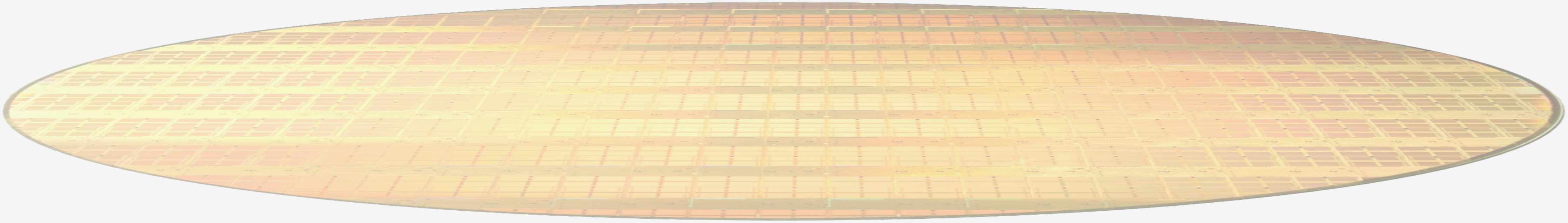
TBD

Jim Keller, Intel Senior Vice President
General Manager, Silicon Engineering Group

To be determined

Will complexity stop us?

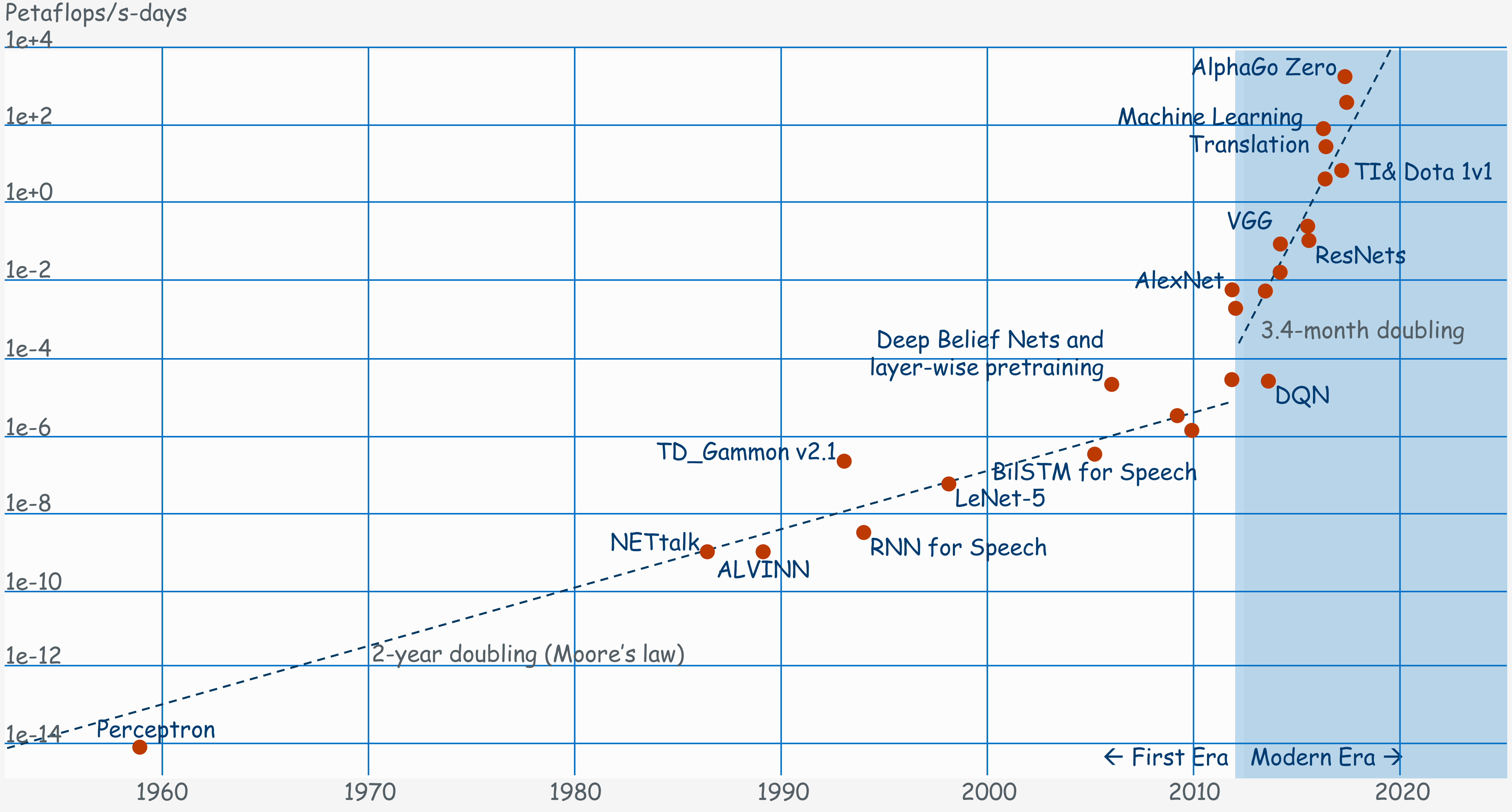
Will technology pessimism win?



> "The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law."

Rich Sutton, The Bitter Lesson, March 2019

TWO DISTINCT ERAS OF COMPUTER USAGE IN AI



(Source: https://openai.com/blog/ai-and-compute/)

"Everything that can be invented has been invented"
Charles H. Duell
US Patent Commissioner
1899

"Moore's law won't work at feature sizes less than a quarter of a micron"
Erich Bloch – Head of IBM Research, later Chairman of NSF
1988

"There is nothing new to be discovered in physics now"
Lord Kelvin
1900

Intelligent Machines

The End of Moore's Law?

The current economic boom is likely due to increases in computing speed and decreases in price. Now there are some good reasons to think that the party may be ending.

2000

by Charles C. Mann May 1, 2000

Moore Sees 'Moore's Law' Dead in a Decade

By Mark Hachman on September 18, 2007 at 5:12 pm 1 Comment 2007

In a "fireside chat" with NPR "Tech Nation's" Moira Gunn, Intel co-founder and chairman emeritus Gordon Moore said he sees his famous law expiring in 10 to 15 years.

MIT Technology Review

Computing / Microchips

Moore's Law Is Dead. Now What?

Shrinking transistors have powered 50 years of advances in computing—but now other ways must be found to make computers more capable.

2016

by Tom Simonite May 13, 2016

Report: IBM researcher says Moore's Law at end

IBM Fellow Carl Anderson says at a conference this week that Moore's Law is hitting a ceiling, according to a report.

2009

1,901 views | Apr 29, 2010, 01:37pm

Life After Moore's Law

By Bill Dally

Bill Dally is the chief scientist and senior vice president of research at NVIDIA and the Willard R. and Inez Kerr Bell Professor of Engineering at Stanford University.

2010

SCI-TECH

Moore's Law limit hit by 2014?

The high cost of semiconductor manufacturing equipment is making continued chipmaking advancements too expensive, threatening Moore's Law, according to iSuppli.

2009

BY BROOKE CROTHERS | JUNE 16, 2009 12:18 PM PDT

Theoretical physicist: Moore's Law has just 10 years to go

The age of silicon will come to a close but nobody knows when. Well, almost nobody.

2012

Death of Moore's Law Will Cause Economic Crisis

"Around 2020 or soon afterward, Moore's law will gradually cease to hold true and Silicon Valley may slowly turn into a rust belt unless a replacement technology is found," says Kaku in an extract published on Salon.com website.

