GPU TECHNOLOGY CONFERENCE

HETEROGENEOUS HPC, ARCHITECTURAL OPTIMIZATION, AND NVLINK

STEVE OBERLIN

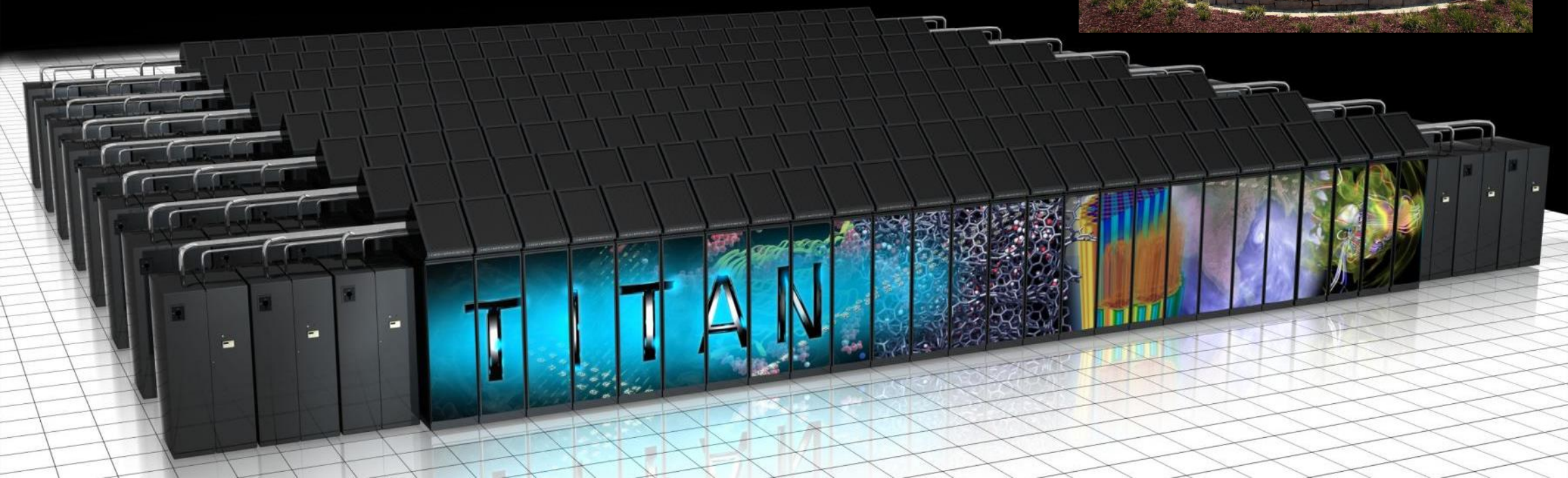
CTO, TESLA ACCELERATED COMPUTING

NVIDIA

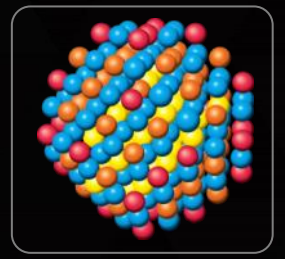
STATE OF THE ART 2012

18,688 Tesla K20X GPUs

27 PetaFLOPS

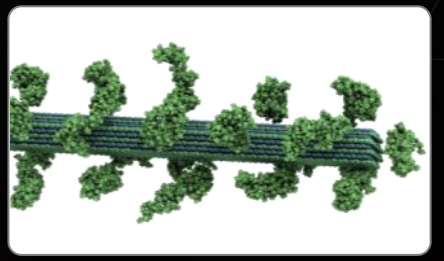


FLAGSHIP SCIENTIFIC APPLICATIONS ON TITAN



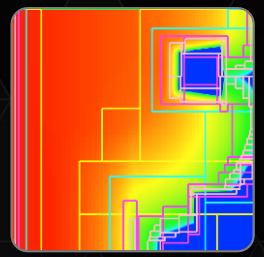
Material Science (WL-LSMS)

Climate Change (CAM-SE)

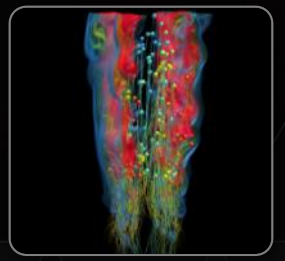


Biofuels (LAMMPS)

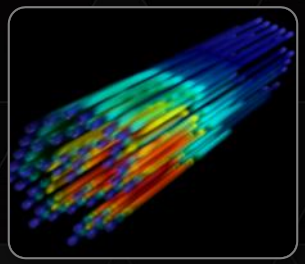
Astrophysics (NRDF)



Combustion (S3D)



Nuclear Energy (Denovo)



STATE OF THE ART 2017

CORAL Summit System
5-10x Faster
1/5th the Nodes,
Same Energy Use as Titan

Earth Simulator
40.96 TF
Top500 #1 for 3 years

Just 1 Node in Summit
Is the same performance as the
Earth Simulator in 2002

AGENDA: 3 STORIES

- Heterogeneous HPC
  - HPC System Evolution
  - Architecture and Technology Basis for Heterogeneous Processing
- Architectural Optimization and NVLink
  - Tesla Accelerated Computing Platform
  - NVLink
- CORAL Punctuation Mark
  - Optimized Fat node performance projections

ARE YOU CALIBRATED?

.6 MPH

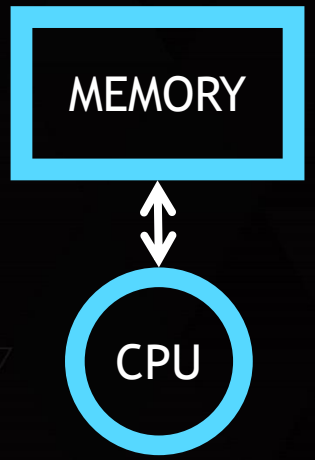
6 MPH

60 MPH

600 MPH

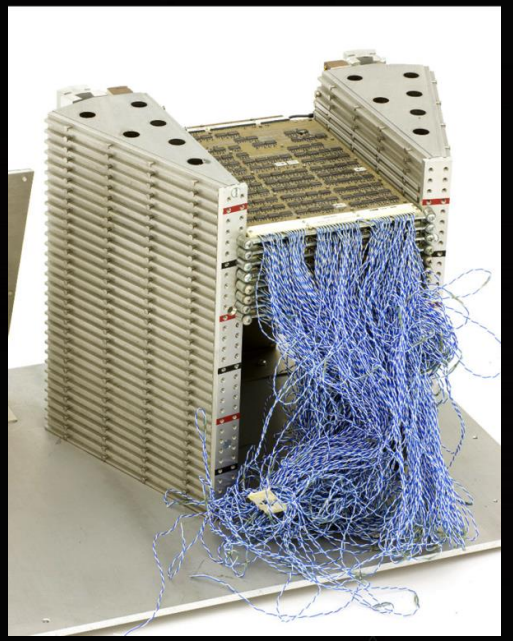
3 ORDERS OF MAGNITUDE

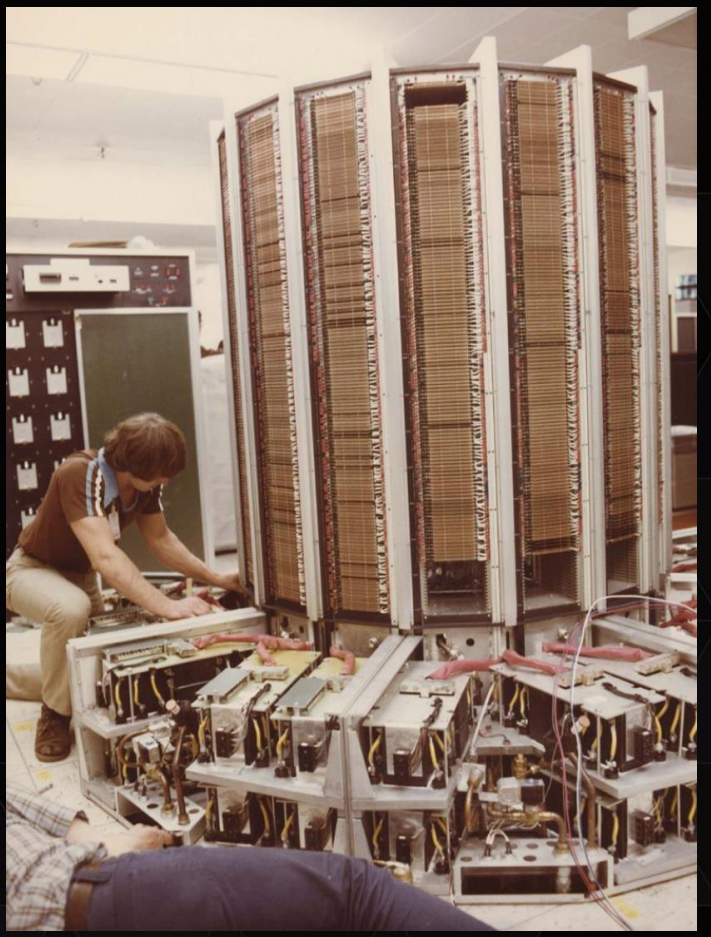
HPC SYSTEM EVOLUTION



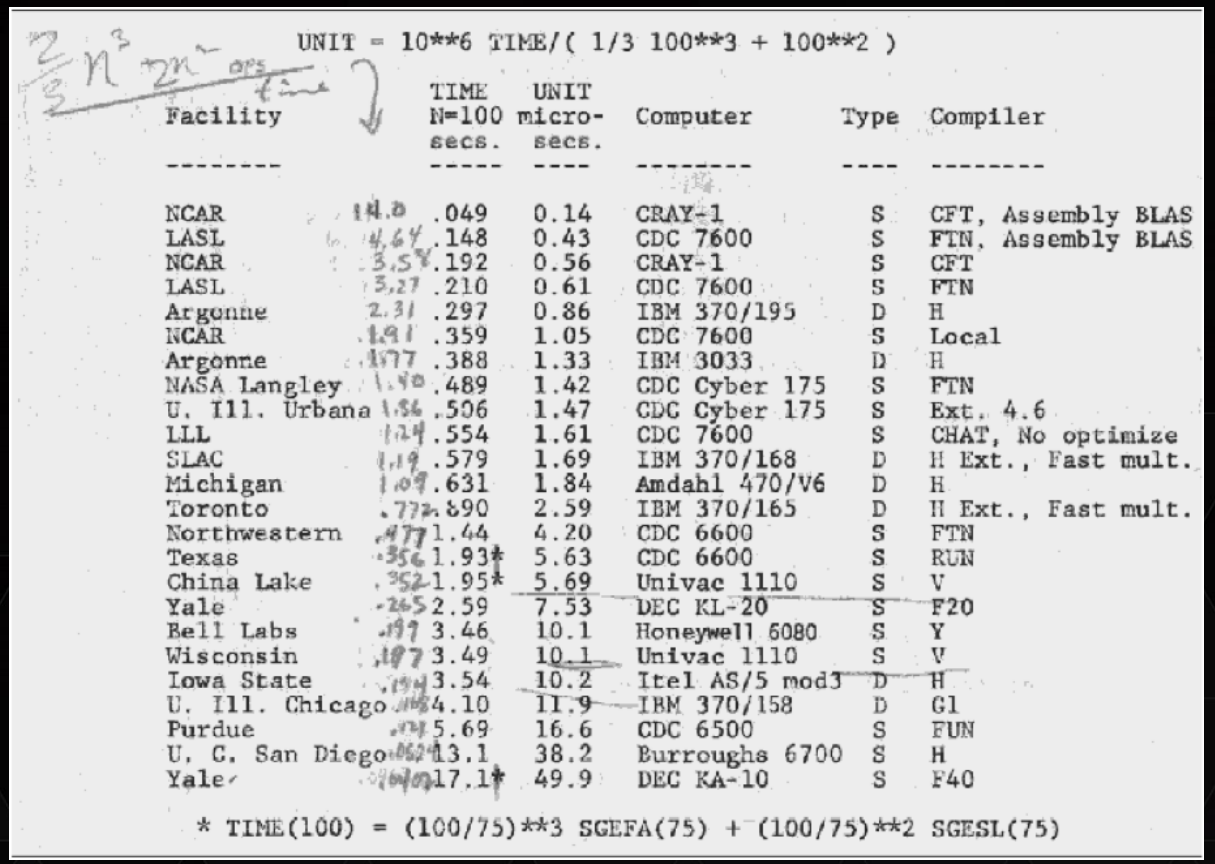
CRAY-1

LATENCY-HIDING SINGLE VECTOR CPU, 160 MFLOPS PEAK





FIRST LINPACK LIST, JANUARY 1979



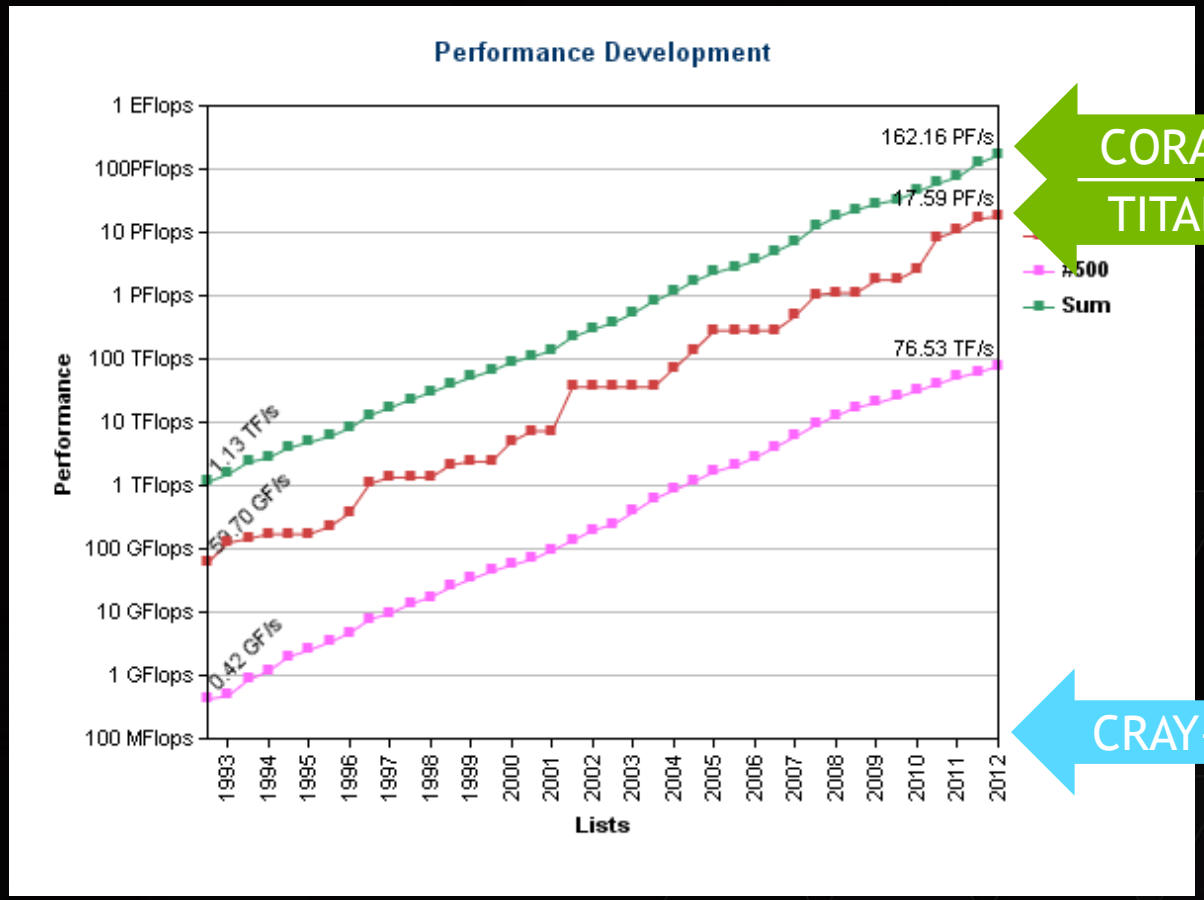
UNIT = 10**6 TIME/(1/3 100**3 + 100**2)

