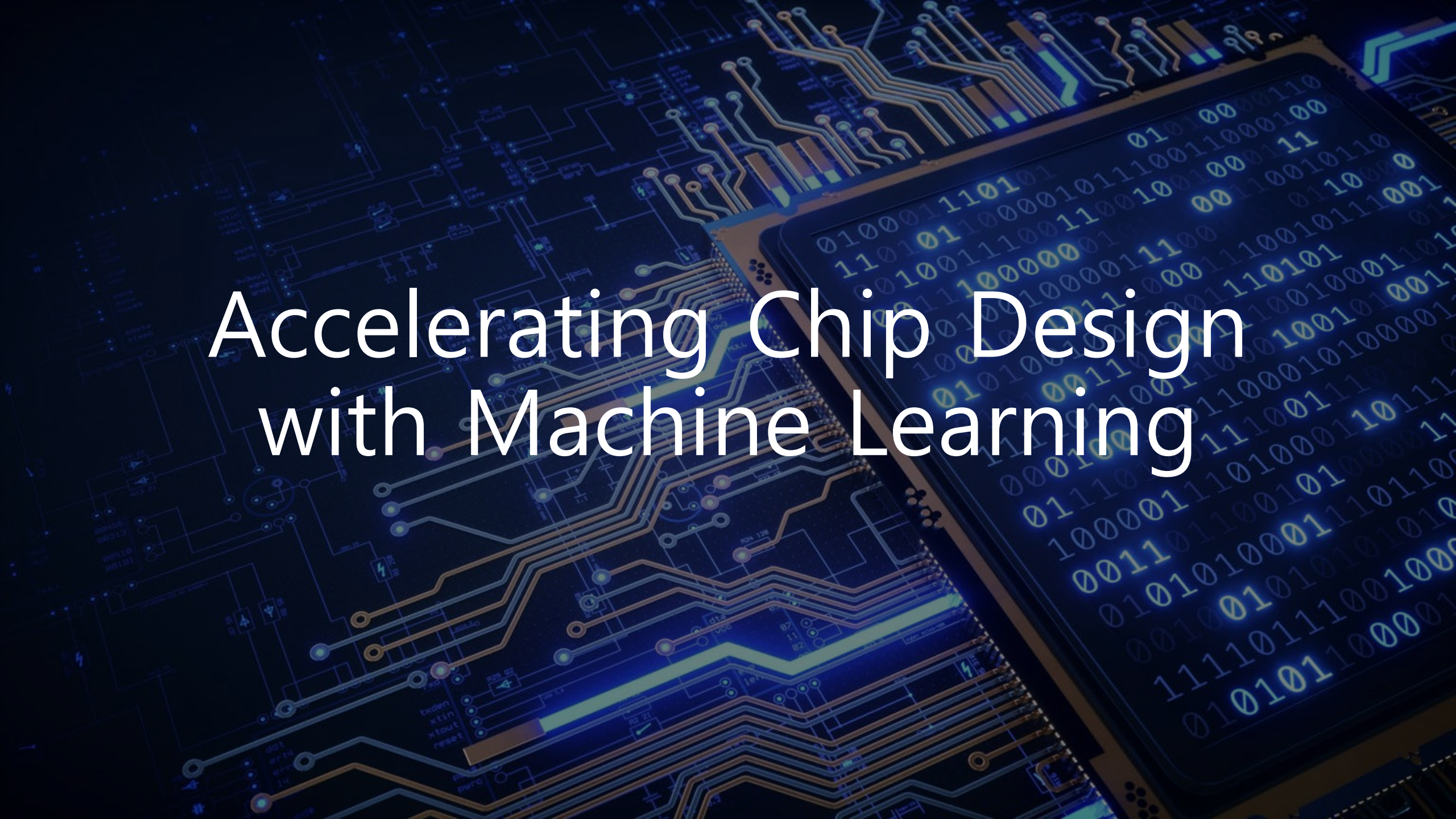

Accelerating Chip Design
with Machine Learning

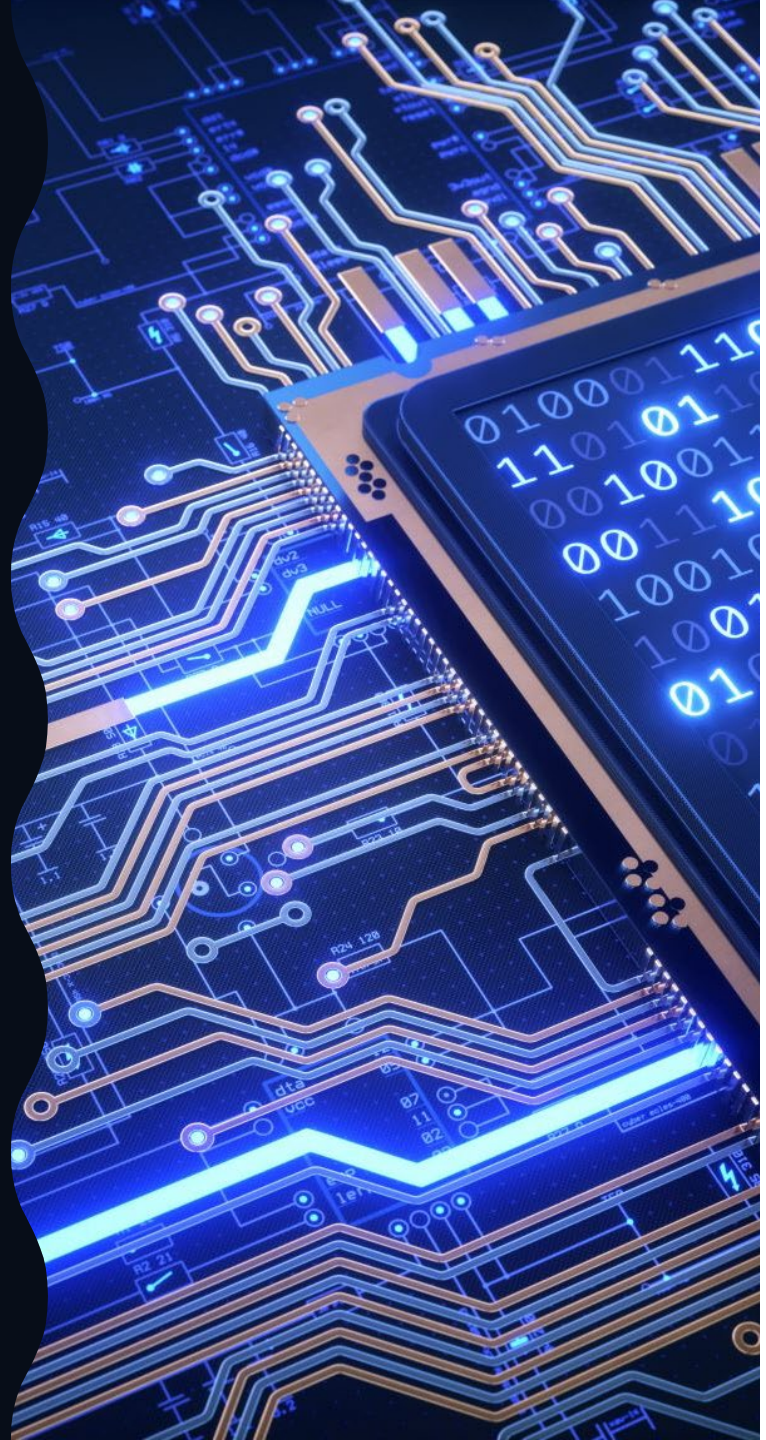
PAPER INFORMATION

- B. Khailany et al., "Accelerating Chip Design With Machine Learning," in IEEE Micro, vol. 40, no. 6, pp. 23-32, 1 Nov.-Dec. 2020, doi: 10.1109/MM.2020.3026231

- **URL**: http://ieeexplore.ieee.org.ssl.sa.skku.edu:8080/stamp/stamp.jsp?tp=&arnumber=9205654&isnumber=9234128

- **Published in**: IEEE Micro (Volume: 40, Issue: 6, Nov.-Dec. 1 2020)

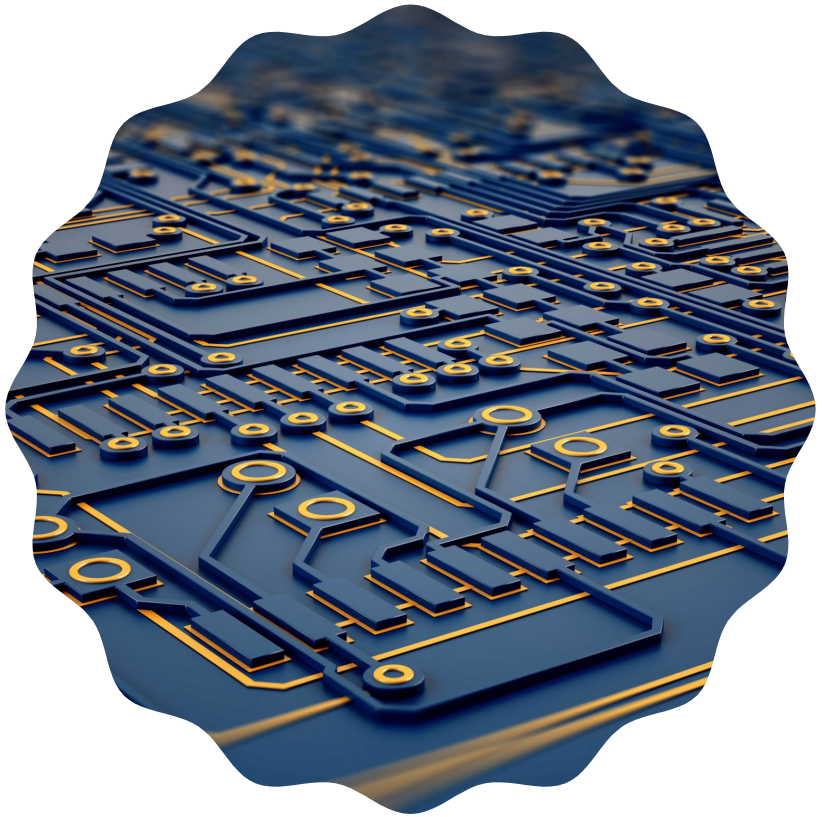
- **Date of Publication**: 24 September 2020



SECTION 1
INTRODUCTION

INTRODUCTION

- Over the last few decades, electronic design automation (EDA) algorithms and methodologies were developed for all aspects of chip design: design verification and simulation, logic synthesis, place-and-route, and timing and physical signoff analysis
- With each increase in automation, total work per chip has increased, but more work was offloaded from manual effort to software
- With machine learning (ML) transforming software in many domains, this trend will continue with ML based automation of electronic design automation (EDA)