By John E Dunn 2011

Techworld.com | MARCH 21, 2011 12:28 PM PT

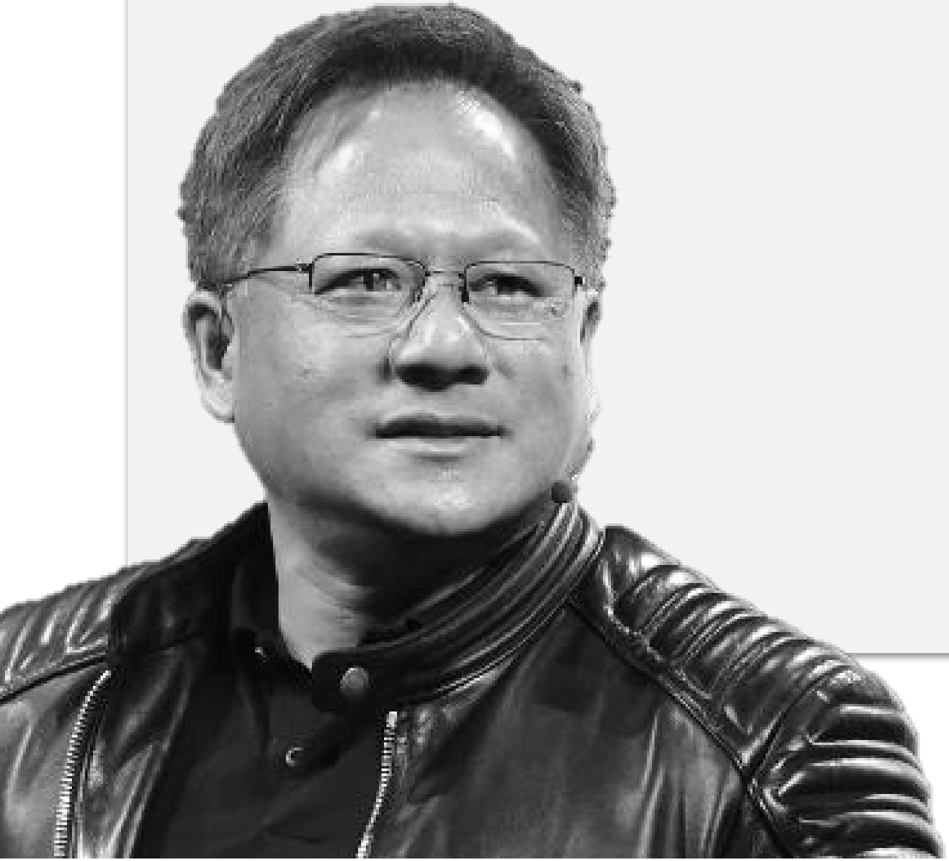
Jon Masters 🏴‍☠️
@jonmasters

The number of industry leaders saying Moore's Law is dead has increased at a rate of roughly a factor of two per Jim Keller talk. Certainly over the short term this rate can be expected to continue, if not to increase.

8:13 AM · Feb 14, 2020 from Cambridge, MA · Twitter for iPhone

SKEPTICS

Moore's Law is Dead
CES 2019: Jensen Huang
Nvidia CEO



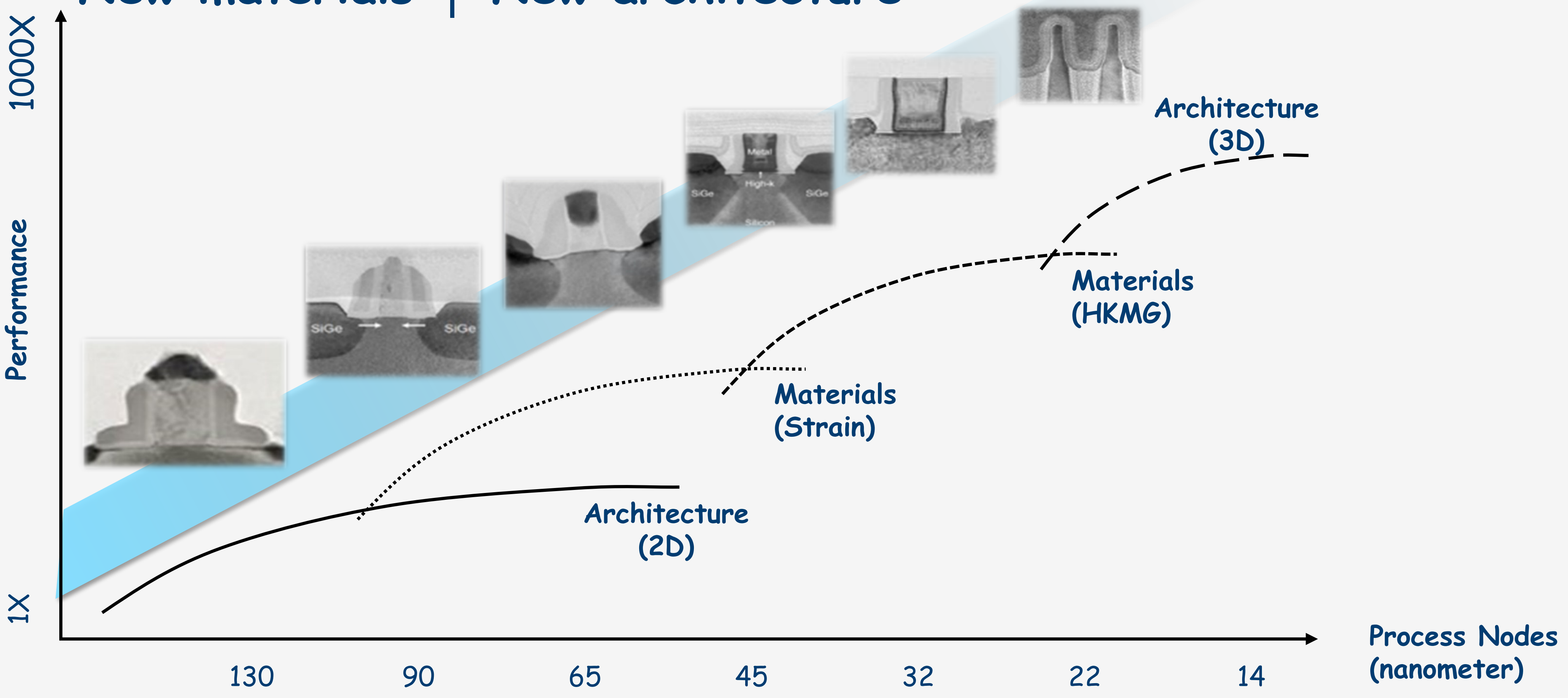
Moore's Law is Over
2018 IEEE Spectrum: David Patterson
UC Berkeley / Google


No I MINE.

