Artificial Intelligence and Machine Learning for the Era of Hyperconnected IoT Devices

Frank Schirrmeister, Sr. Group Director, Solutions & Ecosystem
IEEE IoT @ RWW, January 12th 2022
Agenda

• Hyperconnectivity, Hyperscale Computing & Intelligent Systems
• Why? Show me the Money!
• Engineering Challenges
  o The need for domain specific design
  o Enabling AI/ML chips and systems
  o Using AI/ML to improve development productivity
• Outlook
Who Is This Guy?

Living in
Silicon Valley, CA

Born & raised in
Berlin, Germany

https://www.linkedin.com/in/frankschirrmeister/

HTDV Motion Estimation Chipset

MIDI Software Development

Book Chapter Contributions

Living in Silicon Valley, CA

Born & raised in Berlin, Germany

Who Is This Guy?
Hyperconnectivity?
Hyperscale Computing?
Intelligent Systems?
Intelligent Systems: Ubiquitous Hyperconnectivity

Sensor Edge
Device Edge
Far Edge
Middle Edge
Near Edge
Cloud Data Center

Data
000111 1010 111100000011
001111 1010 011110001000 1110001010 1110001010
011110000011
11 111100000011
011110001000 1110001010
101000011110000
000111 1010 111111001000
011110001000 1110001010 1110001010
11 111100000011
011110001000 1110001010 1110001010
Intelligent Systems: Ubiquitous Hyperconnectivity

Own the Data  Monetize the Data  Serve the Data
Global DataSphere 2025: 175 Zettabyte

Endpoint Data Creation 85% CAGR 2019 to 2025
2.8B 5G Mobile Subscriptions by 2025
164 EB per month, 76% video

Sources: IDC, Seagate, IBS, Ericsson Mobility Report
How much data is in a Zettabyte?
... as much information as there are grains of sand on all the world's beaches ... 

What’s in it for the World?
Show me the Money!
Total World Economy: $87.56 T
GDP by Country 2019

https://howmuch.net/articles/the-world-economy-2019
https://databank.worldbank.org
Total World Economy: $87.56 T
Breakdown to Systems & Semi (2019)

- Global GDP: $87.56 T
  - Services: $56.91 T
  - Industrial & Agriculture: $30.65 T
- Systems & OEM: $2.4 T
- Semiconductor: $429 B
- Technical Software: $51 B
Systems & Semi: A Once-in-a-Lifetime Opportunity!
2025 Estimates

Sources: Omdia, CIMData, Cadence, 2025 estimates
From Building Blocks to Systems
Technical Software & EDA are both enabling and using AI/ML …

Using AI/ML to increase productivity of EDA flows

- IP Selection
  “Reuse the right building blocks”
  Processors, Interfaces, Analog

- HW/SW Verification
  “Is it functionally correct?”
  Hardware/Software, Power, Architecture, Safety, Security

- Chip Implementation
  “Optimized, Advanced Node Implementation”
  Performance, Power, Cost

- Packaging
  “Silicon into packages and parts into Printed Circuit Boards”

- System Analysis
  “Does it work when put together?”
  Electromagnetic, Thermal, Low Power

Optimized flows to enable AI/ML Chips and systems
Design of AI/ML Systems and Chips
AI is Happening Everywhere – Location of Data is Crucial!

**Edge Training**
- Face recognition to unlock phone
- Update with localized data
- ...

**Edge Inferencing**
- Real-time decisions (Vision)
- Phone unlock
- ...

**Datacenter Training**
- High complexity NNs for ADAS
- Analysis to predict our behavior
- ...

**Datacenter Inferencing**
- More complex data - higher latency
- Train and deploy inferenced NN
- ...

Latencies:
- <1ms
- <20ms
- <40ms
- ~100ms

- Google Coral
- Intel Up Squared
- Google Cloud TPUs
- Nvidia GPUs
- Intel Nervana NNP-T
A Complex Engineering Challenge

Application Domains
- HyperScale Computing
- Aero/Defense
- Automotive
- Comms
- Mobile
- Consumer
- Industrial
- Health

