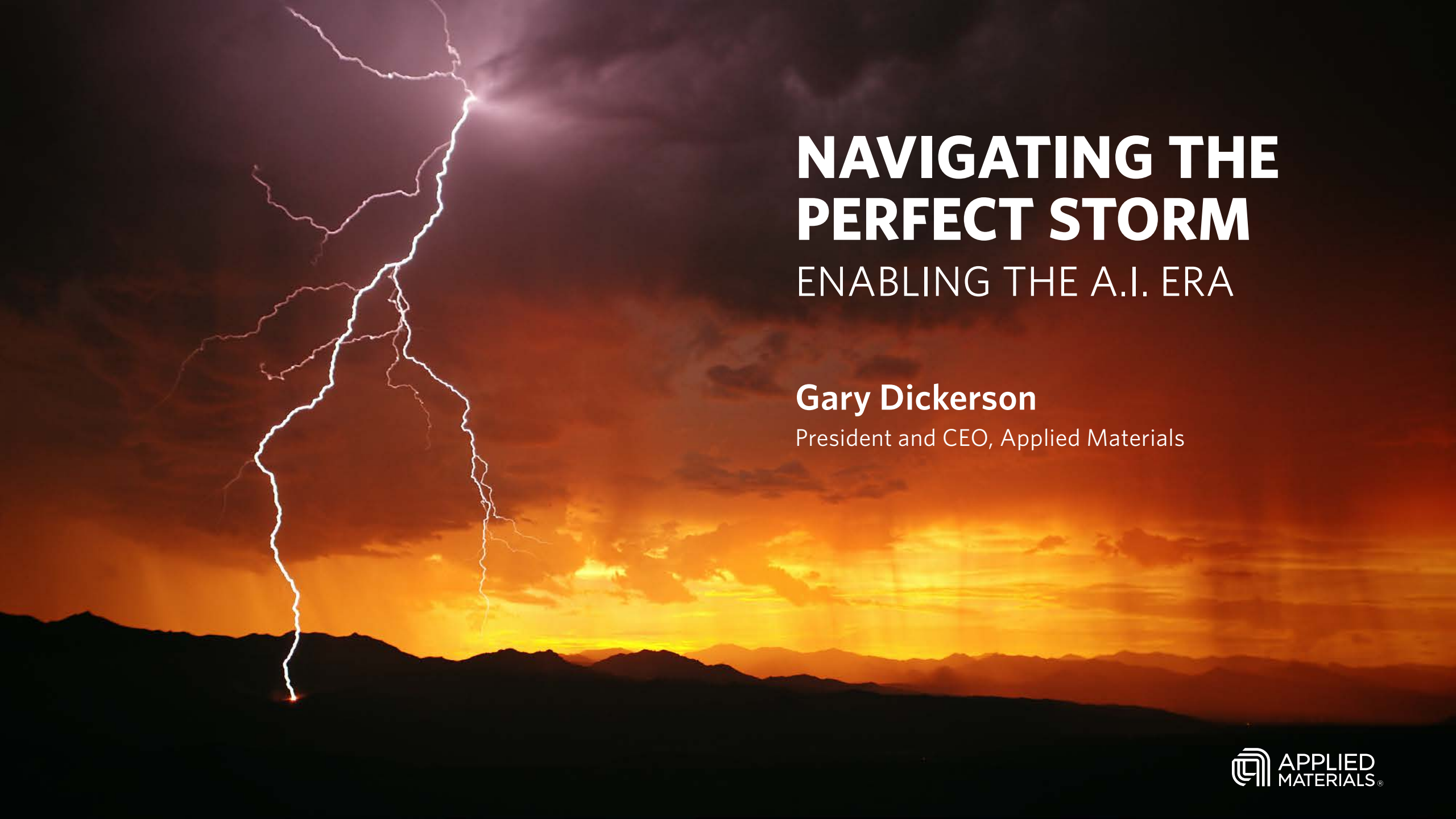


NAVIGATING THE PERFECT STORM

ENABLING THE A.I. ERA

Gary Dickerson
President and CEO, Applied Materials

APPLIED MATERIALS®

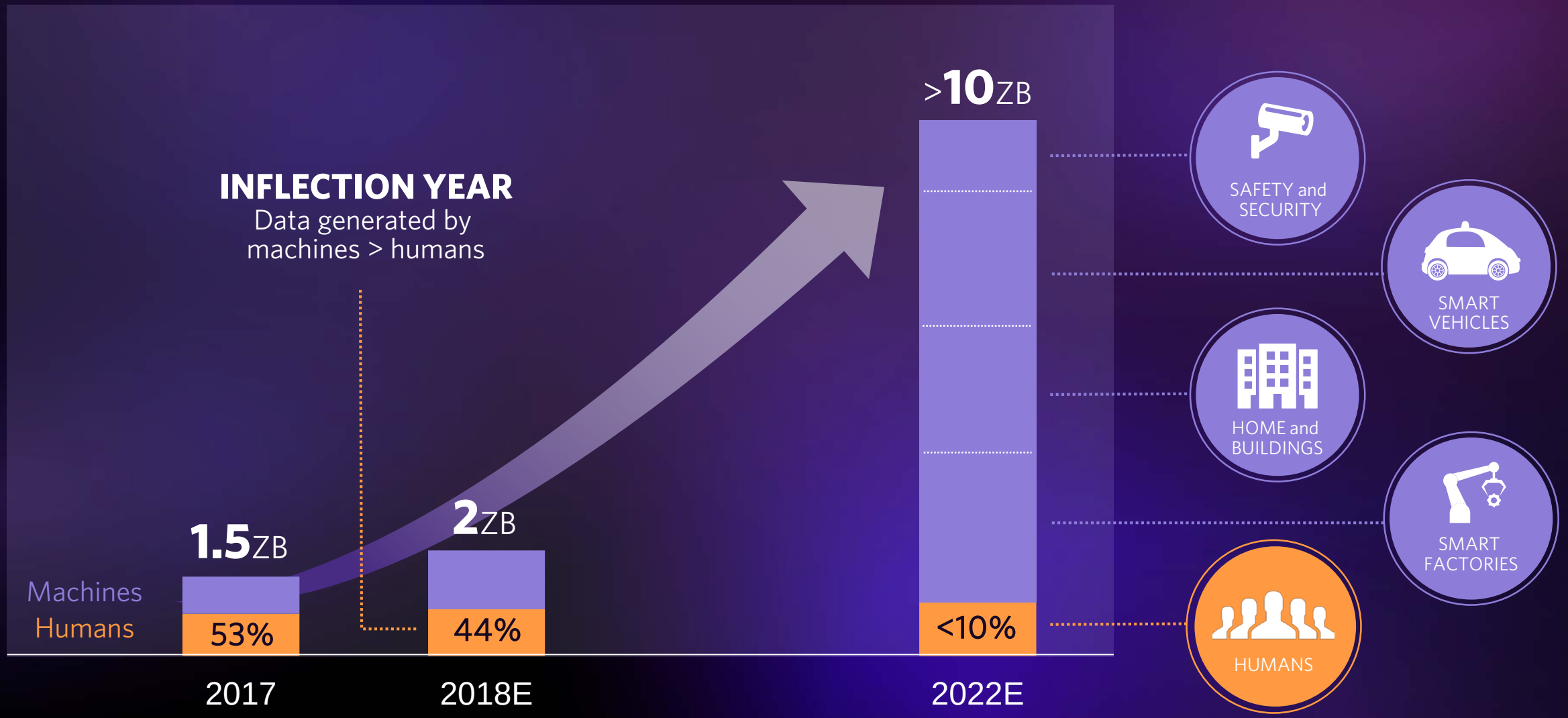
"Data is to this century what oil was to the last one: a driver of growth and change"

