Leveling Up: A Trajectory of OpenROAD, TILOS and Beyond

Andrew B. Kahng
UCSD CSE and ECE Departments

abk@ucsd.edu
http://vlsicad.ucsd.edu

Agenda

- **Leveling Up: A Trajectory**
- **The OpenROAD Project**
- **The TILOS AI Institute**
- **And Beyond**

Leveling Up

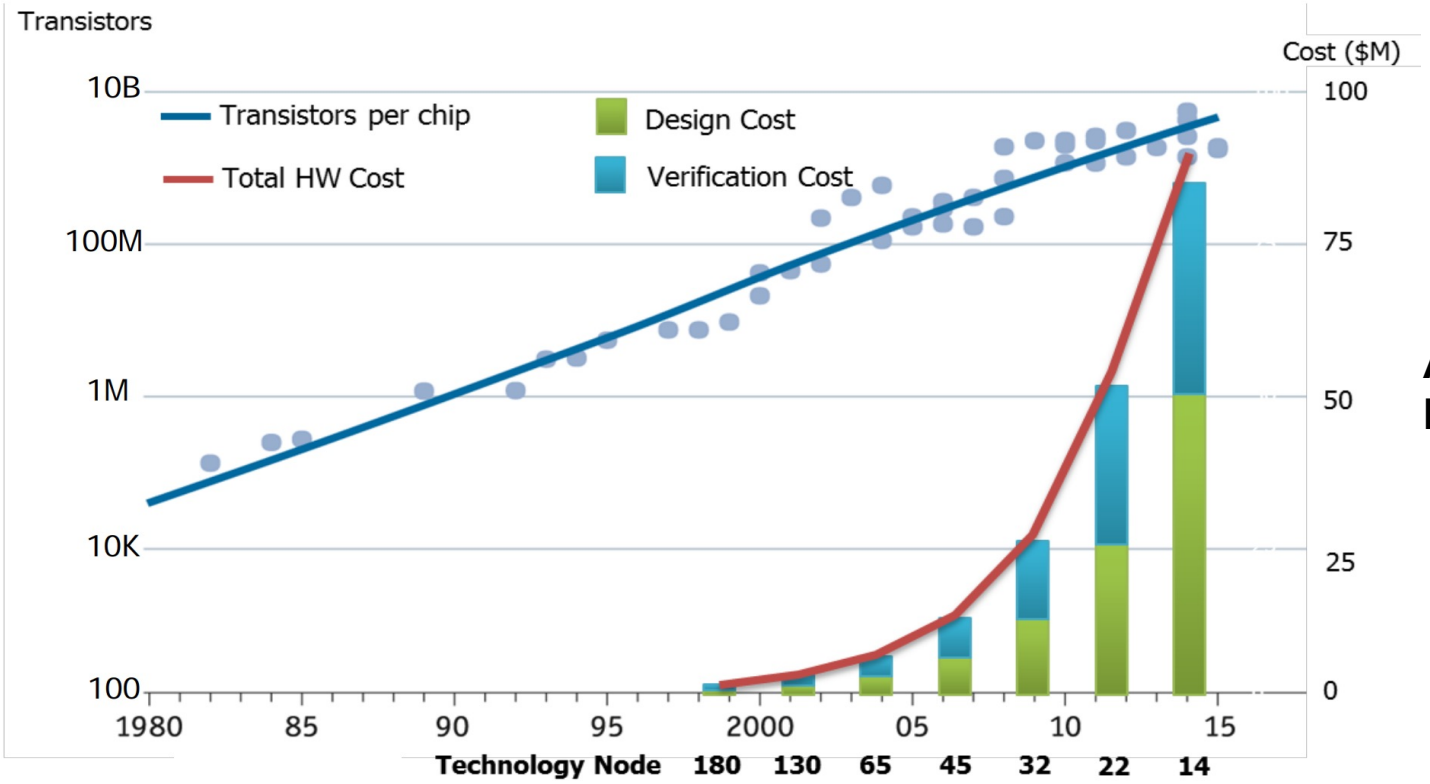
- **A trajectory of openness, infrastructure and culture change**
  - Open source, open data, benchmarking, roadmapping
- **Scaling and empowerment of people and a future workforce**
- **Culture change at academia-industry research interface**
- **Two waypoints:** OpenROAD 24h TILOS

Agenda

- Leveling Up: A Trajectory
- The OpenROAD Project
- The TILOS AI Institute
- And Beyond

The Crisis of Hardware Design

- ASIC design in advanced nodes: barriers of Cost, Expertise, Risk



A. Olofsson, ISPD-2018

- Innovators can't evaluate "SWaP", "PPA" of their design ideas
- "Foundations and **R**ealization of **O**pen, **A**ccessible **D**esign"

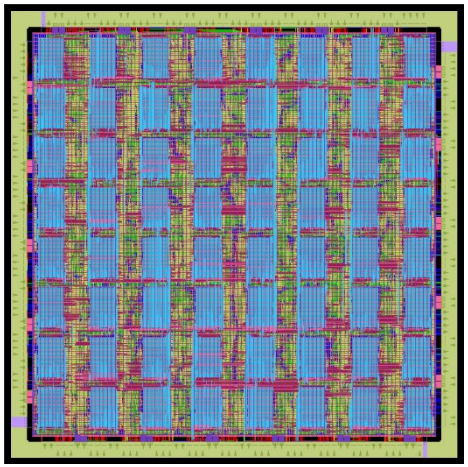
OpenROAD: No Humans, 24 Hours

- **FOCUS: Ease of use and runtime**

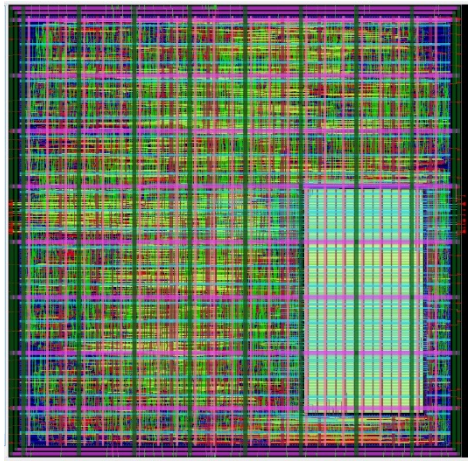
- **Directly attack the crises of design and innovation**
  - **Schedule barrier: RTL-to-GDS in 24 hours**
  - **Expertise barrier: No-human-in-loop, tapeout GDS**
  - **Cost barrier: Open source (and, runs in 24 hours)**

- **Unleash system innovation and design innovation**
- **Enable tool customization to system, application needs**

