

GPU ACCELERATORS AT JSC SUPERCOMPUTING INTRODUCTION COURSE

22 November 2023 | Andreas Herten | Forschungszentrum Jülich



Member of the Helmholtz Association

Outline

GPUs at JSC JUWELS JUWELS Cluster JUWELS Booster JURECA DC GPU Architecture Empirical Motivation Comparisons GPU Architecture Summary Programming GPUs Libraries Directives CUDA C/C++ Performance Analysis Advanced Topics Advanced Topics





JUWELS Cluster – Jülich's Scalable System

- = 2500 nodes with Intel Xeon CPUs (2 \times 24 cores)
- 46 + 10 nodes with 4 NVIDIA Tesla V100 cards (16 GB memory)
- 10.4 (CPU) + 1.6 (GPU) PFLOP/s peak performance (Top500: #86)





JUWELS Booster - Scaling Higher!

- 936 nodes with AMD EPYC Rome CPUs (2 \times 24 cores)
- Each with 4 NVIDIA A100 Ampere GPUs (each: FP64TC: 19.5 FP64: 9.7 TFLOP/S, 40 GB memory)
- InfiniBand DragonFly+ HDR-200 network; 4×200 Gbit/s per node







Top500 List Nov 2020:

- #1 Europe
- #7 World
- #4* Top/Green500



JUWELS Booster – Scaling Higher!

- = 936 nodes with AMD EPYC Rome CPUs (2 \times 24 cores)
- Each with 4 NVIDIA A100 Ampere GPUs (each: FP64TC: 19.5 FP64: 9.7 TFLOP/S, 40 GB memory)
- InfiniBand DragonFly+ HDR-200 network; 4 \times 200 Gbit/s per node





JURECA DC – Multi-Purpose

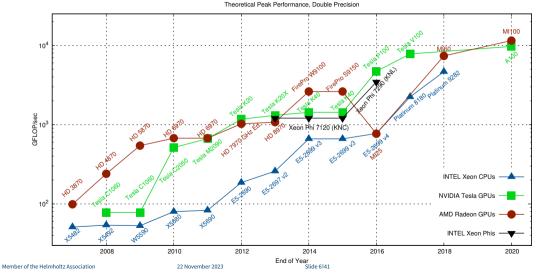
- 768 nodes with AMD EPYC Rome CPUs (2 \times 64 cores)
- 192 nodes with 4 NVIDIA A100 Ampere GPUs
- InfiniBand DragonFly+ HDR-100 network



GPU Architecture

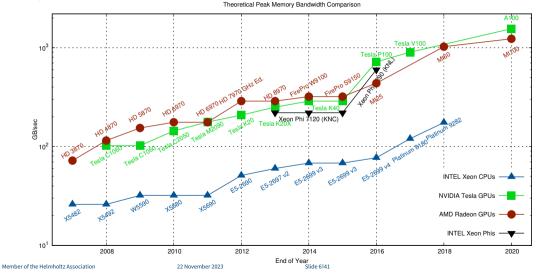
Status Quo Across Architectures

Performance



Status Quo Across Architectures

Memory Bandwidth





CPU vs. GPU

A matter of specialties







CPU vs. GPU

A matter of specialties



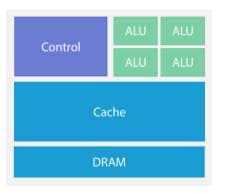
Transporting one



Transporting many



CPU vs. GPU _{Chip}

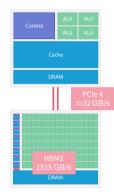






GPU optimized to hide latency

- Memory
 - GPU has small (40 GB), but high-speed memory 1555 GB/s
 - Stage data to GPU memory: via PCIe 4 bus (32 GB/s)

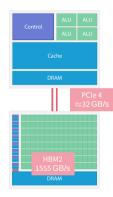


Host



GPU optimized to hide latency

- Memory
 - GPU has small (40 GB), but high-speed memory 1555 GB/s
 - Stage data to GPU memory: via PCIe 4 bus (32 GB/s)

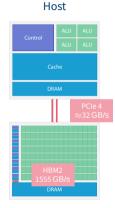


Host



GPU optimized to hide latency

- Memory
 - GPU has small (40 GB), but high-speed memory 1555 GB/s
 - Stage data to GPU memory: via PCIe 4 bus (32 GB/s)
 - Stage automatically (Unified Memory), or manually