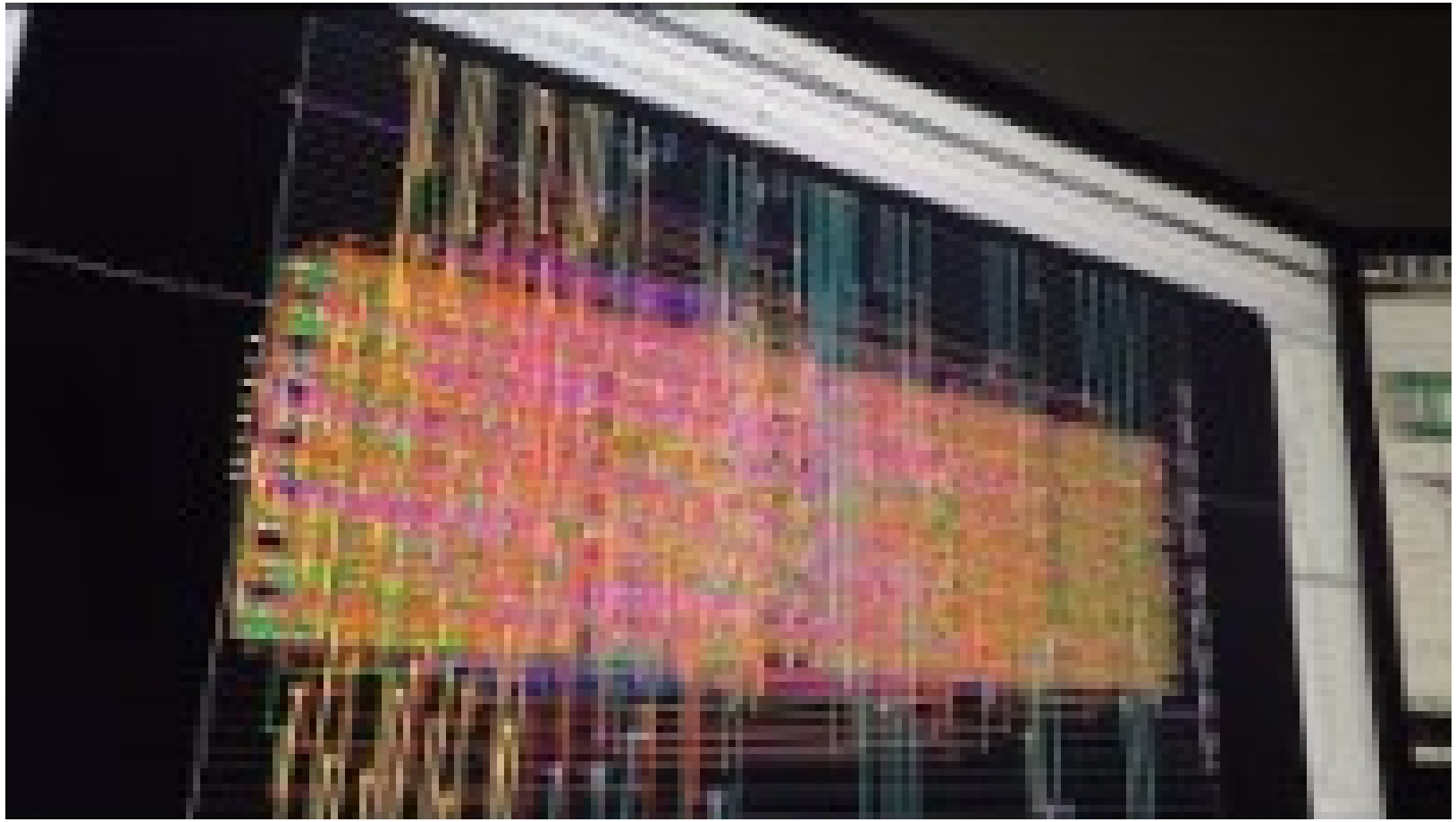
<u>This paper highlights the followings</u>

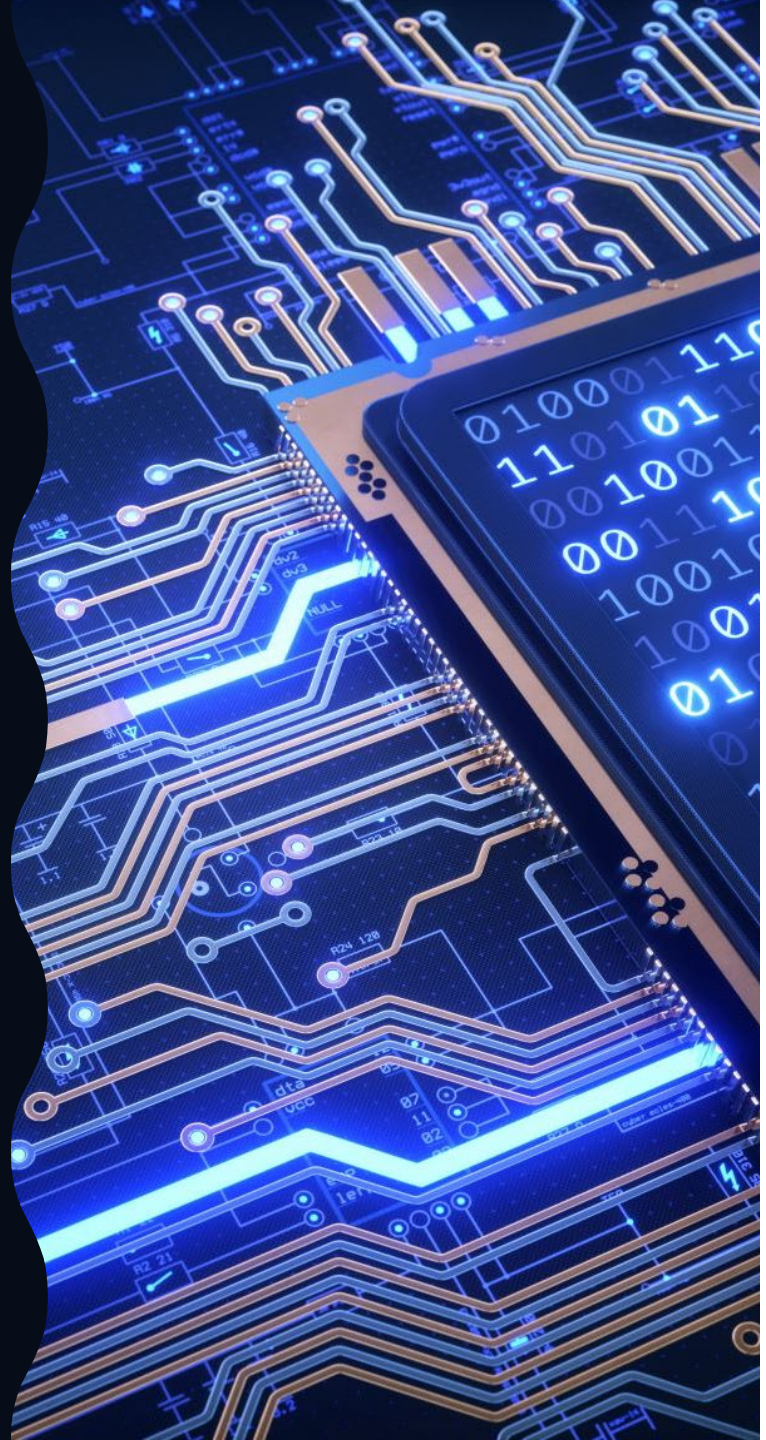
1. selected work from NVIDIA research group and the community applying ML to chip design tasks
2. present a vision for a future AI-assisted design flow, where GPU acceleration, neural-network predictors and reinforcement learning techniques combine to automate the VLSI design

DEEP REINFORCEMENT LEARNING: DEFINITION

- Deep reinforcement learning (deep RL) is **a subfield of machine learning** that combines reinforcement learning (RL) and deep learning.
- Deep RL incorporates deep learning into the solution, allowing agents to make decisions from unstructured input data without manual engineering of the state space.