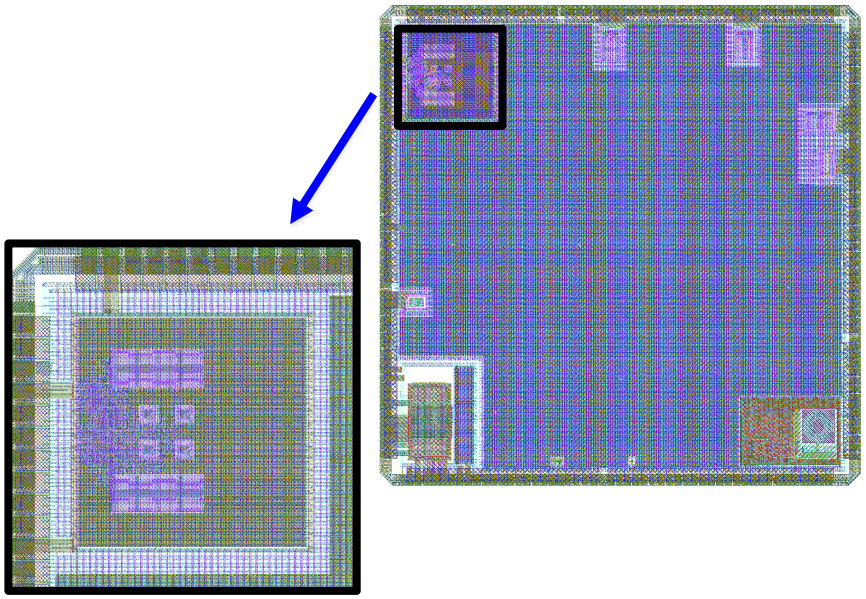
OpenROAD Usage Today



GF55 AI platform



GF12LP AI tile



OpenTitan SoC

Army Research Labs
GF55, GF12LP

U. Michigan / FASoC
GF12LP

OpenROAD Usage Today

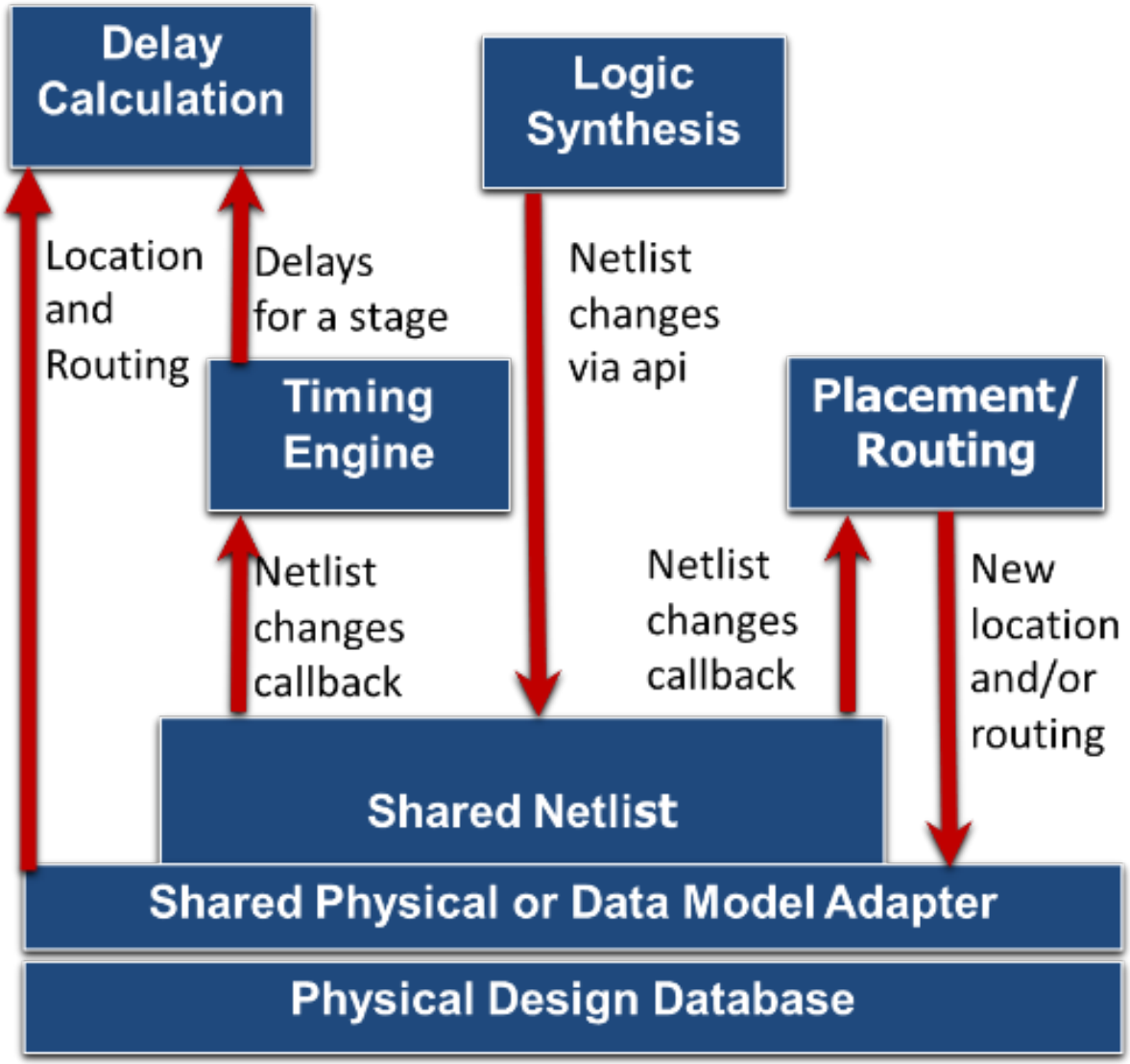
- **220+ tapeouts in SKY130**
  - Google-funded shuttles
- **Efabless "chipIgnite" commercial offering**
- **Matt Venn: From Zero to ASIC**
- **VLSI System Design (VSD)**
  - Total enrollment > 2500 in OpenLane / OpenROAD courses and workshops
  - Students: 85 countries, 26 languages

Growing User Community

- Users: novice to expert
- Applications: Trust, 3DIC, AI/ML, …
- Community metrics all growing
  - 2-week period in July 2021
  - Git clones (downloads), visitors, views

Industry-Strength Incremental Architecture: Built to Last

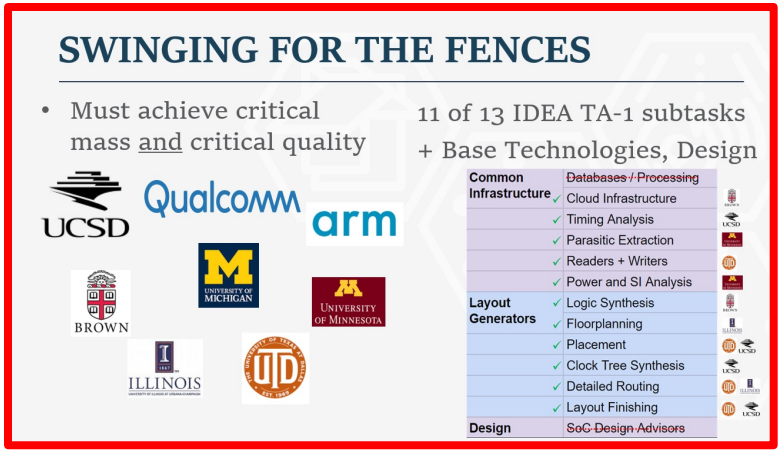
Further notes: link link link



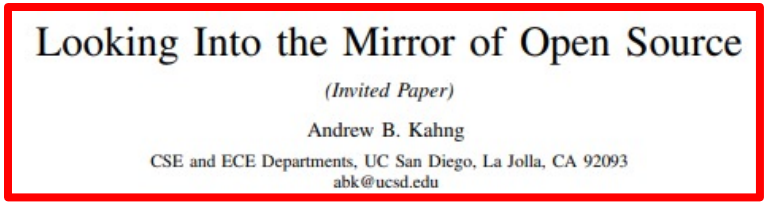
Leveling Up: “Seeding an Ecosystem”

Pathways to Enable Open-Source Ecosystems (POSE)
PROGRAM SOLICITATION
NSF 22-572
National Science Foundation

- ERI Summit 2018: Critical mass and critical quality



- ICCAD 2019: Open Source is a mirror



- VLSI-SoC 2020: If we build it, who will come?

Leveling Up: “Not Research As Usual”

- An **academic** research project, delivering **tapeout-clean** layout generation for **commercia**l FinFET nodes in **permissive open source**, per **contract** with the U.S. Government

New: OpenROAD RTL-to-GDS v1.0 Expectations

30 Nov 2019

- → expectations, names, “signoff”, …

NEW: OpenROAD “Safe Names” Conventions, v1.0

10 Dec 2019

- **“Not Research As Usual”**

A post-route timing evaluation flow with OpenRCX is available !

14 Dec 2020

IEEE CEDA DATC Robust Design Flow (RDF)

- **Academic reference design flow**
  - Initiated 2016; latest release 2021
  - Design flow built upon academic tools
  - Also supports complete, industrial format-based RTL-to-GDS flow with OpenROAD
- **GitHub:** https://github.com/ieee-ceda-datc/datc-rdf
- **High-level goals:**
  - Preserve leading research codes, and integrate them into design flow
  - Trigger design flow-scale and cross-stage optimization research

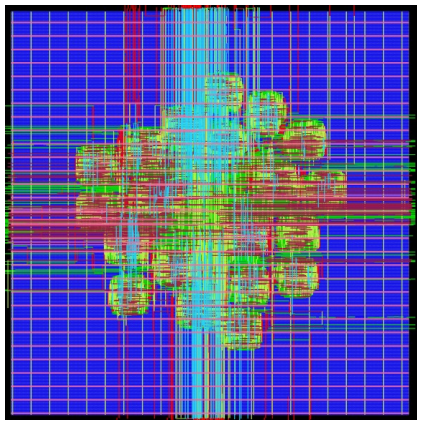
Component	Tool
RTL generator	Chisel/FIRRTL
RTL obfuscation	ASSURE
Logic synthesis	Yosys+ABC
DFT insertion	Fault
Floorplanning	TritonFP
Global placement	RePlAce, FZUplace, NTUPlace3, ComPLx, Eh?Placer, FastPlace3-GP
Detailed placement	OpenDP, MCHL, FastPlace3-DP
Flip-flop clustering	Mean-shift, FlopTray
Clock tree synthesis	TritonCTS
Global routing	FastRoute4-lefdef, NCTUgr, CUGR
Detailed routing	TritonRoute, NCTUdr, DrCU
Layout finishing	KLayout, Magic
Gate sizing	Resizer, TritonSizer
Parasitic extraction	OpenRCX
STA	OpenSTA, iTimerC
Integrated P&R app	OpenROAD
Database	OpenDB
Libraries/PDK	NanGate45, SKY130, ASAP7, NCTUcell
Benchmark conversion	RosettaStone

Jinwook Jung, IBM Research

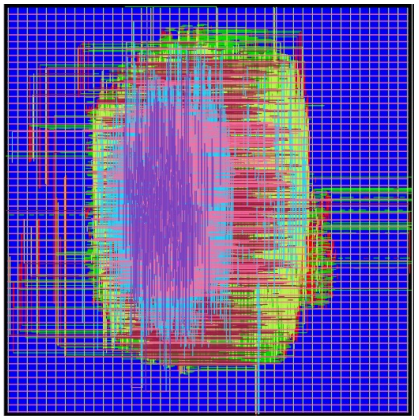
Calibrations Repository

Link: https://github.com/ieee-ceda-datc/datc-rdf-calibrations

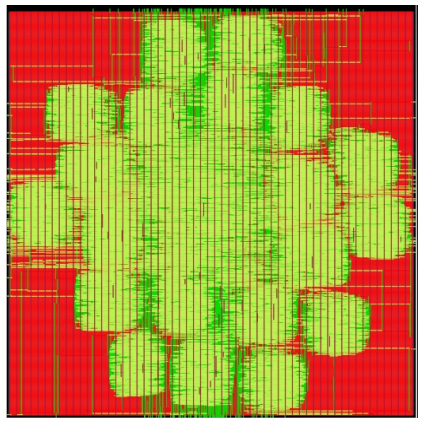
- **Reference analysis data for STA, RCX, and IR drop** obtained from **DRV-free** routed testcases using OpenROAD in open enablements
- Support academic research on analyses and verifications
- Routed *.v, *.def, *.sdc, and *.spef as well as *.json data
  - Two JSON formats for timer calibration (5-worst paths, worst slacks at endpoints)



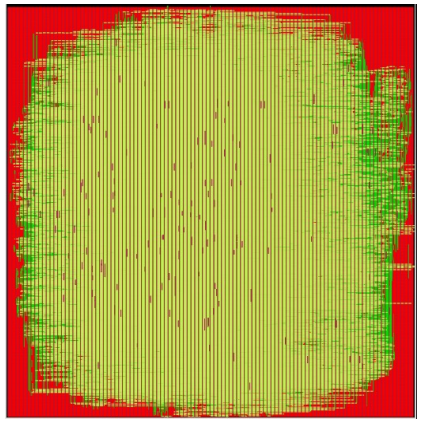
NanGate45-AES



NanGate45-JPEG



SKY130-AES

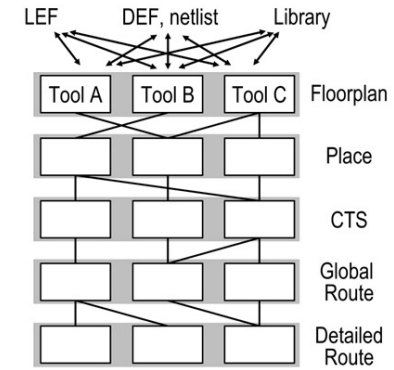


SKY130-JPEG

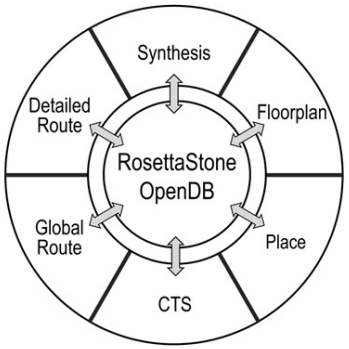
OpenDB as RosettaStone

Link: https://github.com/ABKGroup/RosettaStone

- **How to connect** Past, Present and Future of PD research?
  - Answer: **Data model** (e.g., LEF/DEF 5.8) + in-memory DB