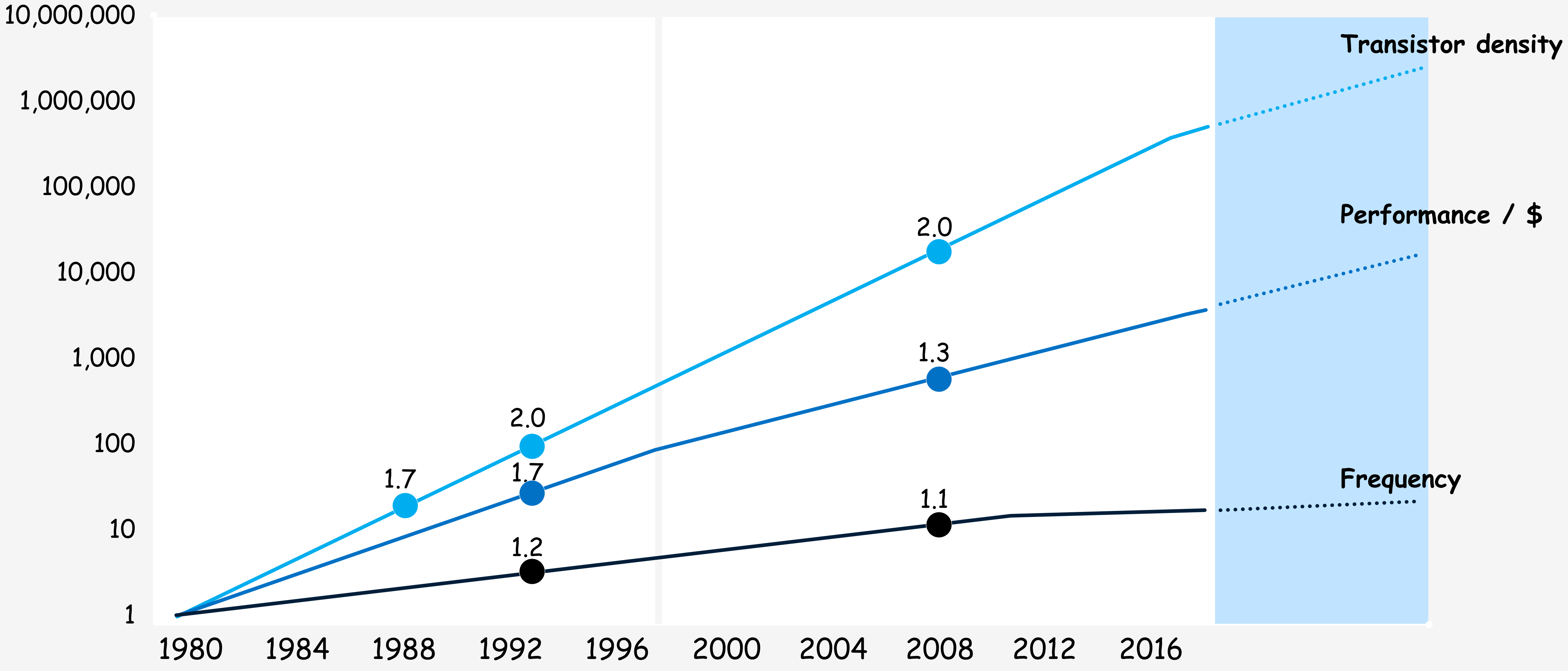
MOORE'S LAW TRANSISTORS

1000X reduction in feature size

New materials | New architecture

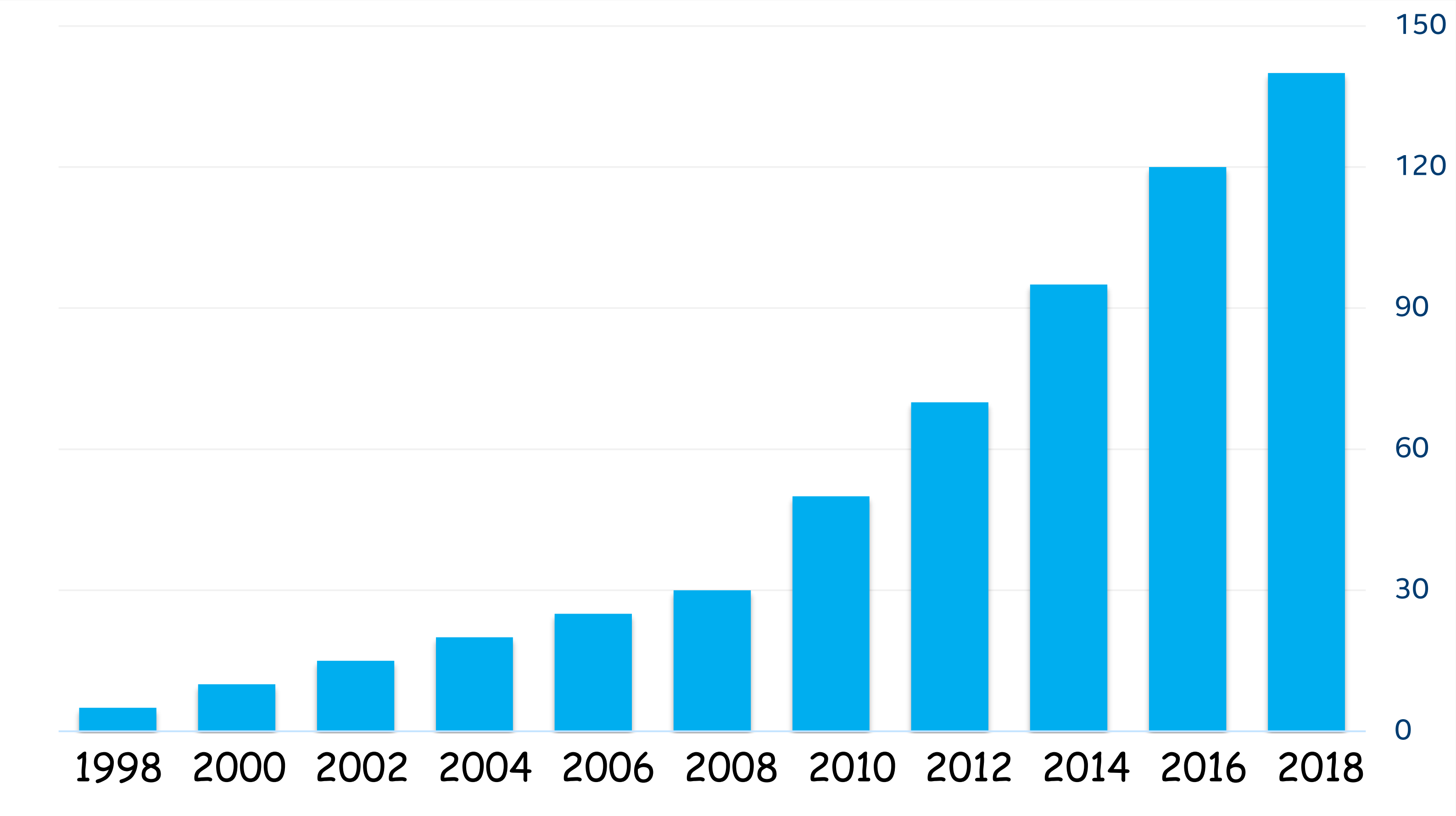


TRANSISTOR SCALING



PARADIGM SHIFT

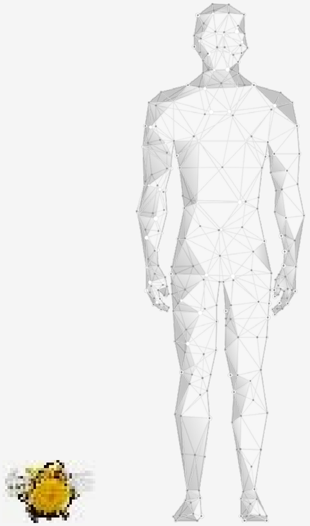
Average number of IPs in an SOC



(Source: Semico Research)

1000 SCALARS

	1980	2020	Scalar
Transistors per core	100,000	100,000,000	1,000X
Frequency of operation	5 MHz	5 GHz	1,000X
# Processing steps	~100	~10000	100X
Wafer diameter (inch) Printed die per wafer	6 inches 1X	12 inches 4X	2X 4X
# Mask layers	~10	~100	10X
Transistors on a chip	100,000	30,000,000,000	30,000X
Minimum feature size	3 microns	< 5 nanometers	600,000X
Transistors / mm^2	1,000	100,000,000	100,000X
Cost per transistor (cents)	0.1 cents	0.00000001 cents	10,000,000X
Memory Latency	4 cycles	400 cycles	100X
Fab cost	$1M	$10B	10,000X
Power dissipation	< 1W	> 200W	200X
Instructions per cycle	0.3	3	10X
Operating voltage	5 Volt	0.65 Volt	5X
# Personal computing devices	< 10 million	> 10 billion	1,000X



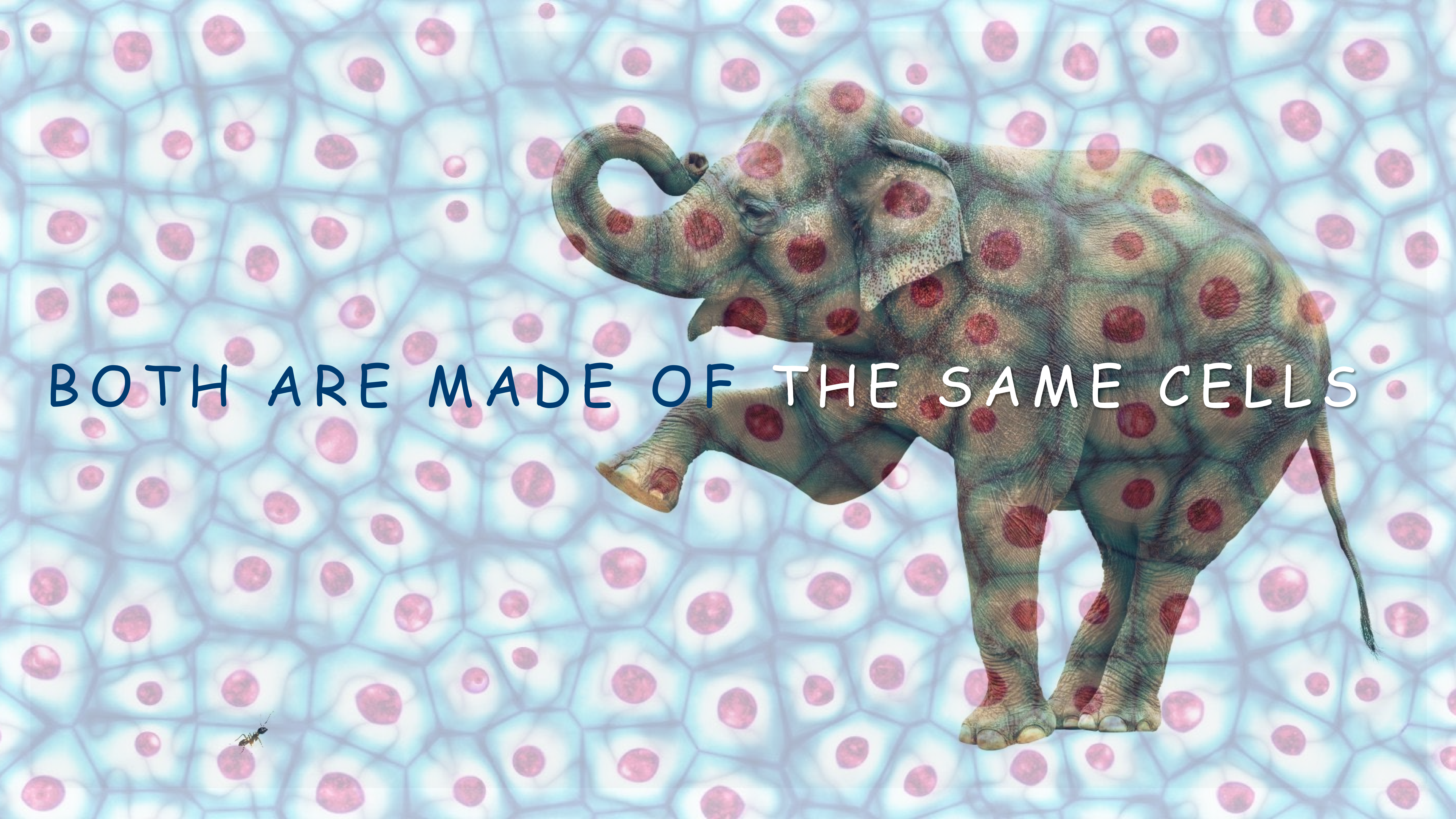
1 WATT GENERATOR

1 GIGAWATT GENERATOR

(Image source: www.rediff.com)

BIG COMPUTERS… WHY NOT?

(Image source: Cray, Inc.)