Compute Domains
- Sensor Edge
- Device Edge
- Far Edge
- Middle Edge
- Near Edge
- Data Center

Requirements
- Latency
- Power/Thermal
- Performance
- Scale
- Cost
Example: Where to plan to do the compute

- **Sensor Edge**: Limited local compute close to sensor, always connected to Device Edge.
- **Device Edge**: Display close to sensor, some local compute.
- **Far Edge**: Detailed analysis in cloud data center, visualization on Device Edge.
- **Middle Edge**: Detailed sleep analysis calculated at the Device Edge.
- **Near Edge**: Limited local compute close to sensor, always connected to Device Edge.
- **Cloud Datacenter**: General overview.

Lively Ecosystems for Hardware and Software
Domain Specific Design is Key!

Source: mattturck.com/data2021
Technical Software & EDA Enables the Era of Hyperconnectivity

IP Selection
“Reuse the right building blocks”
DSPs, Interfaces, Analog

HW/SW Verification
“Is it functionally correct?”
Hardware/Software, Power, Architecture, Safety, Security

Chip Implementation
“Optimized, advanced-node implementation”
Performance, Power, Cost

Packaging
PCB Integration
“Silicon into packages and parts into PCBs and 3D-IC heterogeneous integration”

System Analysis
“Does it work when put together?”
Electromagnetic, Thermal, Low Power, Computational Fluid Dynamics

Optimized flows to enable chips and systems (for users)
Datacenter/Edge - Design and Verification Challenges

• Diverse requirements
  o Training – High throughput, big designs
  o Inference – Flexibility, lowest power

• Verification challenges
  o Significant software content
  o Big designs – Emulation throughput, debug
  o Physical and virtual interfaces, virtualization

• Physical design challenges
  o Complex SystemVerilog descriptions
  o 1st floorplan uncertainty
  o Advanced node foundry limits and closure

• System design challenges
  o Mix of older and advanced nodes
  o New system design innovations with 2.5D/3D

Synergy of tools and IP offerings
Example: Tensilica® AI on Protium™ Platform
NNA Processor IP
Comprehensive IP platform for edge to on-device AI

AI Max
Multi-Core System (Security, Memory)
- AI Engines
- AI ISA Extensions
- Tensilica® Scalar/Vector Processor

AI Boost
- AI Engines
- AI ISA Extensions
- Tensilica DSP

AI Base
- AI ISA Extensions
- Tensilica DSP

Popular Open-Source AI Framework
- TensorFlow
- PyTorch
- Caffe2
- TensorFlowLite
- ONNX
- mxnet

Common Tensilica AI Software Platform

*PyTorch, Caffe2, the PyTorch logo, and the Caffe2 logo, are any related marks are trademarks of Facebook, Inc.*
*TensorFlow, TensorFlowLite, the TensorFlow logo and TensorFlowLite logo, and any related marks are trademarks of Google, Inc.*
*ONNX, OpenInference, and the OpenInference logo are open standards software.*
*Copyright © 2017–2019, The Apache Software Foundation. Apache, MXNet, Apache, the Apache feather, and the Apache MXNet project logo are either registered trademarks or trademarks of the Apache Software Foundation.*
Example: AI Max – NNA 110 Single Core

- Single-core neural network accelerator
  - 0.5 to 4 TOPS
  - optimized for ML inference applications

- Scalable, Configurable

- Neural Network Compiler (XNNC) Toolchain

- Mixed-Precision Support in HW & SW

- True Sparse Compute Engine

- Tensor Compression

- Optimized TOPS, TOPS/Watt, TOPS/mm2
High-Level Synthesis for AI/ML on the Edge

Key challenges when designing IP for machine learning

Explore multiple ML algorithms

Ability to quantify PPA+
Accuracy trade-offs

Rapid design of production silicon

Stratus™ High-Level Synthesis Solution

- Integrates with ML modeling environments
- Turnkey PPA analysis via logic synthesis for multiple implementations
- Fast path from C++ algorithm to high-quality implementation

SystemC for System Modeling

- Build a full System Model in SystemC
  - SystemC fixed-point types give us exact HW behavior
  - Drive performance modeling and architecture at the same time
- Take advantage of Object-oriented programming to build up system
- Unified testbenches can drive