Facility	TIME N=100 secs.	UNIT micro-secs.	Computer	Type	Compiler
NCAR	.049	0.14	CRAY-1	S	CFT, Assembly BLAS
LASL	.148	0.43	CDC 7600	S	FTN, Assembly BLAS
NCAR	.192	0.56	CRAY-1	S	CFT
LASL	.210	0.61	CDC 7600	S	FTN
Argonne	.297	0.86	IBM 370/195	D	H
NCAR	.359	1.05	CDC 7600	S	Local
Argonne	.388	1.33	IBM 3033	D	H
NASA Langley	.489	1.42	CDC Cyber 175	S	FTN
U. Ill. Urbana	.506	1.47	CDC Cyber 175	S	Ext. 4.6
LLL	.554	1.61	CDC 7600	S	CHAT, No optimize
SLAC	.579	1.69	IBM 370/168	D	H Ext., Fast mult.
Michigan	.631	1.84	Amdahl 470/V6	D	H
Toronto	.890	2.59	IBM 370/165	D	H Ext., Fast mult.
Northwestern	1.44	4.20	CDC 6600	S	FTN
Texas	1.93*	5.63	CDC 6600	S	RUN
China Lake	1.95*	5.69	Univac 1110	S	V
Yale	2.59	7.53	DEC KL-20	S	F20
Bell Labs	3.46	10.1	Honeywell 6080	S	Y
Wisconsin	3.49	10.1	Univac 1110	S	V
Iowa State	3.54	10.2	Itel AS/5 mod3	D	H
U. Ill. Chicago	4.10	11.9	IBM 370/158	D	G1
Purdue	5.69	16.6	CDC 6500	S	FUN
U. C. San Diego	13.1	38.2	Burroughs 6700	S	H
Yale	17.1*	49.9	DEC KA-10	S	F40

* TIME(100) = (100/75)**3 SGEFA(75) + (100/75)**2 SGESL(75)

TOP500 SUPERCOMPUTER LIST 1993-2012

9 ORDERS OF MAGNITUDE

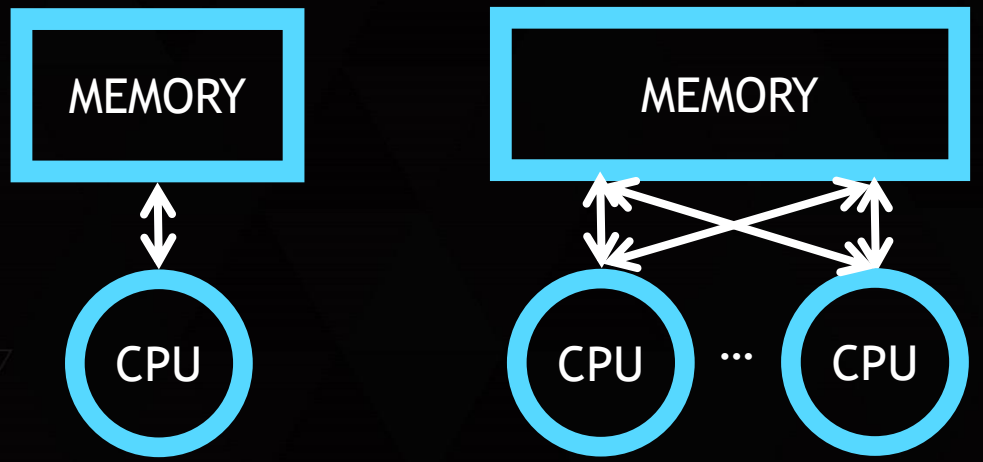


CORAL

TITAN

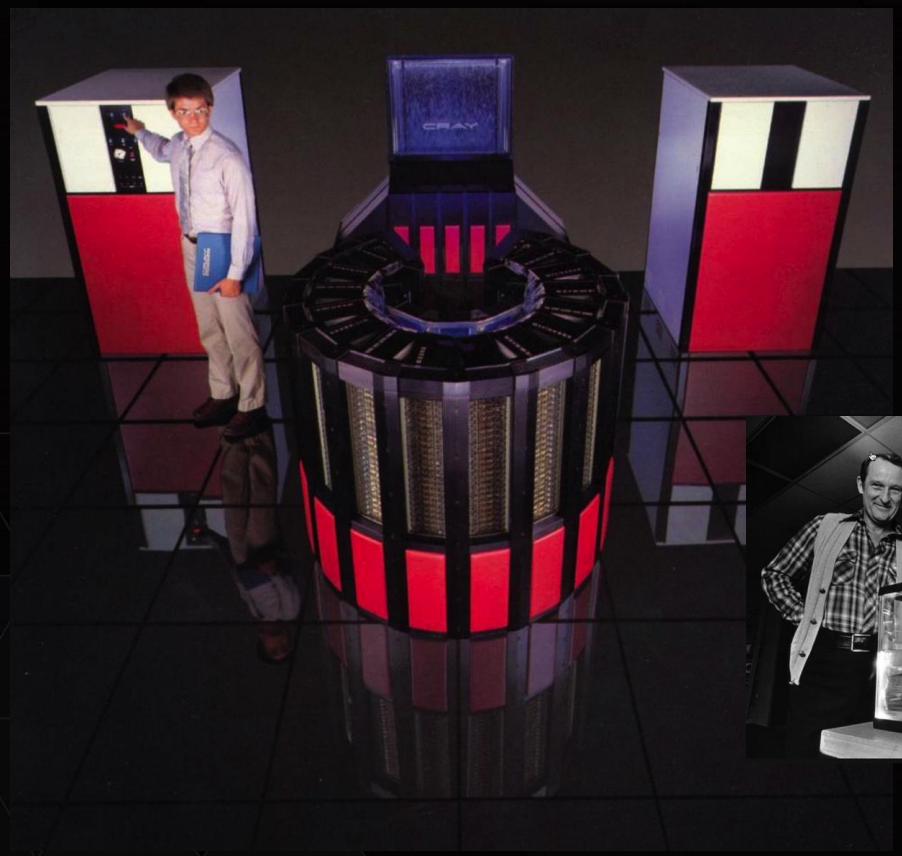
CRAY-1

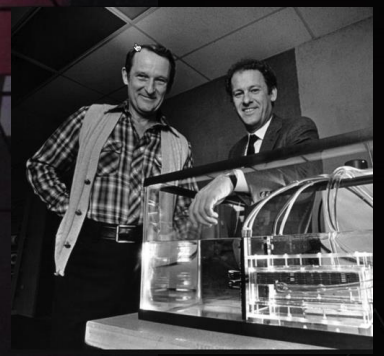
HPC SYSTEM EVOLUTION



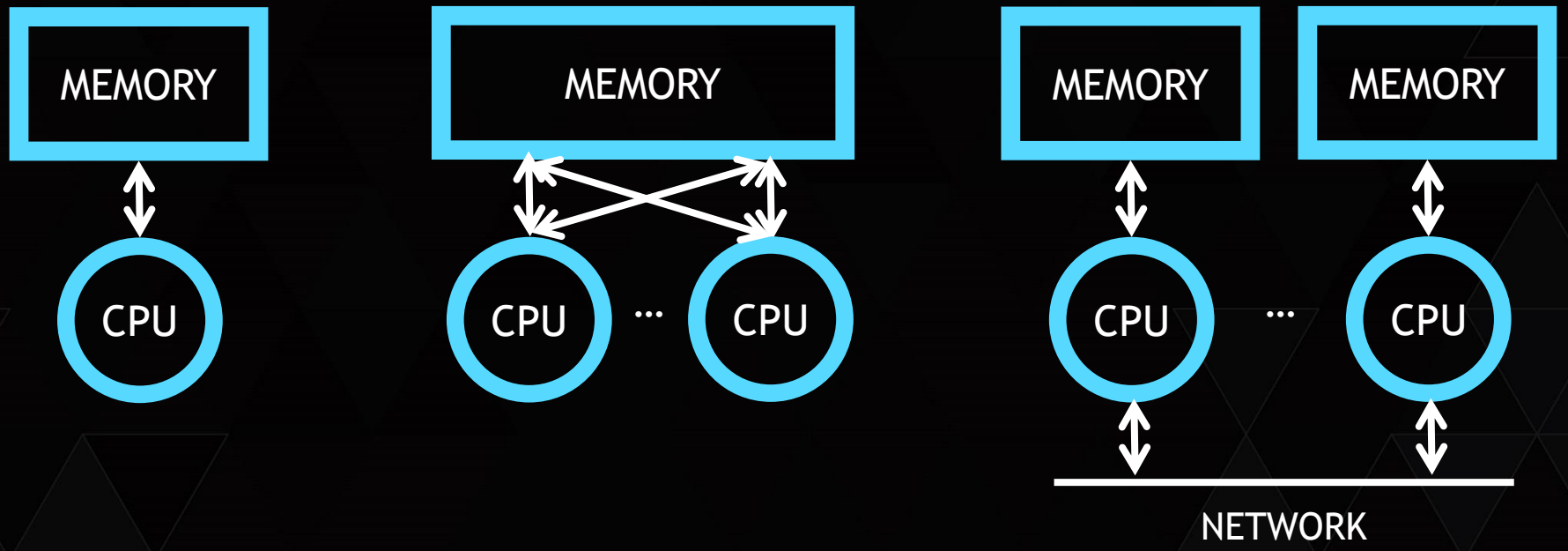
CRAY-2

4 LATENCY-HIDING VECTOR CPUS, 2 GFLOPS PEAK, 1985





HPC SYSTEM EVOLUTION



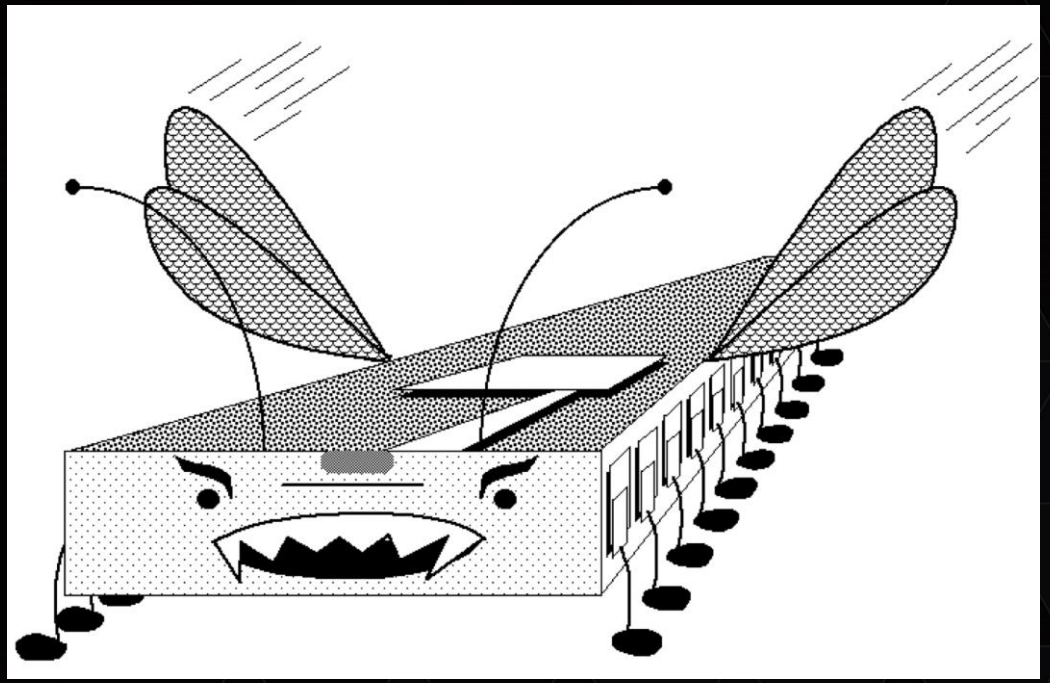
“ATTACK OF THE KILLER MICROS”

