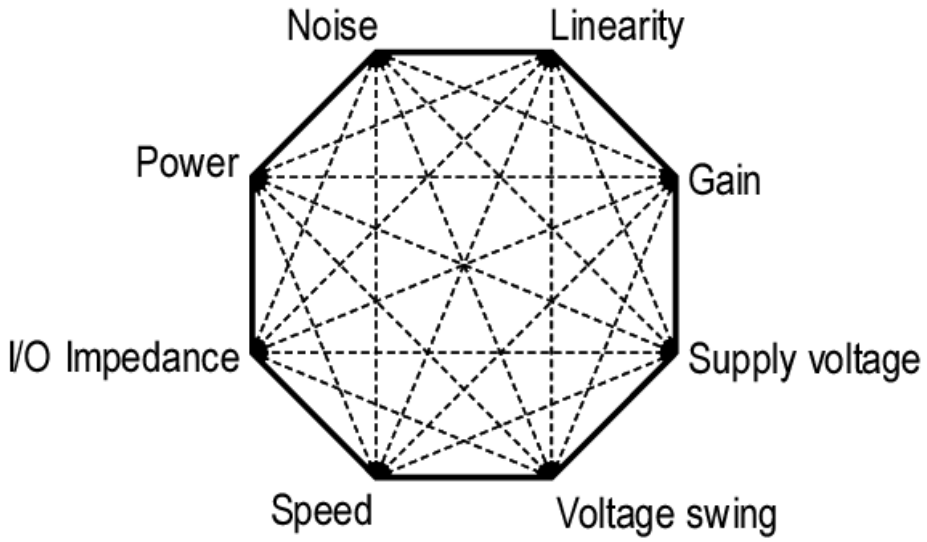
- Maybe model routing in a fluidic, non-Manhattan, continuous/ differentiable manner, yet still manufacturable (even better yield)?
- If so, maybe even simultaneous P & R & DTCO

What Can We Learn from Applying AI/ML Techniques to Analog/MS Design?

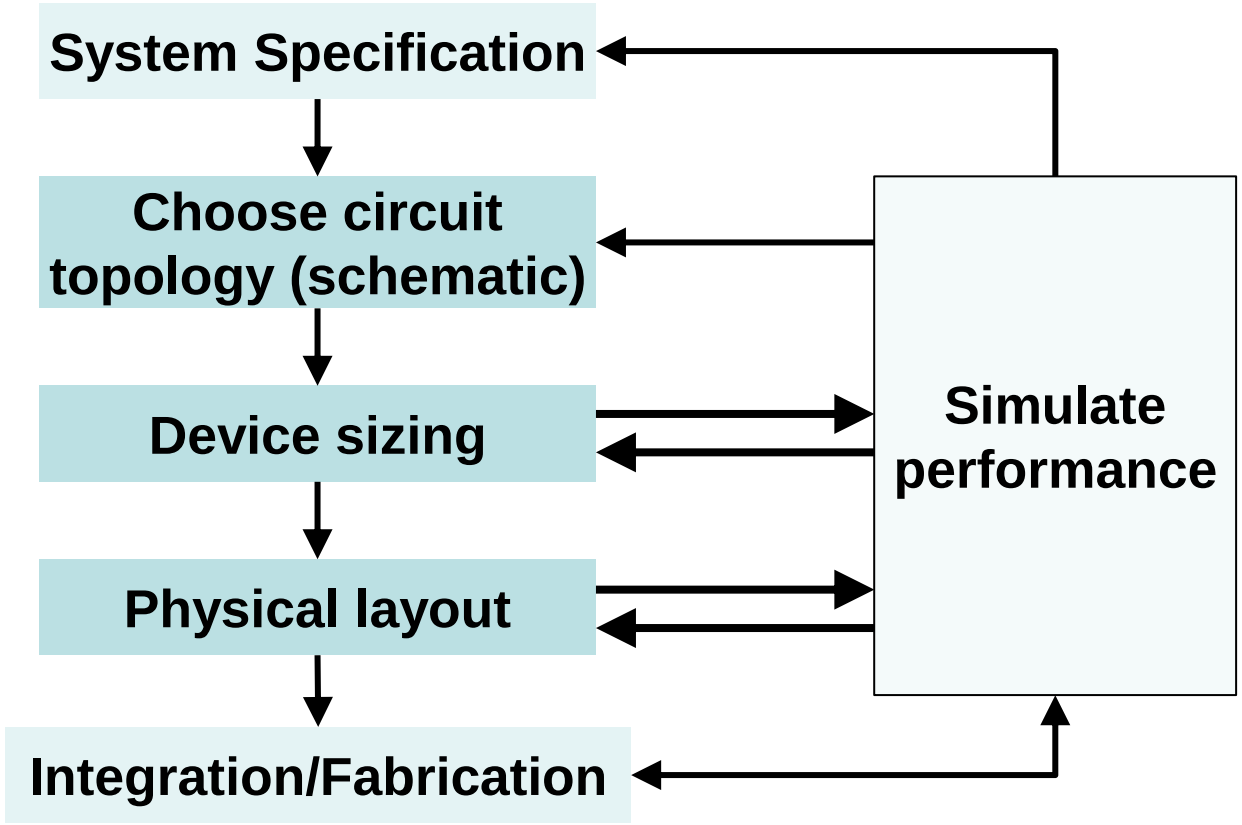
David Z. Pan

University of Texas at Austin

Analog IC Design is Hard



[Razavi, Design of Analog IC]



- Many design specs to juggle
- Heavily rely on designer experience
- Tons of simulations

