How AI and Semiconductors Will Drive Innovation and Productivity
SIA Panel

Stelios Diamantidis, Distinguished Architect
Center for Generative AI
22 June 2023
Techonomic ‘Pull’: Era of Smart Everything

Exponential Growth in Data (ZettaBytes)

2018 Human < Machine

App-centric  Data-centric


Automotive
Home IOT
Industrial

Human

Courtesy: Applied Materials
Models Outpacing Moore’s Law

AI Training Compute CAGR 2012-2021: 694%

Exponential Model Size

- Perceptron
- ALVINN
- NETtalk
- LeNet-5
- BiLSTM for Speech
- ResNets
- T17 Dota 1v1
- AlphaGoZero
- Neural Machine Translation
- Deep Belief Nets and layer-wise pretraining
- RNN for speech
- TD-Gammon v2.1
- FIRST ERA
- MODERN ERA

PETAFOIP/S-DAYS

1e+0 1e+2 1e+4 1e+6 1e+8 1e+10 1e+12 1e+14

1e-2 1e-4 1e-6 1e-8 1e-10 1e-12 1e-14

2-year doubling (Moore’s Law)

1e-0 1e-2 1e-4 1e-6 1e-8 1e-10 1e-12 1e-14

Petaflop/s - Days

© 2022 Synopsys, Inc.

Courtesy: https://openai.com/blog/ai-and-compute/
Generative AI Further Pushing Limits of Compute

All Models Excluding Transformers:  
8X / 2 years

Transformer AI Models:  
275X / 2 years

Source: https://blogs.nvidia.com/blog/2022/03/25/what-is-a-transformer-model/
Accelerated Evolution of Compute

Exponential Chip Complexity

https://en.wikipedia.org/wiki/Transistor_count
Exponential Growth in Energy Requirements

Exponential AI Compute Emissions

- Rail Car of coal
- Yearly home energy
- Barrel of oil
- Gallon of gasoline
- Mile drive
- Phone charge

Model:
- BERT finetune
- BERT LM
- 6B transf
- Dense 121
- Dense 169
- Dense 201
- VIT tiny
- VIT small
- VIT base
- VIT large
- VIT huge
Mandating New Innovations

Global CO₂ Levels (1010-2022)

Most Important
Exponential

Carbon Dioxide (CO₂ ppm)

Source: 2 Degrees Institute
Techonomic ‘Push’ Engines

I/O

Transistors

DENSITY

1990 2000 2010 2020

<10 IO/mm²

<100 IO/mm²

<1,000 IO/mm²

<10,000 IO/mm²

<10,000 - 10M+ IO/mm²

<10pJ/bit

<1.5pJ/bit

<.5pJ/bit

<.1pJ/bit

<.05pJ/bit

10,000 - 10M+

IO/mm²

10μm - 28nm

22nm - 3nm

2nm

14Å

10Å

<10Å

Multi-Die Era

Angstrom Era

SysMoore

PLANAR

FINFET

Nanosheet

CFET
CAD to EDA to EDA.ai

Log Productivity vs Time

- 1970: Circuit Simulation
- 1980: Layout & Digital Simulation
- 1990: Synthesis & Place & Route
- 2000: IP Reuse
- 2010: Fusion
- 2020: Design Space Optimization
- 2030: AI

Technologies:
- QOR
- TTR
- COR

EDA to EDA.ai

Design Space Optimization

- Machine Learning (ML)

Timeline:
- 1970
- 1980
- 1990
- 2000
- 2010
- 2020
- 2030
DSO.ai AI-Grade Productivity

Best known result found by designers

Best DSO result

QOR
9-13% less total power
30% less leakage vs. manual

TTR
2-5X faster convergence
learning design to design

COR
Single engineer
vs. team of experts

Leakage Power vs. Timing (TNS)

9-13% less total power
30% less leakage vs. manual

2-5X faster convergence
learning design to design

Single engineer
vs. team of experts
Synopsys’ AI Journey

2020
Synopsys introduces DSO.ai, world’s first AI application for chip design

DSO.ai named Innovative Product of the Year by EETimes

2021
World productivity record:
10 blocks by a single engineer

Wired: World’s first AI-designed chip

2022
Record adoption:
9 of Semi Top-10
100% Better Results

2023
Industry’s first AI driven EDA suite
Design, test, verification, manufacturing

200+ Commercial Tape-outs with DSO.ai
VSO.ai
Verification Space Optimization

TTR: 3X Fewer Tests

QOR: 100% constrained-random
+8% Extended Coverage

1200+ passes to reach Target

Coverage

Test Seeds/Passes

0 100 200 300 400 500 1200+

33%

VSO.ai
TSO.ai: Test Space Optimization

ATPG Test Pattern Reduction

Results:
43% fewer patterns
Thank You