High-Level Synthesis (HLS) - Tight Inner Loop

- You need SPEED, AREA, and POWER trade-offs to make decisions
- Stratus drives an integrated environment to run experiments
- Joules for synthesized power estimates driven from real sims

Dave Garrett, VP Hardware, Syntiant
DAC 2018
Design Examples

**Habana Labs**
Billion-gate class inference processor: Goya

**Many IPs**
- Standard I/O: Wi-Fi, USB, PCI Express®, etc.
- System infrastructure: Interconnect, interrupt control, UART, timers…
- Differentiators: custom accelerators, modem…

**Many cores**
- Both symmetric and asymmetric
- Both homogeneous and heterogeneous

**Lots of software**
- Part of core functionality: Communication stack, DSP software, GPU microcode…
- User application software infrastructure: Android, Linux…

**Fujitsu**
Billion-gate class HPC/AI Processor: A64FX

Source: Hotchip Conference
Addressing AI/ML Verification Challenges

**Formal**
- Best Capacity
- Regression improvements
- SAT solver inferencing

**Simulation**
- Fast simulation for high-activity designs
- UVM randomization
- Fast elaboration for replicated structures
- Coverage metrics

**Emulation**
- Billion-gate designs
- Parallel partition compiler
- INT8 to INT64
- Power/Performance
- Memory models: HBMx
- Sensor model: MIPI CSI

**Prototyping**
- Scaling to large designs
- Unified frontend
- ICE test suite: Faster, data-extended (DL) regression
- SW driven validation – deep learning data training refinement
INT64 to INT8 Computational Efficiency

Fujitsu
Billion-gate class HPC/AI Processor: A64FX

- For certain ML applications, accuracy may be traded off for faster and more power efficient implementation methods.
- Depending on the target applications, design may scale down (INT8) or scale-across (INT 8 to INT 64) architecturally to ensure computational efficiency.

Source: HotChips Conference
Technology Optimized for Implementation of AI Silicon 1/2

Genus-Innovus iSpatial flow
- Better predictability for RTL designers
- Better full flow PPA for power reduction of key repeated hierarchies

Stratus HLS
- Integrates with ML modeling environments
- PPA analysis for multiple implementations
- Fast path from C++ to high-quality implementation

Genus with new Compus technology
- Analytical ROI tradeoff for datapaths
- Modern SystemVerilog for AI designs
- Area and Power improvement focus

Genus Synthesis
Joules RTL Power
SystemC/C++
Techlib

Constraints
RTL
HLS Reports
Timing/Area Reports
Power Reports
Gates

Elaborated Design
Advanced Structuring Optimization
Control DataFlow Transforms
RTL and Datapath Transforms

CPU1
CPU2
CPU<n>
Analytical RoI costing engine

Correlation with Signoff
Incr GigaOpt CCOpt NanoRoute

Correlation with Physical

Innovus
Tempus
Volitus
Pegasus

RTL Compile/Map
iSpatial
Innovus Mixed Placer

- Gigaplace™ - XL
- Concurrent Macro and Std Cell placement

- Significant wirelength reduction
- Total power reduction
- Huge productivity gain with design convergence

Innovus Rapid FP Implementation

- Quick map with physical sizes
- Rapid exploration
- PPA estimation

- Fast block netlist modeling without real netlist
- Automatic floorplan synthesis and hierarchical partition generation

Tempus Distributed STA

- Interface
- CPU1
- CPU2
- Bus
- Memory

- For performance challenges of large AI chips
- Server Farm with distributed STA
- Distributes workload across smaller memory footprint machines
Optimize based on critical workloads

System power exploration

- Analyze and explore software/hardware changes impact on power
- Genus™/Joules™/Palladium® integration

<table>
<thead>
<tr>
<th>10M Cycles Workload</th>
<th>TAT to Power Feedback (Palladium + Joules)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SW/workload changes</td>
<td>~1 hour</td>
</tr>
<tr>
<td>RTL/HW change</td>
<td>~Overnight regressions</td>
</tr>
</tbody>
</table>