– The Economist

CONNECTIVITY

DATA GENERATION	IoT and Industry 4.0 driving an **explosion of data**
DATA STORAGE	More data needs to be processed and stored – **Storage alone is not sufficient or economical**
COMPUTE	**New compute models** to turn data into value **New compute architectures** to process data at edge and in cloud at right performance / watt

Explosion of Data Generation



SOURCE: Applied Materials model based on forecasts published by Cisco, Intel, Western Digital

McCarthy & Minsky's

INITIAL A.I. RESEARCH PAPER

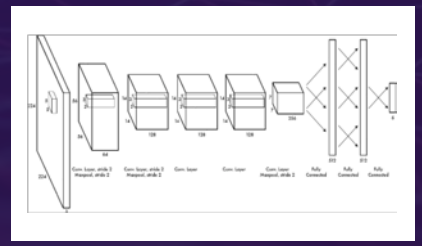
A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We

Yann LeCun's

CONCEPT OF CONVOLUTION



Quoc Le's

IDEA OF BRUTE FORCE

ACCELERATION OF A.I. FUELED BY

1. Very large, accessible **data** sets
2. Affordable **high performance computing** to turn data to $$$

1955 — 1982 — 2012 — TODAY

MACHINE LEARNING = relentless classification of data to make determinations or predictions

SOURCE: Historic references based on New Street Research, May 2018

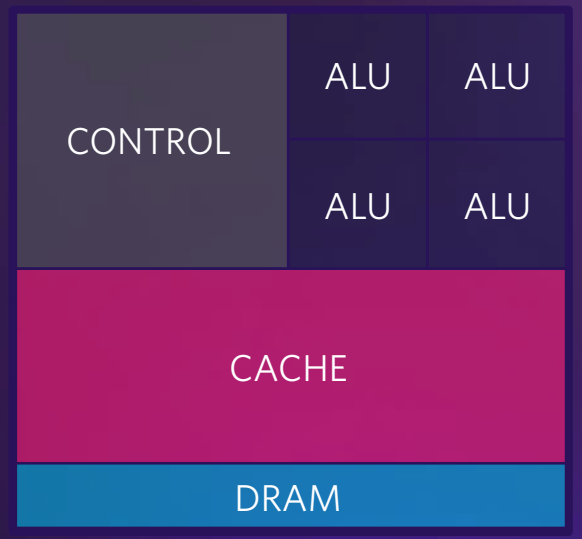
A.I. WORKLOADS NEED

A lot of memory
(because there's a lot of data)

Parallel computing
(for throughput)