The Attack of the Killer Micros

E. D. Brooks III
Massively Parallel Computing Initiative
Lawrence Livermore National Laboratories

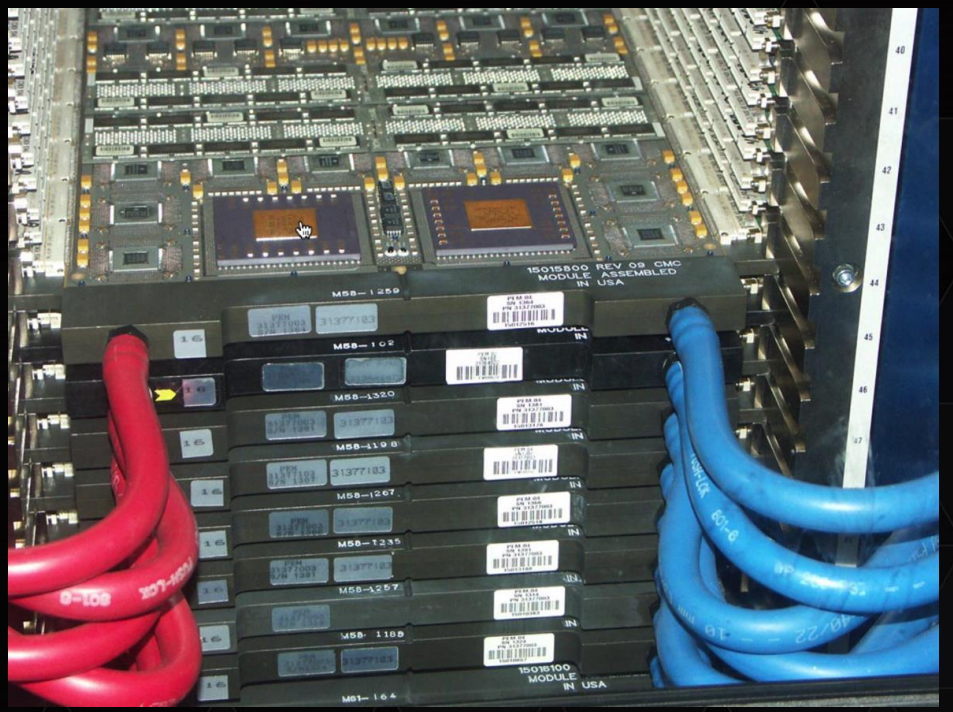
brooks@{maddog.llnl.gov, maddog.uucp}

Presented at: Supercomputing ’89
Reno, Nov 13-17 1989

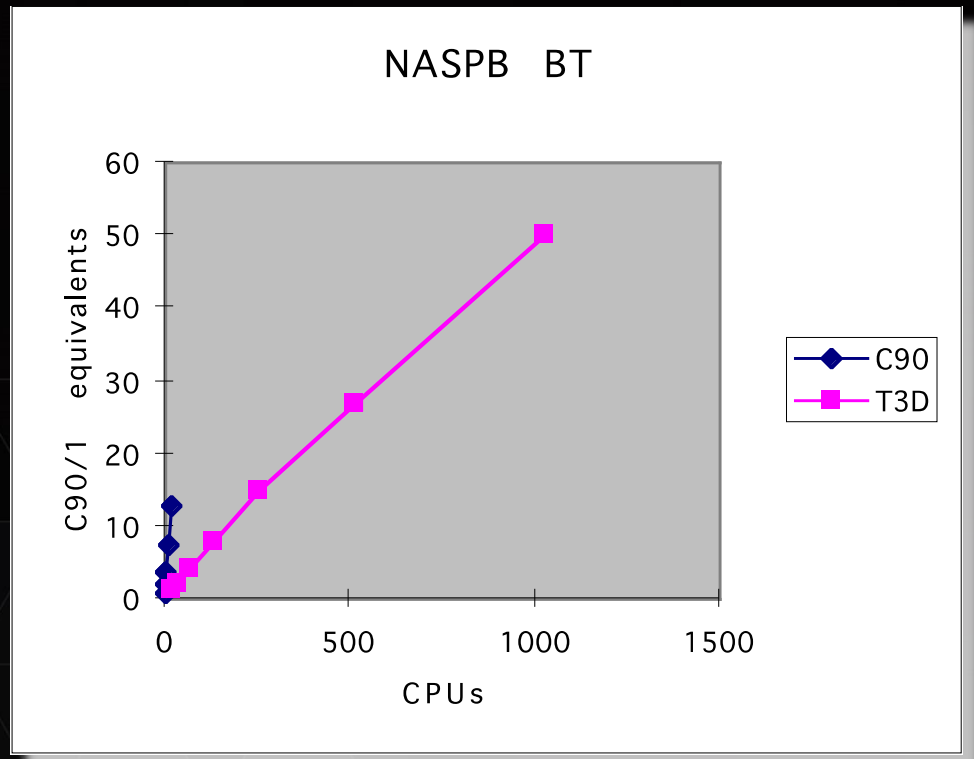


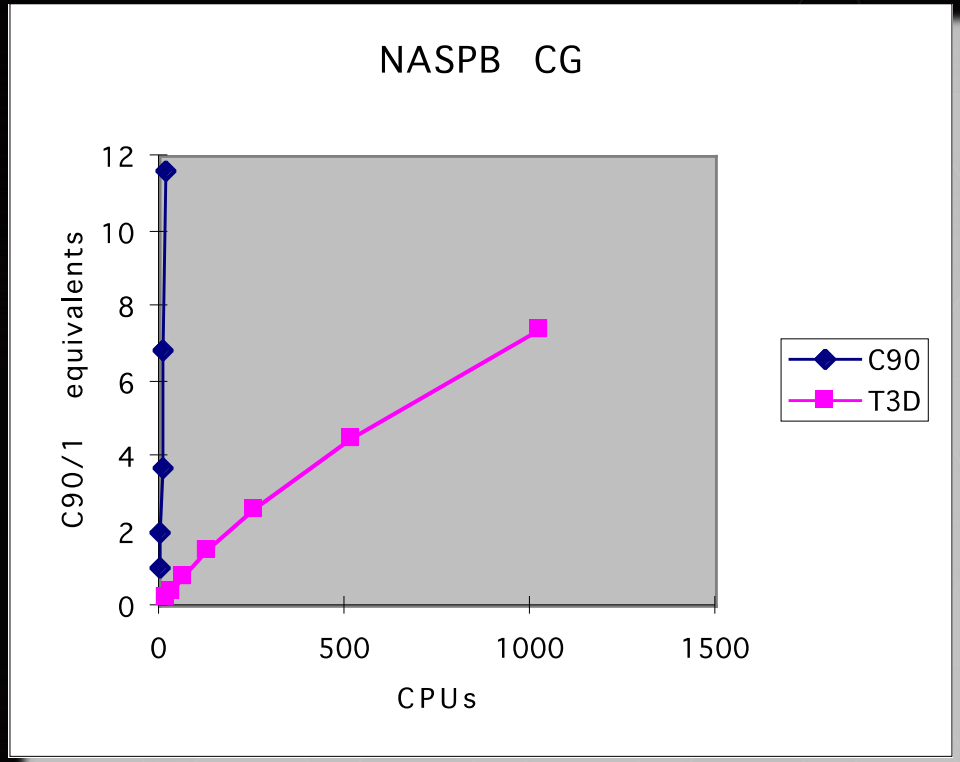
CRAY T3D

DEC ALPHA EV4 MICROPROCESSORS, 1 TFLOPS PEAK, 1993



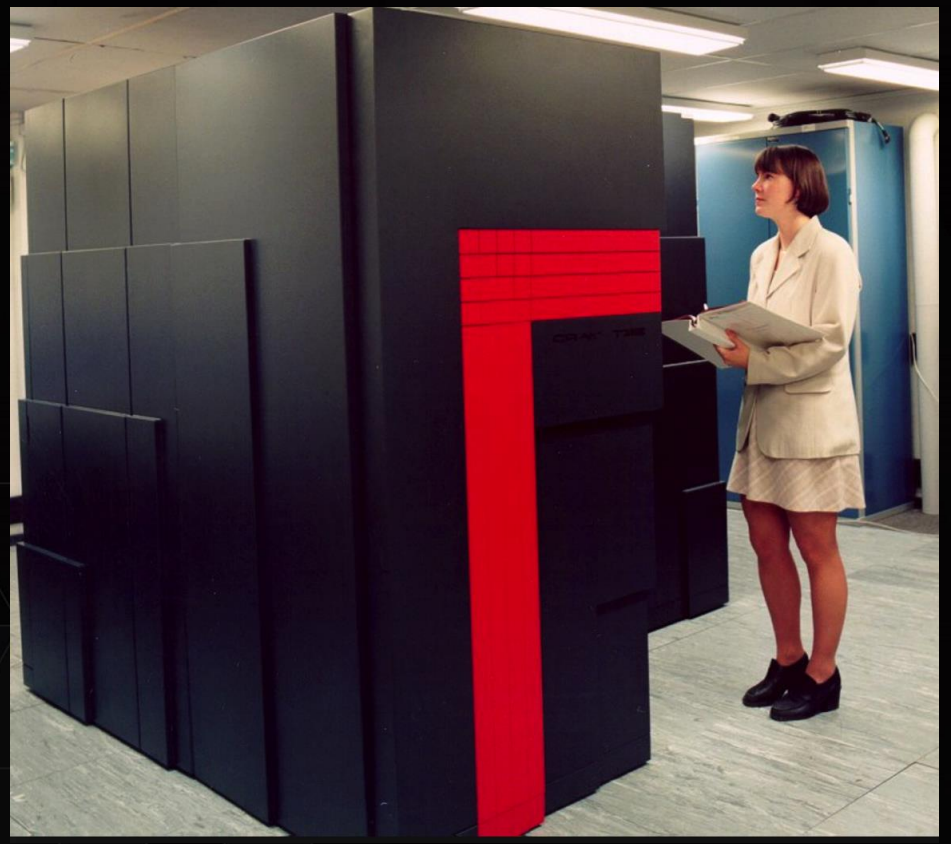
HETEROGENEOUS RESULTS

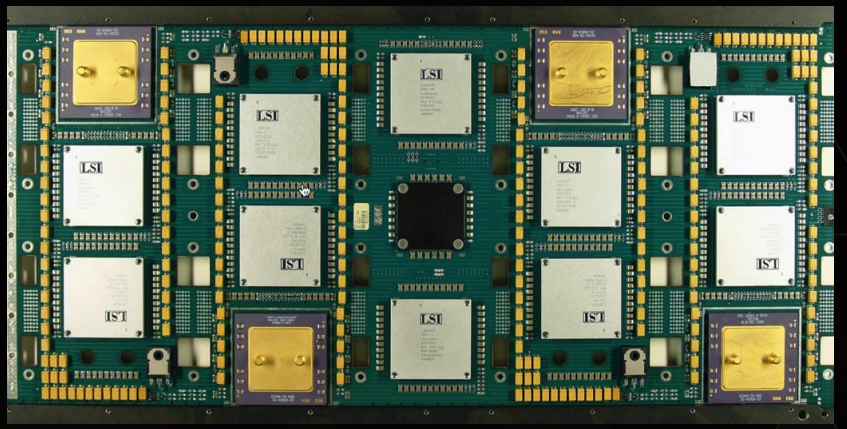


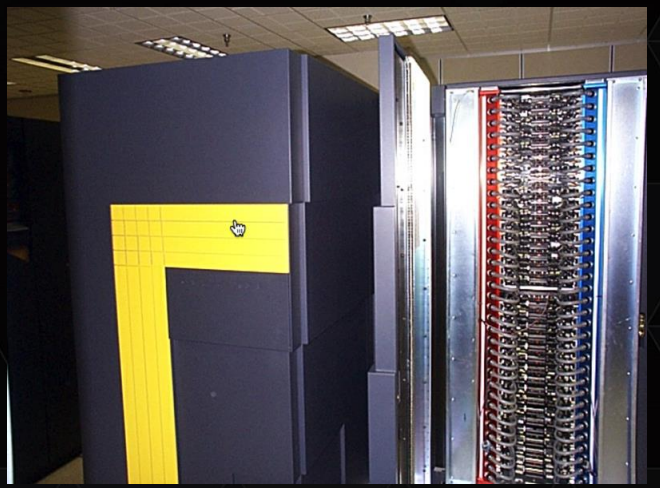


CRAY T3E

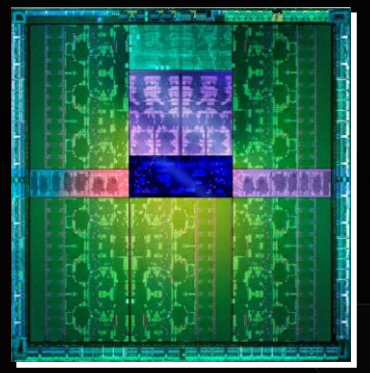
DEC ALPHA EV5 MICROPROCESSORS, 1 TFLOPS SUSTAINED, 1995







FUTURE SHOCK



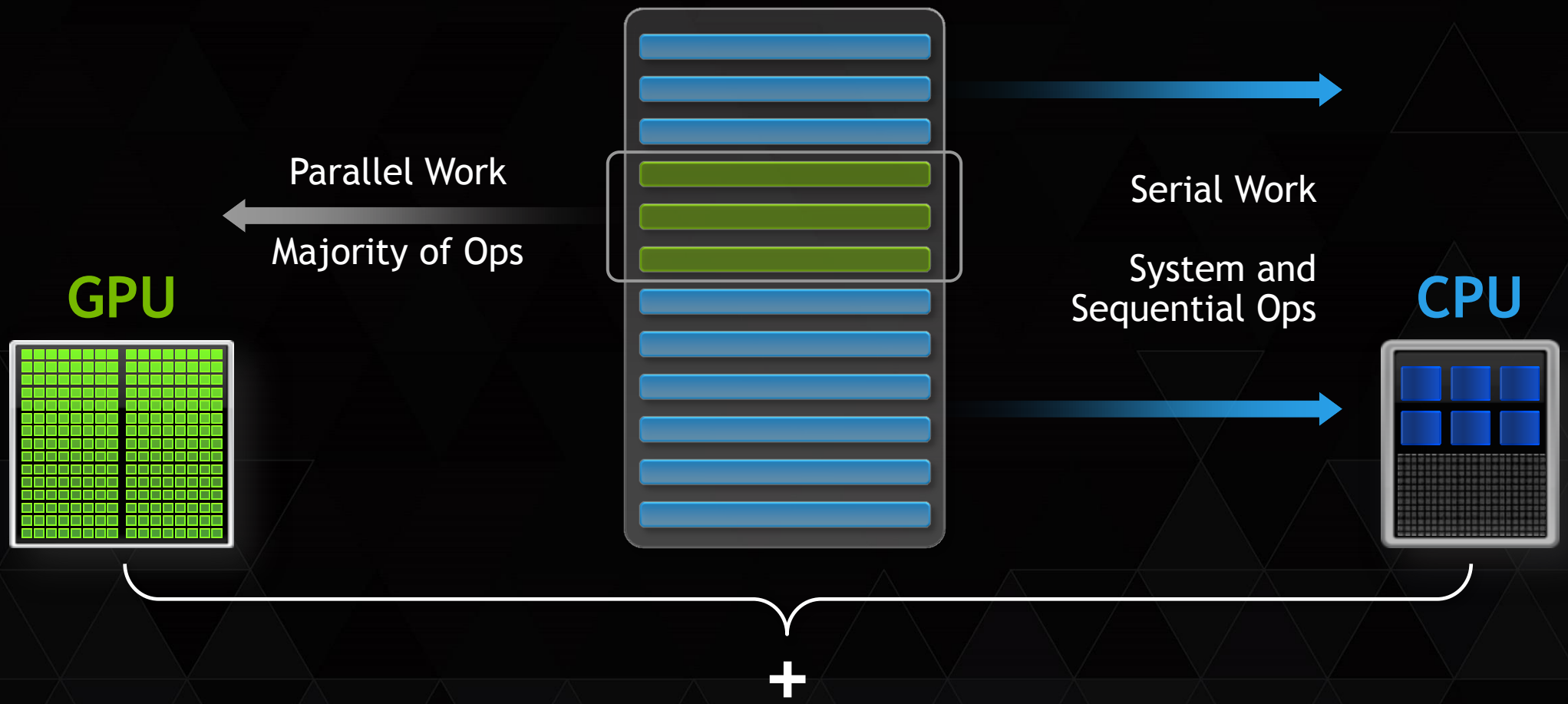
"There's hasn't really been anything truly new for 20 years… except GPUs."

--Thomas Schulthess, Director, CSCS

COMING SOON TO A NODE NEAR YOU

- Cray T3E
  - 2.4 TFLOPS peak performance (2K processors)
  - ~128 GB/s bisection bandwidth
- CRAY T90
  - 56 GFLOPs peak performance (32 processors)
  - 1 TB/s bisection bandwidth
- *Pascal GPU (1H16) has >T3E peak with T90 memory BW*