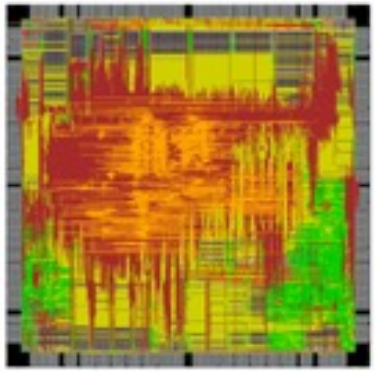
Tool chains passing files



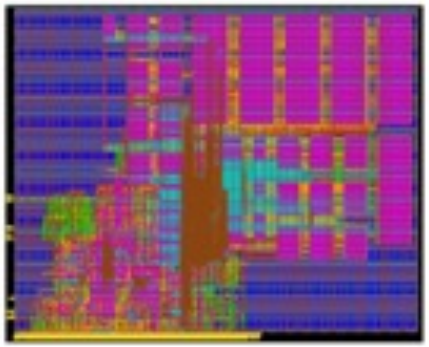
OpenDB as RosettaStone

- Modern tools can run on old benchmarks
- Old academic tools can run on modern designs, enablements

Global placement:	NTUPlace4h
Detailed placement:	NTUPlace4h
Clock Tree Synthesis:	TritonCTS
Global routing:	FastRoute
Detailed routing:	TritonRoute

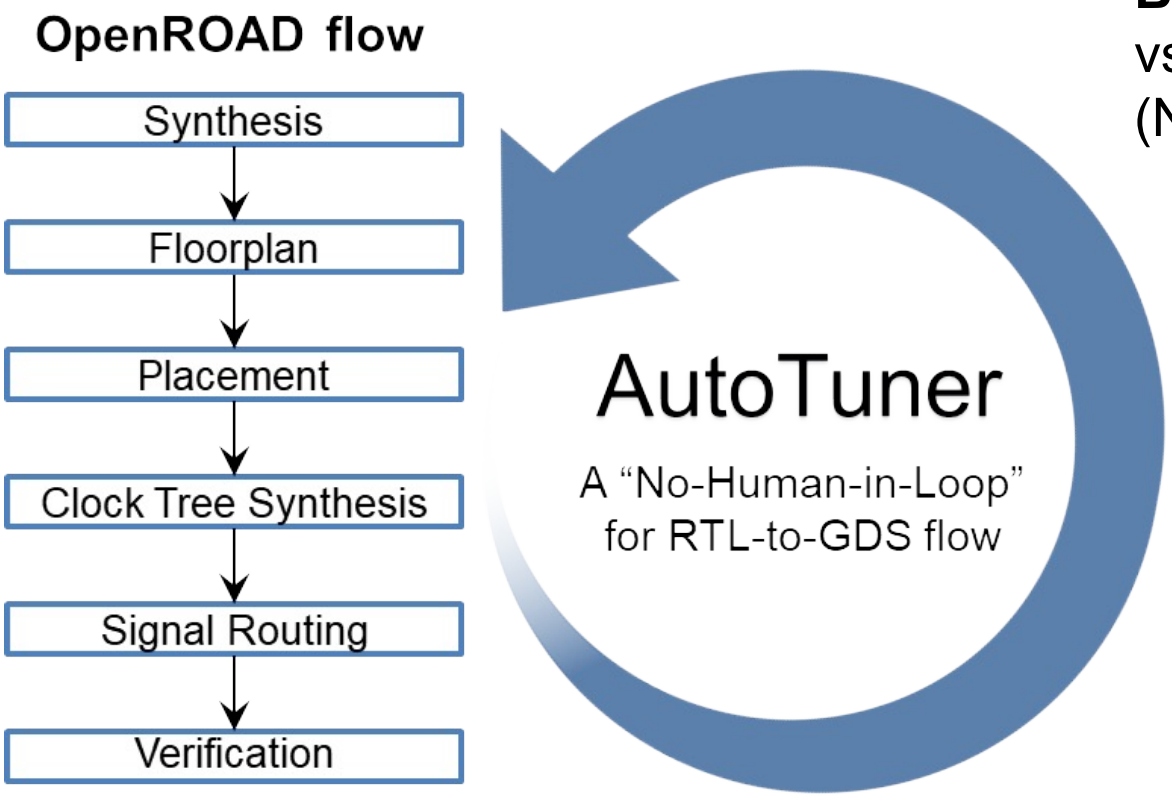


adaptec1 (ISPD05)
on SKY130HD



swerv_wrapper (RISC-V)
on Commercial 12nm

AutoTuner: True No-Human-in-Loop Operation



Best human-tuned flow
vs. **AutoTuned** flow
(No-Human-In-Loop)

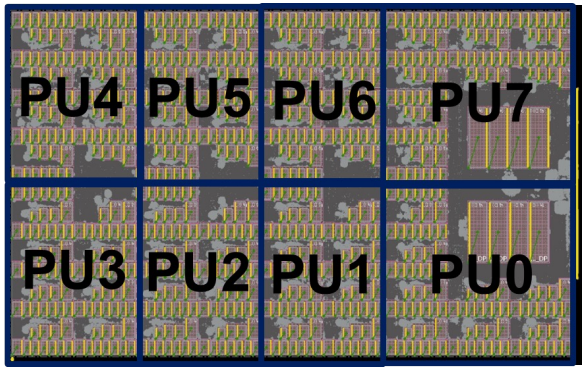
23% Faster
45% Less Power

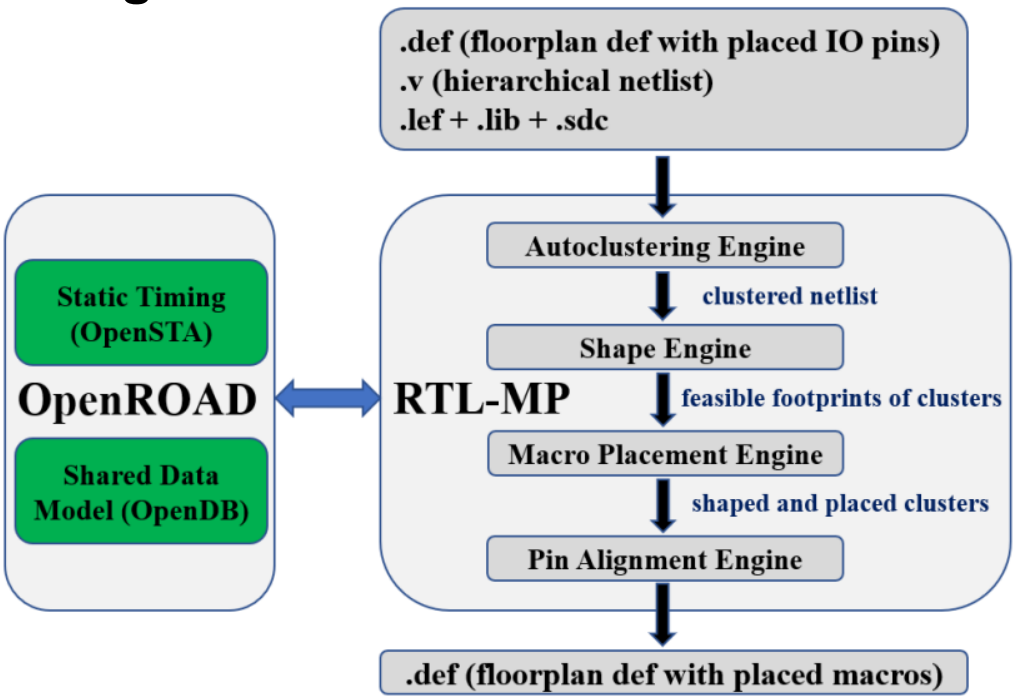
lowRISC RISC-V core (ibex),
SKY130HD

RTL-Macro Placer

RTL-MP: Toward Practical, Human-Quality Chip Planning and Macro Placement

- **"Mimic" behavior of human experts**
- **Capture dataflow defined by RTL designers**
  - Exploit logical hierarchy
  - Extract *connection signatures* between modules
- **Understand key constraint types**
  - Macro placement blockages
  - Preferred locations for macros
- **Next: Hierarchical RTL-MP**