BOTH ARE MADE OF THE SAME CELLS

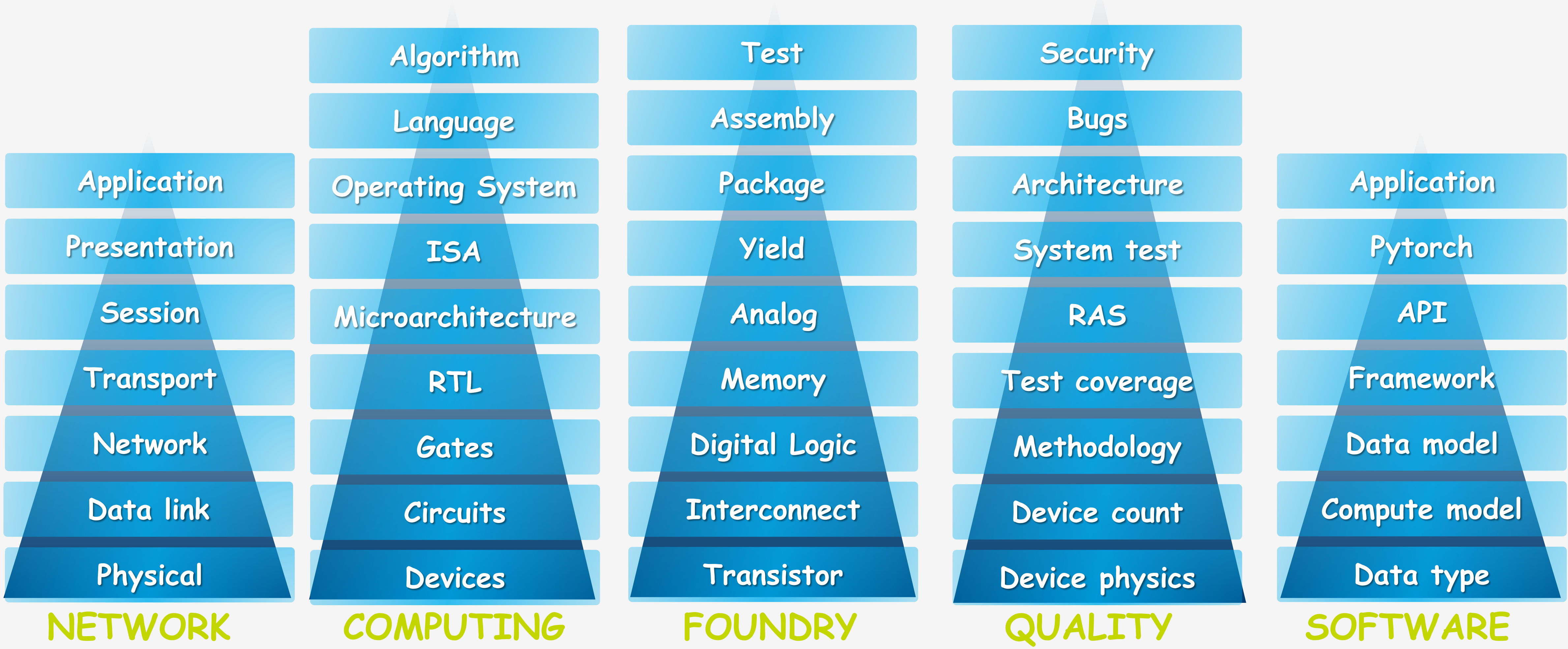
KEY PROBLEMS

> How to program 1 million computers ?

> Address >2^{60} objects ?

> Make 1 peta look-ups into 1 petabyte of data?

ABSTRACTION LAYERS



COMPUTING

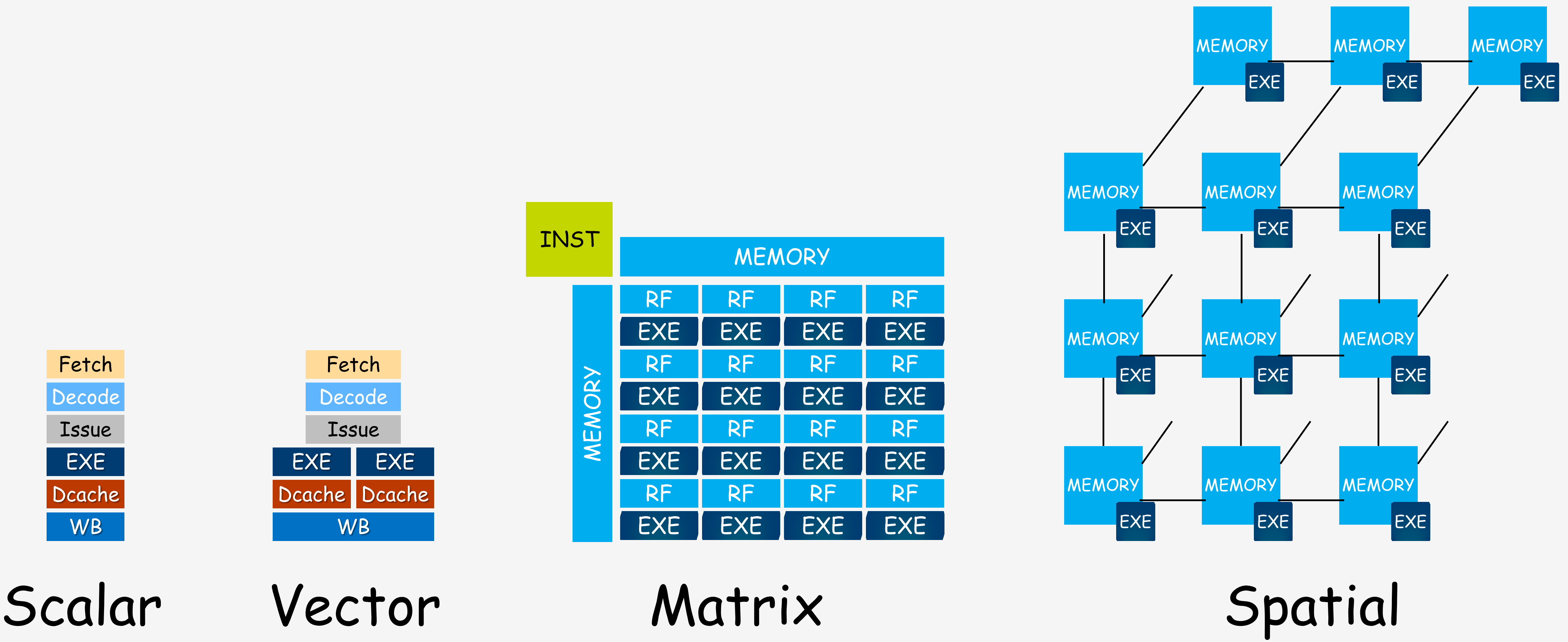
$$A = (B + C) \cdot D$$

$$A_{[i]} = \left(B_{[i]} + C_{[j\cdot]}\right) \cdot D_{[k]}$$

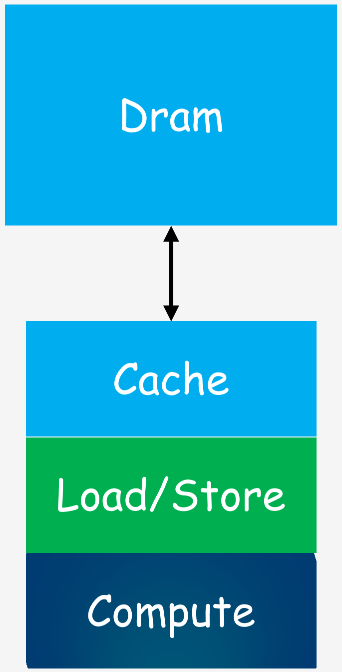
$$A_{[i,j]} = \left(B_{[i,k]} + C_{[k,j]}\right)$$

$$A_{[fma(i,j),f(j\cdot k)]} = \left(B_{[fma(i,k),f(j,k)]} + C_{[fma(k,j),f(k,j)]}\right)$$

Compute models



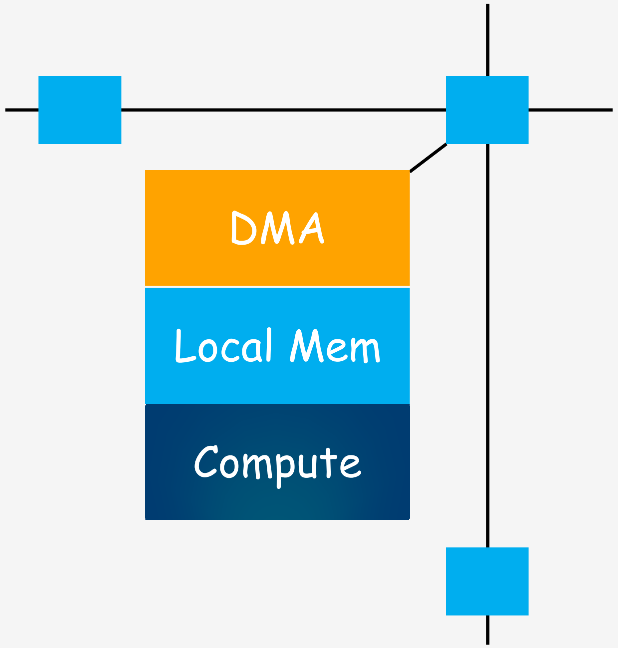
Data models



Global Memory

Caching

Load/Store
Arch



Local Mem

DMA

Networked Memory

(Source: Software Data Movement – SW DM)