OPTIMIZING SERIAL/PARALLEL EXECUTION

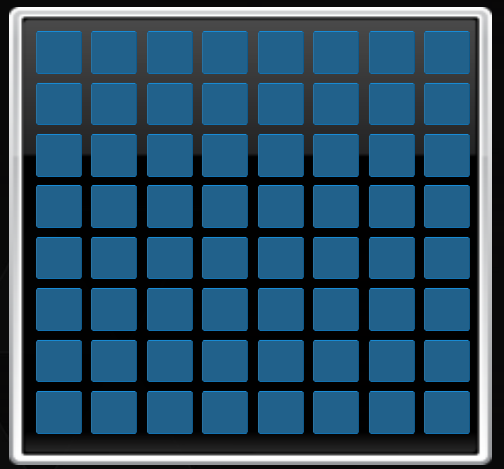


TWO COMPUTING MODELS FOR ACCELERATORS

Many-Weak-Cores (MWC) Model

Single CPU Core for Both Serial & Parallel Work

Xeon Phi (*And Others*)
Many Weak Serial Cores

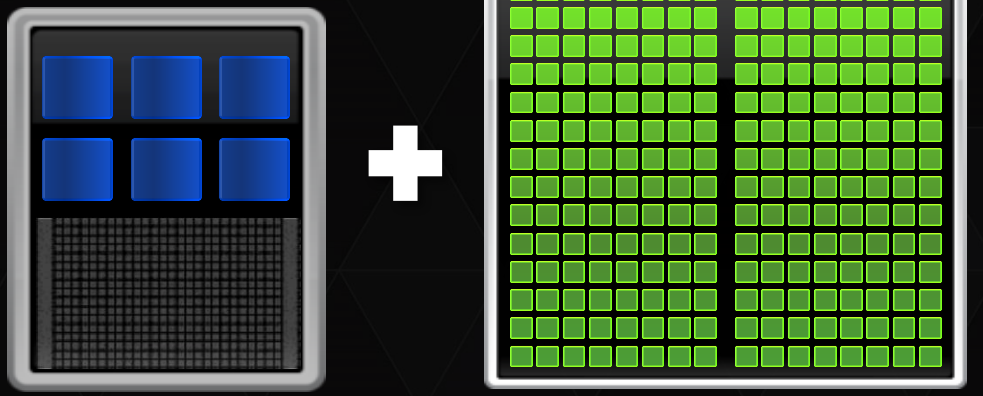


Heterogeneous Computing Model

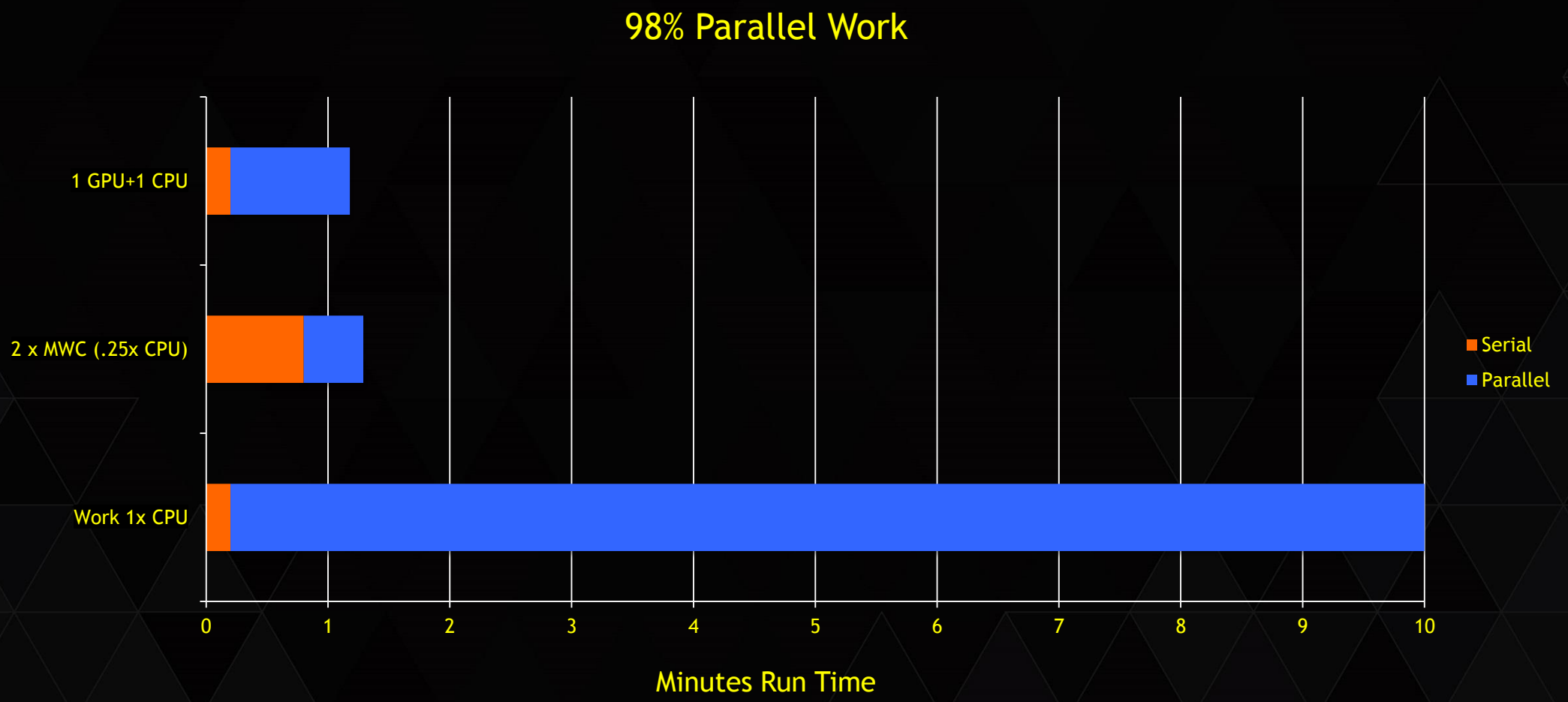
Complementary Processors Work Together

CPU
Optimized for
Serial Tasks

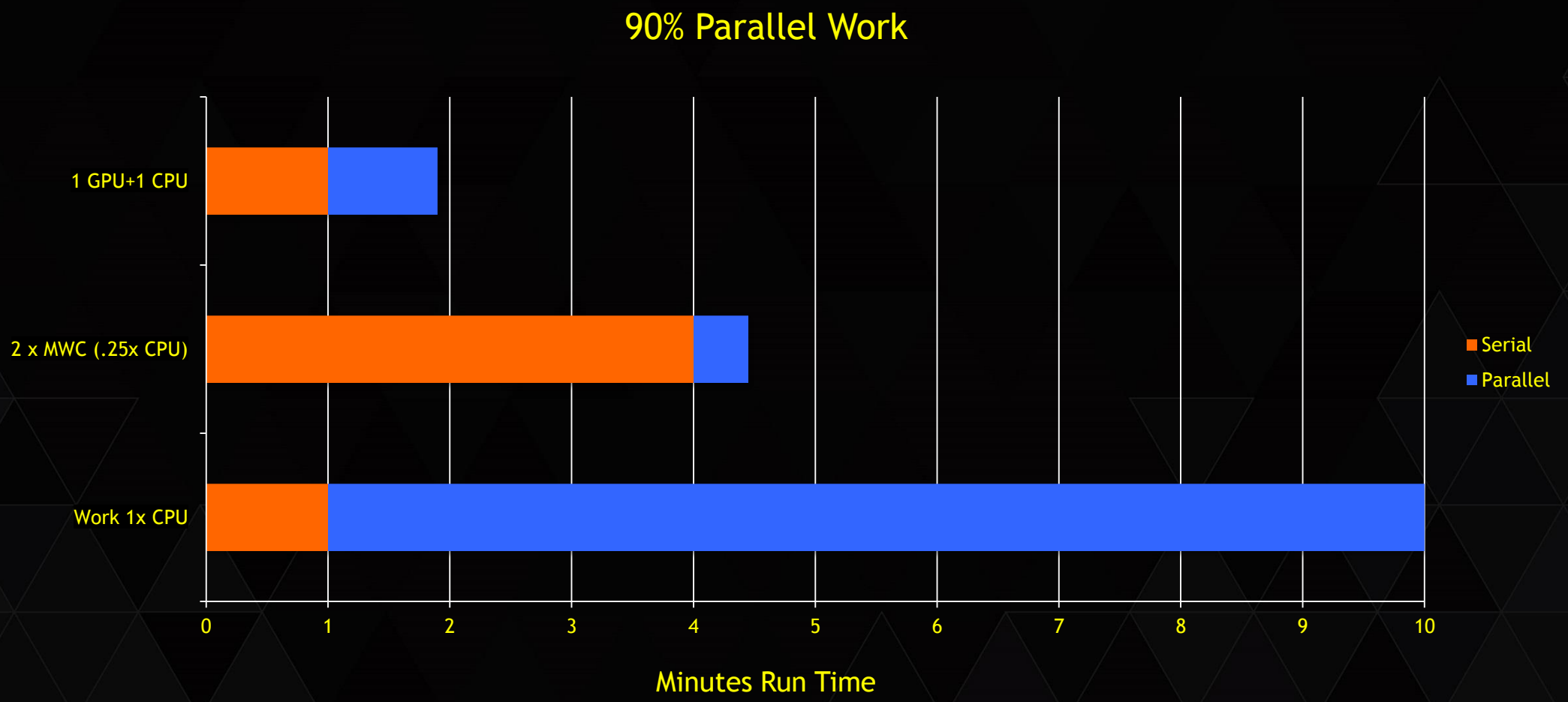
GPU Accelerator
Optimized for
Parallel Tasks