ELECTRONIC DESIGN AUTOMATION: DEFINITION





SECTION II
MACHINE LEARNING OPPORTUNITY

MACHINE LEARNING OPPORTUNITY

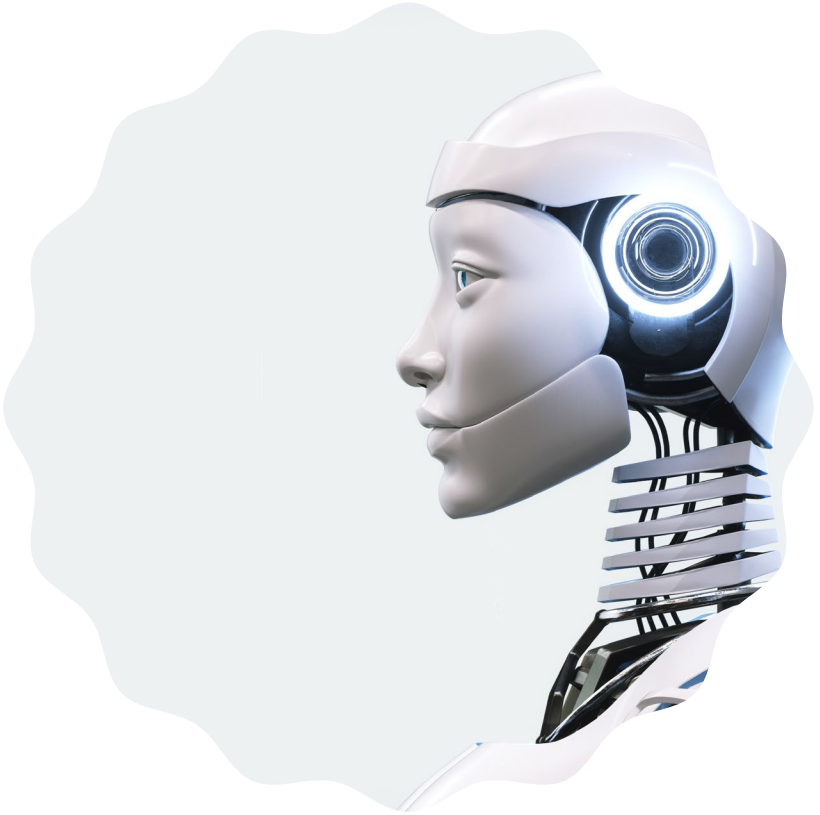
- Deep Learning has enabled tremendous advances in many application areas, including computer vision, image processing, and natural language processing.

- **Key drivers for these successes:**

  1. Large available datasets for training
  2. Large and increasing model sizes
  3. Accelerated computing platforms such as GPUs

- With Machine Learning, rather than implementing hand-tuned algorithms and heuristics for each application area, researchers train models using general-purpose techniques to learn application-specific functions directly from millions of examples

- The goal is to deploy a trained model in a production flow to make accurate inferences on previously unseen data



MACHINE LEARNING OPPORTUNITY

LEARNING TECHNIQUES

- Supervised learning is commonly used in computer vision and other fields
- In EDA, Supervised Learning has often been proposed as a fast approximation for complicated, time-consuming tools
- Unsupervised Learning
- Bayesian Optimization
- Deep Reinforcement Learning (DRL)

MACHINE LEARNING OPPORTUNITY

MODELS

- Conventional Machine Learning Models
- Complex Deep Learning models based on Artificial Neural Networks
  - Multilayer Perceptrons (MLPs)
  - Convolutional Neural Networks (CNNs)
  - Graph Neural Networks (GNNs)