RF/mmWave IC Design is HARDER

- Working with a "poorer" device – Johnson's limit & lossy passives

$$V_m \cdot f_T = \frac{E_m v_s}{2\pi} \Longrightarrow$$

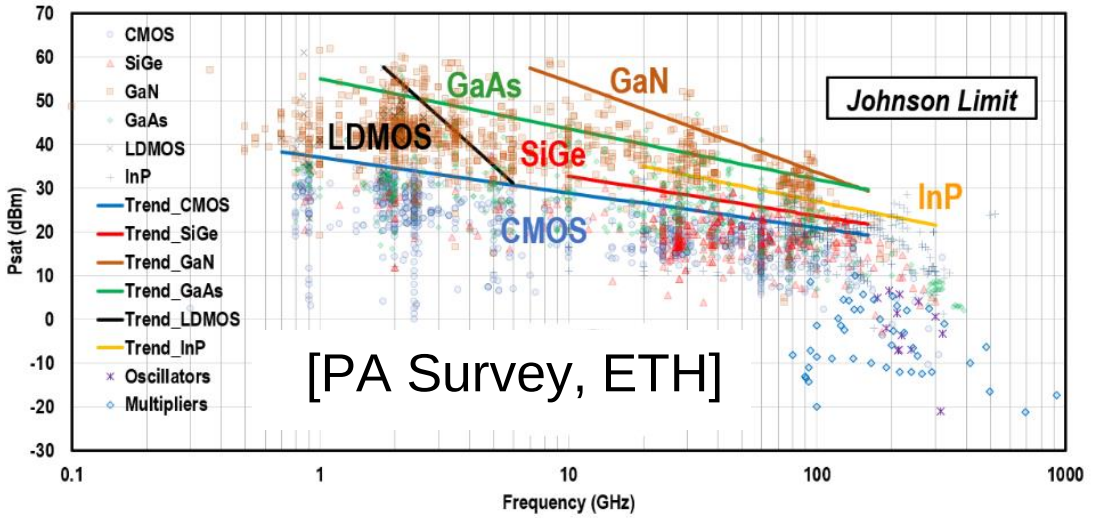
A faster device with inherently poorer power generation capability

Lower quality factors for passives (e.g., ind, cap, var., etc.)

- Everything becomes "sensitive" – parasitic elements, distributive effects, crosstalk

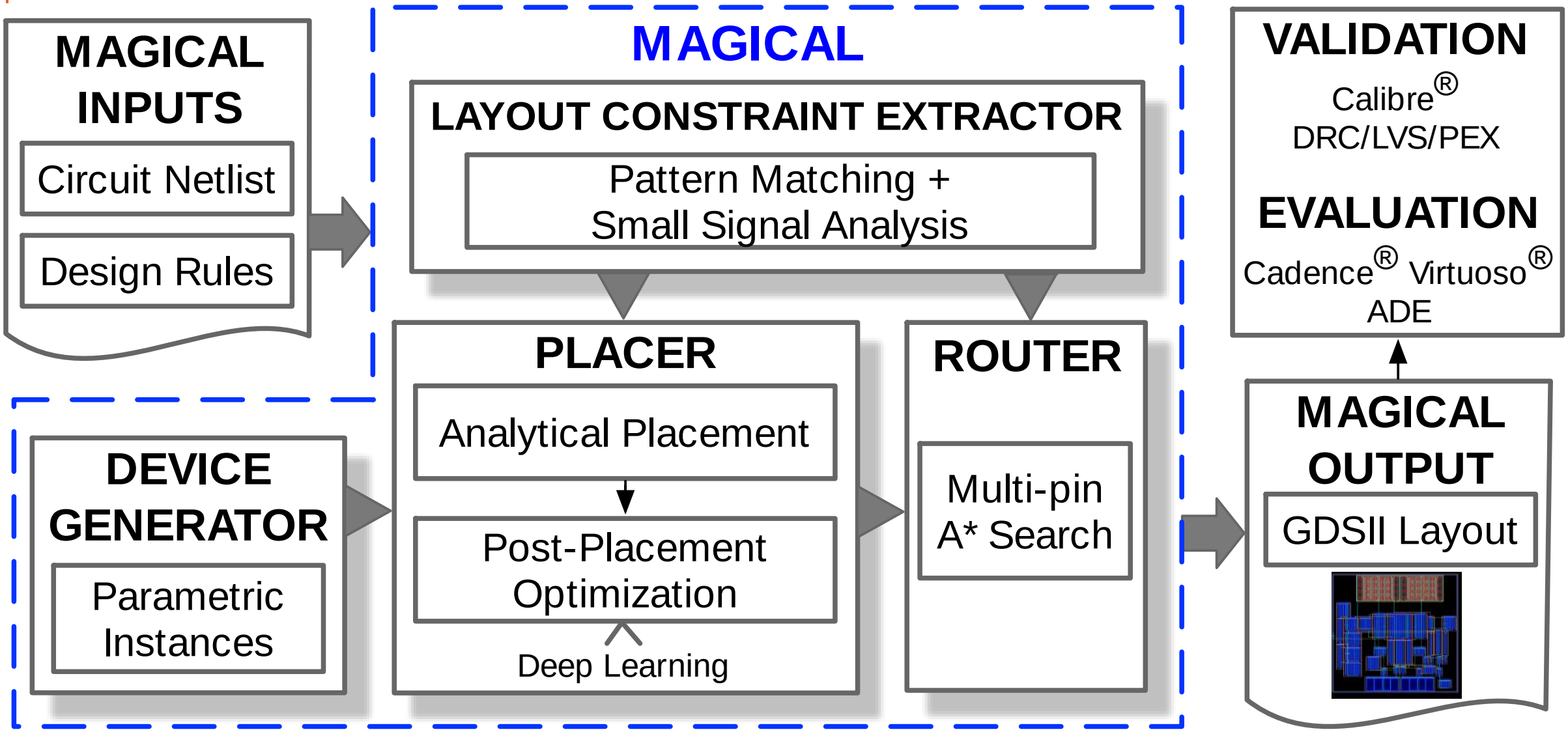
- Deteriorated channel/propagation conditions – higher path loss, multi-path effect, easy blockage

- More than circuits: **EM** effects!