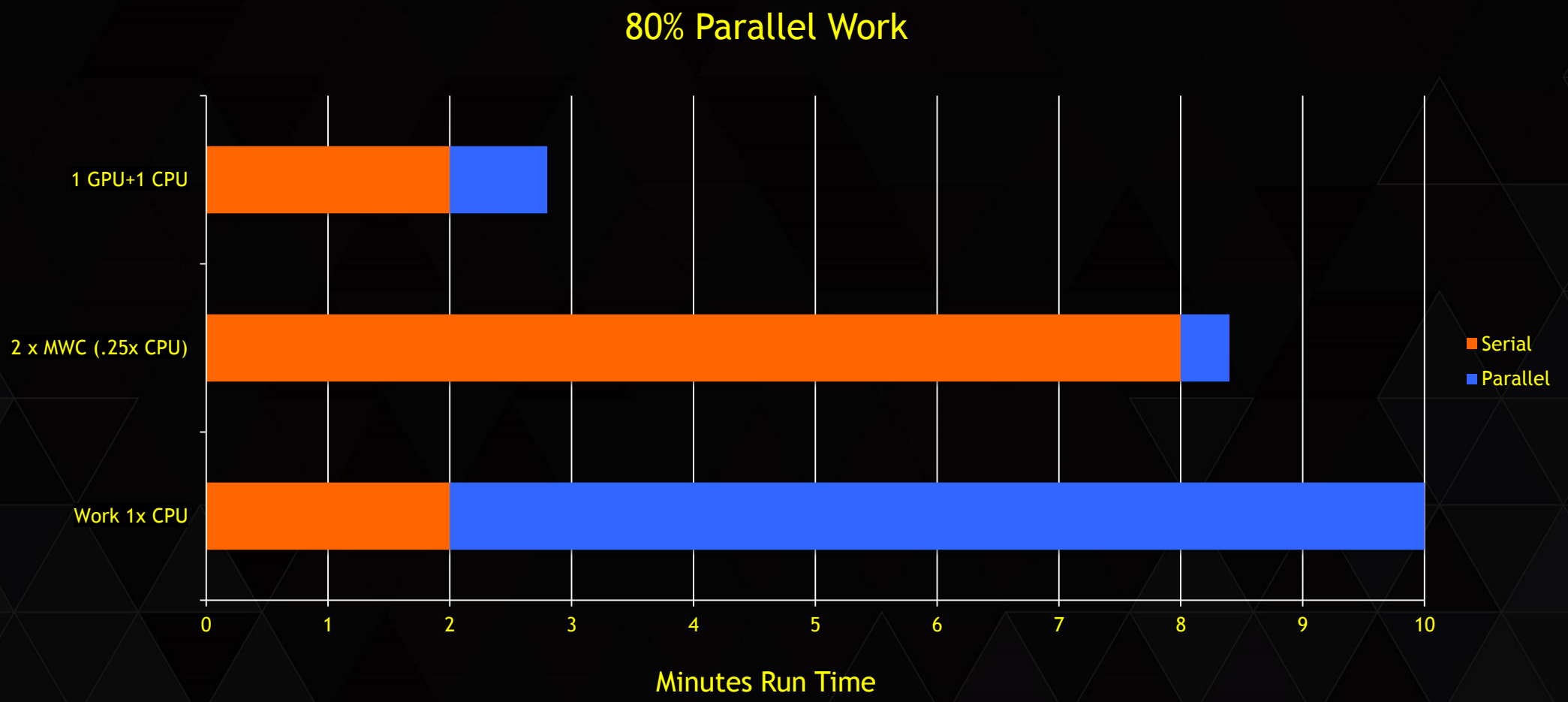
AMDAHL'S LAW ANALYSIS



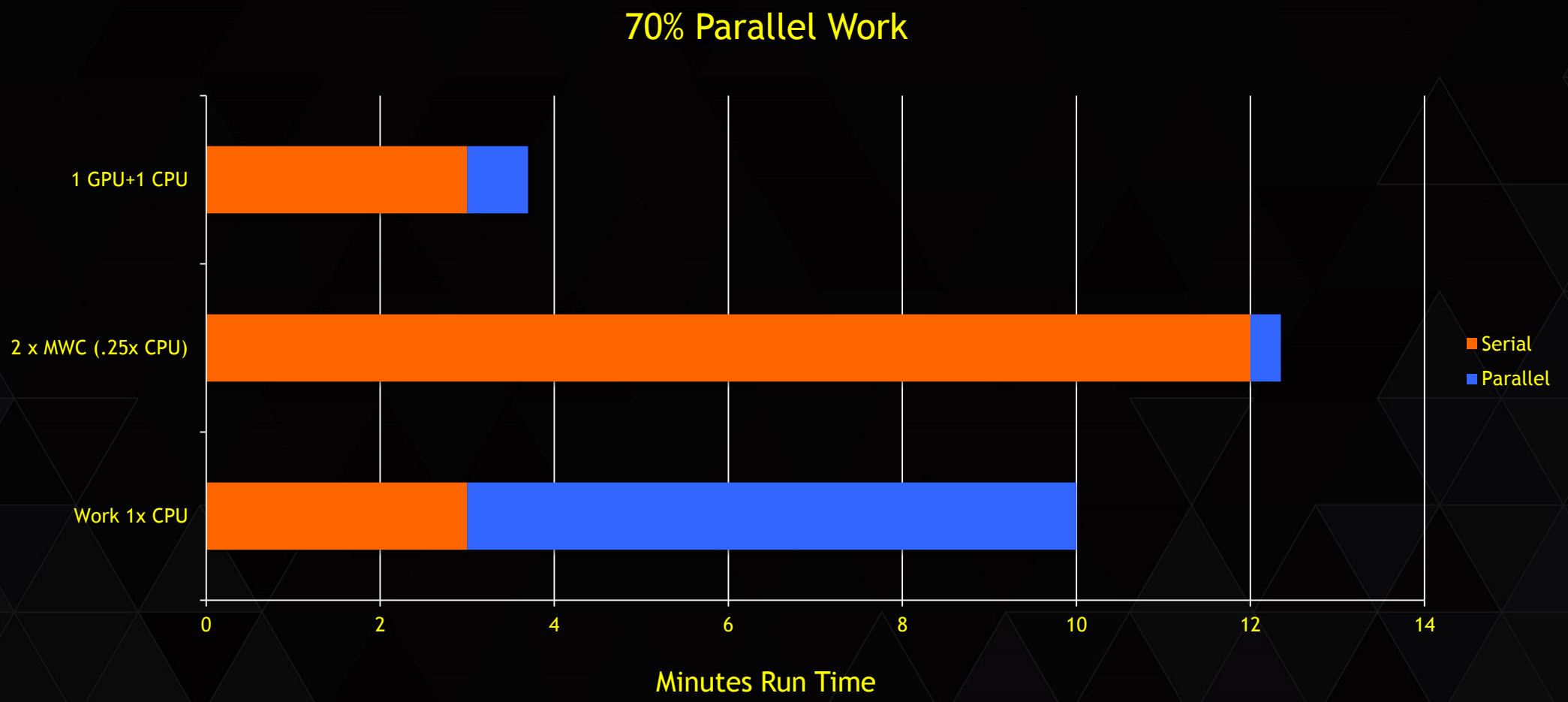
AMDAHL'S LAW ANALYSIS



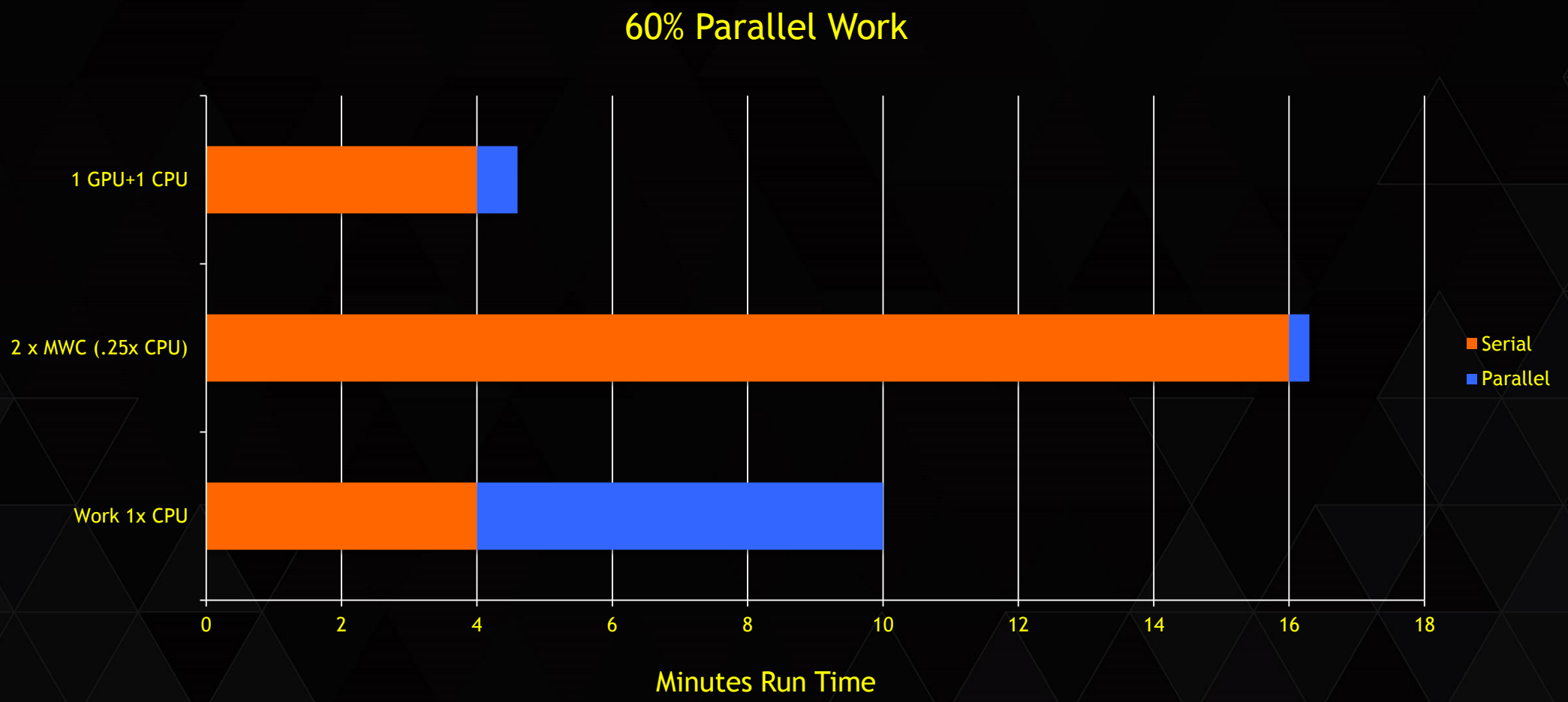
AMDAHL'S LAW ANALYSIS



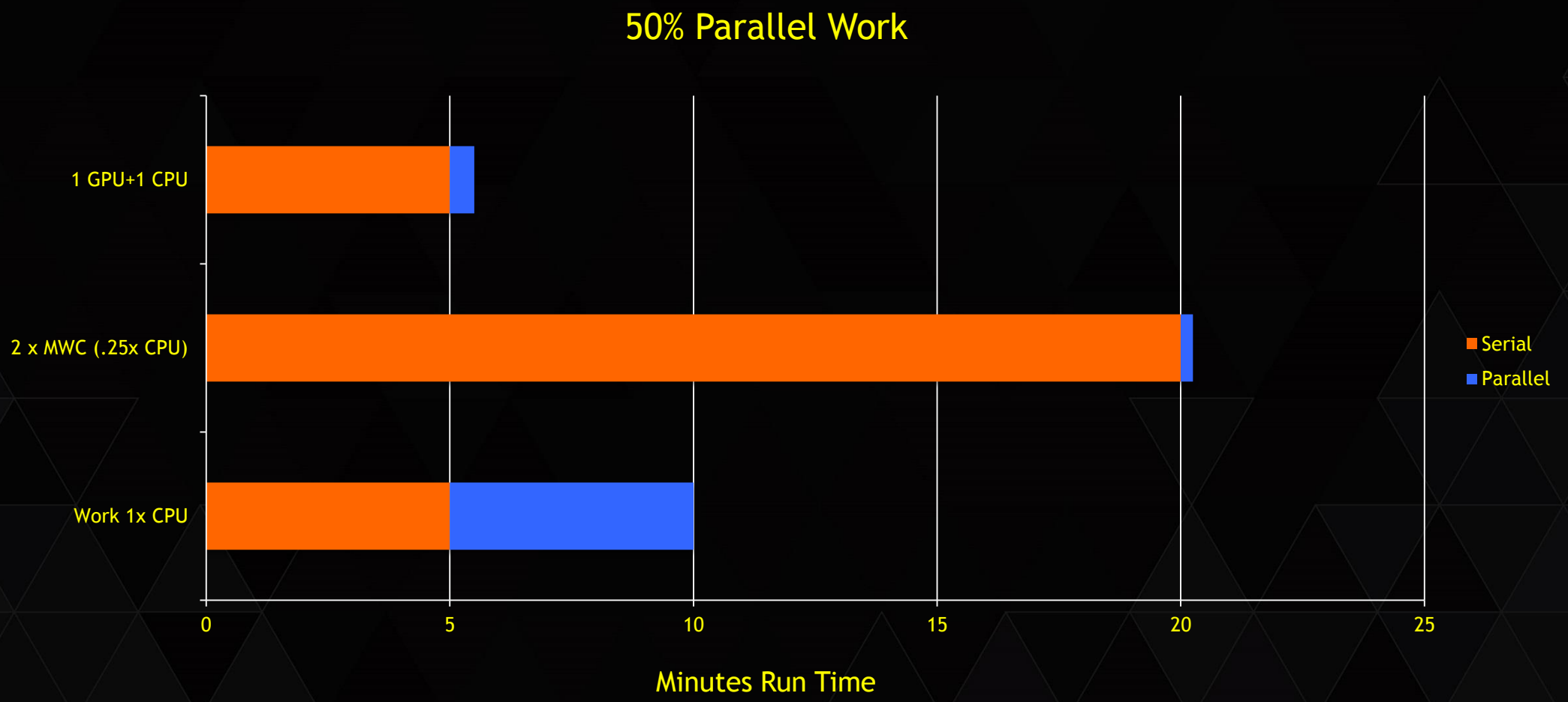
AMDAHL'S LAW ANALYSIS



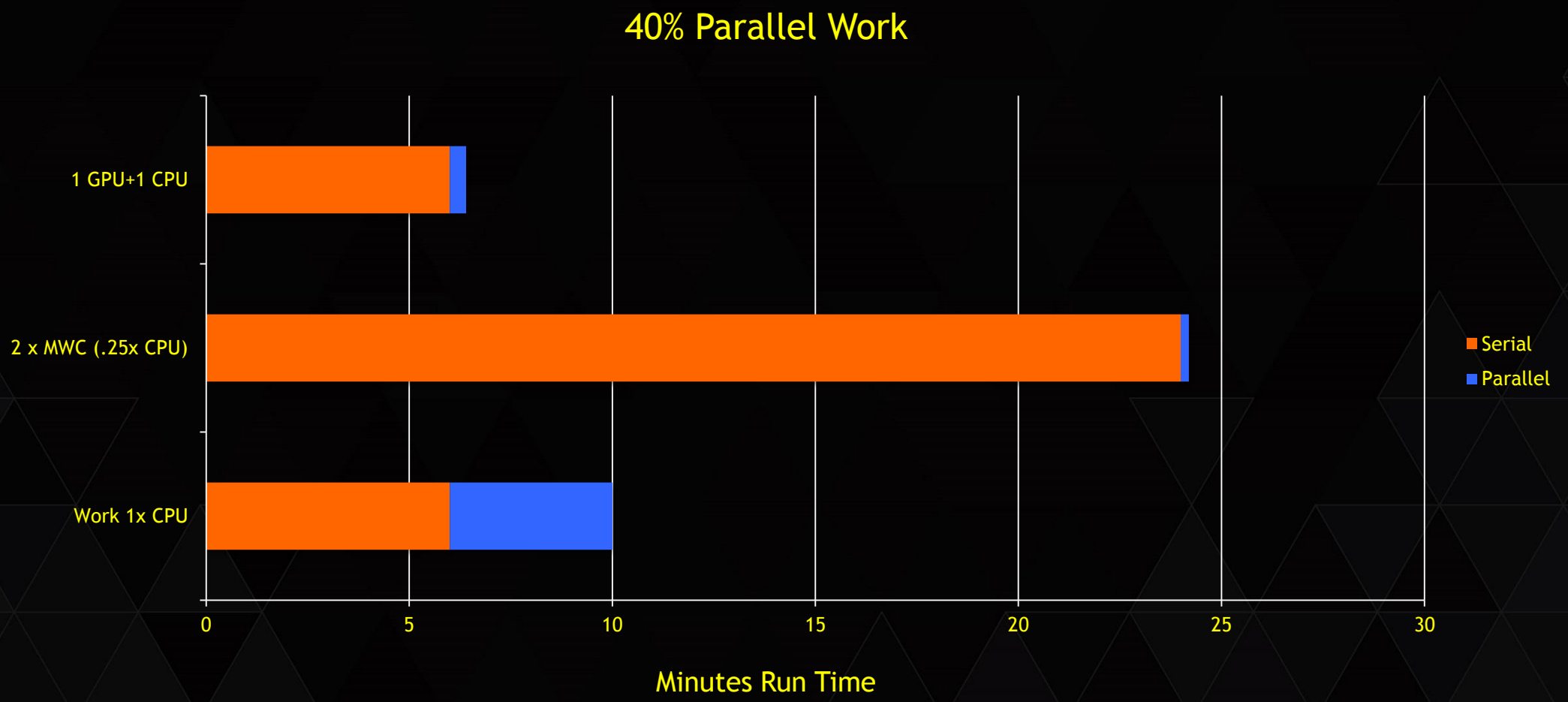
AMDAHL'S LAW ANALYSIS



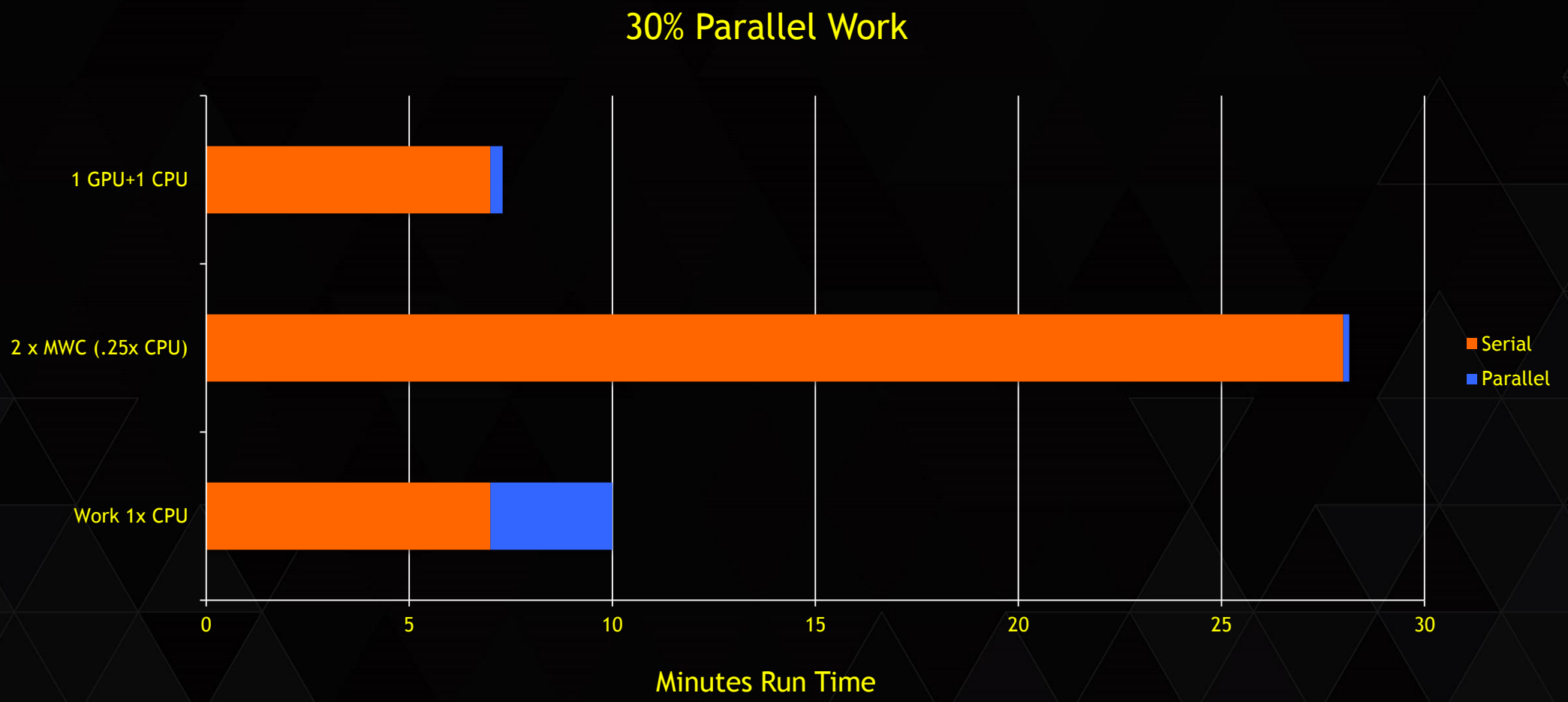
AMDAHL'S LAW ANALYSIS



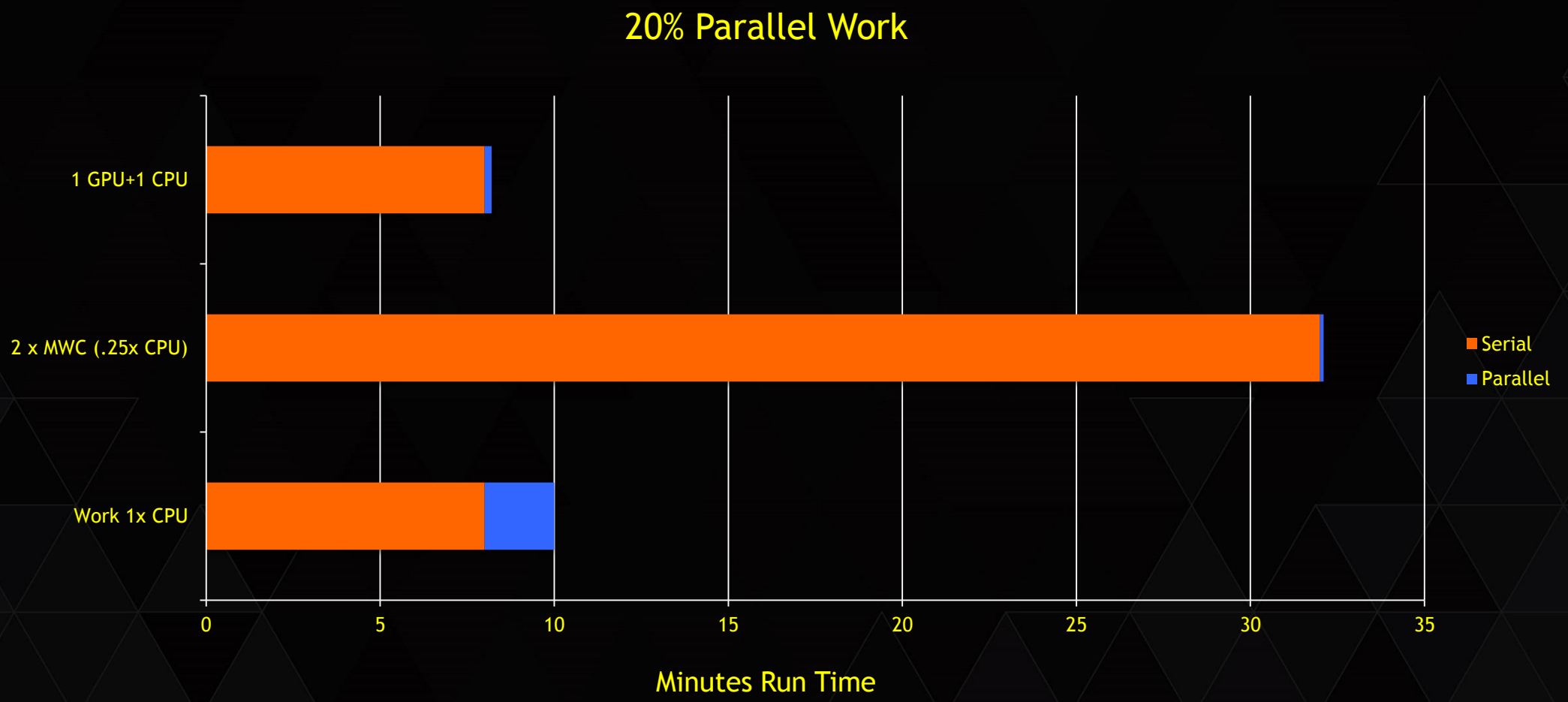
AMDAHL'S LAW ANALYSIS



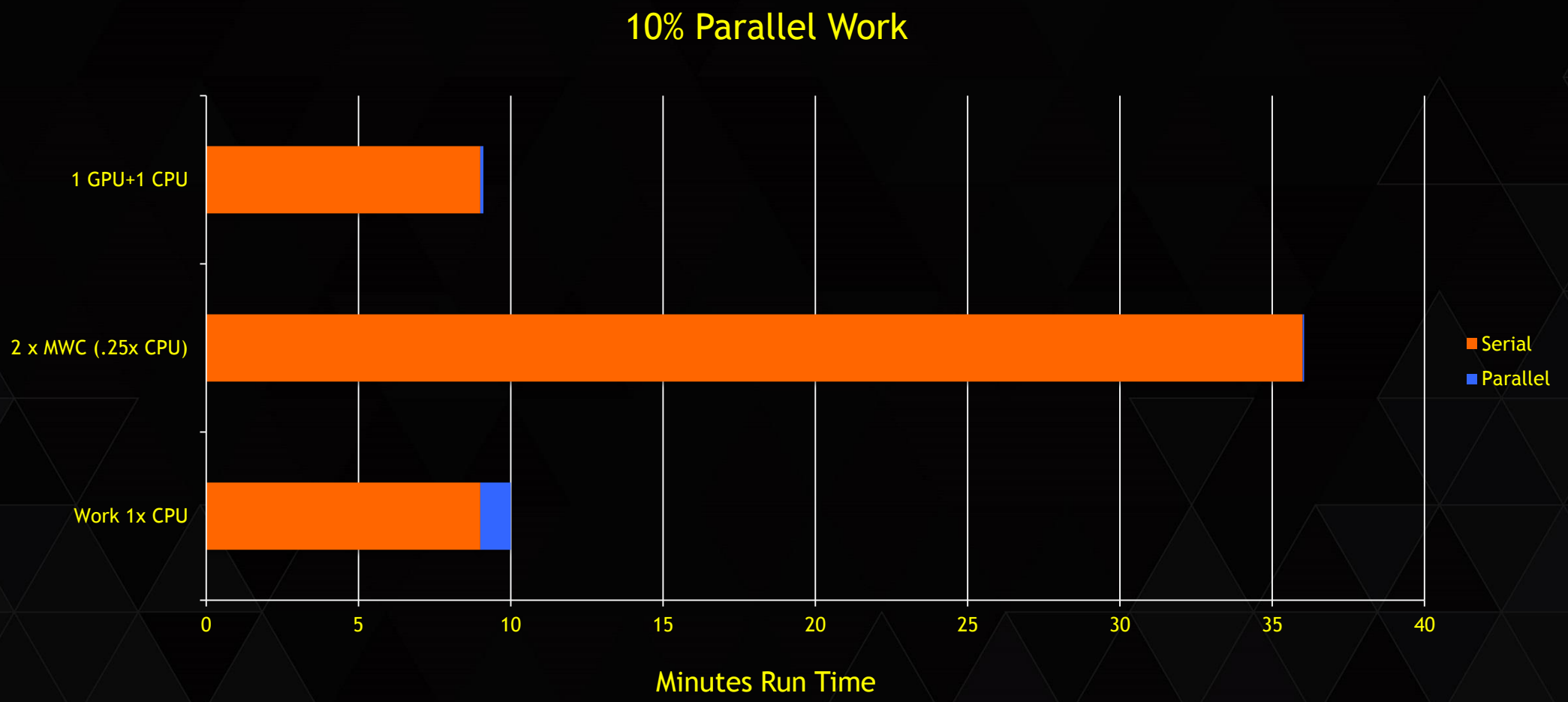
AMDAHL'S LAW ANALYSIS



AMDAHL'S LAW ANALYSIS

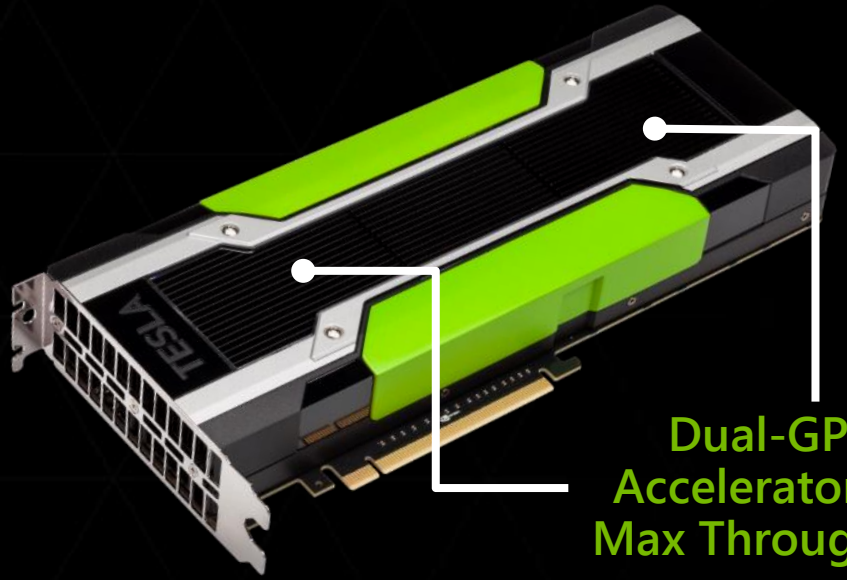


AMDAHL'S LAW ANALYSIS



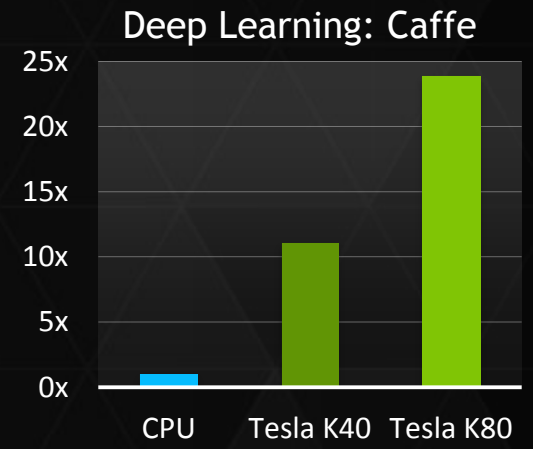
TESLA K80

WORLD'S FASTEST ACCELERATOR FOR DATA ANALYTICS AND SCIENTIFIC COMPUTING