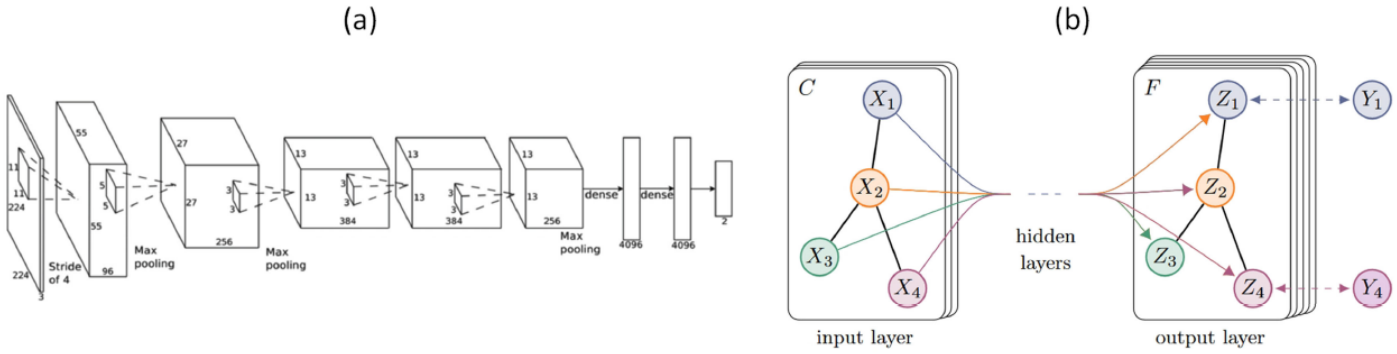
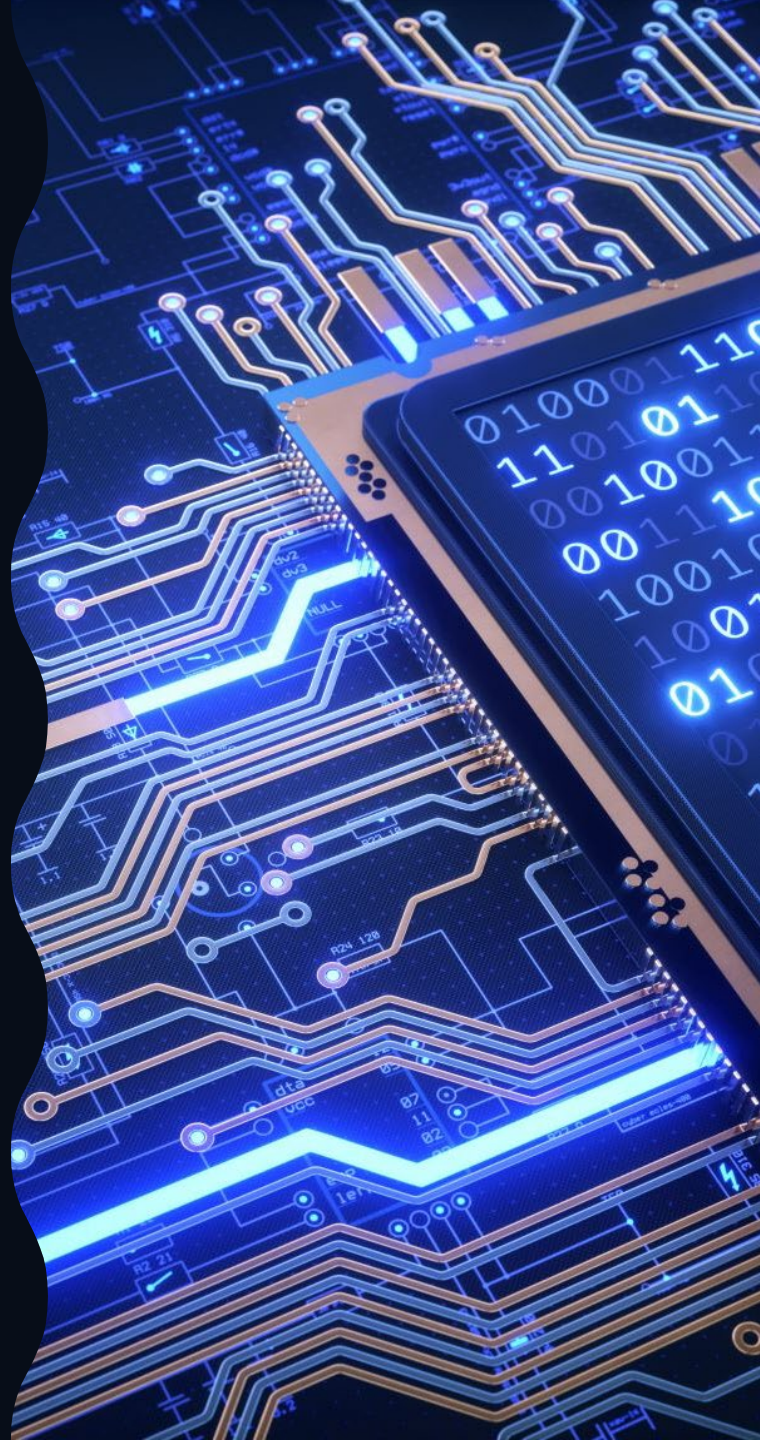


Figure 1. Neural network models. (a) CNN computation in AlexNet.[5] (b) Graph convolutional networks (GCNs).[6]



SECTION III
ML-BASED PREDICTORS FOR THE CHIP DESIGN

ML-BASED PREDICTORS FOR THE CHIP DESIGN

A. TYPICAL CHIP DESIGN

A typical chip design spec-to-layout flow is an iterative process. Large, complex SoCs are split into dozens of units. Each unit proceeds through the following steps,

1. Microarchitecture
2. RTL design and verification
3. Logic synthesis,
4. Timing analysis
5. Power analysis
6. Floor planning
7. Placement
8. Clock tree synthesis (CTS)
9. Routing and final sign off

ML-BASED PREDICTORS FOR THE CHIP DESIGN

B. Trained Machine Learning Models

- Trained machine learning models can help speed time to tapeout by predicting downstream results in a chip design flow, which can reduce time per iteration or improve the quality of results (QoR) such as performance, power, or area
- Rather than waiting hours or days for exact results, predictions can be provided in seconds
- Four ways predictions can be used are:

1. Microarchitectural design space exploration
2. Power analysis
3. VLSI physical design
4. Analog design

ML-BASED PREDICTORS FOR THE CHIP DESIGN

C. MICROARCHITECTURAL DESIGN SPACE EXPLORATION

- Parametrized RTL
- Development of MAGNet
- Machine learning Predictors (MLP)
- Models R^2 accuracy on layers from real image classification CNNs

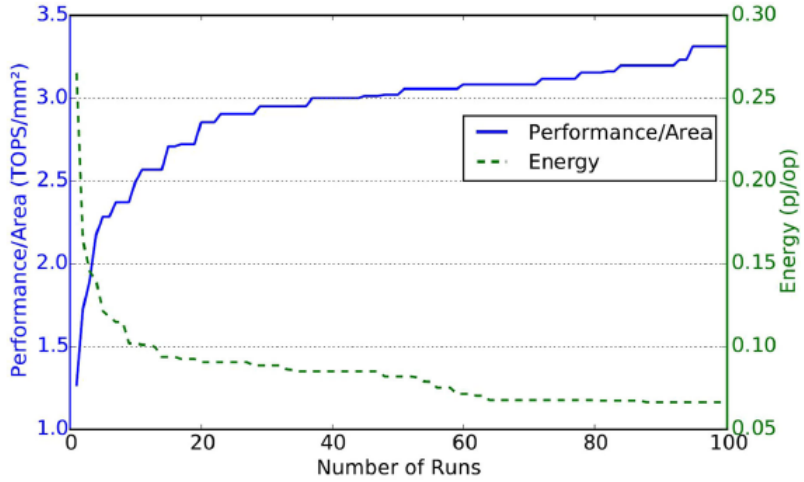


Figure 2. MAGNet tuning via Bayesian optimization.[7]