RTL power estimations within 11% of signoff

Activity-driven optimization

- Joules replay – Cycle-based accuracy from RTL vectors
- Integrated into Genus/Innovus™ optimization

<table>
<thead>
<tr>
<th></th>
<th>w/o Replay</th>
<th>Replay Flow</th>
<th>Diff %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leakage</td>
<td>56.8</td>
<td>58.3</td>
<td>2.6%</td>
</tr>
<tr>
<td>Internal</td>
<td>138.5</td>
<td>132.3</td>
<td>-4.5%</td>
</tr>
<tr>
<td>Switching</td>
<td>86.1</td>
<td>66.9</td>
<td>-22.3%</td>
</tr>
<tr>
<td>Total Power</td>
<td>281.3</td>
<td>257.6</td>
<td>-8.4%</td>
</tr>
</tbody>
</table>

Up to 10% total power reduction
Moore’s Law Slowing and Die Size Reaching Lithography’s Reticle Limit

Moore’s Law Slowdown

Die Size Trend
Die Size Increases Over Time in Server CPUs and GPUs

Advanced Nodes Reaching Physical Transistor Size Limit
Die Sizes Increasing at an Unsustainable Rate

Source: Delivering the Future of High-Performance Computing, AMD keynote, Dr. Lisa Su, Hotchips, August 2019
Solution: Go to the Third Dimension

- Shorter Wire
- Less Power
- Higher Performance
- Higher Bandwidth
- Smaller Profile
- Better Yield

2D SoC

Long global wire
Shorter wire

Replaced by

3D-IC
Top Considerations for a Full 3D-IC Design Flow

- System Planning & Partitioning
  - Implementation and System PPA
  - Thermal and Power (Early Analysis and Signoff)
  - Integration (Analog and Package co-design)
  - Multi-Die Timing Signoff
  - Multi-Die Physical Verification
  - Testing for 3D & 2.5D Die Stacks (IEEE 1838)

- Package
  - Design & Verification

- Multi-Chips
  - Design & Verification

- System Verification
Integrity 3D-IC Platform

3D design planning, implementation and system analysis in a single, unified cockpit

Flow Manager

3D Analysis and Signoff
- Thermal
  - Thermal System Analysis
  - Voltus™/Clarity™/Celsius™
  - EM/IR-3DEM/SiPI/Thermal
  - Tempus™
  - Signoff STA
  - Pegasus™
  - Physical Verification
- Power
- Timing
- DRC/LVS

Co-design

Virtuoso® Platform
- Virtuoso Schematic Editor
- Virtuoso Layout Suite
- Analog Design Environment

Analog/Custom IC Design

Allegro® Technologies
- BGA Package Layout
AI/ML and EDA?
AI/ML is Enabling EDA!

Using AI/ML to increase productivity of EDA flows

Optimized flows to enable AI/ML chips and systems

IP Selection
“Reuse the right building blocks”
DSPs, Interfaces, Analog

HW/SW Verification
“Is it functionally correct?”
Hardware/Software, Power,
Safety, Security

Chip Implementation
“Optimized, advanced-node implementation”
Performance, Power, Cost

Packaging
“Silicon into packages and parts into PCBs and 3D-IC heterogeneous integration”

System Analysis
“Does it work when put together?”
Electromagnetic, Thermal, Low Power,
Computational Fluid Dynamics
EDA Keeps Design Cost in Check

1970s

1980s

1990s

2000s

2010s

Platform-Based Design

Multicore

Automation (EDA)
Technology Changes: Processor Data

Hennessy and Patterson, Turing Lecture 2018, overlaid over “42 Years of Processors Data”
https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp
Next in EDA: “More than Moore”
EDA Keeps Design Cost in Check – Commonalities

1970s
1980s
1990s
2000s
2010s

Platform-Based Design

Multicore

Automation (EDA)

Model

Capture

Analyze, Optimize

Abstract

Simulate

Automate
Automating “High Effort” Aspects of the Design Flows

Effort for Advanced Designs

- Prototype Validation
- Prototype
- Software
- Physical
- Verification
- Architecture
- IP Qualification

Source Data: IBS July 2020, Cadence

Model

Abstract

Capture

Simulate

Analyze, Optimize

Automate

Board and Package

Systems

Implementation

Verification

IP and Subsystems
So You Think Chess & Go Were Complex?