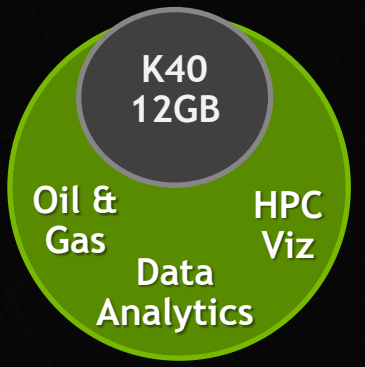
2x Faster

2.9 TF| 4992 Cores | 480 GB/s



Double the Memory

Designed for Big Data Apps



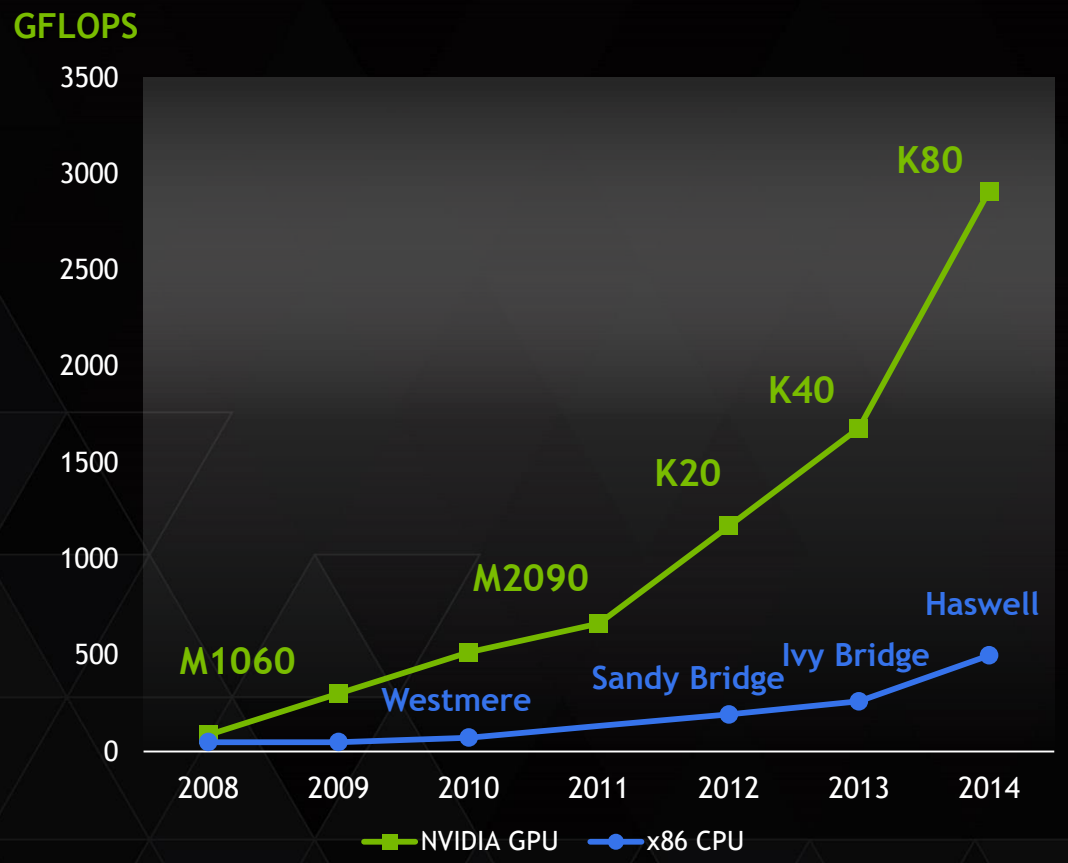
Maximum Performance

Dynamically Maximize Performance for Every Application

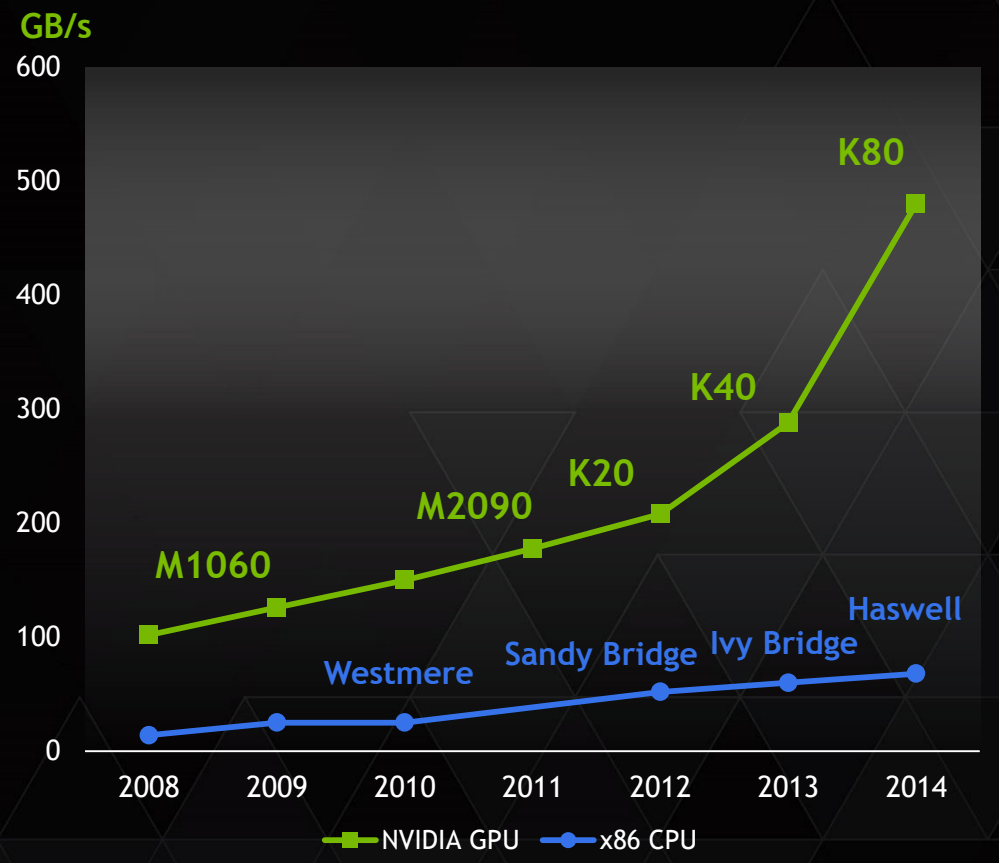
Caffe Benchmark: AlexNet training throughput based on 20 iterations, CPU: E5-2697v2 @ 2.70GHz. 64GB System Memory, CentOS 6.2

FOCUS ON THROUGHPUT PERFORMANCE

Peak Double Precision FLOPS

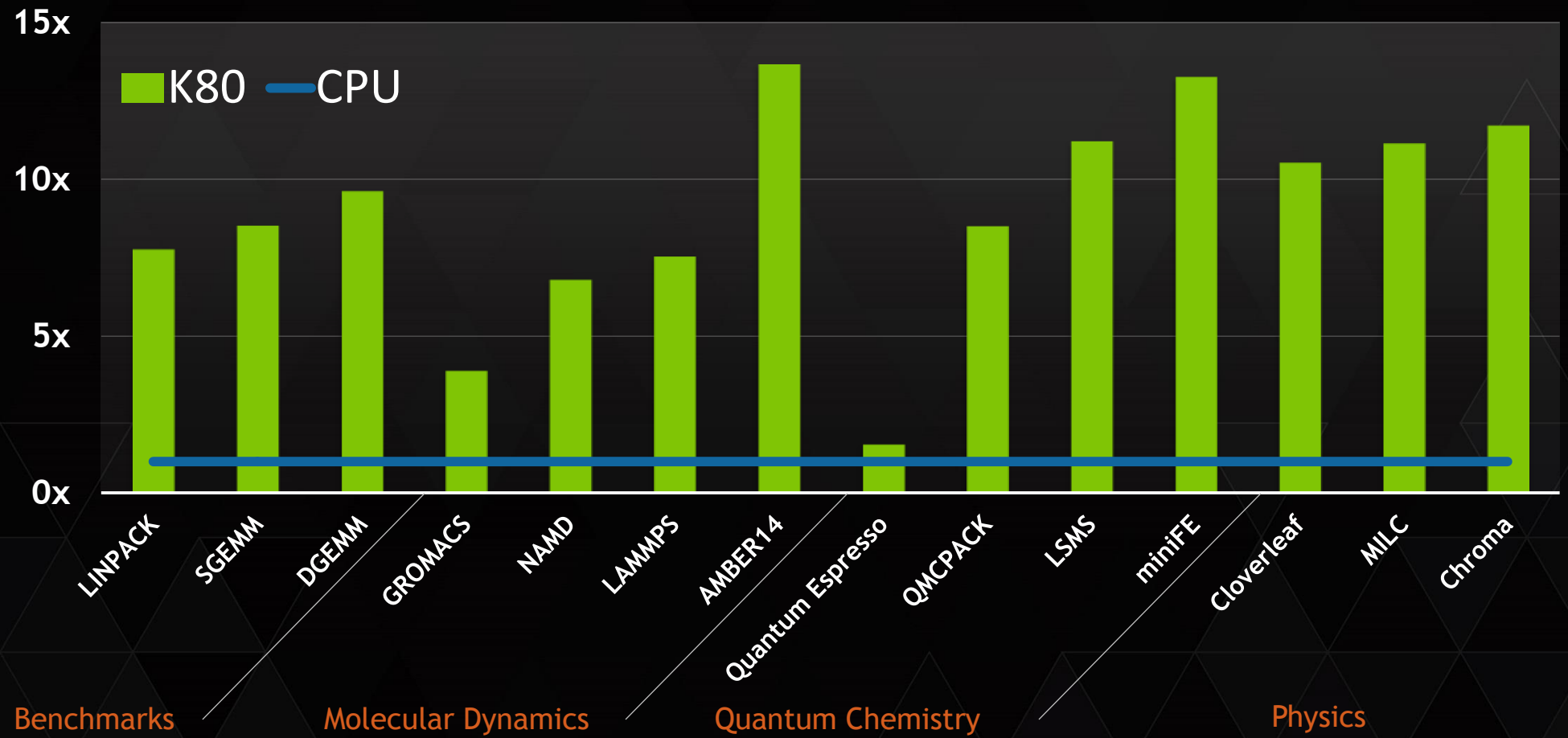


Peak Memory Bandwidth



GPU TECHNOLOGY CONFERENCE

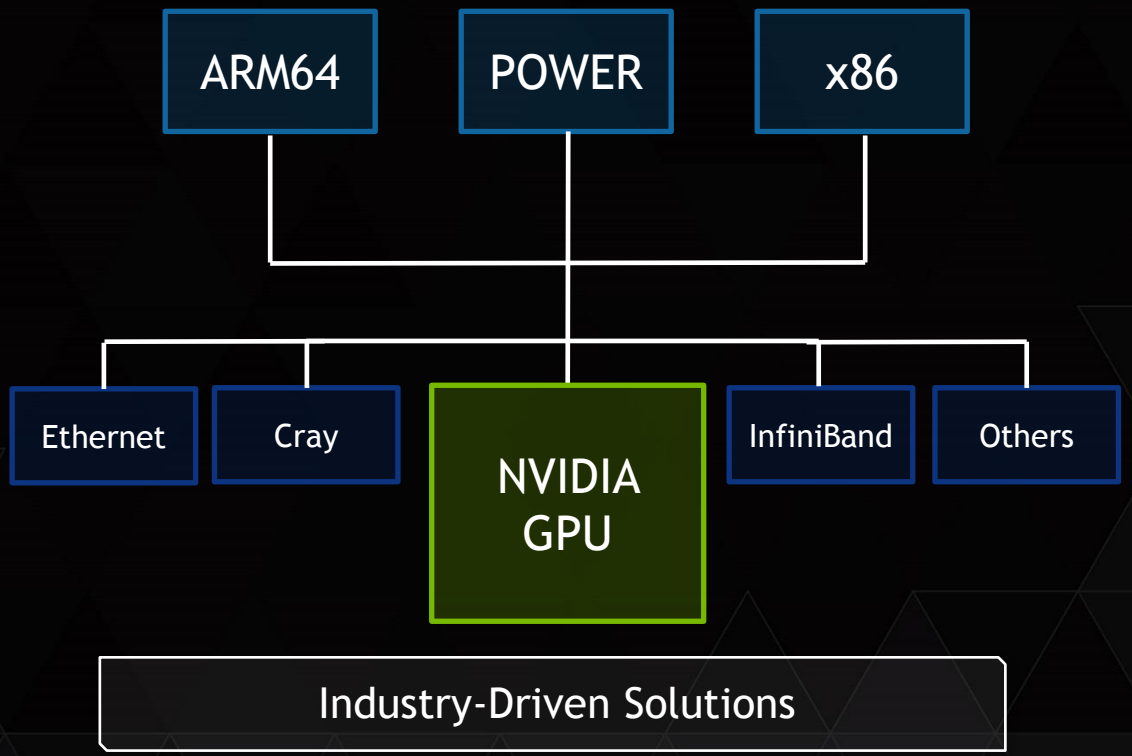
10X FASTER THAN CPU ON APPLICATIONS



CPU: 12 cores, E5-2697v2 @ 2.70GHz. 64GB System Memory, CentOS 6.2
GPU: Single Tesla K80, Boost enabled