ML-BASED PREDICTORS FOR THE CHIP DESIGN

D. POWER AND IR DROP ANALYSIS

- Pre-tapeout power estimation and optimization is a critical aspect of all chip design flows today
- Accurate analysis requires running logic simulations on gate-level netlists with annotated capacitive parasitics
- However, these simulations are very slow, typically running 10-1000 clock cycles per second, depending on activity factor and design size
- To provide faster yet accurate power estimates, Machine Learning models can predict dynamic power by estimating the propagation of switching activity factors through a logic netlist without simulation (PRIMAL, GRANNITE)
- Once design proceed to physical design, IR drop analysis becomes a critical part of power signoff (PowerNet)
    1. Excessive IR drop prevents circuits from running at targeted speeds
    2. Accurate analysis requires solving large systems of linear equations to obtain the voltage of every node in a circuit, which can take days on large designs
    3. Multiple iterations of analysis and mitigation are desirable during a VLSI flow, but conventional analysis is too low to be used in this way

ML-BASED PREDICTORS FOR THE CHIP DESIGN

E. PHYSICAL DESIGN

- Detailed routing is perhaps the most time-consuming stage of a modern physical design flow. It is difficult to determine exactly whether a synthesized or placed design will be DRC violation free while meeting timing constraints after performing detailed routing

- Image-based DL models provide an excellent opportunity to predict routability compared to prior heuristic-based approaches

1. Fully Convolutional Network (FCN)

I. **RouteNet :** leverages a fully convolutional network (FCN) to predict postdetailed-route DRCs from post-placement global routing results

II. **CongestionNet:** Routability problems can be identified even earlier in the flow, after logic synthesis. CongestionNet leverages a graph attention network to estimate routing congestion based only on the circuit graph.

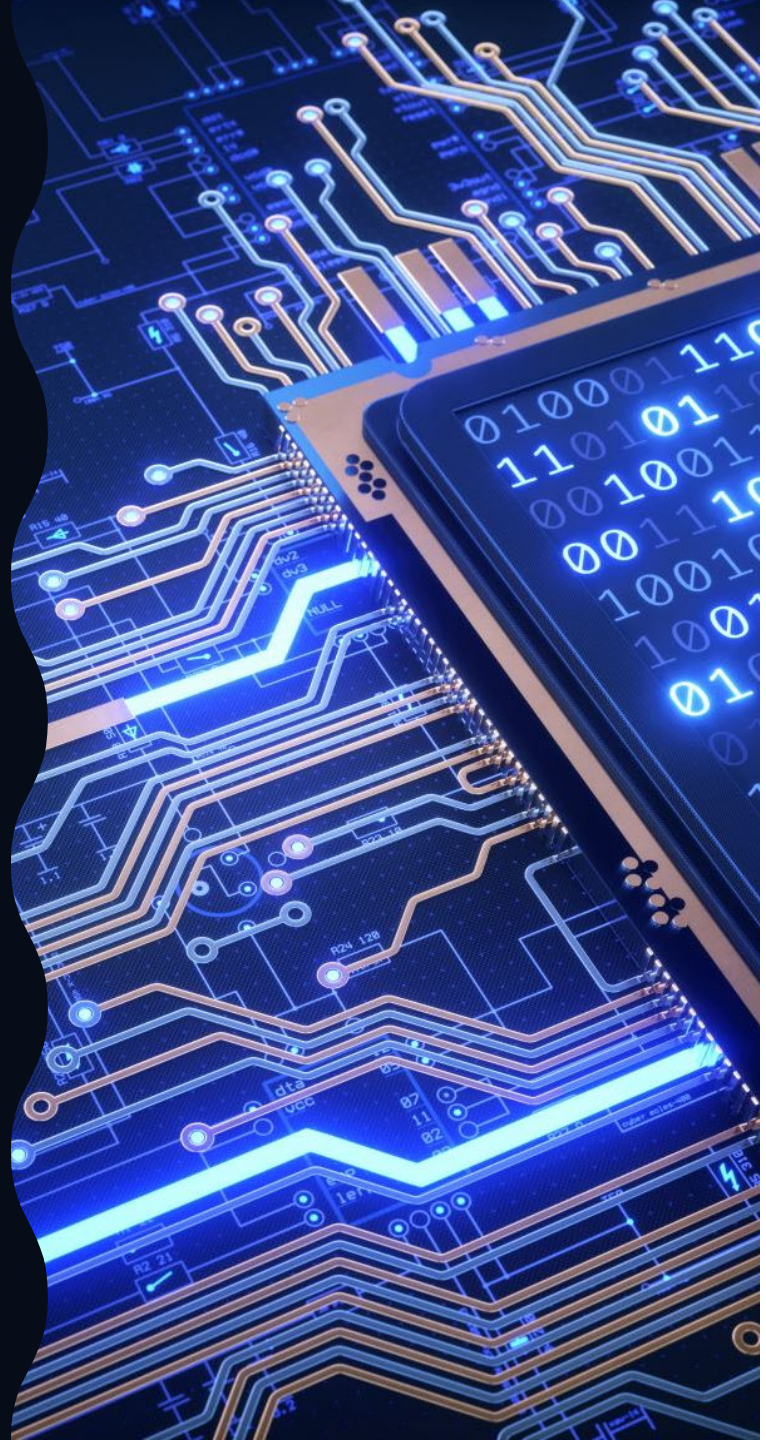
ML-BASED PREDICTORS FOR THE CHIP DESIGN

F. ANALOG LAYOUT

- Analog design, and especially analog layout, is a labor intensive process still lacking high-quality design automation tools
- One difficulty is the larger degree of freedom compared to digital design, which is constrained to standard cell rows. Furthermore, analog circuit QoR metrics are design dependent, whereas digital QoR metrics such as timing, area, and power are universal
- The advancement of AI provides a good opportunity to automate the analog design process.