Extremely high logic ⇆ memory bandwidth

CPU

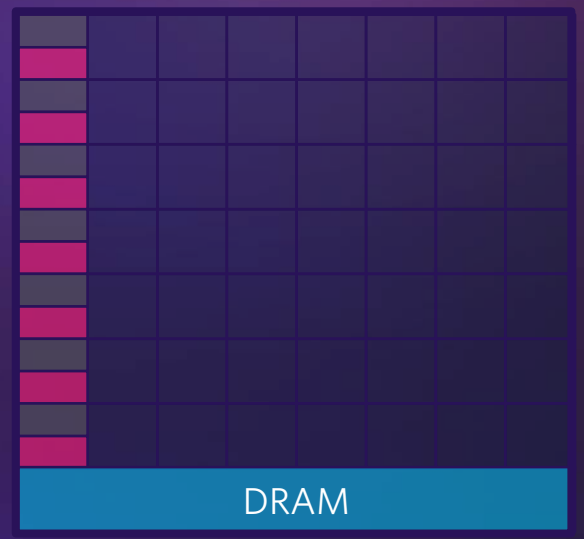


Sequential computing

General purpose

Good at many things

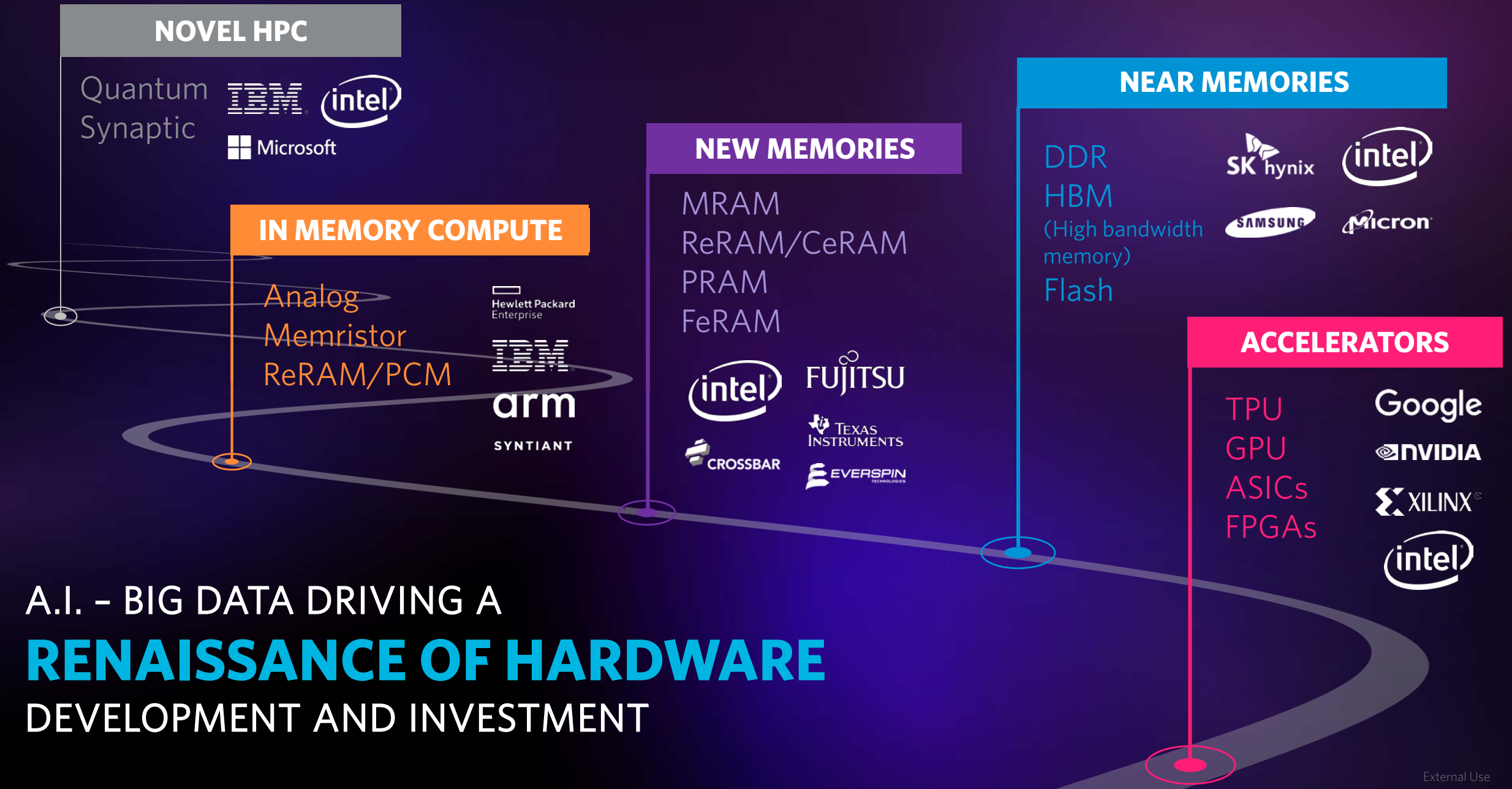
TPU / GPU



✓ Parallel computing

✓ Designed for throughput

✓ Great at specialized tasks



A.I. – BIG DATA DRIVING A

RENAISSANCE OF HARDWARE

DEVELOPMENT AND INVESTMENT

1,000x

IMPROVEMENT IN COMPUTE PERFORMANCE / WATT NEEDED

SOURCE: DARPA, Intel, NVIDIA, A.I. startups

A.I. NEEDS EDGE AND CLOUD INNOVATIONS...

STORAGE
abundant low cost,
high performance,
low power data storage

+

HPC
orders of magnitude
improvement in performance,
energy efficiency and cost

WHILE AT THE SAME TIME...

MOORE'S LAW CHALLENGED
as classic 2D feature shrink slows

THE PERFECT STORM

OR

THE PERFECT OPPORTUNITY?

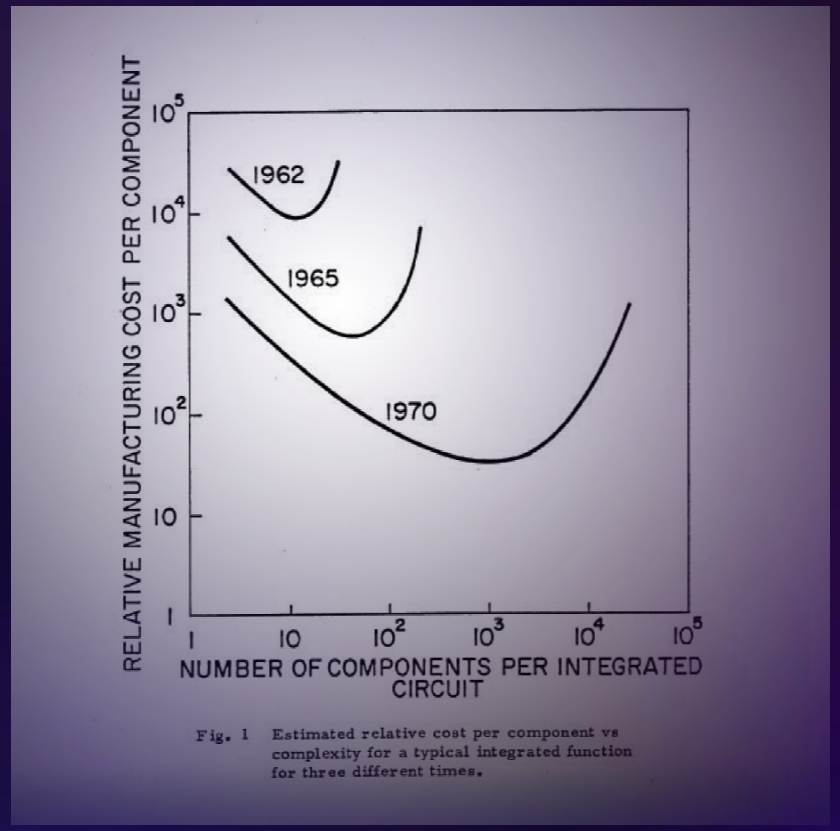


Fig. 1 Estimated relative cost per component vs complexity for a typical integrated function for three different times.

FIGURE 1

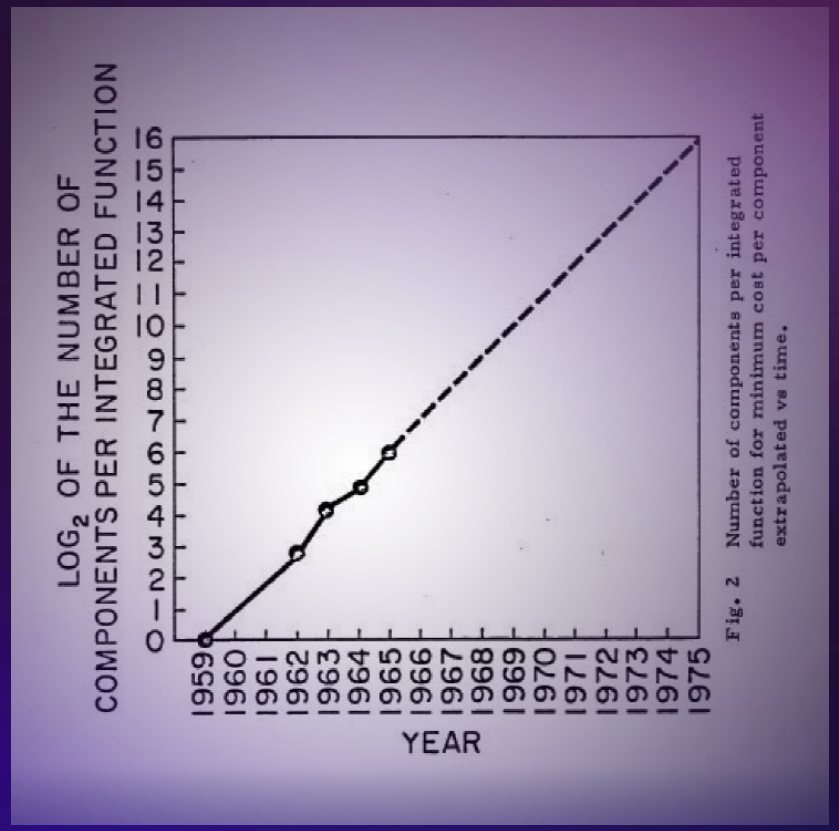


Fig. 2 Number of components per integrated function for minimum cost per component extrapolated vs time.