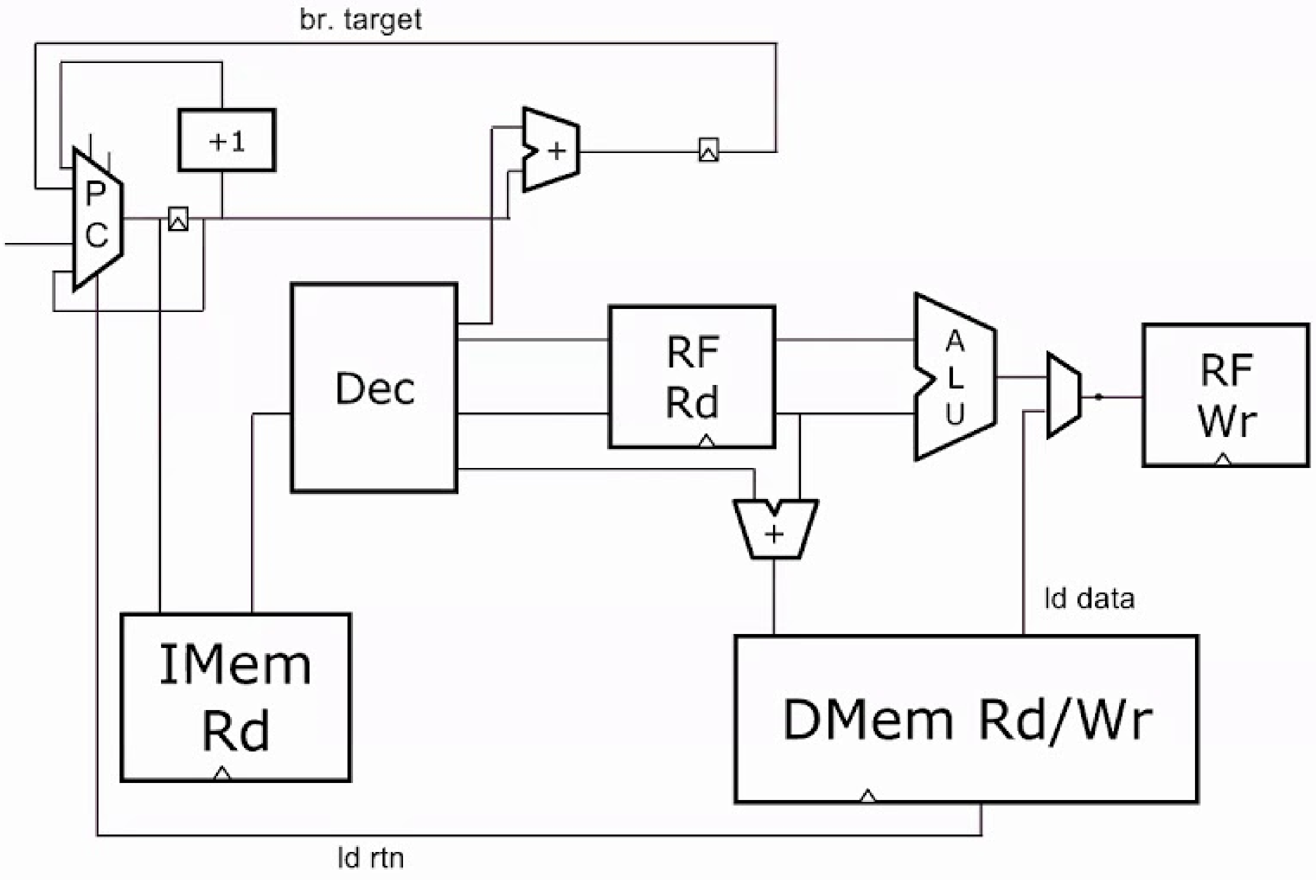
HelloWorld - Notepad

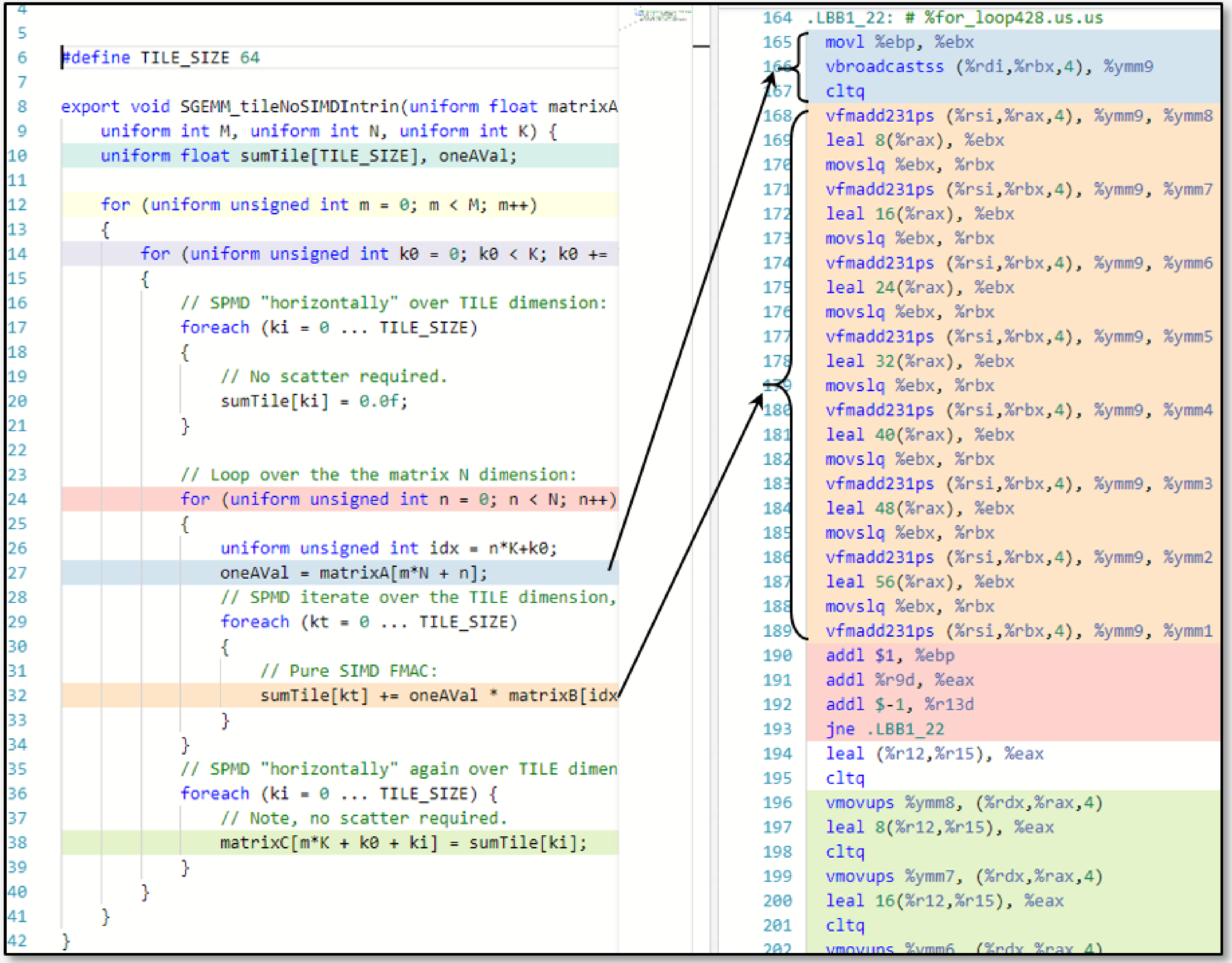
File Edit Format View Help

```
// Simple C program to display "Hello World"

// Header file for input output functions
#include <stdio.h>

// main function -
// where the execution of program begins
int main()
{

    // prints hello world
    printf("Hello World");

    return 0;
}
```