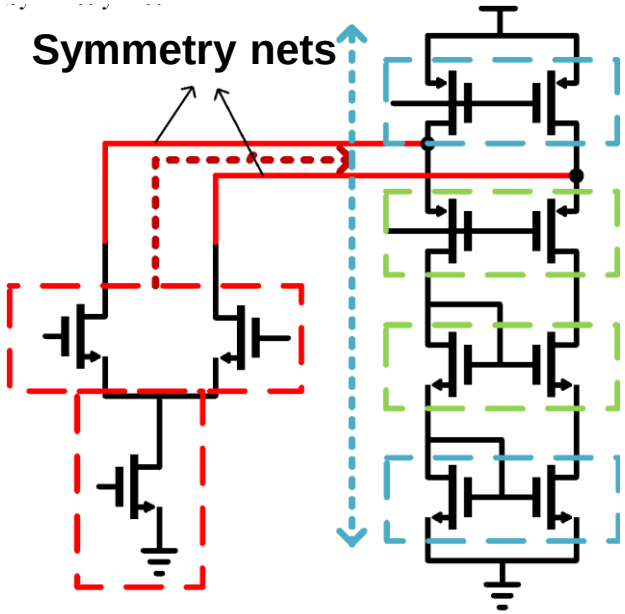
[PA Survey, ETH]

MAGICAL Layout Automation System

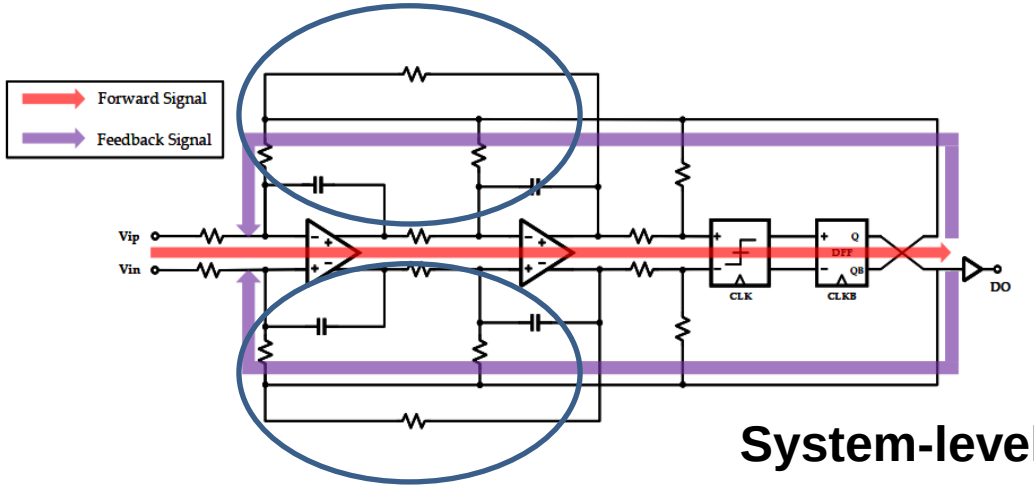


MAGICAL Constraint Generation

- Both block level and system level constraints
- Pattern detection [Xu+, ICCAD'19]
- Graph similarity [Liu+, ASPDAC'20]
- **Graph neural network** [Chen+, DAC'21]
- MAGICAL can take designer-provided constraints too, e.g., signal-flow constraints



Building block symmetry



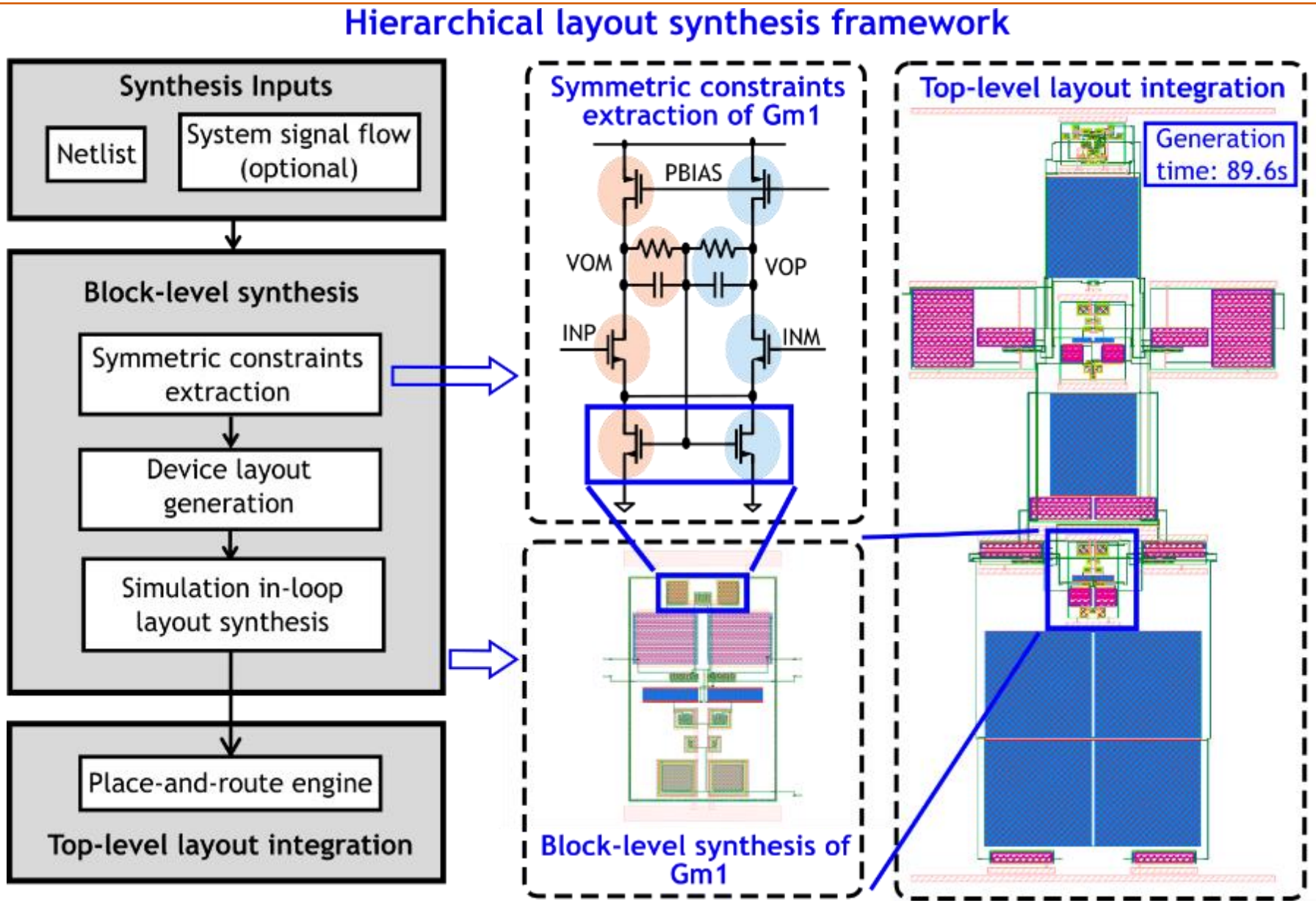
System-level constraints

MAGICAL Place & Route Engines

- Not just **AI/ML**, but also leverage proven non-AI techniques
- Nonlinear global placement, with some tweaks [Zhu+, ICCAD'20]
- Linear programming based detailed placement
- Well generation using **GAN** [Zhu, ASP-DAC'22]