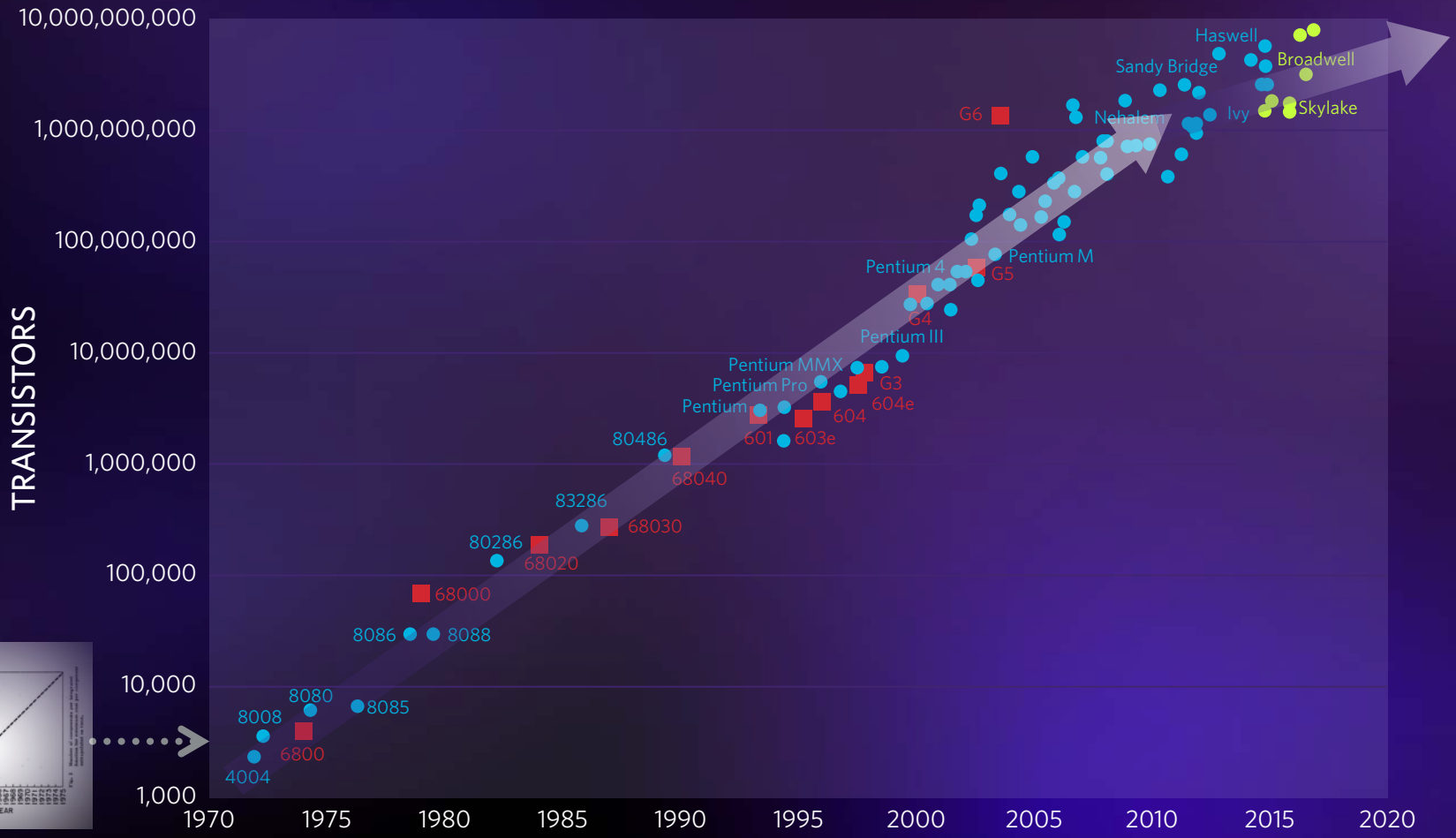
FIGURE 2

1965: **MOORE'S THESIS BASED ON FIVE DATA POINTS**

~1975: **ESTIMATE UPDATED TO 'DOUBLING EVERY 2 YEARS'**

SOURCE: Electronics, Volme 38, Number 8, April 19, 1965

Projection Held For 40 Years...



Recent data points suggest

~2x more every 5 years

1970 - 2010

$\sqrt[40]{1{,}000{,}000} \cong 1.413$

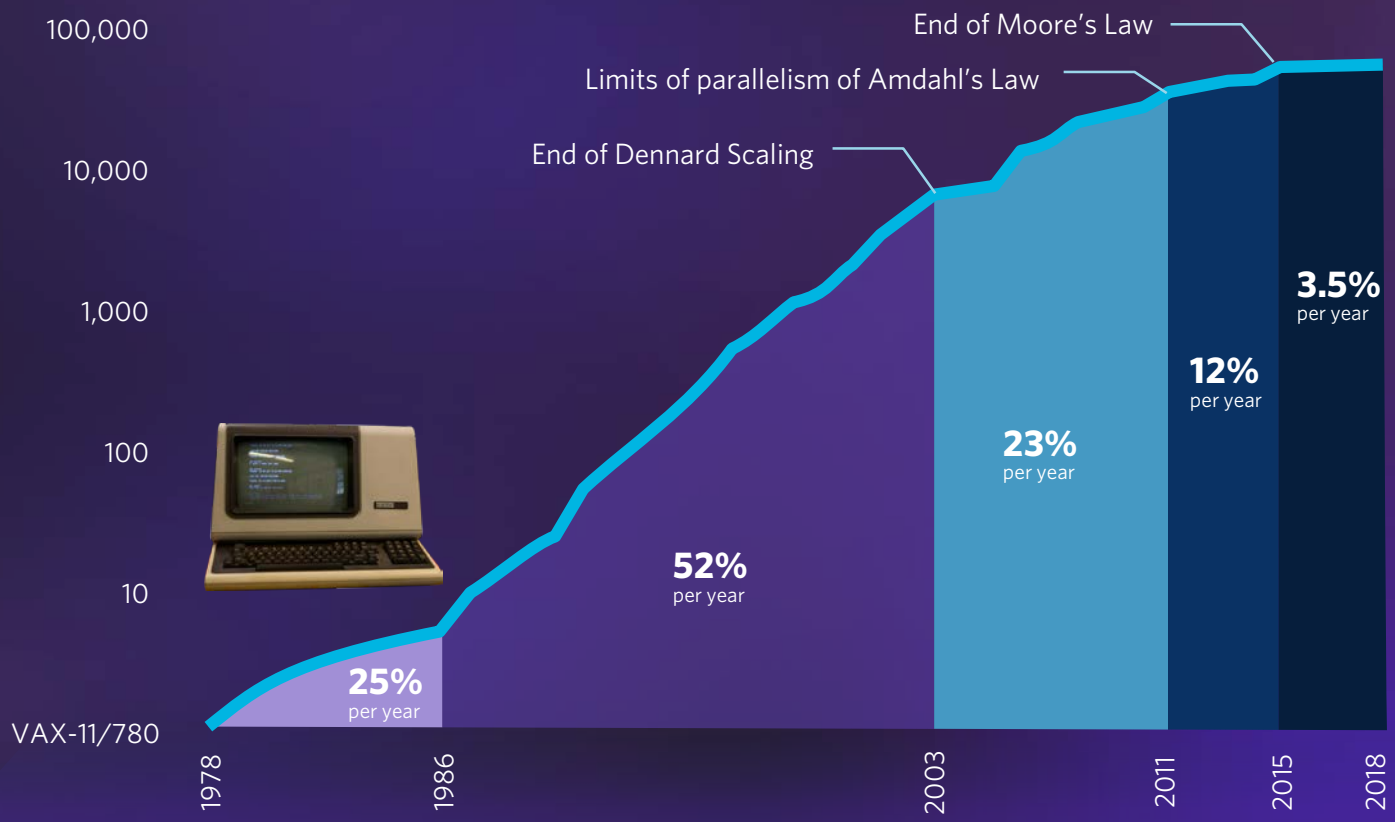
$(1.413)^2$ more every 2 years

~2x more every 2 years

CLASSIC 2D FEATURE SCALING SLOWING

SOURCE: University of Wisconsin

PERFORMANCE IMPROVEMENTS OVER TIME (VS. VAX-11/780)