Ln 15, Col 2 | 100% | Windows (CRLF) | UTF-8

Example RISC-V Block Diagram

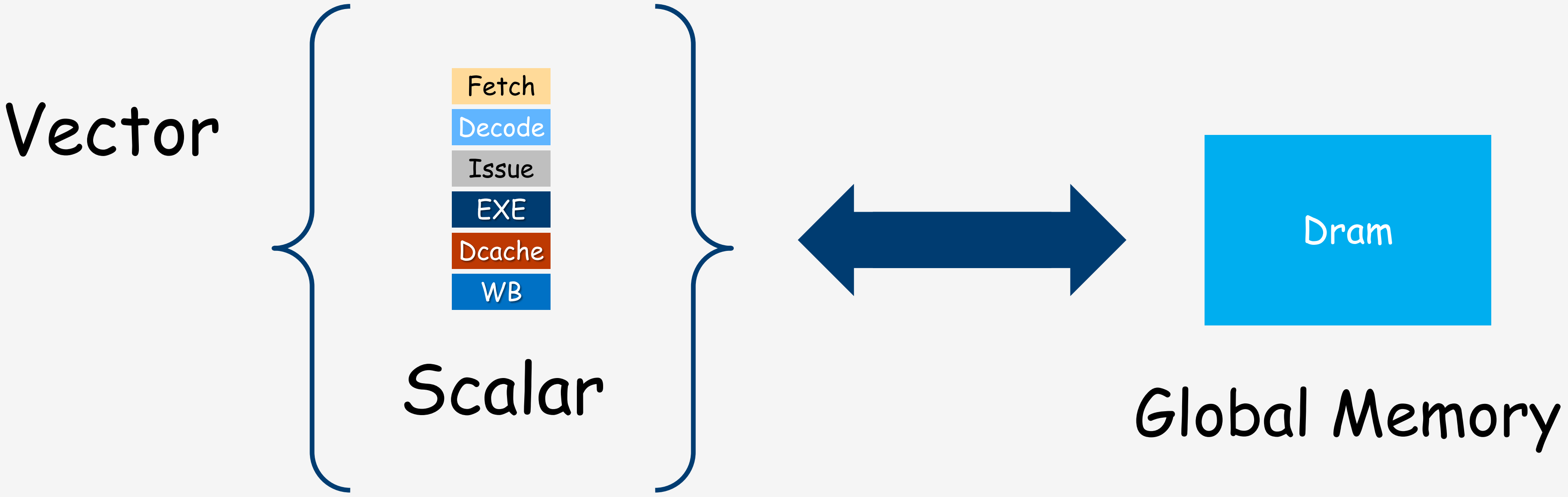


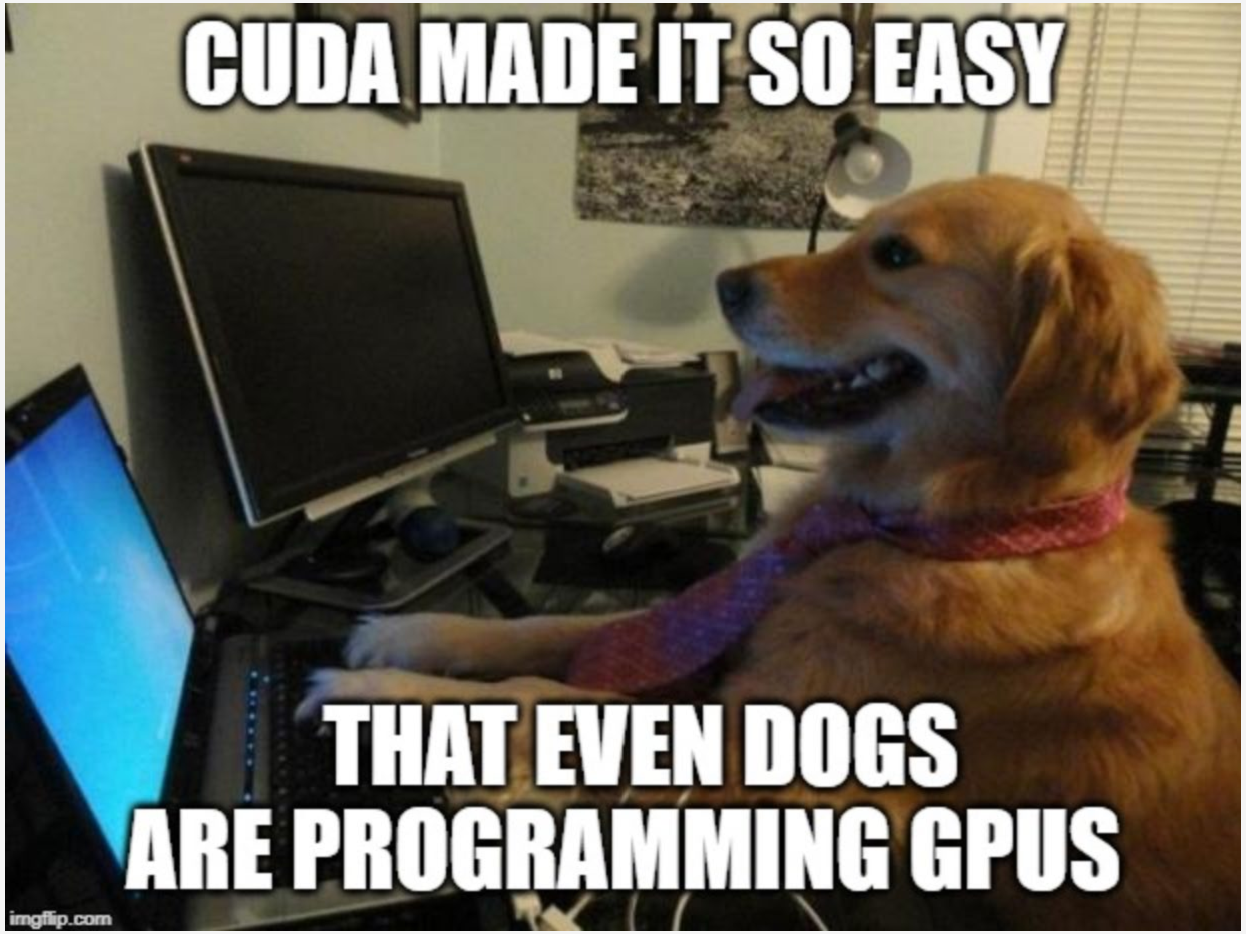


#define TILE_SIZE 64
export void SGEMM_tileNoSIMDIntrin(uniform float matrixA
uniform int M, uniform int N, uniform int K) {
uniform float sumTile[TILE_SIZE], oneAVal;
for (uniform unsigned int m = 0; m < M; m++)
{
for (uniform unsigned int k0 = 0; k0 < K; k0 +=
{
// SPMD "horizontally" over TILE dimension:
foreach (ki = 0 ... TILE_SIZE)
{
// No scatter required.
sumTile[ki] = 0.0f;
}
// Loop over the the matrix N dimension:
for (uniform unsigned int n = 0; n < N; n++)
{
uniform unsigned int idx = n*K+k0;
oneAVal = matrixA[m*N + n];
// SPMD iterate over the TILE dimension,
foreach (kt = 0 ... TILE_SIZE)
{
// Pure SIMD FMAC:
sumTile[kt] += oneAVal * matrixB[idx
}
}
// SPMD "horizontally" again over TILE dimen
foreach (ki = 0 ... TILE_SIZE) {
// Note, no scatter required.
matrixC[m*K + k0 + ki] = sumTile[ki];
}
}
}
}
.LBB1_22: # %for_loop428.us.us
movl %ebp, %ebx
vbroadcastss (%rdi,%rbx,4), %ymm9
cltq
vfmadd231ps (%rsi,%rax,4), %ymm9, %ymm8
leal 8(%rax), %ebx
movslq %ebx, %rbx
vfmadd231ps (%rsi,%rbx,4), %ymm9, %ymm7
leal 16(%rax), %ebx
movslq %ebx, %rbx
vfmadd231ps (%rsi,%rbx,4), %ymm9, %ymm6
leal 24(%rax), %ebx
movslq %ebx, %rbx
vfmadd231ps (%rsi,%rbx,4), %ymm9, %ymm5
leal 32(%rax), %ebx
movslq %ebx, %rbx
vfmadd231ps (%rsi,%rbx,4), %ymm9, %ymm4
leal 40(%rax), %ebx
movslq %ebx, %rbx
vfmadd231ps (%rsi,%rbx,4), %ymm9, %ymm3
leal 48(%rax), %ebx
movslq %ebx, %rbx
vfmadd231ps (%rsi,%rbx,4), %ymm9, %ymm2
leal 56(%rax), %ebx
movslq %ebx, %rbx
vfmadd231ps (%rsi,%rbx,4), %ymm9, %ymm1
addl $1, %ebp
addl %r9d, %eax
addl $-1, %r13d
jne .LBB1_22
leal (%r12,%r15), %eax
cltq
vmovups %ymm8, (%rdx,%rax,4)
leal 8(%r12,%r15), %eax
cltq
vmovups %ymm7, (%rdx,%rax,4)
leal 16(%r12,%r15), %eax
cltq

PCI Express 3.0 Host Interface
GigaThread Engine
GPC
Raster Engine
TPC
PolyMorph Engine
SM
L2 Cache
Memory Controller