- GeniusRoute: **GAN**-based routing guidance [ICCAD'19]
- Grid-based detailed routing, A* search, DRC handling [Chen+, ICCAD'20]
- Various symmetry handling and special power/ground routing

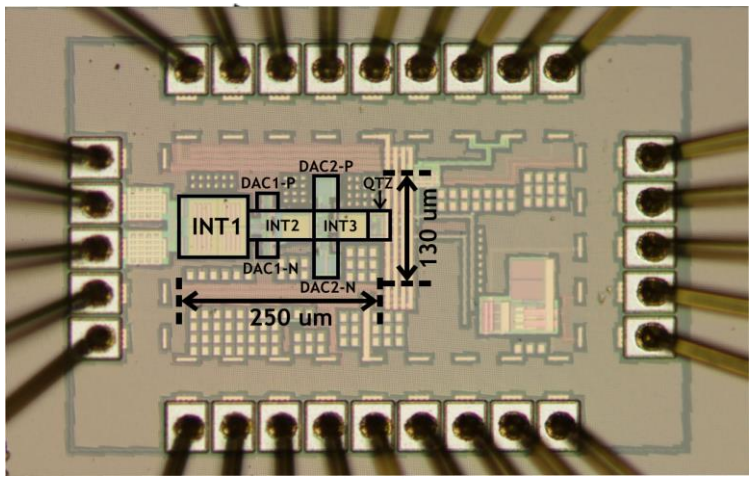
MAGICAL Hierarchical Framework [Chen+, CICC'21]

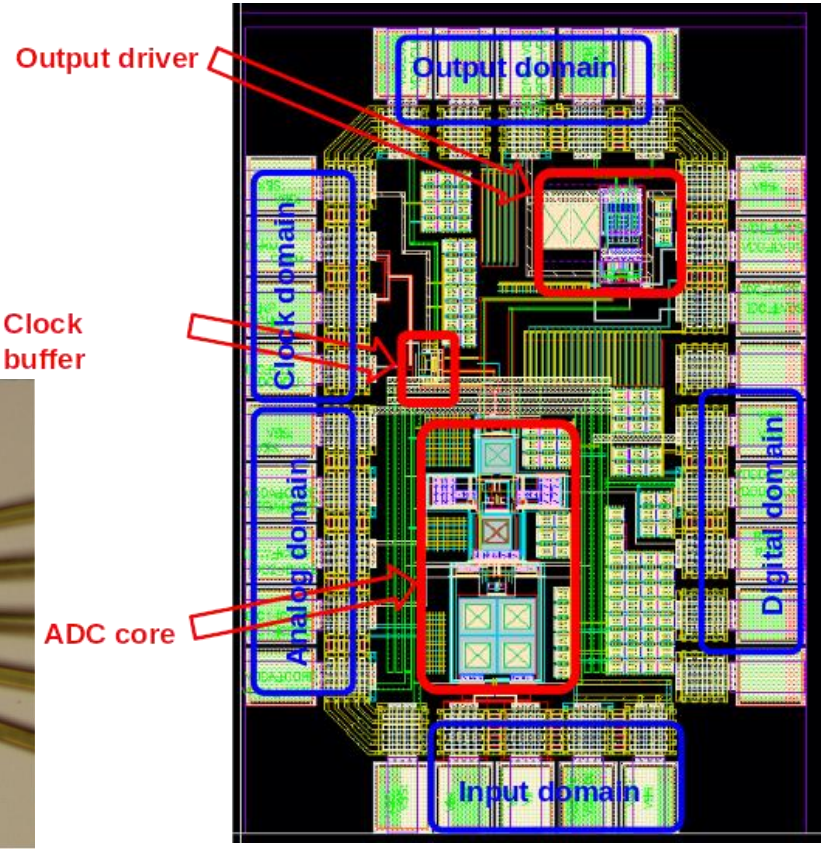


MAGICAL Tapeout Proven

[Chen+, CICC'21]

- 1GS/s 3rd-order high-performance continuous time ΔΣ modulator
- State-of-the-art performance, cf. original design [SSC-L'20]
- Include various sub-block types
  - Three integrators: One passive, two active
  - Two FIR-based feedback DACs
  - One comparator + Digital logic
- TSMC 40nm
- Core: 22,000 um^2
- Chip: 800um x 550um
- O(month) ➔O (min) for layout

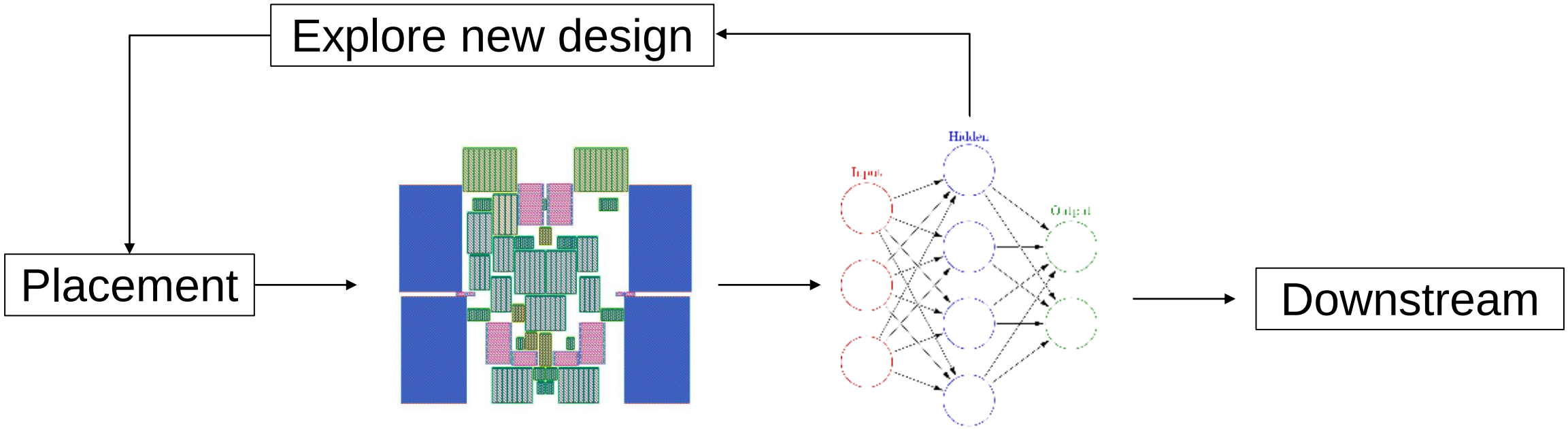




Placement Quality Prediction

[Liu, Zhu+, DATE'20]