1. ParaGraph

1. One recent example is ParaGraph, a GNN that predicts layout parasitics and devise parameters directly from circuit schematics
2. Postlayout parasitic prediction is key to automating analog layout generation since it can help with schematic and layout convergence, floorplan feasibility, or QoR estimates



SECTION IV
AI-ASSISTED CHIP DESIGN

AI-ASSISTED CHIP DESIGN

A. AI-DRIVEN FLOORPLANNING AND PLACEMENT

- Augmenting today's semi-automated VLSI flows with ML-based predictors can provide immediate benets
- Looking to the future, we expect even larger breakthroughs to come from deploying AI to directly optimize chip designs without human intervention and to augment or replace existing best known algorithms in EDA tools
- Several promising research directions in these areas are highlighted
    - **AI-driven Floorplanning and Placement,** enabled by three key technologies
        1. Dee Reinforcement Learning (DRL) for physical constraint optimization
        2. Fast VLSI placers running on accelerated computing platforms such as GPUs (DREAMPlace, ABCDPlace)
        3. DL based deep QoR predictors to classify postplacement results Finally, DRL needs an accurate QoR predictor: fast reward functions to evaluation candidate placements based on predictions of downstream routability and QoR

AI-ASSISTED CHIP DESIGN

A. AI-DRIVEN FLOORPLANNING AND PLACEMENT

- A central challenge in a baseline physical design flow (see Figure 4) is that the backend VLSI tools – placement, Clock Tree Synthesis (CTS), and routing must produce final design-rule and timing clean designs in deeply scaled modern process technologies

- To address this, we envision an AI-driven physical design flow (Figure 4) to intelligently explore the design space of potential physical floorplans, timing and tool constraints, and placements

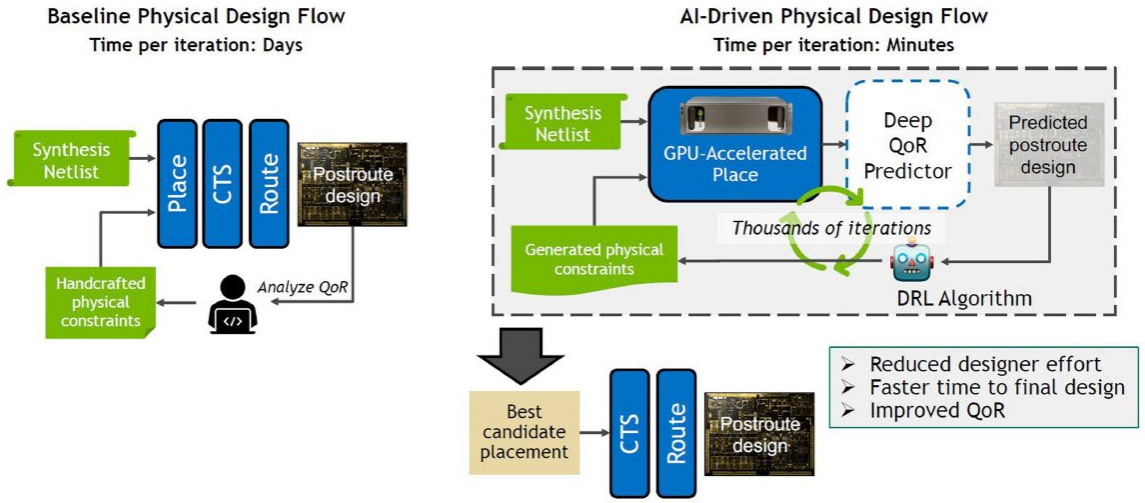
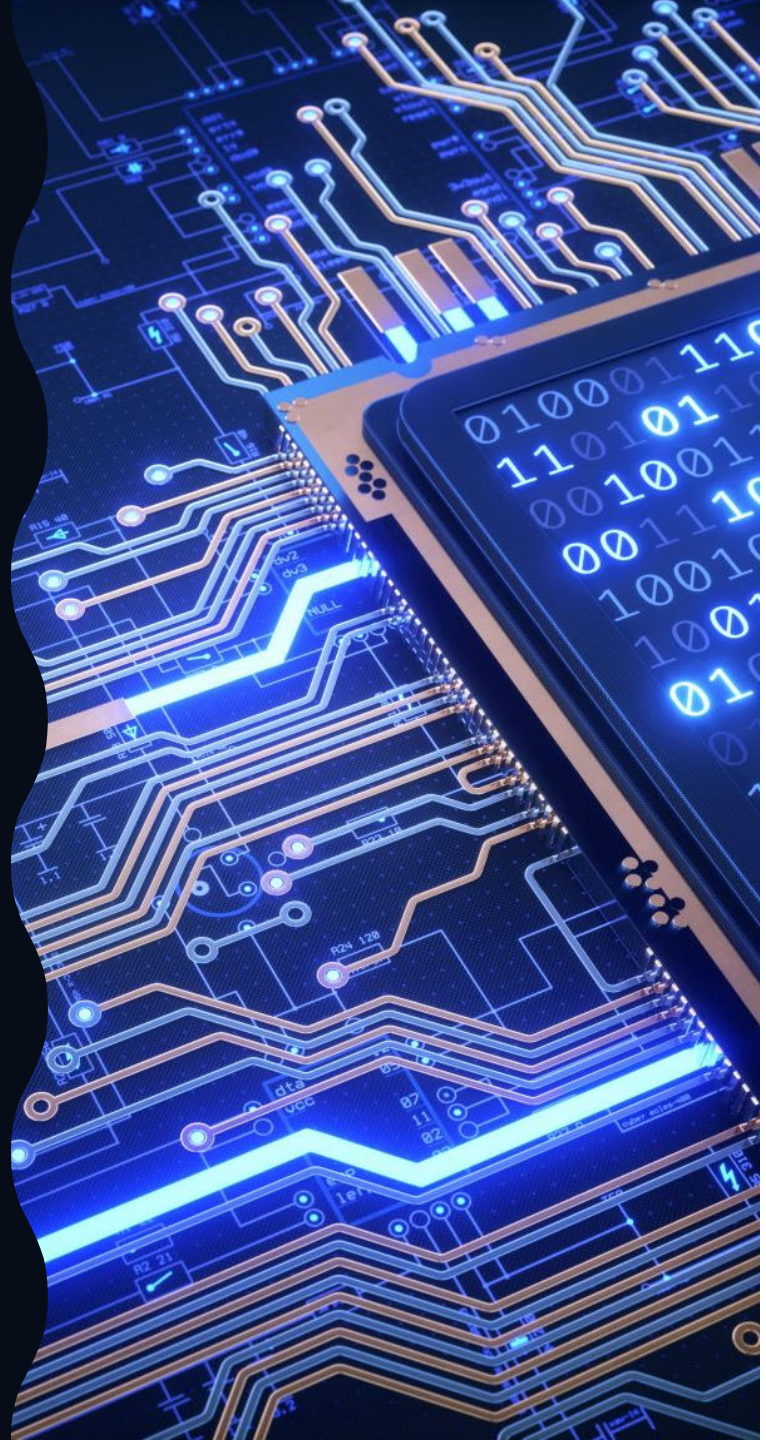


Figure 4. VLSI physical design implementation flows.