ML in Electronic Design Automation

EDA Tool

Input: Design Data

Output: Design Objectives

• Options
• Preferences
• Constraints
• Commands

ML Inside

Hyperparameters
• Learning Rate
• Regularization

ML Outside

Trained Parameters

Algorithm

Model

Output: Design Objectives
## AI/ML Opportunities in Technical Software & EDA

<table>
<thead>
<tr>
<th>Category</th>
<th>Opportunities</th>
</tr>
</thead>
</table>
| **Functional Verification**     | • **Simulation** smarter and faster regressions  
                                | • **Formal verification** formal proof orchestration                         |
| **Digital Implementation**      | • Digital implementation “ML inside” delay prediction  
                                | • Digital implementation “ML outside” flow optimization                      |
| **Library Characterization**    | • **Library** characterization with ML-based prediction                       |
| **Custom IC Implementation**    | • **SPICE simulation** ML for accurate response surface models  
                                | • **Custom implementation** Layout and analog placement                       |
| **Design for Manufacturing**    | • **ML-DFM** Predicting unknown yield limiting hotspots                       |
| **PCB Synthesis**               | • **PCB design** placement, via, routing, power delivery, analysis            |
| **System Design and Analysis**  | • **System analysis “Smart Sweep”** Predict eye openings                      |
Digital Implementation
Implementation as Part of Chip Development Efforts

Percentage of Project Effort

- Specification Development
- RTL Development
- RTL Verification
- Qualification of IP
- Design Management
- OS Support
- Utility Software Development
- Application Software Development

Elapsed time as percentage of time from RTL development to tapeout

Source Data: IBS, Cadence
ML for Better Performance and Productivity

**ML Delay Prediction**
- Innovus™ pre-route delay engine
- Better PPA
- Faster, more predictable results
- ML optimizes delay prediction
- All optimization done on customer data, at customer site

**Intelligent Chip Explorer**
- Automated smart design flows
- Productivity improvements
- ML optimizes flow (SI congestions)
- Adjusting tool and library options, constraints, parallel runs in cloud
- All optimization done on customer data, at customer site
ML Inside: Enhanced Delay Prediction

“Learns” to predict actual results from model inputs

<table>
<thead>
<tr>
<th>Node</th>
<th>WNS</th>
<th>TNS</th>
<th>Pwr</th>
</tr>
</thead>
<tbody>
<tr>
<td>8nm</td>
<td>23%</td>
<td>67%</td>
<td>3.6%</td>
</tr>
<tr>
<td>7nm</td>
<td>5%</td>
<td>77%</td>
<td>3.4%</td>
</tr>
<tr>
<td>5nm</td>
<td>25%</td>
<td>49%</td>
<td>1.1%</td>
</tr>
<tr>
<td>12nm</td>
<td>38%</td>
<td>22%</td>
<td>1%</td>
</tr>
<tr>
<td>5nm</td>
<td>18%</td>
<td>51%</td>
<td>2.5%</td>
</tr>
</tbody>
</table>

Improved Design Performance

Initial Design

Training Data Extraction

Training

Model

Prediction Accuracy

Actual

Machine Learning Training

Innovus™

Machine Learning Inference Model

More Accurate Model
ML Outside Objective: Improve the PPA/Productivity Curve

Cerebrus™

PPA

Engineering Effort

PPA

20%

Improve PPA more quickly

10X

Use CPU time instead of engineering time

Manual Flow Development

Productivity
Cerebrus: ML for Better PPA and Full Flow Productivity

Cerebrus™
Flow

- Design Input (Spec)
- Genus™ RTL Synthesis
- Innovus™ Implementation System
- Tempus™ Signoff STA

Knowledge Graph
Reinforcement Learning

Machine Learning-Driven Flow Automation

Result
Optimized PPA

Cerebrus Improvements vs Baseline

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Improvement</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>420MHz</td>
<td>14%</td>
</tr>
<tr>
<td>Leakage power</td>
<td>26mW</td>
<td>7%</td>
</tr>
<tr>
<td>Total power</td>
<td>62mW</td>
<td>3%</td>
</tr>
<tr>
<td>Density</td>
<td></td>
<td>5%</td>
</tr>
</tbody>
</table>