- Machine learning dataset and models
  - Auto-generated 16,000+ layouts for each circuit with **MAGICAL**
  - **UT-AnLay dataset** with post layout simulations: **Open-sourced**
- Prune "bad layouts" directly early in the design stage



CNN + transfer learning

OpenSAR Tapeout

[Liu+, ICCAD'21, SSCL'22]

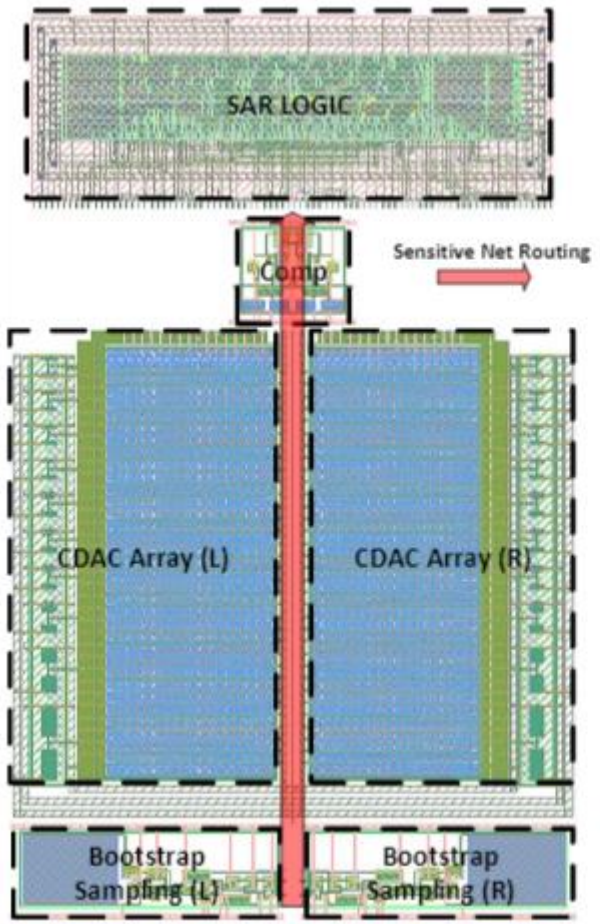
- End-to-end SAR ADC compilation

Digital APR

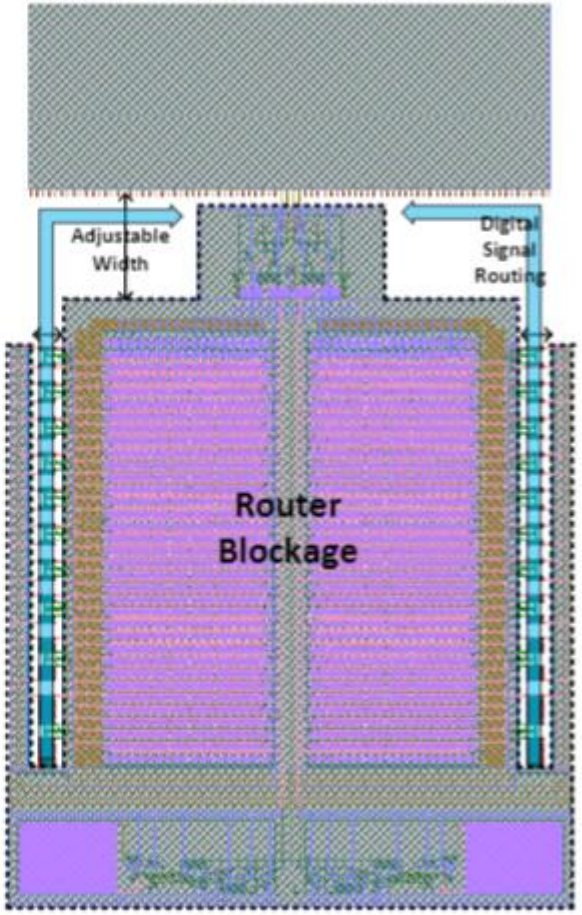
MAGICAL

Template-based Generation

MAGICAL

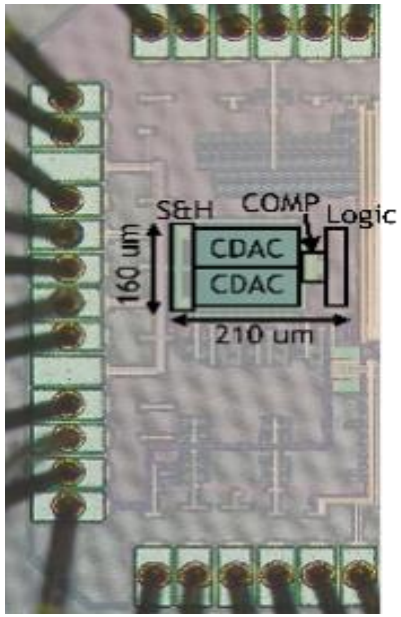


Floorplan

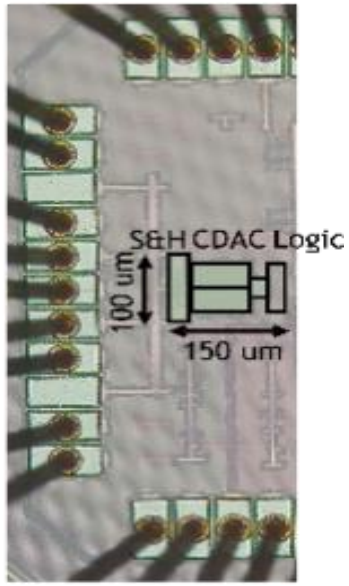


Route planning

Tape-out validated
TSMC 40nm



12-bit 1MS/s



10-bit 100MS/s

MAGICAL for FinFET: AutoCRAFT

- Targeting custom layout generation for advanced **FinFET**
- Exploding base layers DRC ➔ **Tech-agnostic FinFET layouts**
  - Use region-based FinFET layout style
  - Handle complex geometrical/electrical constraints to satisfy performance
  - SMT-based placement, etc.
- E.g., a high-speed 4-stage VCO
  - 2 regions, 110 primitives, 71 nets
  - Various placement geometrical constraints and pin-to-pin resistance matching constraints
  - TSMC 5nm