SOURCE: Computer Architecture: A Quantitative Approach, Sixth Edition, John Hennessy and David Patterson, December 2017

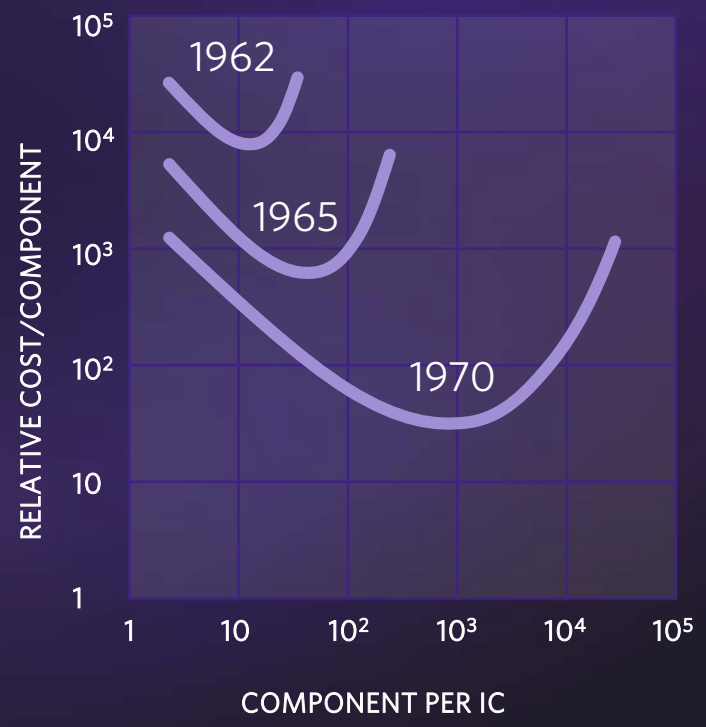
TIME BETWEEN LOGIC NODES IN YEARS

SOURCE: Bernstein

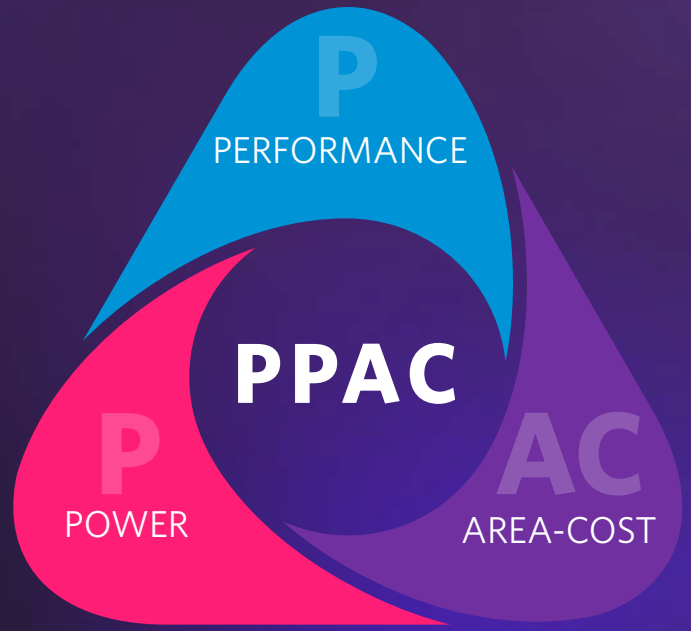
PERFORMANCE IMPROVEMENTS SLOWING

In the Past…

MOORE'S LAW



PPAC



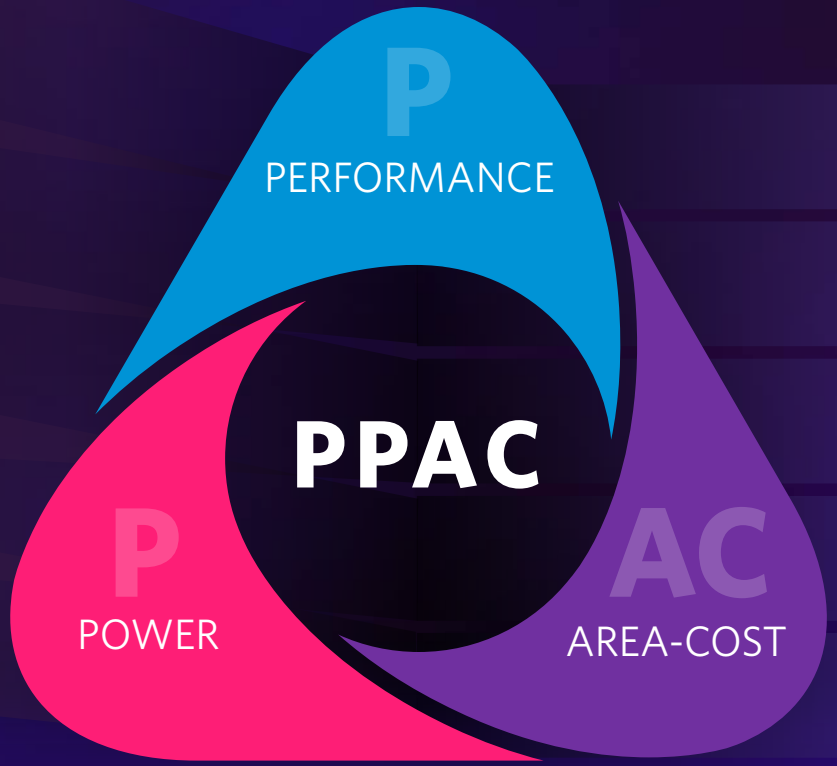
ENABLED BY

"Classic" 2D feature shrinking

+

materials engineering to improve power and performance

In the Future...