Cerebrus automatically improved PPA of 5nm mobile CPU, using 30 parallel jobs
Within 10 days, converged on improved flow

Manual Flow Development

Start
1 Engineer
10 days

Many Engineers
Many Months

CPU

<table>
<thead>
<tr>
<th>Process</th>
<th>5nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>3.5GHz</td>
</tr>
</tbody>
</table>

© 2022 Cadence Design Systems, Inc. All rights reserved.
Customer wanted to achieve 2GHz on latest CPU implementation
Cerebrus optimized floorplan and implementation flow concurrently

• Floorplan can be automatically resized in any direction
• Innovus™ mixed placer used to find optimal macro location in resized floorplan

Cerebrus Improvements vs Baseline

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>+200MHz</td>
</tr>
<tr>
<td>Total failing timing</td>
<td>83%</td>
</tr>
<tr>
<td>Leakage power</td>
<td>17%</td>
</tr>
</tbody>
</table>
To efficiently maximize the performance of new products that use emerging process nodes, digital implementation flows used by our engineering team need to be continuously updated. **Automated design flow optimization is critical for realizing product development at a much higher throughput.** Cerebrus, with its innovative ML capabilities, and the Cadence RTL-to-signoff tools have provided automated flow optimization and floorplan exploration, improving design performance by more than 10%. Following this success, the new approach will be adopted in the development of our latest design projects.

*Satoshi Shibatani, director, Digital Design Technology Department, Shared R&D EDA Division, Renesas*
Functional Verification
Verification as Part of Chip Development Efforts

Percentage of Project Effort

- Specification Development
- RTL Development
- RTL Verification
- Qualification of IP
- Design Management
- Implementation to Tapeout
- Utility Software Development
- Application Software Development

Elapsed time as percentage of time from RTL development to tapeout

Source Data: IBS, Cadence
ML in Functional Verification

Baseline

CPU1

T1.1 | T1.2 | T1.3 | T1.4 | ...

CPU2

T2.1 | T2.2 | ...

CPUm

Tm.1 | ...

...

...

Time

Regression finished

Machine Learning

Regression finished faster and with same coverage

Test run from regression specification generated from ML model

Machine Learning
ML in Functional Verification

Accelerates verification throughput by reducing simulation cycles with matching coverage on randomized test suites.
Xcelium-ML Up to 5x efficiency at same coverage

Regression CPU Cycles

Coverage

Example 1
3X
99%

Example 2
4X
99%

Example 3
4.5X
99%
Kioxia has effectively utilized Xcelium simulation for a variety of our designs, and it addresses our ever-growing verification needs. With the new Xcelium ML, we’ve seen a **4X shorter turnaround time** in our fully random regression runs **to reach 99% function coverage of original**, and plan to use this technology in production designs to shorten the time to market for Kioxia’s business.”

*Kazunari Horikawa*

Senior Manager, Design Technology Innovation Division

Kioxia Corporation
Smart Proof in Formal Verification
Optimize Resources, Reproduce Previous Runs

Reinforcement learning dynamically adjusts engine selection, time limits, parallel threads, etc.
Smart Proof in Formal Verification
Computational logistics technology for formal throughput

Inside Cadence® R&D

Proof Algorithm 1 ... Proof Algorithm N
Parameter 1 ... Parameter N

Training set of proof problems

Knowledge Graph and Machine Learning

Smart Proof Solver

Released Product

Proof Apps

Smart Proof Solver

Design

Testcase | Baseline | Smart Proof | Gain
-------|----------|-------------|-----
A      | 50%      | 59%         | 1.2X
B      | 69%      | 69%         | 1.0X
C      | 12%      | 25%         | 2.1X
D      | 44%      | 83%         | 1.9X
E      | 57%      | 94%         | 1.6X
F      | 68%      | 69%         | 1.0X
Total  | 53%      | 71%         | 1.3X

* Average results from typical regressions; actual results may vary
Smart Proof in Formal Verification