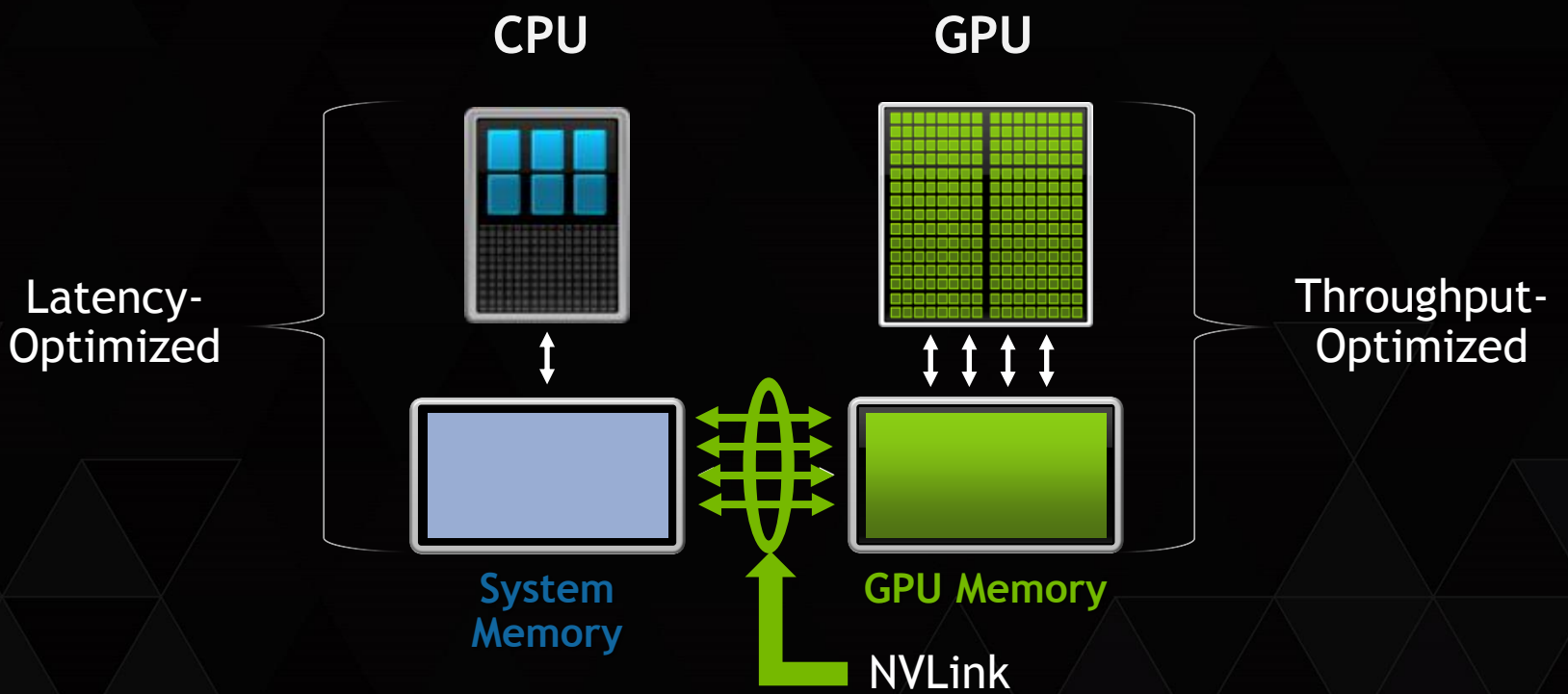
TESLA PLATFORM ENABLES OPTIMIZATION

Ecosystem Industry Standard CPUs and Interconnects



LOGICAL VS. PHYSICAL INTEGRATION

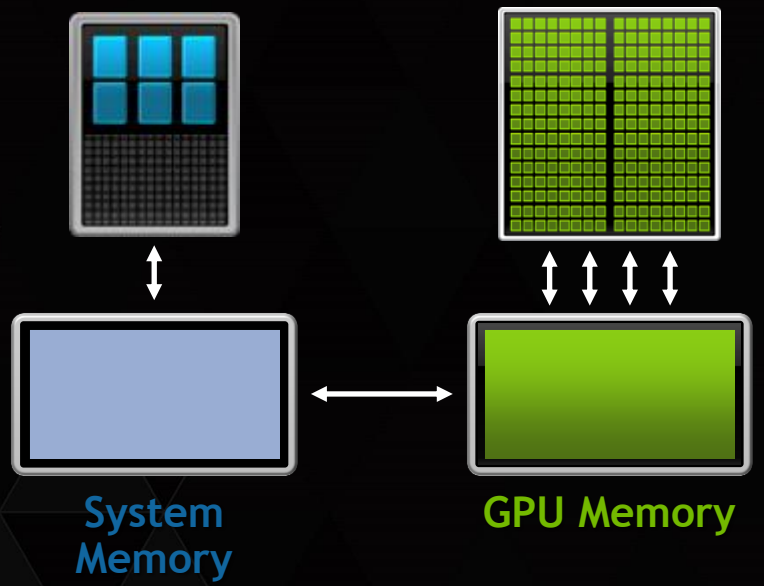
OPTIMIZING FOR EFFICIENCY + FLEXIBILITY



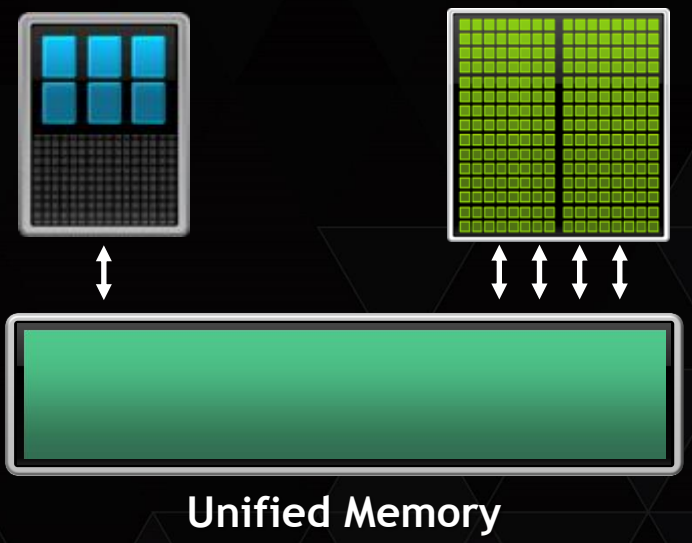
UNIFIED MEMORY

DRAMATICALLY LOWER DEVELOPER EFFORT

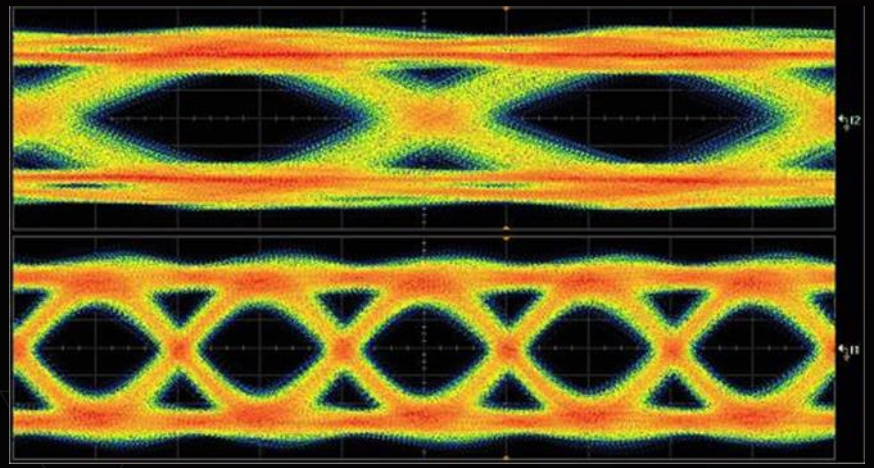
Exposed Developer View
(Pre-CUDA 5.5)