ENABLED BY

- New **architectures**
- New **structures / 3D**
- New **materials**
- New ways to **shrink**
- Advanced **packaging**

FOUNDATION IS MATERIALS ENGINEERING

3D NAND

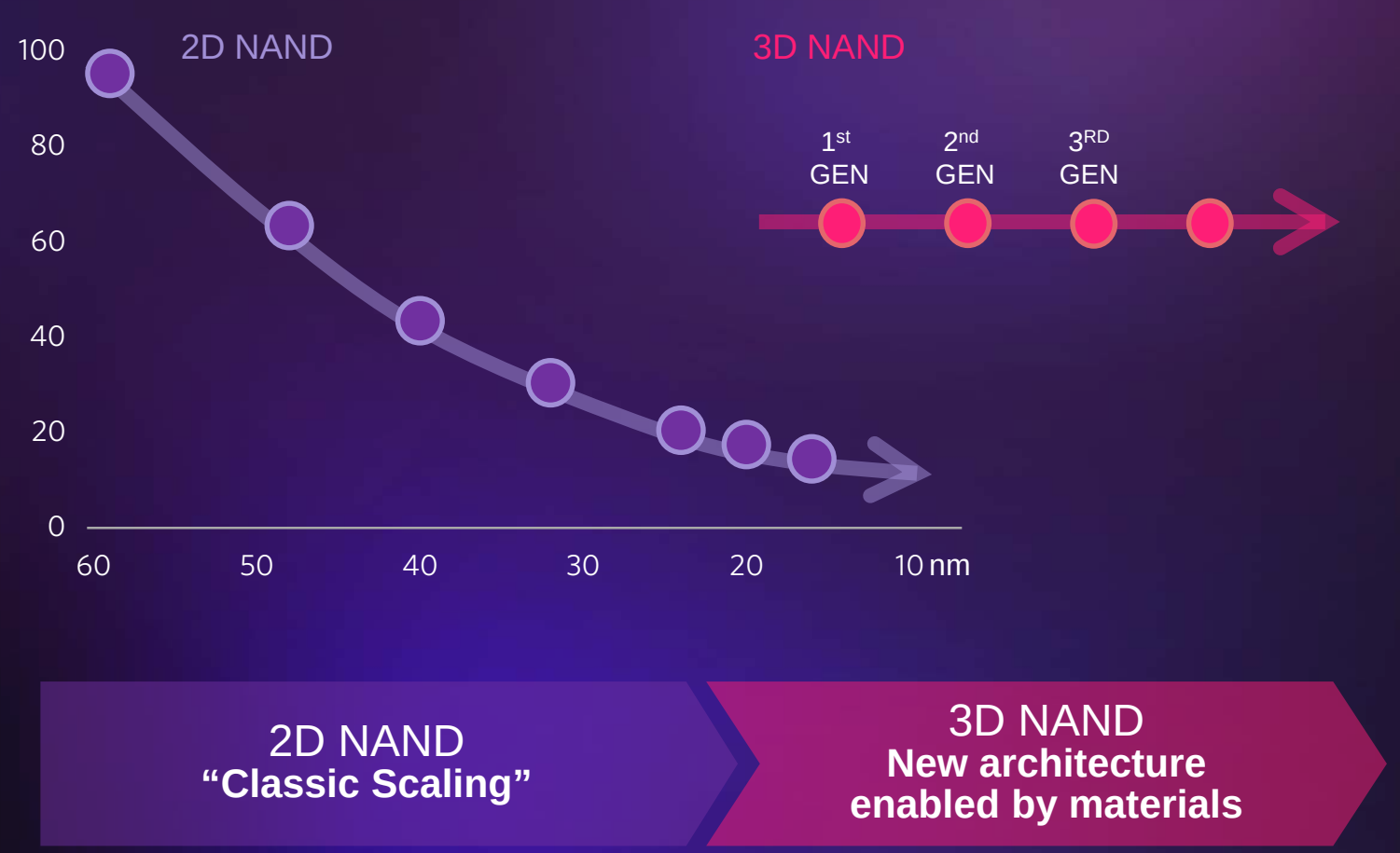
shows **power of architecture inflections**