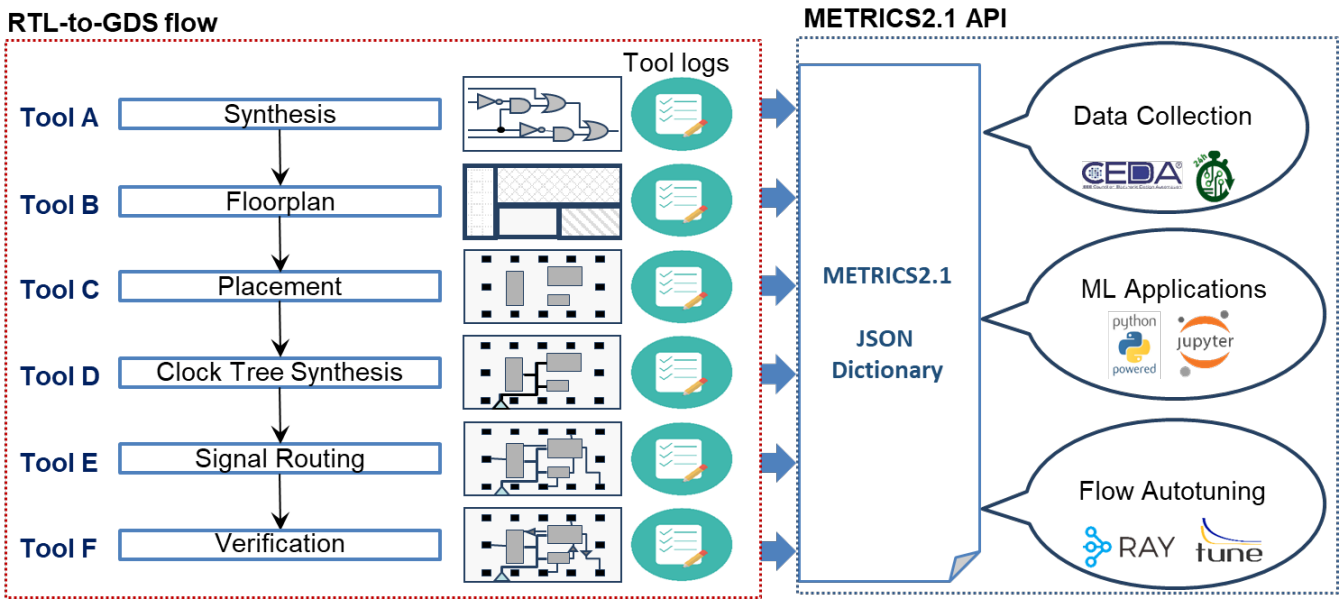




Overview of RTL-MP flow

METRICS2.1 https://github.com/ieee-ceda-datc/datc-rdf-Metrics4ML

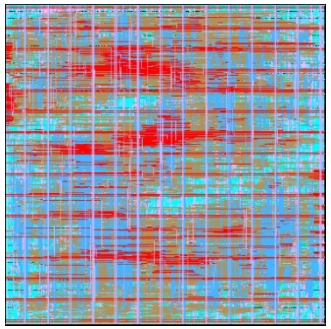
- "Measure, to Improve": enables collection and sharing of **public** datasets for research on ML in EDA
  - Sample Jupyter notebooks to visualize and build simple prediction models
- Feeds the score function evaluations used by AutoTuner



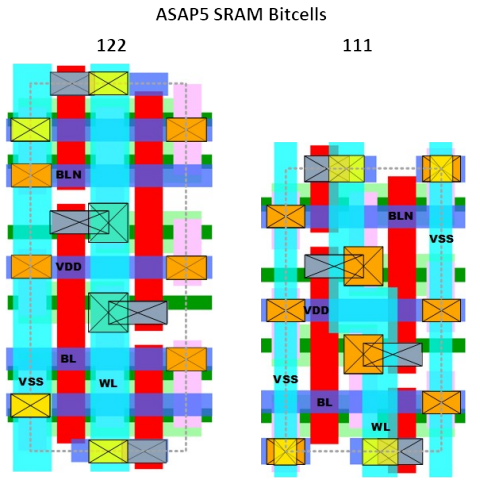
- Guiding principles
  - General and extensible
  - No ambiguity: 1-1 mapping between measurements and names
  - Free, open and frictionless; agnostic to EDA provider

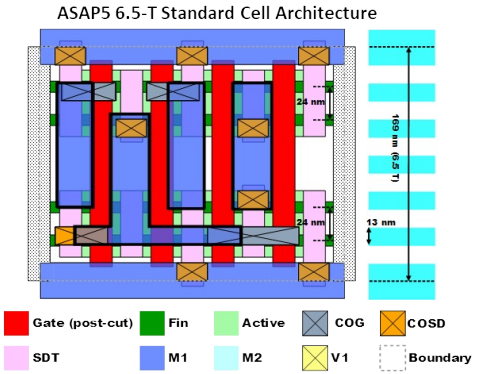
“Base Technologies” Task: ASAP7 / ASAP5

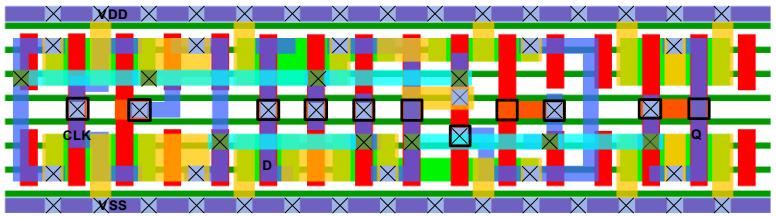
- **ASAP7: 7nm FinFET predictive PDK**
  - https://github.com/The-OpenROAD-Project/asap7
  - **7.5-track: 212 cells × four Vt’s**
  - **6-track: 242 cells x four Vt’s**
  - Cell CDL, GDS (RVT only), LEF, LIB (NLDM and CCS), Verilog, and parasitic extracted CDL views
- **ASAP5: horizontal nanowire transistors**
  - 3-D TCAD based compact models
  - Reflects scaling boosters: single diffusion breaks, contact over-active gate, denser cross-overs
  - **Near release**; ASAP5 PDK journal paper in review



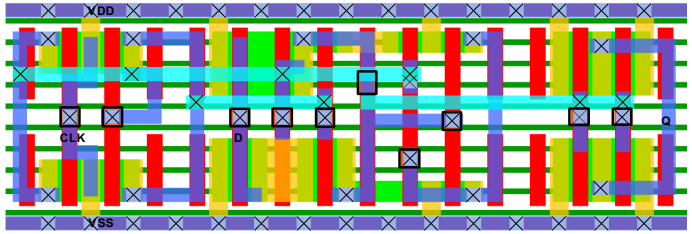
Pipelined TMR AES





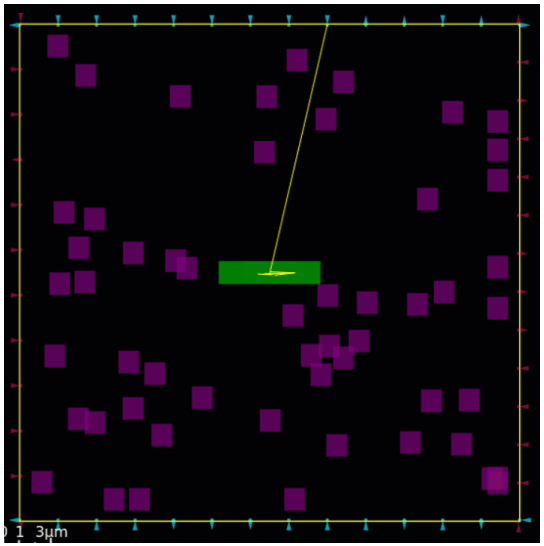


D-Latch (6-track)

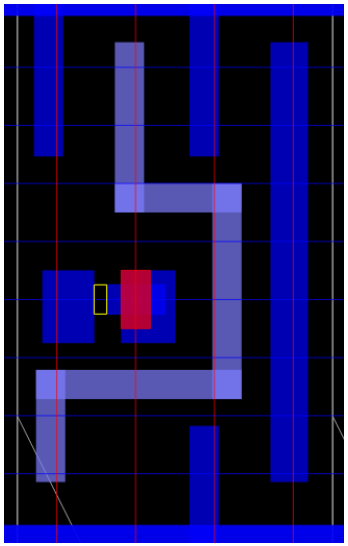


D-Latch (7.5-track)

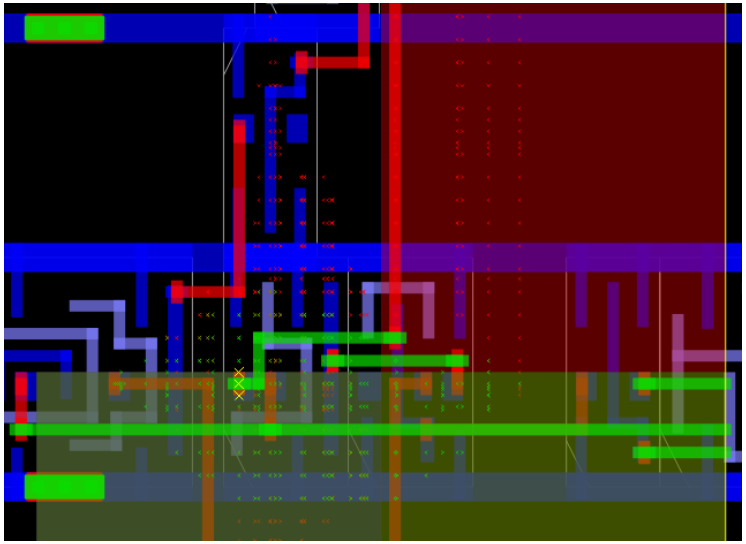
Many GUI Features for Algorithm Developers