AI-ASSISTED CHIP DESIGN

B. LEARNING EDA ALGORITHMS WITH DRL

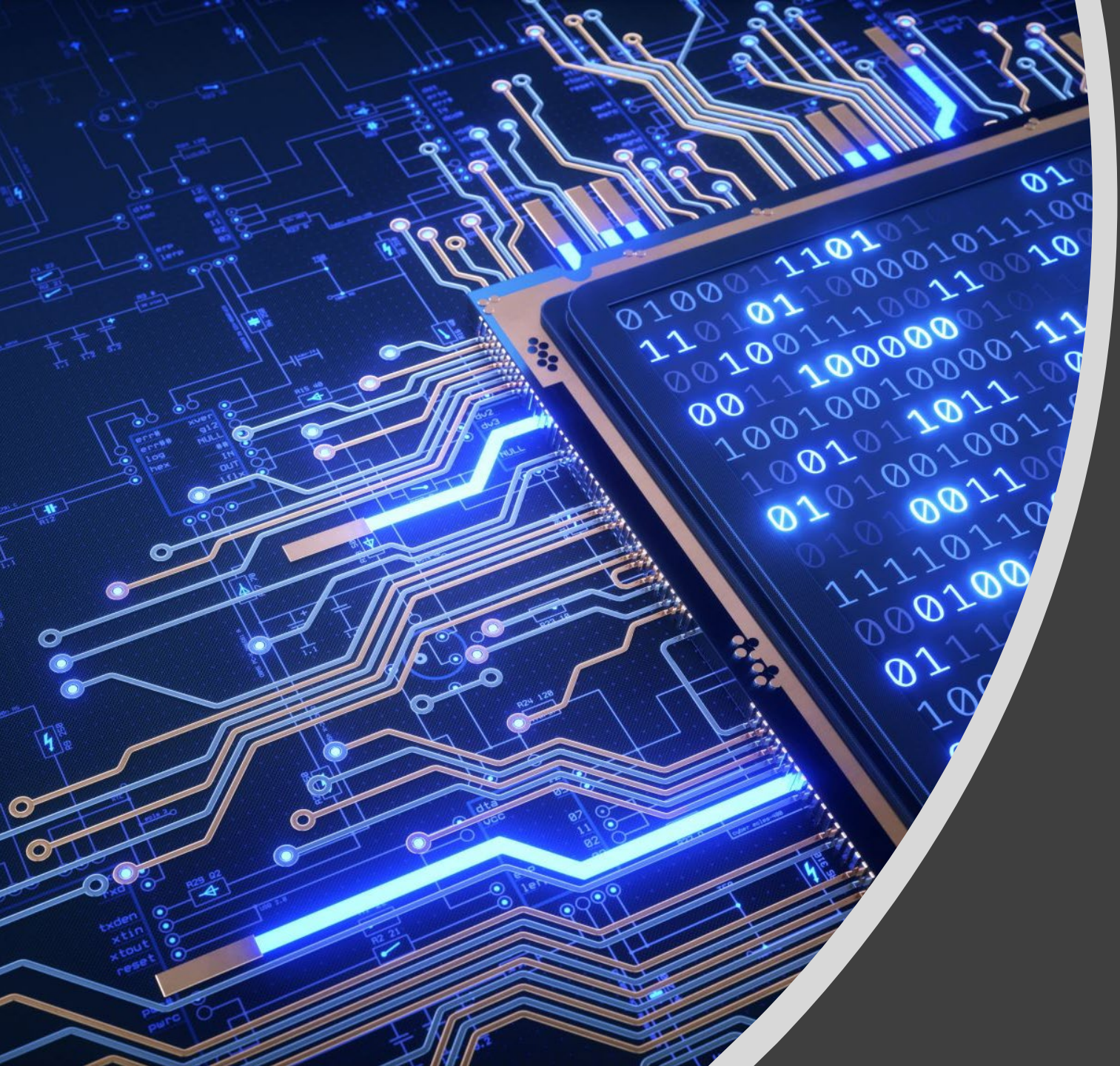
- Deep Reinforcement Learning (DRL) might also be able to reduce runtime for timing optimization in VLSI backend tools
- With today's tools, it can take up to a week to run placement, Clock Tree Synthesis (CTS), and routing on designs with millions of cells
- Much of that runtime is spent on repeatedly and incrementally optimizing the same cells along with costly static timing analysis (STA) updates
- Deep Reinforcement Learning (DRL) might be able to learn optimal policies that avoid repetitive optimizations on similar cells and reduce the total number of STA updates



SECTION V
CONCLUSION

CONCLUSION

- Initial research applying Machine Learning predictors for VLSI has shown the potential of AI in chip design flows
- In the future, we expect Machine Learning based approaches such as Deep Reinforcement Learning (DRL) to be suitable for many EDA optimization problems, especially when modeling the exact objective or constraints is difficult
- Deep Reinforcement Learning (DRL) could become a new general purpose algorithm for EDA, just like simulated annealing, generic algorithms, and linear/nonlinear programming
- Since Deep Learning models are optimized to run efficiently on accelerated computing systems such as GPUs, we expect to see a virtuous cycle of exploiting the world's most powerful computers for designing the next generation of chips, which in turn will improve the performance of future EDA algorithms



THANK YOU