Supply (mV)	Frequency (GHz)		Power (μW)	
	Manual	AutoCRAFT	Manual	AutoCRAFT
650	3.02	**3.08**	304.4	**300.2**
700	3.28	**3.34**	398.8	**392.7**
750	3.49	**3.55**	507.5	**499.6**
800	3.67	**3.73**	632.4	**621.6**
850	3.83	**3.88**	774.6	**758.5**
900	3.96	**4.00**	936.0	**914.4**

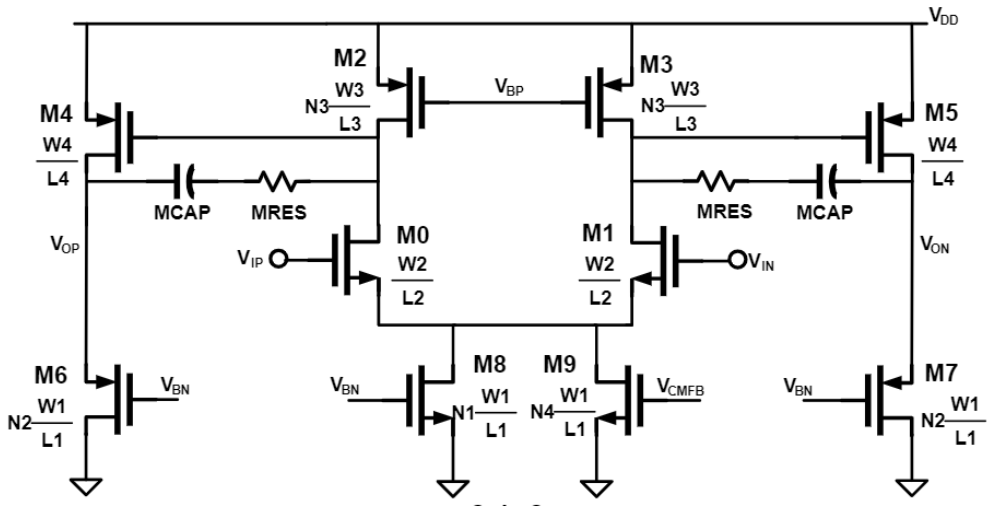
[Chen+, ISPD'22] in collaboration with Nvidia. Open-sourced https://github.com/NVlabs/AutoCRAFT

Analog Sizing: Problem Formulation

minimize $Power$
s.t. DC Gain > 60 dB
CMRR > 80 dB
PSRR > 80 dB
Output Swing > 2.4 V
Output Noise $< 3 \times 10^{-4}$ V_{rms}
Phase Margin > 60 deg
Unity Gain Frequency > 40 MHz
Settling Time $< 3 \times 10^{-8}$ s
Static error < 0.1
Saturation Margin > 50 mV

Specifications

Topology



Miller OTA

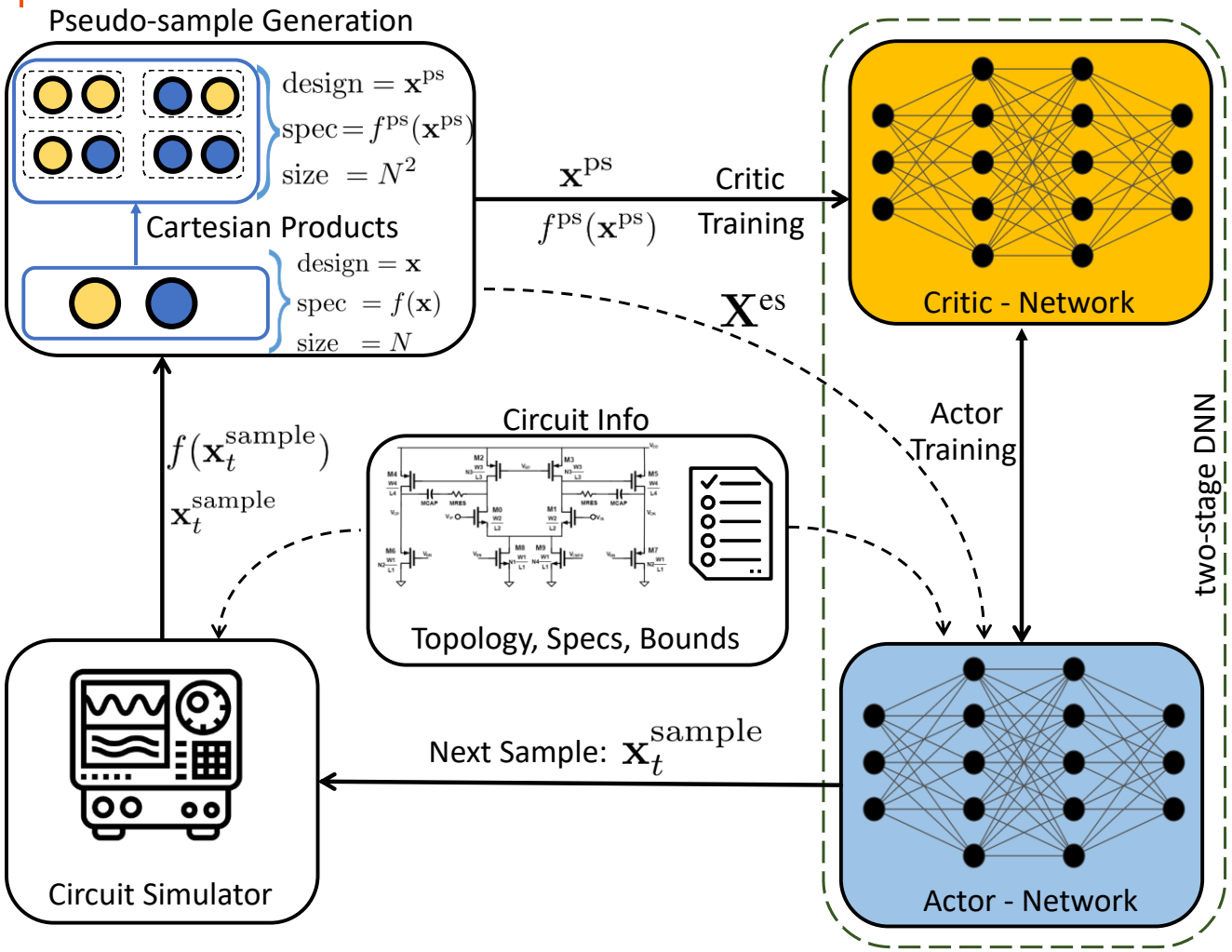
TABLE I: Design parameters and their ranges for Miller OTA