Extended NAND **cost** roadmap by >10 years + better device

Speed **2x**

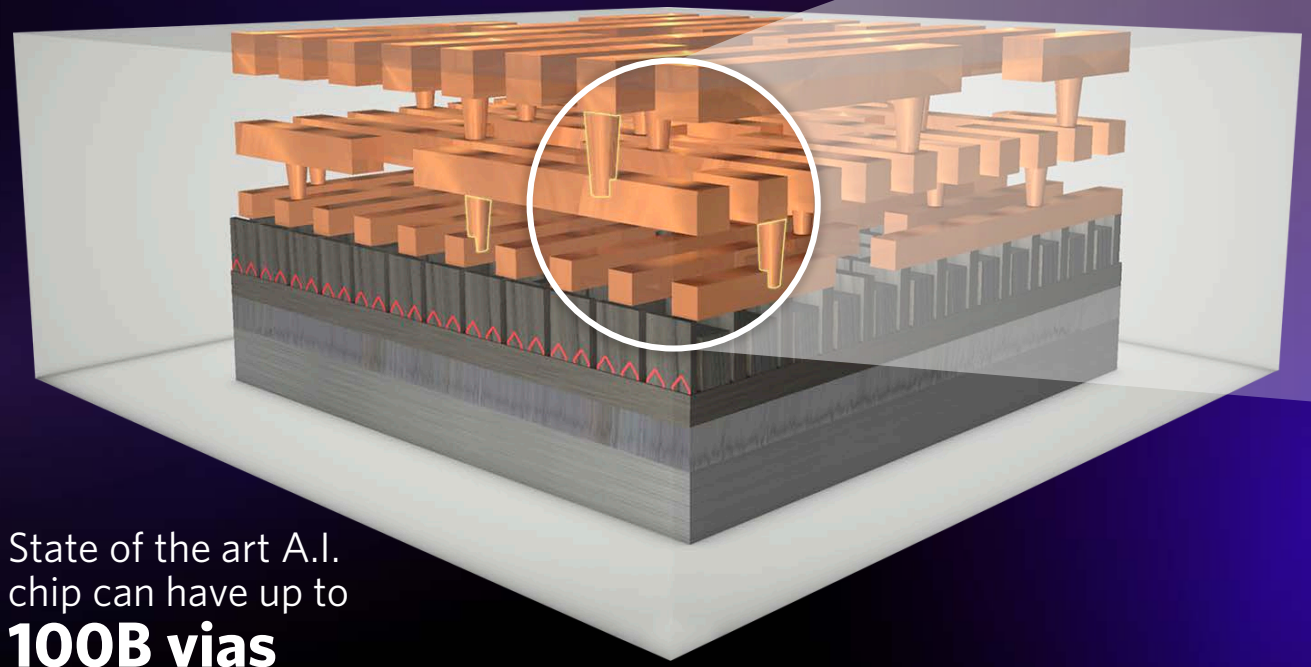
Endurance **10x**

Power efficiency **2x**

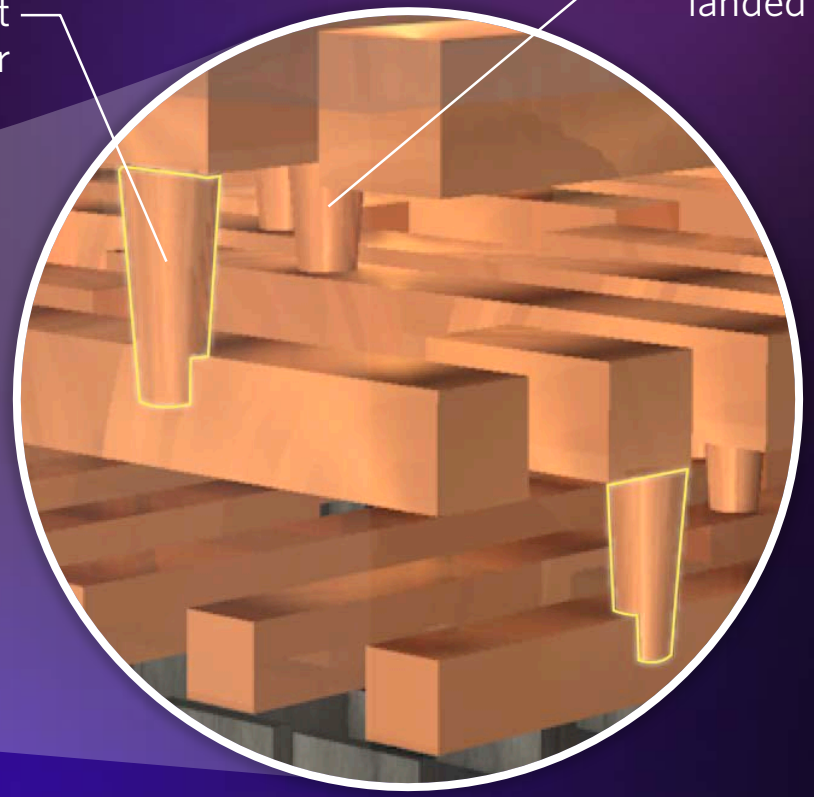


Future 2D shrink is not only limited by resolution, but also

PLACEMENT ERRORS



State of the art A.I. chip can have up to **100B vias**



Can be addressed by self-aligned structures

STEPS
1 - 8

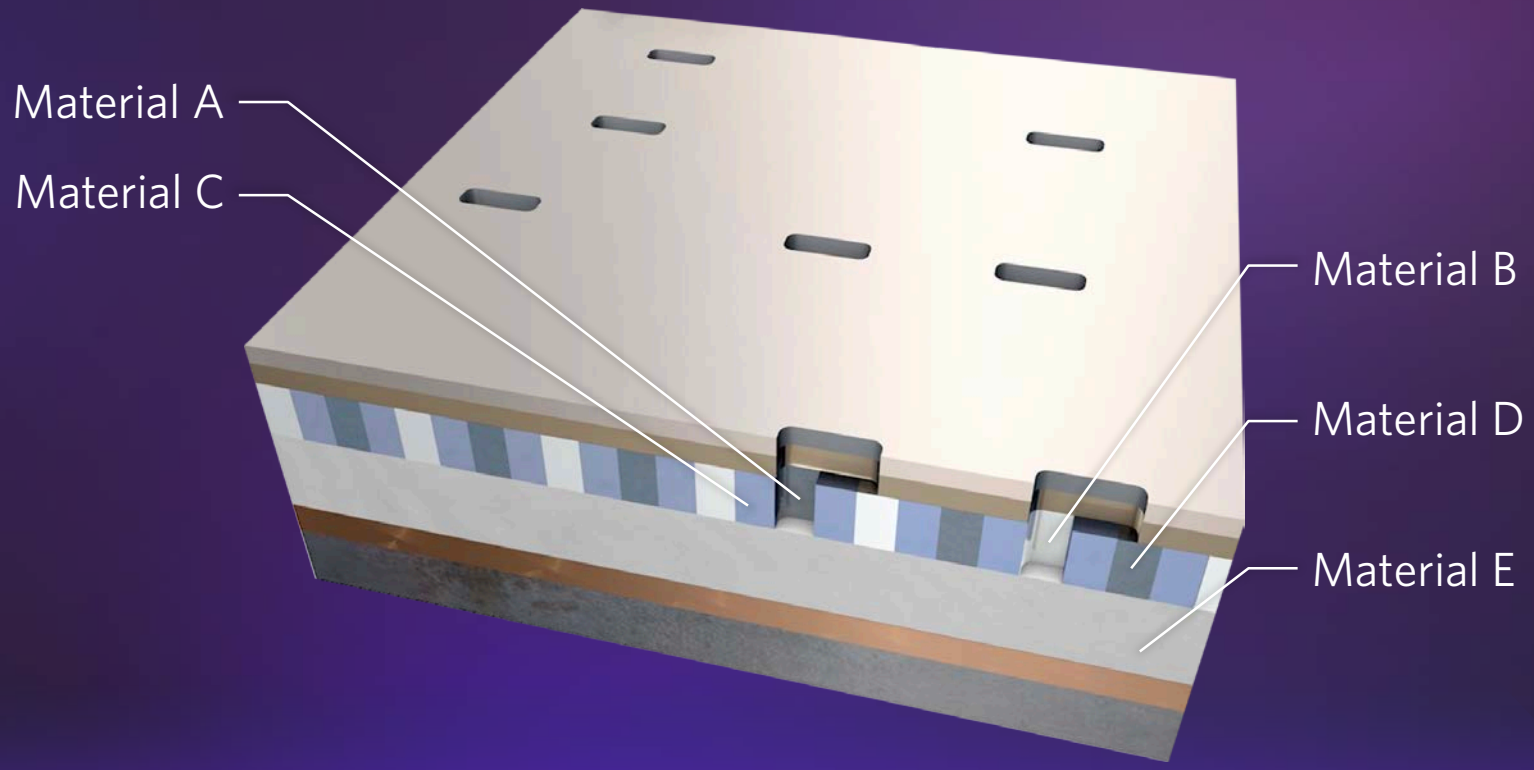
Self-aligned Spacer Patterning:
Litho-spacer Etch

STEPS
9 - 10

Materials Engineering:
Gapfill-CMP

STEPS
11 - 14

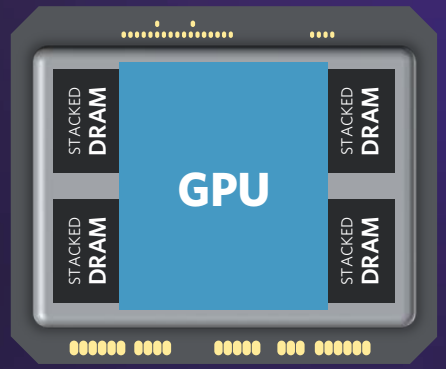
Deposition-Lithography-Etch:
Large Overlay Tolerance