GPUs

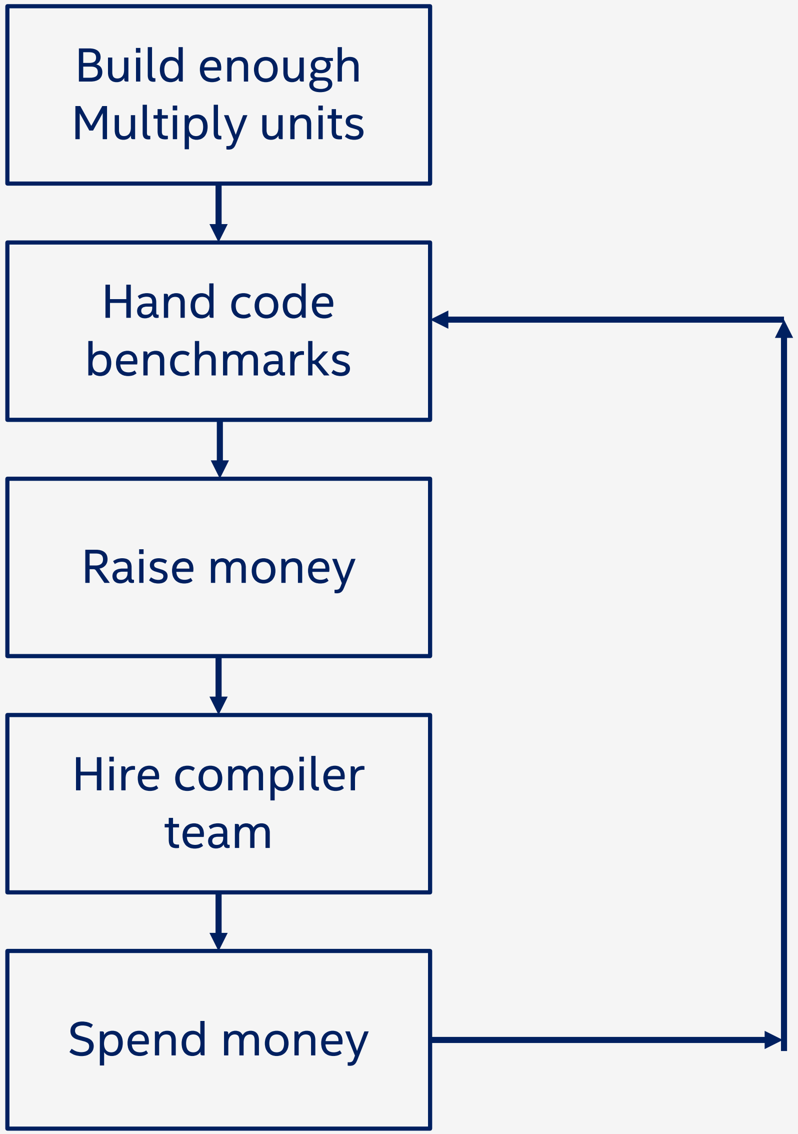

Vector
Fetch
Decode
Issue
EXE
Dcache
WB
Scalar
Dram
Global Memory



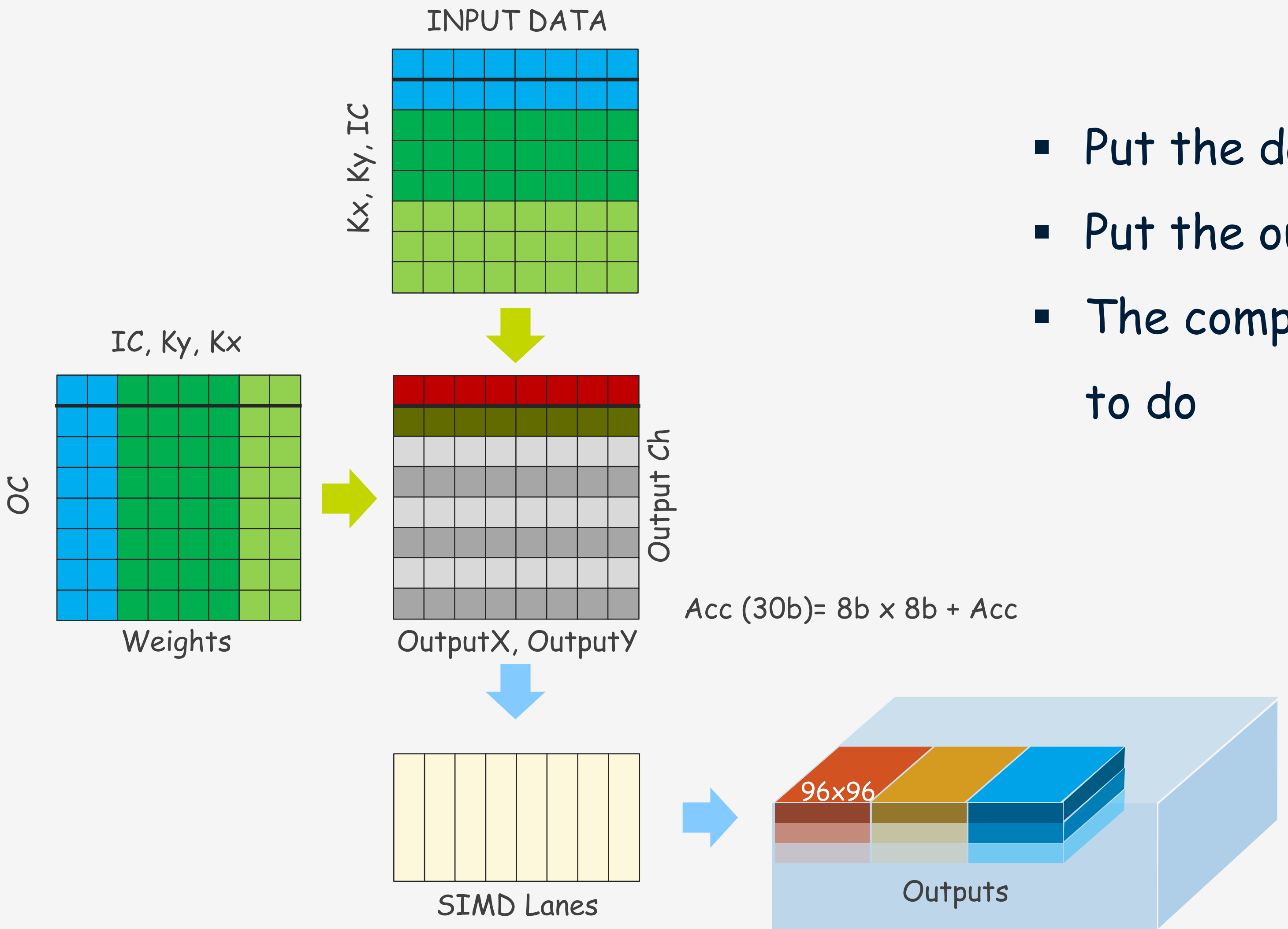
(Source: https://imgflip.com)

AI chips

SIMD
Local memory
no privilege model
hard to program
…….



ACCELERATORS

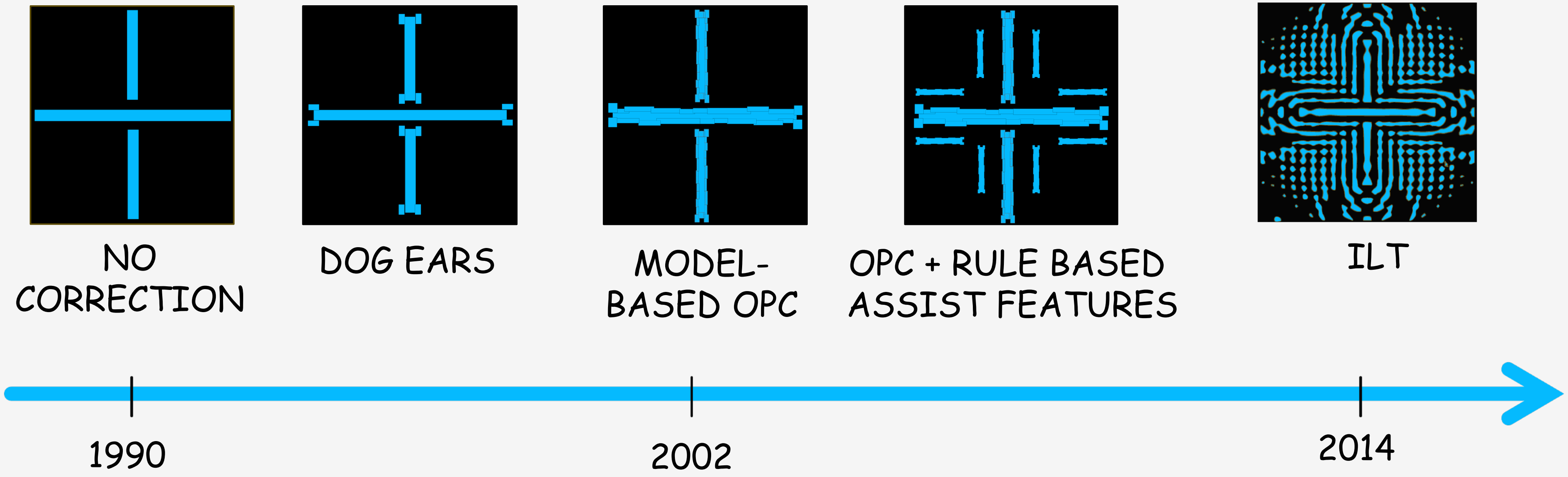


- Put the data in the local ram
- Put the out put of CAFÉ in the control story
- The compute arrays do what the CAFÉ said to do