Watching a cell in global placement



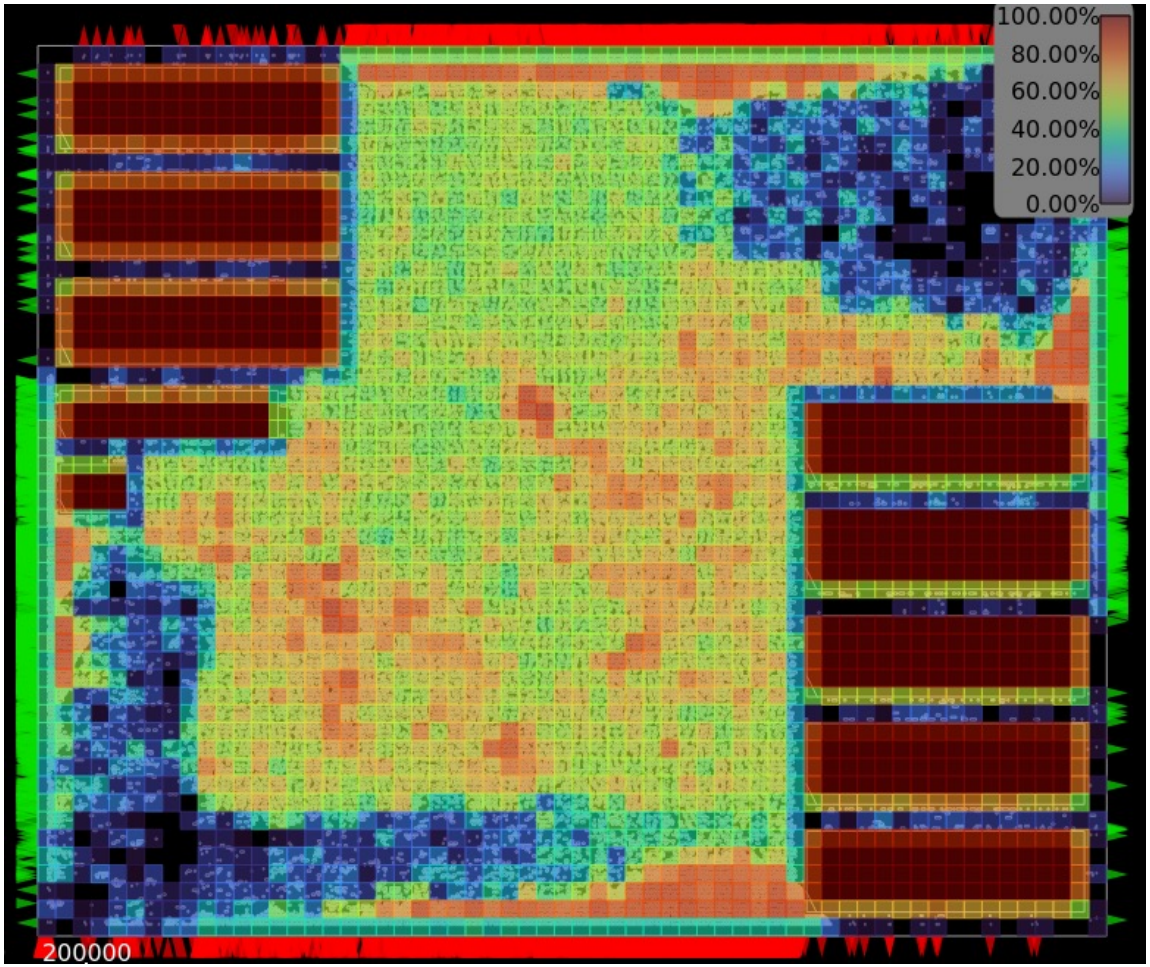
Via-to-pin spacing violation with track pattern overlaid



Grid points searched by the A* maze router, colored by layer

Heatmaps

Heat Maps
- Placement Density
- Power Density
- Routing Congestion
- IR Drop



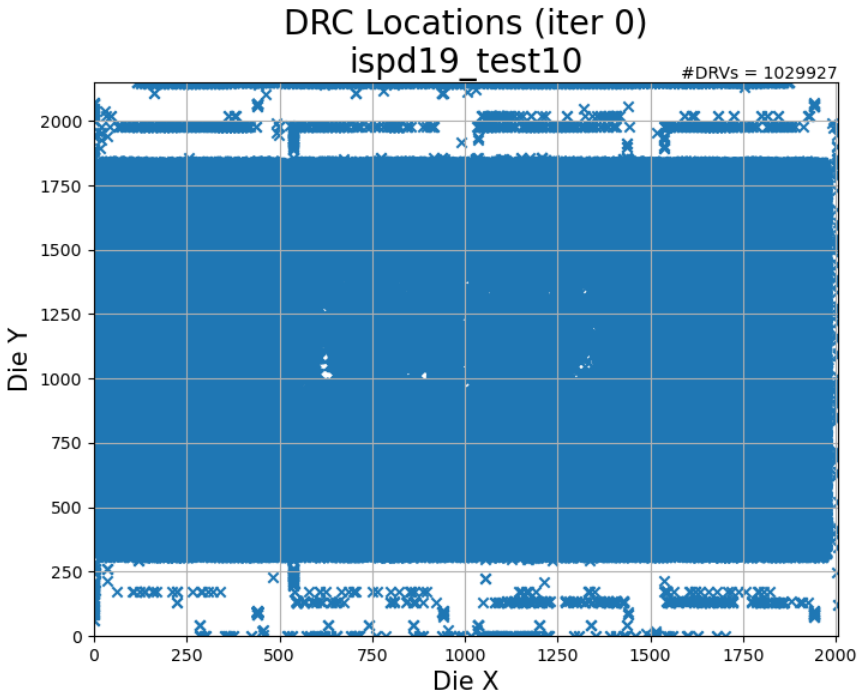
Dump to files for machine learning

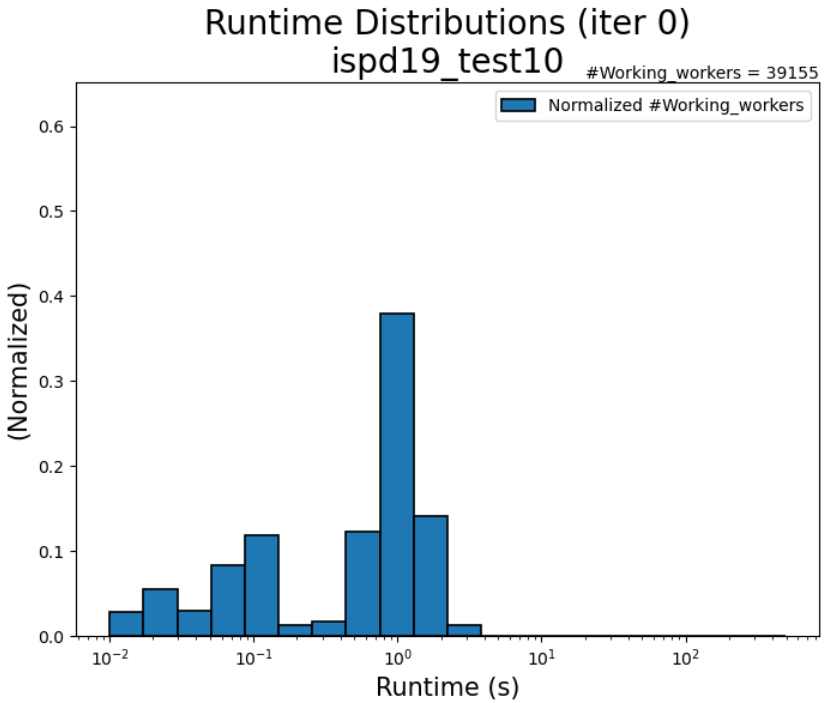
Example:

"gui::dump_heatmap Placement pd.csv" produces:

```
x0,y0,x1,y1,value
0,960,20,980,7.200000e-01
0,940,20,960,2.400000e-01
0,920,20,940,1.200000e+00
0,900,20,920,0.000000e+00
0,700,20,720,1.440000e+00
0,720,20,740,2.520000e+00
0,740,20,760,1.200000e+00
…
```

Real Accessibility of Real ML Challenges





- **ispd19_test10, TritonRoute** [*drt* in OpenROAD]
  - **Left: Tiles with DRVs** (down to 21 at final iteration #64)
  - **Right: Runtimes of workers** (up to > 500s in late iterations)
- **Prediction**
  - Doomed runs? Hotspots, long tails? DRVs from placement, global route?
- **Learning**
  - Costing strategies for workers? Sampling / multi-start strategies for tiles?

Trajectory Impacts of OpenROAD

- "Leveling up" amplifies researchers' efforts, reduce wasted energy
  - More efficient discovery, innovation in physical design
  - More accessible education and training of a future workforce
- Openness and infrastructure
  - Documentation in code of how to architect and build an EDA tool
  - Integration backplane that lowers barriers and improves transparency
- Virtuous cycle of democratization and innovation
  - High-schoolers making tapeouts → disruptive enabler for education
  - Rewarding for project contributors
- Critical mass and critical quality are possible
- Sustainability and translation are first-class concerns
  - Governance, stable funding, organizational structure, roadmap, stakeholders, …
- On the road to "EDA2.0": Intelligence + Cloud
  - **COPILOT**: **C**loud **O**ptimized **P**hysical **I**mplementation **L**everaging **O**penROAD **T**echnology

Agenda

- Leveling Up: A Trajectory
- The OpenROAD Project
- The TILOS AI Institute
- And Beyond

What is TILOS?

NSF National AI Research Institute for Advances in Optimization

Mission: make impossible optimizations possible, at scale and in practice.

5-year grant, $20M total funding from NSF (started November 1st)

Partial support is from Intel Corporation

TILOS: Fishburn and Dunlop, ICCAD85

TILOS: A POSYNOMIAL PROGRAMMING APPROACH TO TRANSISTOR SIZING

J. P. Fishburn[1] and A. E. Dunlop[2]
AT&T Bell Laboratories
Murray Hill, New Jersey 07974

Abstract
A new transistor sizing algorithm, which couples synchronous timing analysis with convex optimization techniques, is presented. Let A be the sum of transistor sizes, T the longest delay through the circuit, and K a positive constant. Using a distributed RC model, each of the following three programs is shown to be convex: 1) Minimize A subject to $T < K$. 2) Minimize T subject to $A < K$. 3) Minimize AT^K. The convex equations describing T are a particular class of functions called posynomials. Convex programs have many pleasant properties, and chief among these is the fact that *any point found to be locally optimal is certain to be globally optimal.* TILOS (TImed LOgic Synthesizer) is a program that sizes transistors in CMOS circuits. Preliminary results of TILOS's transistor sizing algorithm are presented.

1. Introduction

Given a synchronous MOS circuit of the form shown in Figure 1 with N transistors of sizes (channel widths) x_1, x_2, ..., x_N, the following question is considered: How can the circuit's performance be improved by adjusting the x_i? Two figures of merit are of special interest. T is defined to be the minimum

John P. Fishburn, PhD

SACO – John P. Fishburn, PhD, 69, of Sa
away peacefully Saturday, April 24, 2021.
as "Jack" by his friends and family, was bo
Muscatine, Iowa on May 21, 1951, the son
Susan (Wooldridge) Fishburn.