Parameters	LB	UB	Parameters	LB	UB
L1(μm)	0.18	2	MCAP(fF)	10	2000
L2(μm)	0.18	2	MRES(Ω)	100	100k
L3(μm)	0.18	2	N1 (integer)	1	10
L4(μm)	0.18	2	N2 (integer)	1	10
L5(μm)	0.18	2	N3 (integer)	1	10
W1(μm)	0.22	150	N4 (integer)	1	10
W2(μm)	0.22	150	N5 (integer)	1	10
W3(μm)	0.22	150	N6 (integer)	1	10
W4(μm)	0.22	150	NC (integer)	1	10
W5(μm)	0.22	150	NR (integer)	1	10

W:width; L=length; UB:upper bound; LB:lower bound

Design Parameters & Ranges

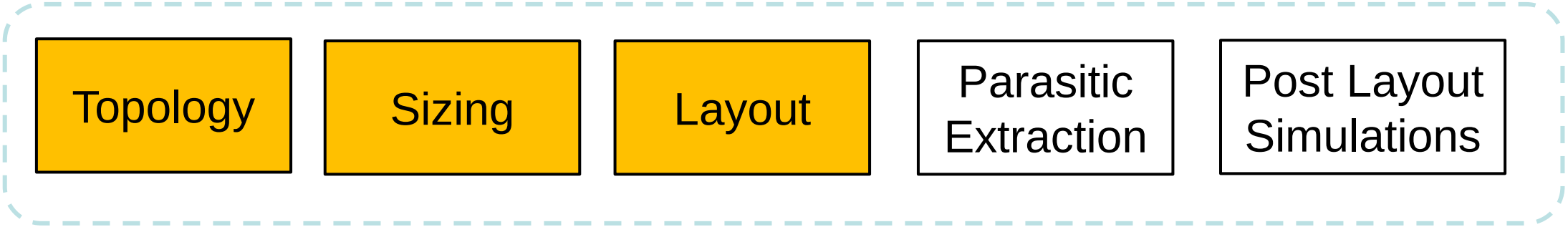
RL View → DNN-Opt [DAC'21 BPA Candidate]



- **State:** Design parameter vector as the state representation
- **Action:** Amount of change in the design parameters
- **Reward:** FoM of the resulting design performance
  - Weighted sum of the design objective and performance constraint violations

[Budak+, DAC'21] in collaboration with Intel

End to End Analog Design Automation?

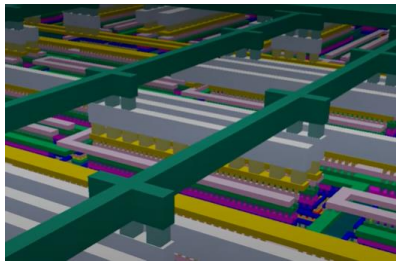
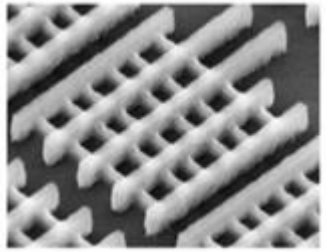
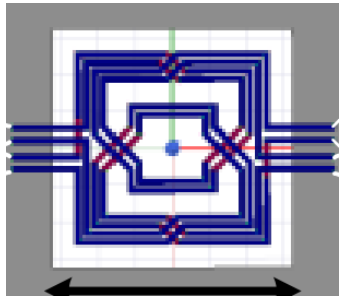
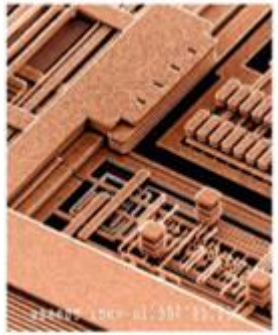


- The overarching goal: an end-to-end analog DA flow?
- Analog “S&PR”, like RTL to GDSII for digital?