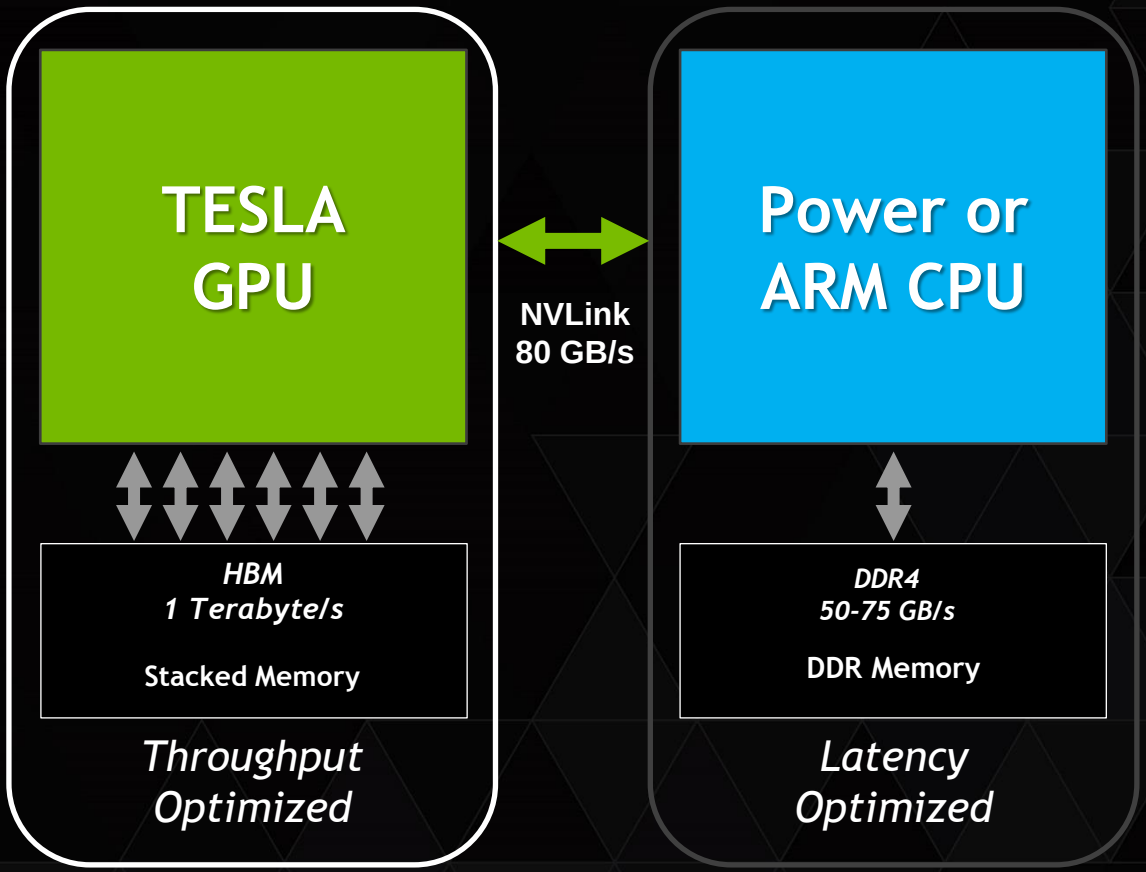
Developer View With Unified Memory



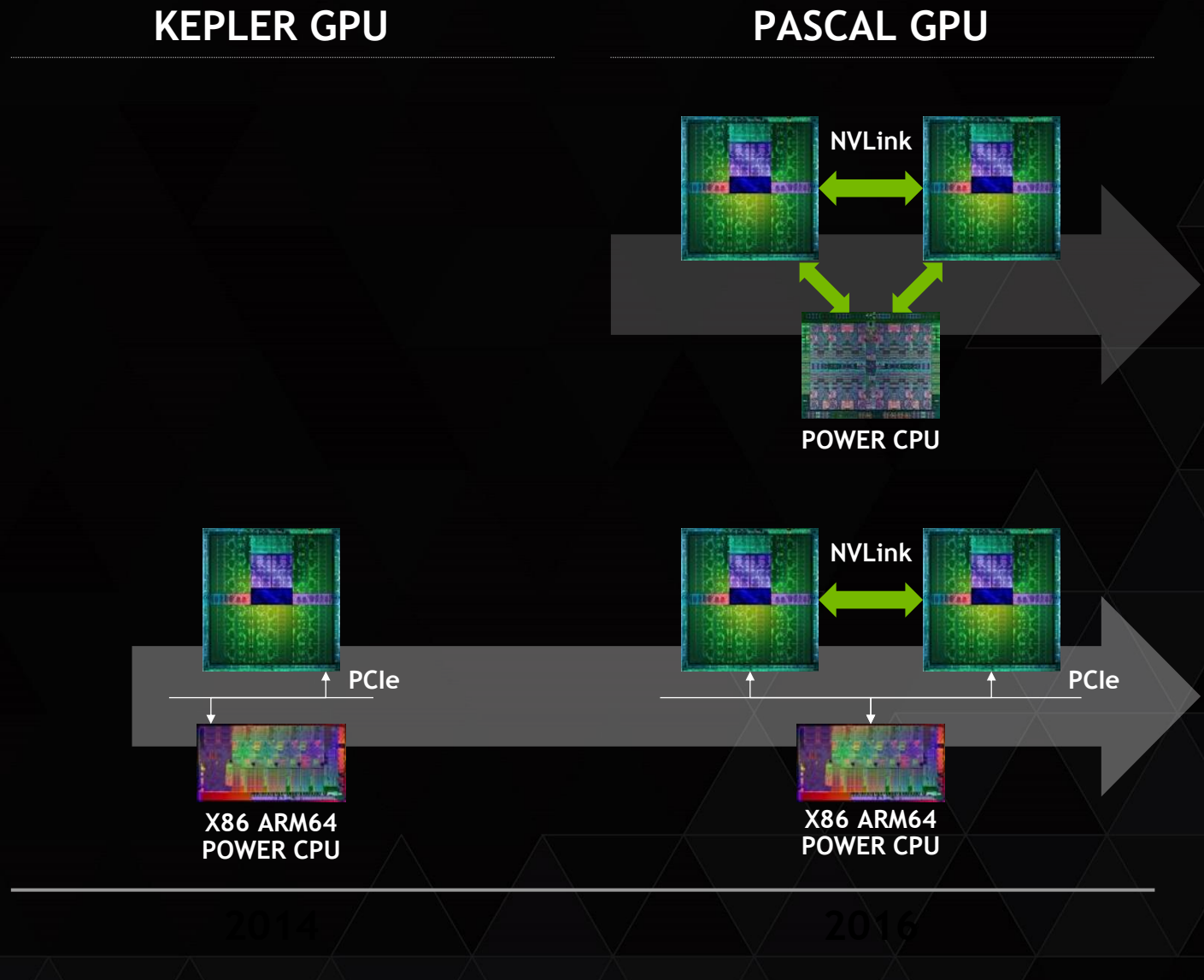
NVLINK: NODE INTEGRATION NETWORK



- 5x PCIe bandwidth
- Move data at CPU memory speed
- 3x lower energy/bit

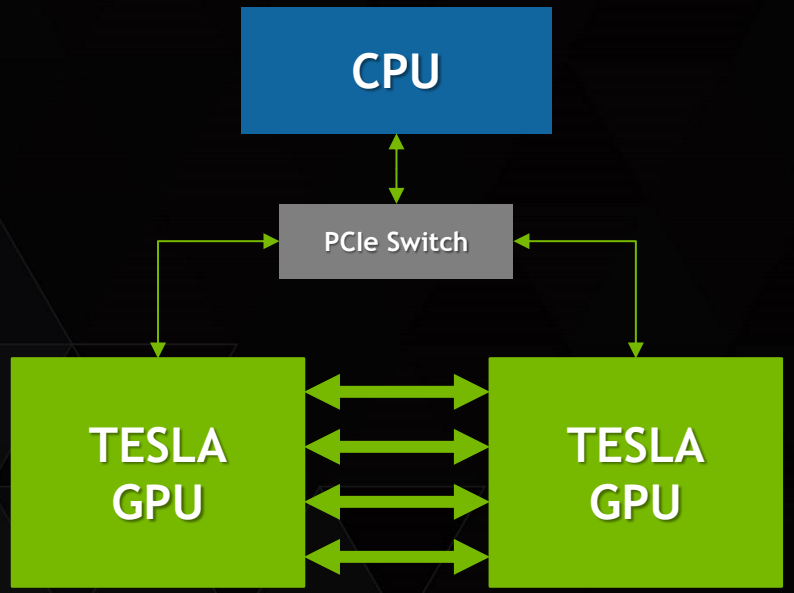


NVLINK
HIGH-SPEED NODE NETWORK



NVLINK MULTI-GPU PERFORMANCE

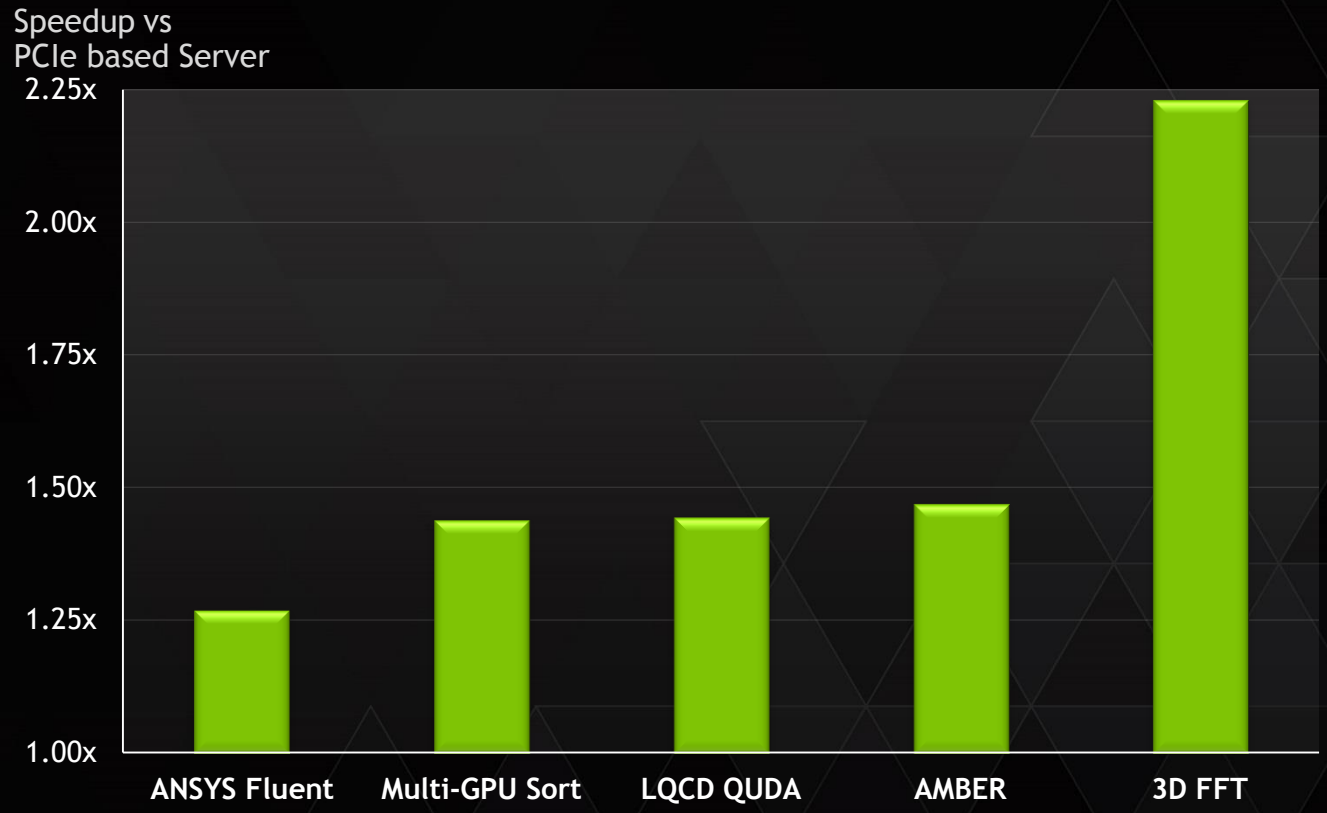
GPUs Interconnected with NVLink



5x Faster than
PCIe Gen3 x16

Over 2x Application Performance Speedup

When Next-Gen GPUs Connect via NVLink Versus PCIe



3D FFT, ANSYS: 2 GPU configuration, All other apps comparing 4 GPU configuration
AMBER Cellulose (256x128x128), FFT problem size (256^3)

TESLA PLATFORM ENABLES OPTIMIZATION

Scalable Nodes, ISA Choice

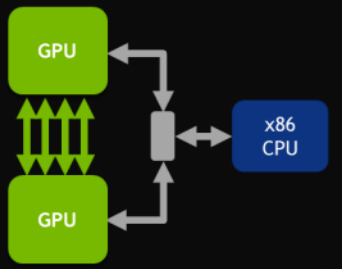
Power™

ARM POWERED

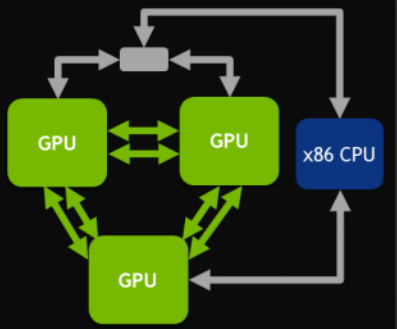
x86

NVLink

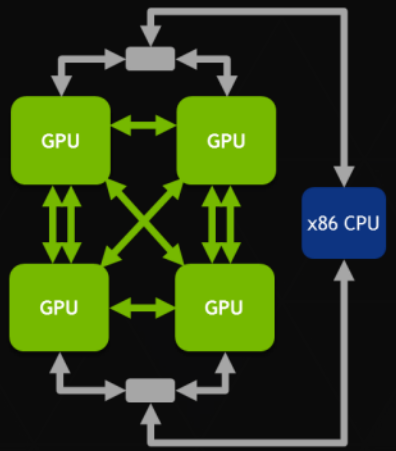
2 GPUs per Node



3 GPUs per Node



4 GPUs per Node

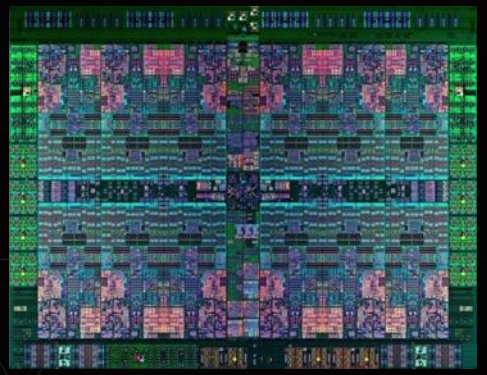


NVLINK 20GB/s

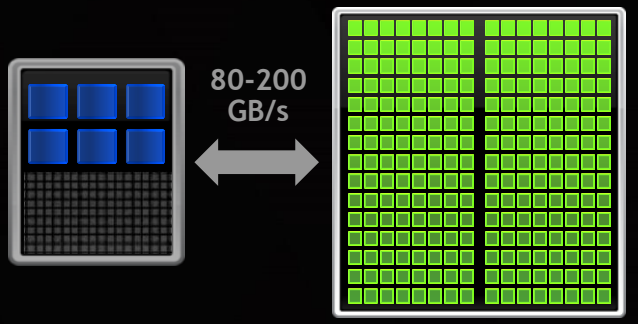
PCIe Gen3 x16

NVLINK-ENABLED HETEROGENEOUS NODE

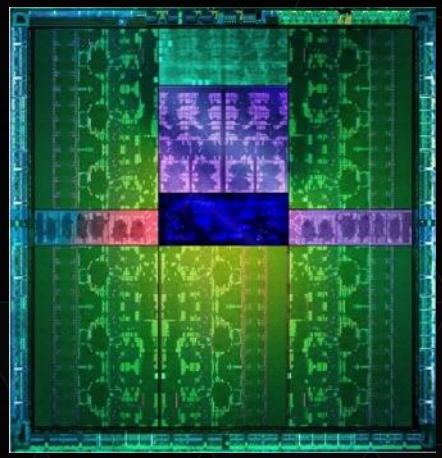
LOGICAL INTEGRATION + FLEXIBILITY + EFFICIENCY



POWER
Most Powerful Serial Processor



NVIDIA NVLink
Node Integration Interconnect



NVIDIA Volta GPU
Most Powerful Parallel Processor

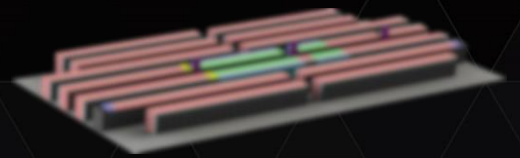
CORAL SCALABLE HETEROGENEOUS NODE

NVLink In Practice

Approximately 3,400 nodes, each with:

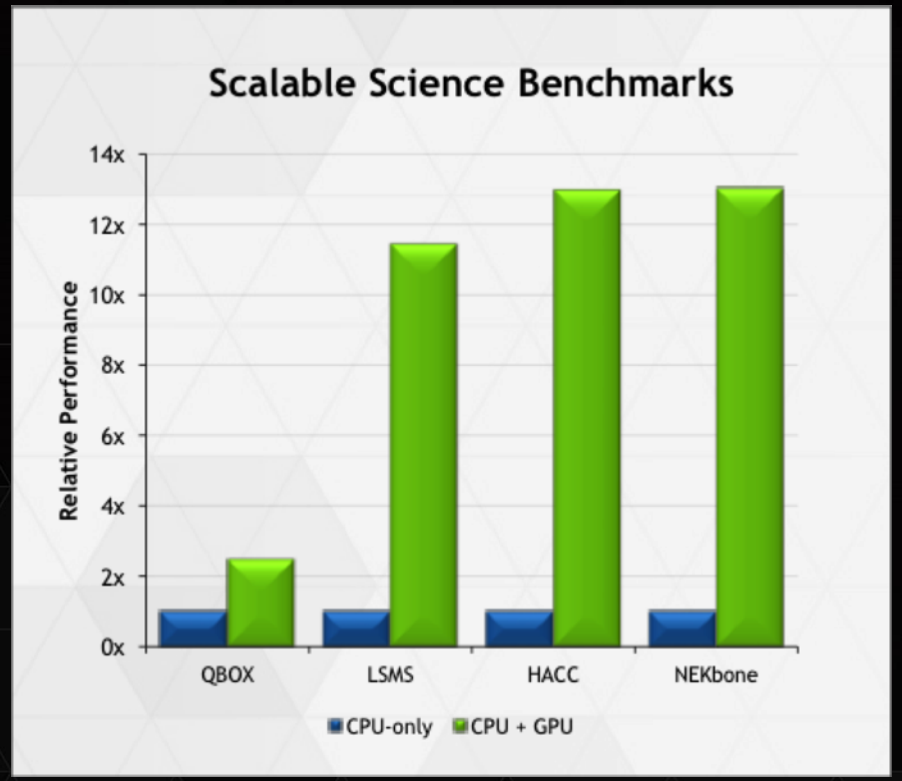
- IBM POWER9 CPUs and multiple NVIDIA Tesla® Volta GPUs
- CPUs and GPUs integrated on-node with high speed NVLink
- Large coherent memory: over 512 GB (HBM + DDR4)
  - All directly addressable from the CPUs and GPUs
- An additional 800 GB of NVRAM, burst buffer or as extended memory
- Over 40 TF peak performance/node(!)

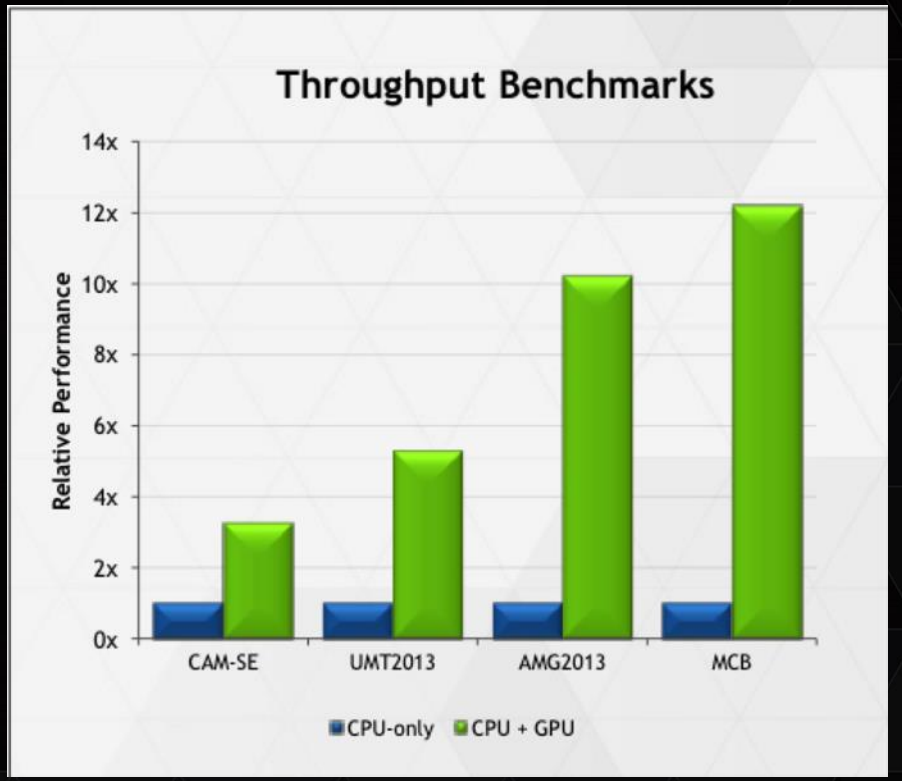
SUMMIT
OAK RIDGE NATIONAL LABORATORY



OPTIMIZED HETEROGENEOUS NODE

CORAL Application Performance Projections





SUMMARY:

- Heterogeneous acceleration is powerful and efficient
  - Latency-optimized cores for serial and system work
  - Throughput-optimized cores for on-node parallel work
  - High-performance integration
- NVLink provides a compelling node integration advantage
  - Flexible configuration of resources in application-driven proportions
  - Scalable performance into the future
- CORAL is awesome, you can buy nodes just like it.
  - OpenPower is surprisingly affordable.
  - It’s just one example of optimized architecture for an application set and budget.

NVIDIA REGISTERED DEVELOPER PROGRAMS

- Everything you need to develop with NVIDIA products
- Membership is your first step in establishing a working relationship with NVIDIA Engineering
  - Exclusive access to pre-releases
  - Submit bugs and features requests
  - Stay informed about latest releases and training opportunities
  - Access to exclusive downloads
  - Exclusive activities and special offers
  - Interact with other developers in the NVIDIA Developer Forums

REGISTER FOR FREE AT: developer.nvidia.com

THANK YOU

JOIN THE CONVERSATION

#GTC15