Jack attended local schools in Muscatine a
his undergraduate degree in Mathematics
University of Iowa. Jack furthered his edu
receiving his PhD from University of Wis
Madison where he defended his thesis on
algorithms including the parallelization o

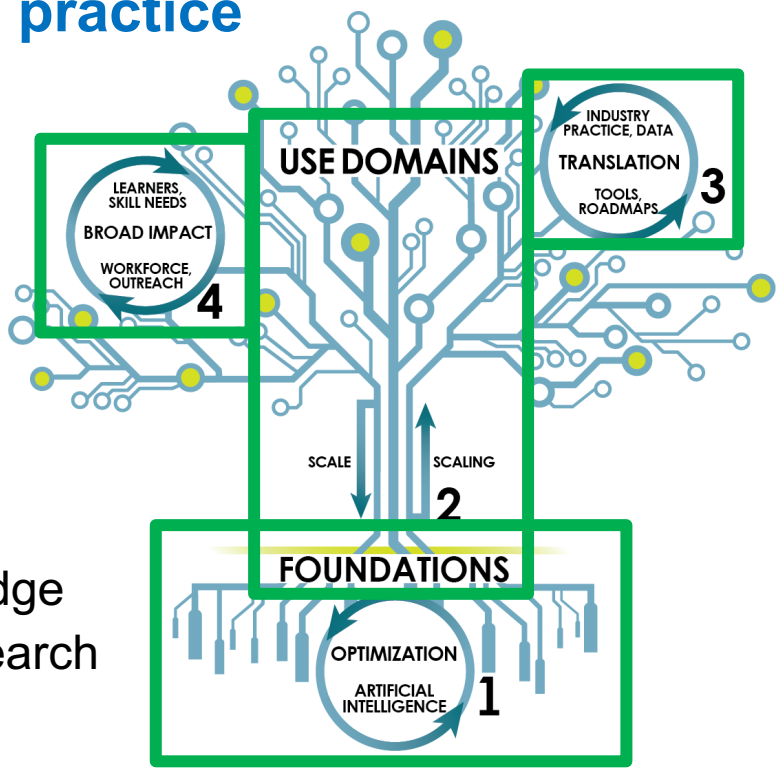
Jack met his wife of 35 years, Lynne, after

The Institute for Learning-enabled Optimization at Scale

Optimization: Find a best-possible solution

Fundamental challenges: scale and complexity

→ Nexus of AI/ML, optimization, use in practice



Vision: Four "virtuous cycles"

1. **Foundations:** AI and Optimization
2. **Scaling:** Foundations and Use Domains
3. **Translation:** Academia and Industry leading edge
4. **Broad Impact:** Education, Outreach, and Research

Learning and Optimization: Foundations

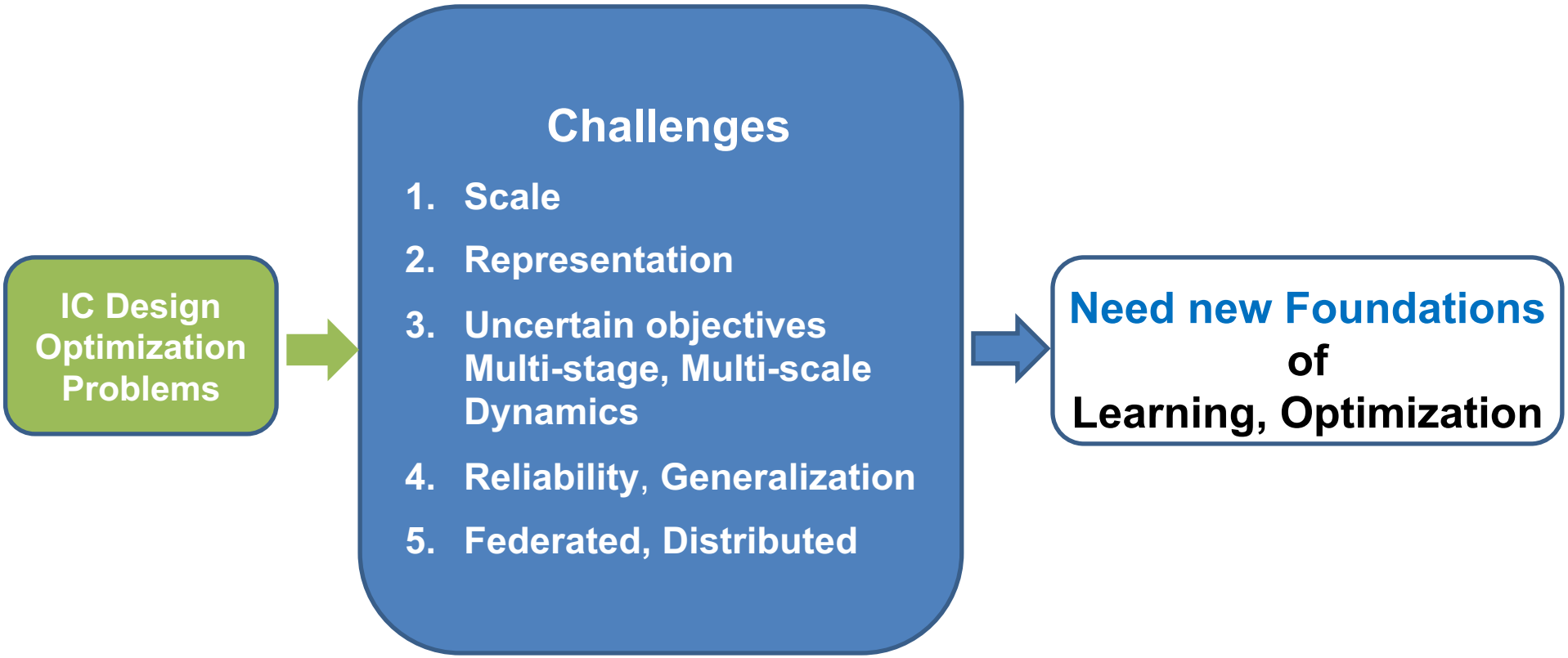
AI advances → new challenges, new tools for optimization

- Bridging Discrete and Continuous
- Distributed, Parallel, and Federated
- Optimization on Manifolds
- Dynamic Decisions under Uncertainty
- Nonconvex Optimization in Deep Learning

New perspectives on classic problems of optimization

Credit: N. Vishnoi, S. Jegelka, D. Spielman, Y. Wang +

IC Design and Design Automation Challenges



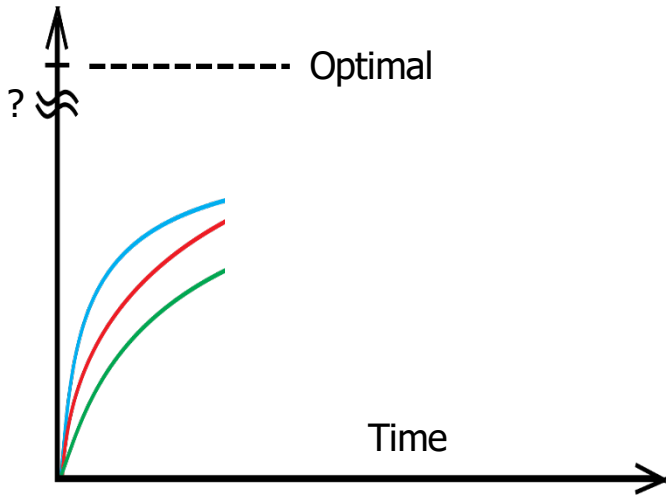
Credit: N. Vishnoi, S. Jegelka, D. Spielman, Y. Wang +

EDA is Optimization and Machine Learning, too …

- **EDA is about optimizations and algorithms**
  - High stakes: performance, power, design closure
- **Discrete, combinatorial formulations at huge scale**
  - **Optimizations:** ILP, MCF, QAP, SAT/SMT, LR, …
  - **Algorithms:** min-cost flow, high-dim DP, …
- **But need an answer overnight**
- Reality under the hood: **metaheuristics**
  - Annealing, multi-start, PSO, NGSA-II, ripup-reroute, greed, …
  - Convenient objectives, customer-/tech-specific tuning, …
  - *… which comes at a cost*

Optimization and Optimality

- **Reality of optimization**
  - **Better, faster, cheaper – pick any two**
- **IC EDA: want all three at once**
  - **"Unfortunately, the runtime of …"**
- **But the world has changed**
  - **Automation, cloud, …**