- Some ongoing efforts on circuit topology generation using ML

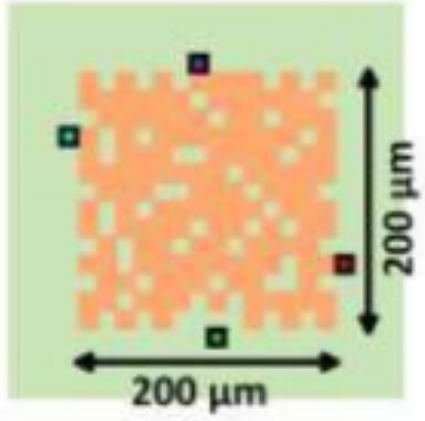
AI/ML for RF/mmWave IC Design

- RFIC = active + passive components



+



Transistors Versatile BEOL

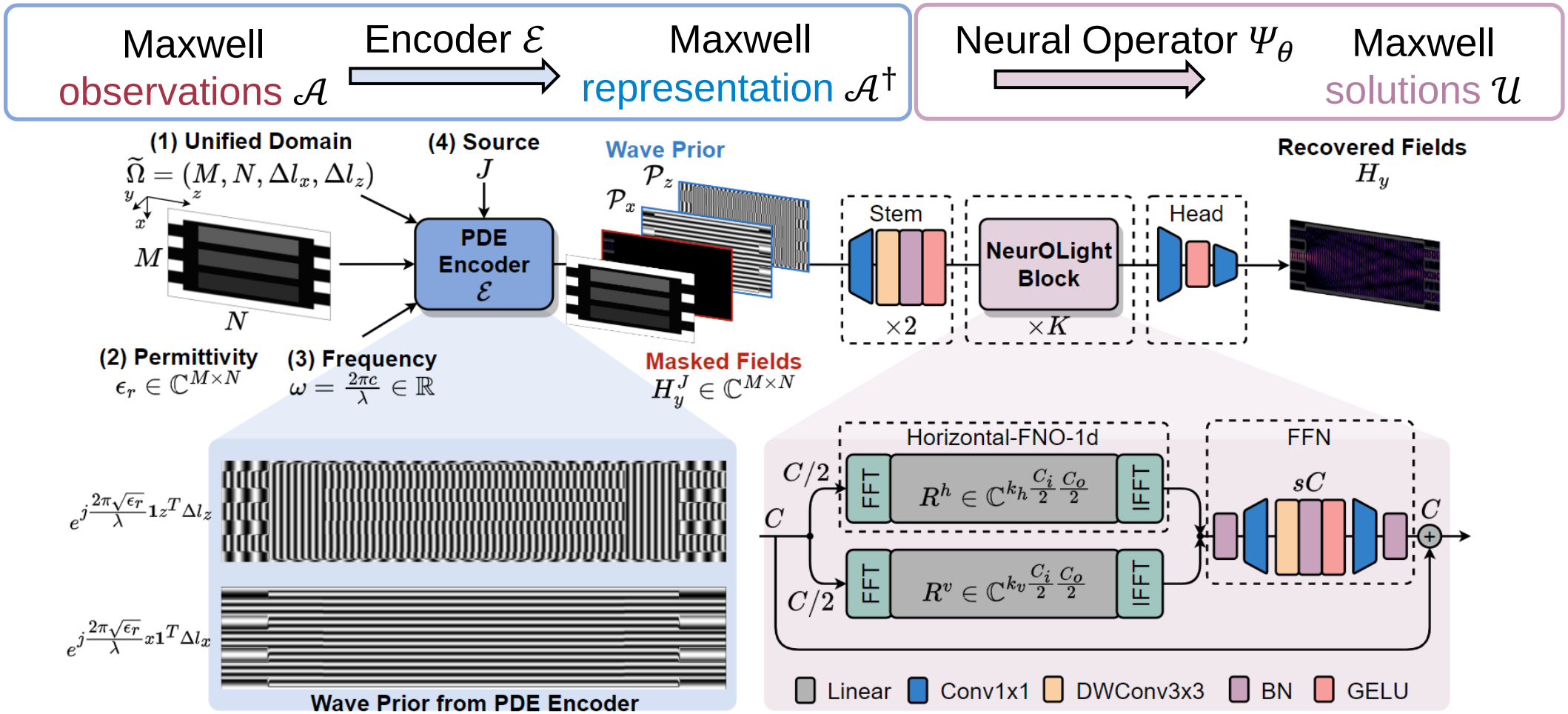
- **Active** circuits: can we leverage analog DA?
  - Yes, but much higher frequencies!
  - Need to develop good surrogate/predictive/ML-assisted models to replace expensive circuit simulators or EM solvers



- **Passive** design synthesis via deep learning
  - **Inverse design** (inspired by photonics and lithography)

NeurOLight [Gu+, NeurIPS'22 Spotlight]

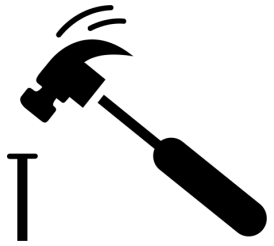
- *New PDE encoding* + *New neural operator model* + *New training recipe*
- **>200×** speedup: FDFD simulation (1-10s) → Fast NN inference (<10 ms)

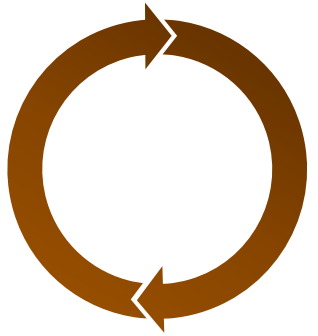


J. Gu, Z. Gao, C. Feng, H. Zhu, R.T. Chen, D.S. Boning, D.Z. Pan, "NeurOLight: A Physics-Agnostic Neural Operator Enabling Parametric Photonic Device Simulation," NeurIPS 2022. (Spotlight)

Conclusion

- AI-assisted chip design for digital and analog/MS ICs
  - Supervised/unsupervised/transfer learning, RL, …
  - Acceleration
  - Prediction
  - Generation
  - ……





- AI/ML can serve as hammers, bridges and optimizers for cross-layer system/design and technology co-optimizations
- DATA, model generalization, transferability, bias, interpretability / explainability, optimality, …

“My dream is to have a silicon compiler which can let people design chips as easily as they can write software”

COMMUNICATIONS OF THE ACM
CACM.ACM.ORG 01/2023 VOL.66 NO.01

ACM Member News

CLOSING THE LOOP BETWEEN AI FOR IC AND IC FOR AI

David Z. Pan is the holder of the Silicon Laboratories Endowed Chair in the Department of Electrical and Computer Engineering at the University of Texas at Austin (UT Austin).

Pan earned his undergraduate degree in physics from Peking University in Beijing, China. He went on to earn both his master's degree and Ph.D. in computer science from the University of California, Los Angeles.

After obtaining his doctorate, Pan became a research staff member at the IBM T.J. Watson Research Center in Yorktown Heights, NY. He spent nearly three years with IBM before joining the faculty at UT Austin in 2003, where he has remained since.

Pan's research interests center on electronic design automation, with a focus on the physical design of integrated circuits (ICs).

"I am trying to close the loop between AI (artificial intelligence) for IC, and IC for AI," Pan says.

He explains that AI for IC leverages artificial intelligence techniques to enable better agile and intelligent integrated circuit design, while IC for AI involves customizing chips for AI applications.

Pan says that as semiconductor technology enters the era of extreme scaling, IC design and manufacturing will become ever more complex, and better IC design technologies will be needed more than ever to optimize factors such as performance, power, manufacturability, and design cost.

In the future, Pan says, he wants to democratize chip design and make it as easy as compiling software.

"My dream is to have a silicon compiler that will let people design chips as easily as they can write software," says Pan.

—John Delaney

AI for Chip Design / EDA