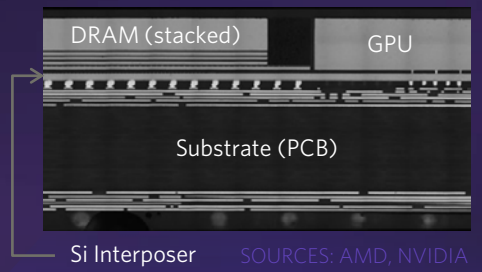


MATERIALS-BASED APPROACHES CAN ELIMINATE PLACEMENT ERRORS

Example: 'Multicolor' = Fully self-aligned multi-material patterning

DRAM ON PCB to STACKED DRAM IN PACKAGE



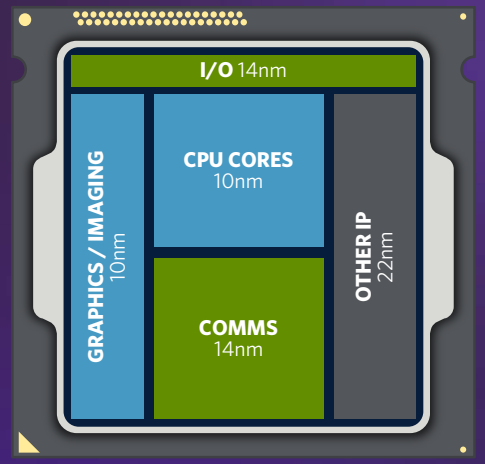


SOURCES: AMD, NVIDIA

3x
Logic ←→ DRAM bandwidth performance

50%
Power savings per bit

HETEROGENEOUS INTEGRATION

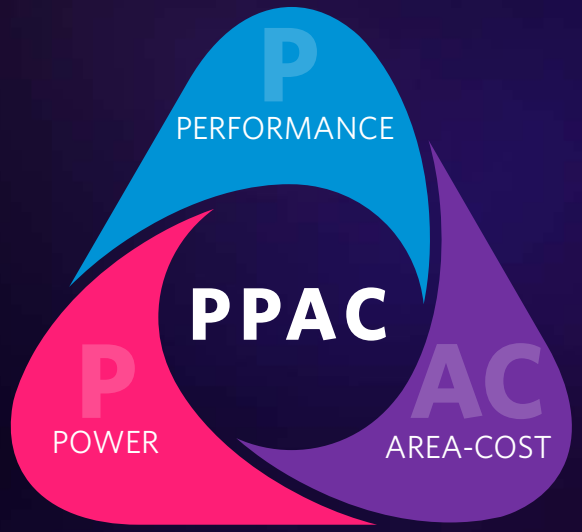


SOURCES: Intel, GLOBALFOUNDRIES

System on Chip to System on Package

Integration of chiplets provides **time**, **cost** and **yield** benefits

ADVANCED PACKAGING
Can Optimize System Level Performance



KEY ISSUES =

Complexity ↑

Integration challenges ↑↑

Time to market ↑↑↑

New **architectures**

New **structures / 3D**

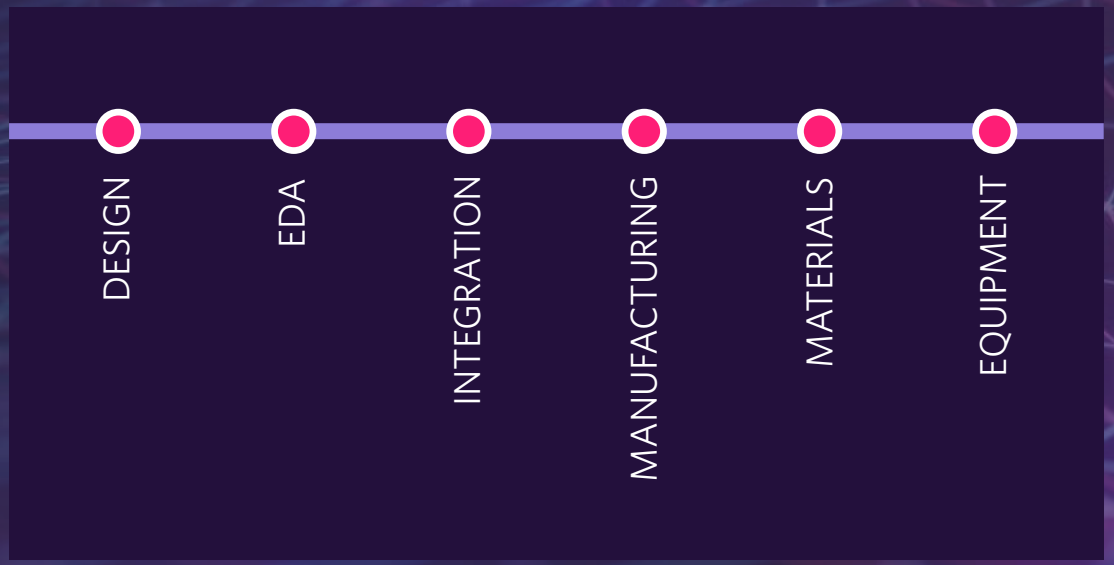
New **materials**

New ways to **shrink**

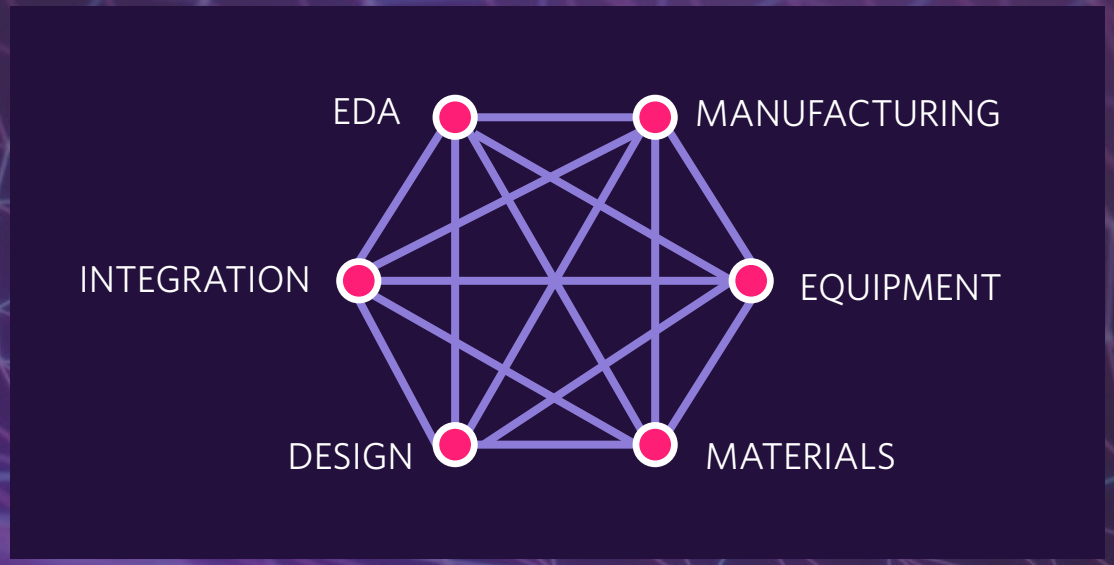
Advanced **packaging**

NEW PLAYBOOK
NEEDED FOR
CONNECTIVITY + SPEED

"Von Neumann" mindset vs. **"Neuromorphic"** mindset



TODAY: Serial / compartmentalized interaction between key parts of eco-system



OPPORTUNITY: Parallel development to get powerful tools to designers faster

CONNECTIVITY TO ACCELERATE INNOVATION

A.I. – Big Data Era

= the biggest opportunity of our lifetimes

UNLOCKED BY

Hardware renaissance

Materials innovation
to enable new architectures, structures, ways to shrink and packaging approaches

New eco-system playbook to drive **connectivity** and **speed**