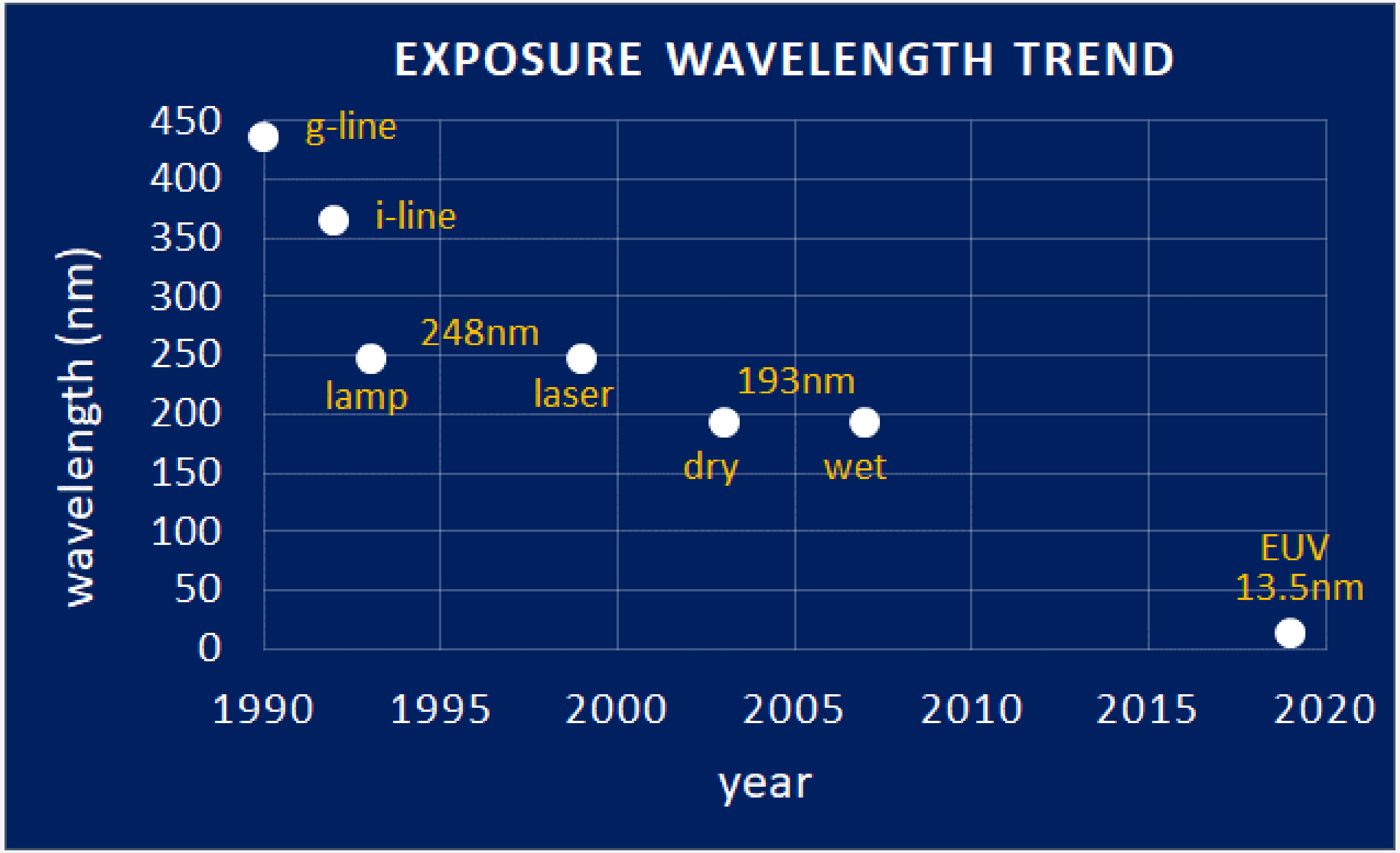
(Source: Peter Bannon, Tesla)

(Source: http://mashable.com/2017/05/07/pennsylvania-coal-miners/)

HOW TO PRINT +

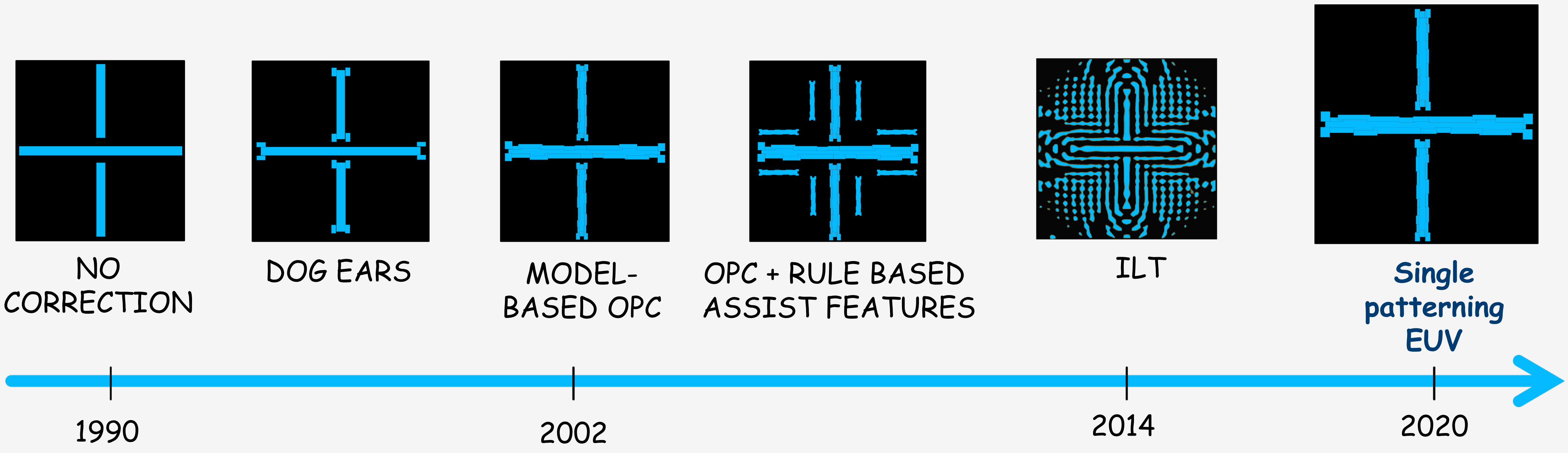

NO CORRECTION
DOG EARS
MODEL-BASED OPC
OPC + RULE BASED ASSIST FEATURES
ILT
1990
2002
2014

EUV: AN INNOVATION EXAMPLE



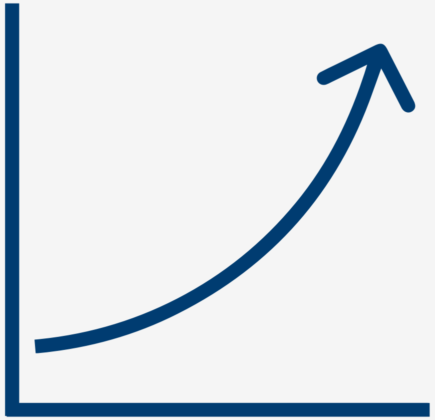
TRANSITION FROM 193NM TO 13.5NM EUV
15 YEARS DUE TO COMPLEXITY

HOW TO PRINT +

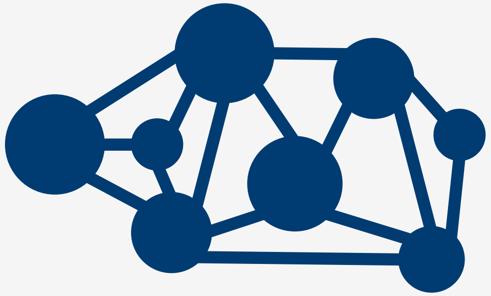


NO CORRECTION
DOG EARS
MODEL-BASED OPC
OPC + RULE BASED ASSIST FEATURES
ILT
Single patterning EUV
1990
2002
2014
2020

TO BE DETERMINED

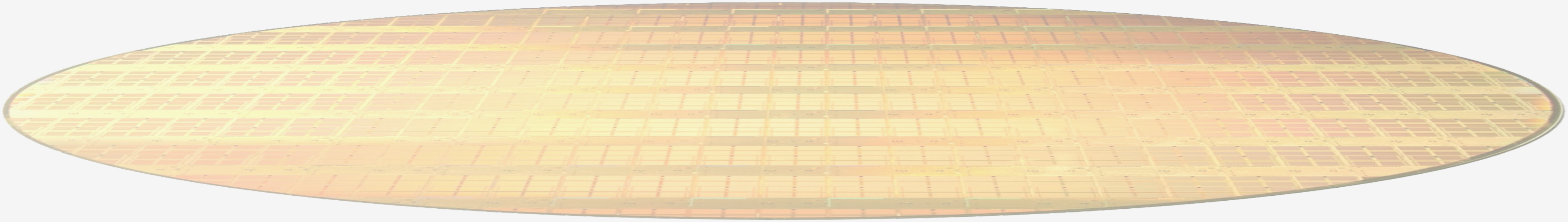


MOORE'S LAW



COMPLEXITY LIMITS

TECHNOLOGY OPTIMISM



intel®