Computational logistics technology for formal throughput

Third-Generation JasperGold® Formal Verification Platform

"We measured 2X faster proofs out of the box, 5X faster regressions, and non-converged properties reduced by 50%”

- Mirella Negro Marcigaglia, digital design verification manager, STMicroelectronics

* Average results from typical regressions; actual results may vary
AI/ML Productivity Improvements - Some Examples

- **Functional Verification**: Up to 5X reduction in simulation cycles (same coverage). Up to 4X (2X avg.) better out-of-the-box proofs.
- **Digital Implementation**: Up 20% better PPA, up to 10X productivity.
- **Library Characterization**: Accelerated library development. Example: 47% of libs interpolated 98%+ pass rate.
- **Custom IC Implementation**: Accurate response surface model of the device or block. Layout group prediction.
- **Design for Manufacturing**: Hotspot prediction. In-design detection and fixing.
- **PCB Synthesis**: Faster design closure. Routability.
- **System Design and Analysis**: Reduction in simulation time.
AI/ML Productivity Improvements - Some Examples

- **Functional Verification**: Up to 5X reduction in simulation cycles (same coverage)
- **Digital Implementation**: Up to 4X (2X avg.) better out-of-the-box proofs
- **Library Characterization**: Up 20% better PPA, up to 10X productivity
- **Custom Implementation**: Accelerated library development
  - Example: 47% of libs interpolated 98%+ pass rate
- **Design for Manufacturing**: Hotspot prediction
  - In-design detection and fixing
- **PCB Synthesis**: Faster design closure
  - Routability
- **System Design and Analysis**: Reduction in simulation time

This is truly just the beginning!
Where is all this going?
Are we there yet?

The future is already here — it's just not very evenly distributed!

"The Science in Science Fiction" on Talk of the Nation, NPR (30 November 1999, Timecode 11:55), William Gibson
Gartner Hype Cycle for Artificial Intelligence, 2021

Source: The 4 Trends That Prevail on the Gartner Hype Cycle for AI, 2021

https://gtnr.it/3eJsdYG
Transformation in Communications

Hyperconnectivity: The Path to 6G

Sources: Ericsson "Mobility Reports" 2015 to 2020, Samsung "The Next Hyperconnected Experience for All", BigStockPhoto
Frank Schirrmeister, Cadence, inspired by an original graph by Prof. Gerhard Fettweis, Barkhausen Institute
* End of arrows indicate peak of smartphone shipments

https://semiengineering.com/hyperconnectivity-and-the-path-to-6g/
Transformation in Datacenters

2008
- Virtual Machines
- CPU centric
- 10 GB/s

2013
- Containerization
- Software Defined Networking
- 10-40 GB/s

2016
- Software Defined Storage
- 25-100 GB/s

2020
- SW Defined Security
- CPU + GPU + Domain Specific
- Smart Storage
- Network Speed
- Latency, Volume
- DSAs/DSLs
- Programmability

Applications
Hypervisor
Compute
Memory
Storage
Network

Building Efficient Data Centers with the Open Compute Project
https://bit.ly/34OkFQT

Facebook announces its open data-center switch
https://bit.ly/34EI5b4

Bloomberg: Cisco Enters Chip Market, Supplying Microsoft, Facebook
https://bloom.bg/2CDHw5u

Amazon Buys Stealthy Israeli Chip Startup Annapurna Labs
https://bit.ly/2EkK2yT

NVIDIA Completes Acquisition of Mellanox
https://bit.ly/34AqX6h
Intelligent Systems: Ubiquitous Hyperconnectivity

Own the Data    Monetize the Data    Serve the Data
New Digital Engineering Challenges Ahead!

- Sustainability
- Verification
- Security
- Safety
- Data Ownership
- Multi-Domain Execution
Electronics Innovation – An Exiting Future Ahead

2020’s
- Edge-to-Cloud and 5G
- Specialized SoCs
- Artificial Intelligence

2020’s
- Parallel and Distributed
- Full-Flow Solutions
- Deep Learning

1960’s
- Moore’s Law
- Semiconductors and Systems

1960’s
- EDA System Design Enablement