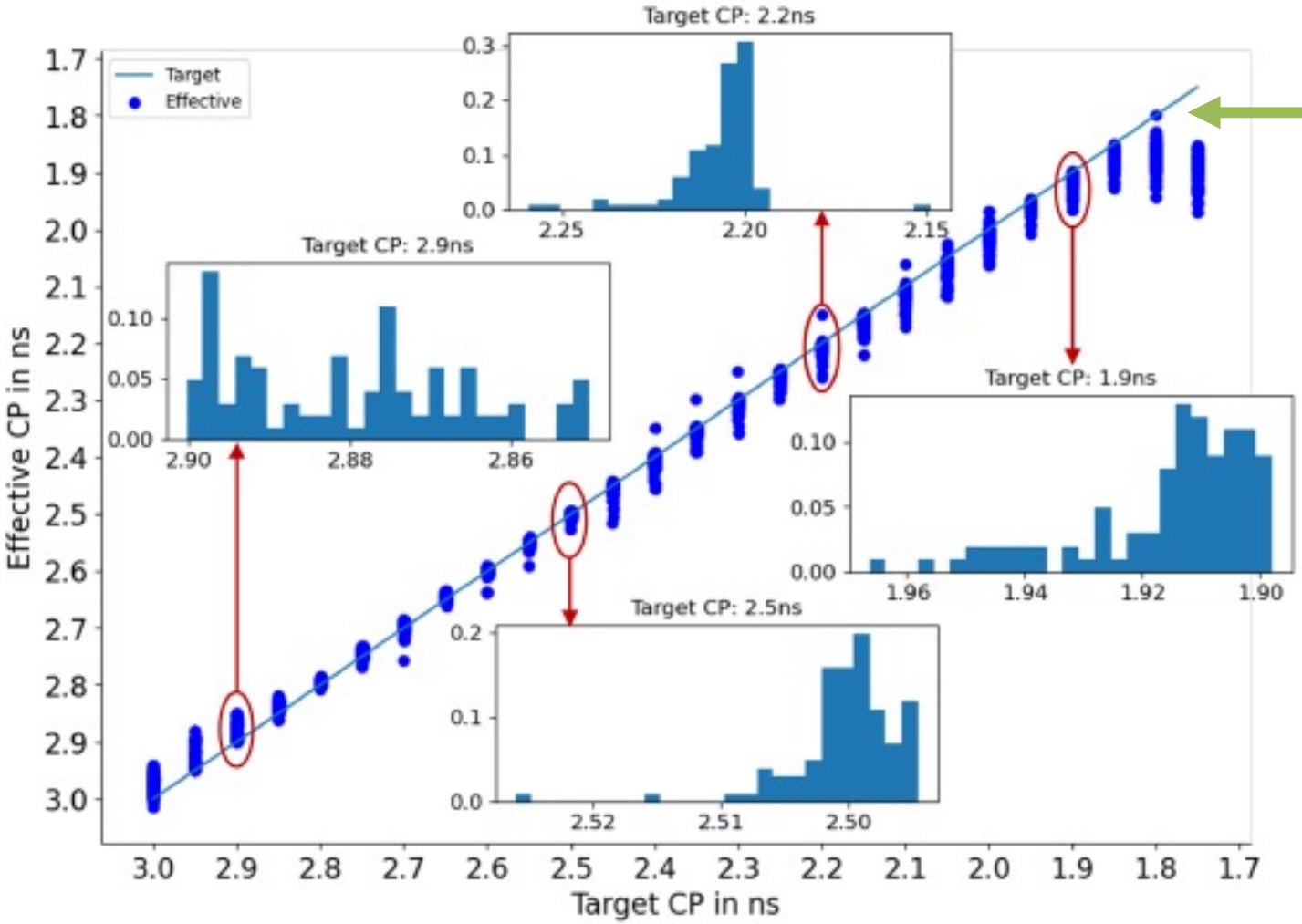


Question: If you give your SP&R flow 3 extra days of runtime, would it know how to use this extra time?

Question: If you could run 10,000 copies of your P&R tool at the same time, what QOR improvement **should** you expect?

“Noise”, “Chaos” → Learning and Sampling

“Actual” vs. “Target” clock period



Best Actual fmax
= what we want

Refocus on Suboptimality [to get closer to Optimality!]

Suboptimality … in what sense?

- **Learn** proper objective/loss functions
- Evaluate in proper context (e.g., full flow, post-detailed route)

Benchmarking considered helpful

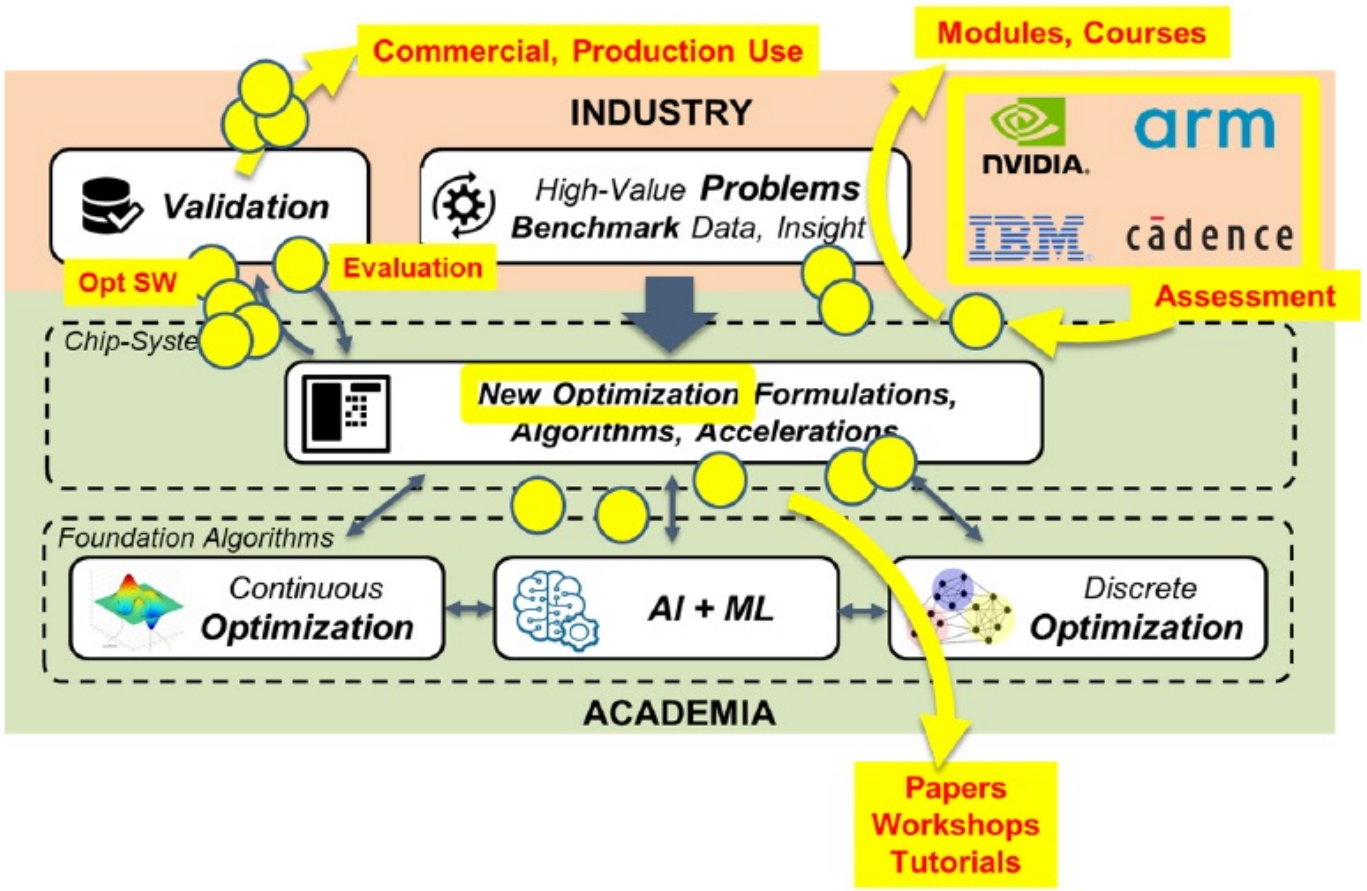
- "Real" benchmarks in EDA are obfuscated, incomplete, non-vertical, old…
- "Artificial" benchmarks wander between realism, known optimal solution quality, scalability …
- **It's time to level up: e.g.,** "Underwriters Laboratories for IC Design Tools"

Modern compute resources *cloud, GPU, accelerators, …*

- EDA Optimization + Learning **naturally** live in the cloud
- **Many EDA optimizer implementations today: "EDA1.0" from the 1980s**
  - **→ Can TILOS help discover new, cloud-scalable "EDA2.0" foundations?**

“Life Cycle” of Research and Translation

Translation is here

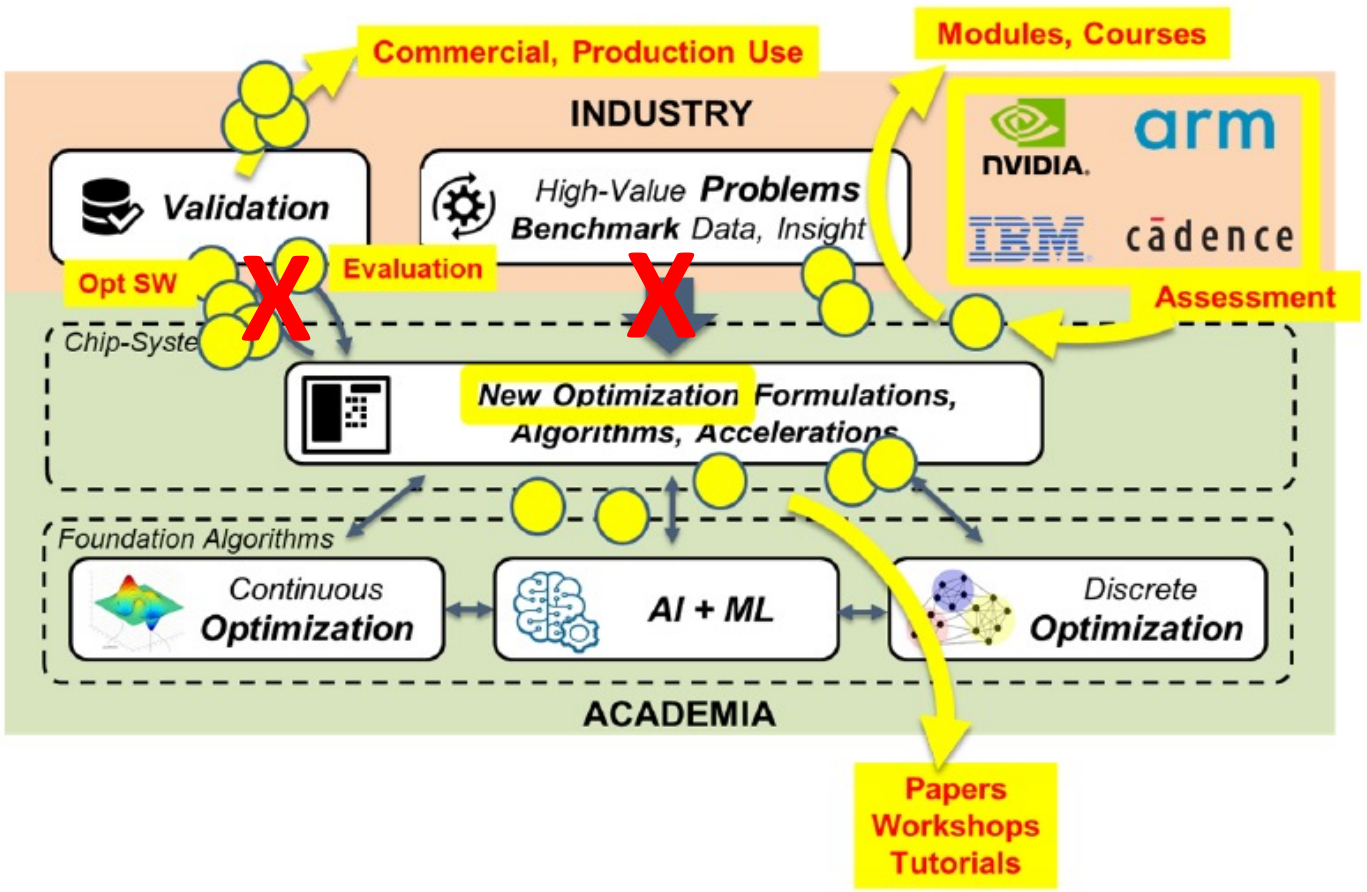


“Life Cycle” of Research and Translation

Translation is here →

Translation requires solutions to the **X’s**:

- data
- benchmarking
- roadmapping



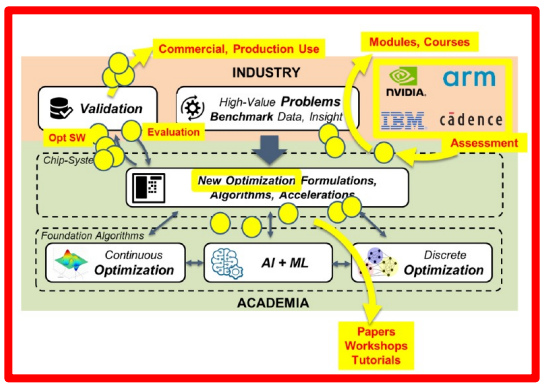
Make the "Third Virtuous Cycle" Real

- Real data is sensitive, artificial data isn't valued
- Multi-way NDAs, export control, …
- Irreproducible research
- Haves vs. Have-Nots
- Risks: bias, ethics, fairness
- …

What we need

- **Democratization and accessibility** of research at the leading edge
- **New norms** for transparency, reproducibility, translation
- **New norms** of benchmarking in a high-stakes use domain
- **Principled** roadmapping to guide investment of time and $

3



Example Research Questions *on the path to Translation*

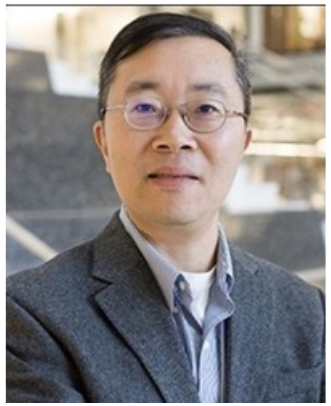
- **Can't access or expose real data**
  - Generate artificial circuit designs that are indistinguishable from real circuit designs *from the perspective of optimizers*
  - Learn from much less data
- **Can't access the best optimizers**
  - Model of "strong optimizer" outcomes based on instance attributes and "weak optimizer" outcomes
- **Can't reveal sources of data**
  - Privacy-preserving anonymization and obfuscation
- **Benchmarking brings risks of misuse**
  - Develop ethical principles and validations to enable fair benchmarking
- **Can't identify the most crucial learning, optimization goals**
  - Roadmaps + Drivers

My Personal Target List

- ***Learning to Optimize (L2O)***
  - Models, predictors and objectives; sampling; RL; hybridized optimizers
- ***Scaling the reach of optimizers***
  - Partitioning, clustering, sparsification; cloud/parallel; multi-{dims,objs}
  - Optimal solvers as well (e.g., optimal peephole/clip P&R&Opt)
- ***System/Arch/SoC PPAC exploration***
  - Learning to cluster+shape+pack+plan; pathfinding with confidence
    - Stack of abstractions: device, circuit, memory, integration fabrics

TILOS Chips Team




- **Layout**
  - Optimal embedding, “modern partitioning”
  - Nexus of sampling, sequential decision-making, learning to optimize
- **Verification**
  - E.g., interior search methods to scale SMT solving
- **Quantifying the cost of “X” (e.g., X = security)**
  - Intrinsic cost of robustness in optimization and learning
  - Data anonymity, data integrity, privacy in federated and distributed settings
- **Data, benchmarking and roadmapping**
  - “Artificial but realistic” benchmarks
  - Principles for ethical and fair benchmarking

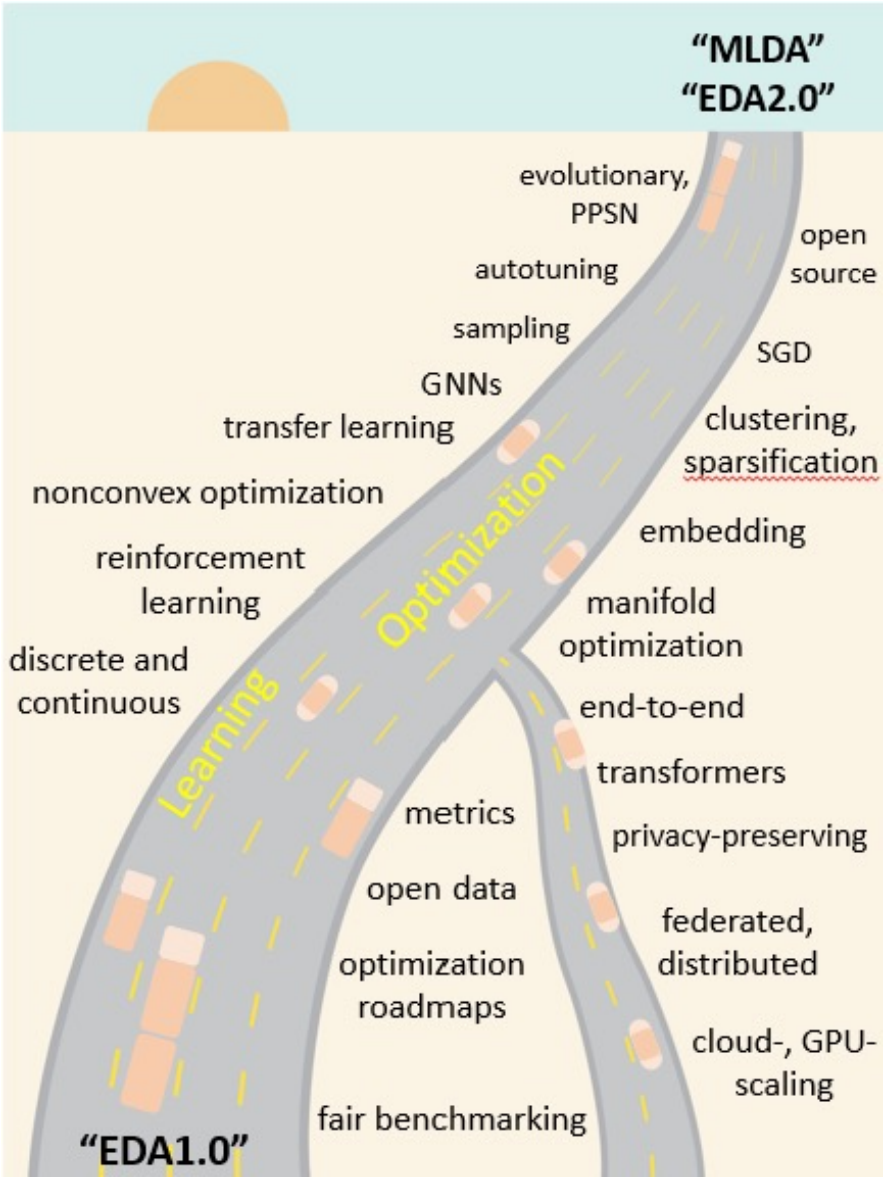
Trajectory Impacts of TILOS

- Cracking the code of translation
  - Unblocking data and benchmarking → transparency, reproducibility, efficiency
  - Roadmapping of core optimization problems → where to invest time and effort
- Improved understanding of suboptimality
  - Modern compute context → federated learning and optimization; sequential decision-making and sampling
- A new nexus of learning, optimization and practice
  - Learning to learn, learning to optimize

Agenda

- Leveling Up: A Trajectory
- The OpenROAD Project
- The TILOS AI Institute
- And Beyond

What Next?



- **"ML-enabled DA"**, **"EDA2.0"** are also waypoints in the trajectory
- **Enabled by learning and optimization technologies**
- Some elements arrive before others
  - E.g., autotuning and sampling before fair benchmarking
- **Can we deliver on this picture within 5-10 years?**

Leveling Up: ISPD-2030

- Did a new nexus of AI, optimization and chip design bring excitement and accessibility, and attract more young people?
- Did we fully leverage advances in optimization and learning?
  - How did we scale the reach of optimizers, e.g., by splitting up problems, or making them smaller or sparser, without losing solution quality?
- Did we advance the most critical optimizations to target with "MLDA" or EDA2.0?
  - E.g., partition-cluster-shape-pack-plan co-optimization to serve design space exploration and pathfinding = essential for pathfinding in the beyond-{Moore, CMOS, von Neumann} era
- Did we make advances in representation to complement advances in learning and optimization?

Links and Acknowledgments

- **More info (papers and talks)**
  - OpenROAD publications and ERI Summit presentations page
  - GOMACTech 2021 “The OpenROAD Project: Unleashing Hardware Innovation” .pdf .mp4
  - DARPA MTO ERI Summit 2021 “OpenROAD: Foundations and Realization of Open, Accessible Design” slides .mp4
  - ICCAD 2021 “METRICS2.1 and Flow Tuning in the IEEE CEDA Robust Design Flow and OpenROAD” .pdf .pptx .mp4
  - Synopsys APUP 2022 “AI/ML, Optimization and EDA in TILOS, an NSF National AI Research Institute” .pptx .mp4
  - UCLA ECE 2022 “The OpenROAD Project: Today and Beyond” .pptx
  - NSF Integrated Circuits Research, Education, and Workforce Development Workshop talk (Oct. 15, 2021) .pptx

- Many thanks are due to many great colleagues in OpenROAD and TILOS, as well as numerous students and collaborators. Research at UCSD is partially supported by NSF CCF-2112665 and CCF-1564302, DARPA HR0011-18-2-0032 and FA8650-20-2-7009, Samsung, Qualcomm, NXP Semiconductors and the C-DEN Center.

https://theopenroadproject.org
https://vlsicad.ucsd.edu
https://tilos.ai

THANK YOU !