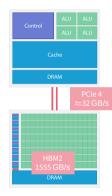




GPU optimized to hide latency

- Memory
 - GPU has small (40 GB), but high-speed memory 1555 GB/s
 - Stage data to GPU memory: via PCIe 4 bus (32 GB/s)
 - Stage automatically (Unified Memory), or manually
- Two engines: Overlap compute and copy





Host



GPU optimized to hide latency

- Memory
 - GPU has small (40 GB), but high-speed memory 1555 GB/s
 - Stage data to GPU memory: via PCIe 4 bus (32 GB/s)
 - Stage automatically (Unified Memory), or manually
- Two engines: Overlap compute and copy



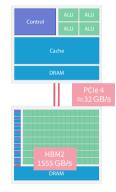
V100 32 GB RAM, 900 GB/s



A100



Host



Device



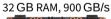
22 November 2023

GPU optimized to hide latency

- Memory
 - GPU has small (40 GB), but high-speed memory 1555 GB/s
 - Stage data to GPU memory: via PCIe 4 bus (32 GB/s)
 - Stage automatically (Unified Memory), or manually
- Two engines: Overlap compute and copy

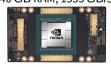


V100

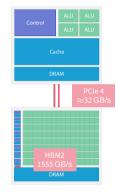




A100 40 GB RAM, 1555 GB/s



Host



Device



22 November 2023



Scalar



CPU:

Single Instruction, Multiple Data (SIMD)





Vector



CPU:

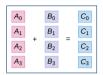
Single Instruction, Multiple Data (SIMD)



SIMT $SIMT = SIMD \oplus SMT$

CPU:

- Single Instruction, Multiple Data (SIMD)
- Simultaneous Multithreading (SMT)



Vector

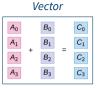






CPU:

- Single Instruction, Multiple Data (SIMD)
- Simultaneous Multithreading (SMT)



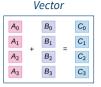






CPU:

- Single Instruction, Multiple Data (SIMD)
- Simultaneous Multithreading (SMT)
- GPU: Single Instruction, Multiple Threads (SIMT)



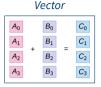






CPU:

- Single Instruction, Multiple Data (SIMD)
- Simultaneous Multithreading (SMT)
- GPU: Single Instruction, Multiple Threads (SIMT)







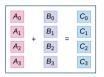




SIMT SIMT = SIMD ⊕ SMT

CPU:

- Single Instruction, Multiple Data (SIMD)
- Simultaneous Multithreading (SMT)
- GPU: Single Instruction, Multiple Threads (SIMT)
 - CPU core \simeq GPU multiprocessor (SM)
 - Working unit: set of threads (32, a warp)
 - Fast switching of threads (large register file)
 - Branching if _____



Vector









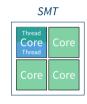
SIMT

$\mathsf{SIMT}=\mathsf{SIMD}\oplus\mathsf{SMT}$



Vector





Graphics: img:amperepictures

SIMT





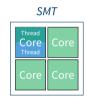
SIMT

$\mathsf{SIMT}=\mathsf{SIMD}\oplus\mathsf{SMT}$



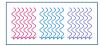
Vector





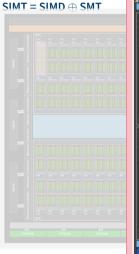
Graphics: img:amperepictures

SIMT



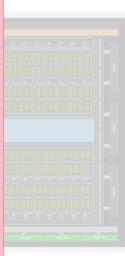


SIMT



Multiprocessor





Vector





SIMT



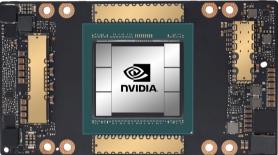


22 November 2023

A100 vs H100

Comparison of current vs. next generation

A100



H100





A100 vs H100

Comparison of current vs. next generation





A100 vs H100

Comparison of current vs. next generation

A100 Reciptor File (16.364 x 32-bit) Register File (16.384 x 32-bit) 11722-11722 10 10 W SFU W W W Dispatch Unit (32 thread/ch) Banjater File 115-384 x 32-bits Benjater File (15 304 x 32-bit) 10112-0012-0 14732 14732 w w w La La La La La SFU

H100





CPU vs. GPU

Let's summarize this!



Optimized for low latency

- + Large main memory
- + Fast clock rate
- + Large caches
- + Branch prediction
- + Powerful ALU
- Relatively low memory bandwidth
- Cache misses costly
- Low performance per watt



Optimized for high throughput

- + High bandwidth main memory
- + Latency tolerant (parallelism)
- + More compute resources
- + High performance per watt
- Limited memory capacity
- Low per-thread performance
- Extension card



Programming GPUs

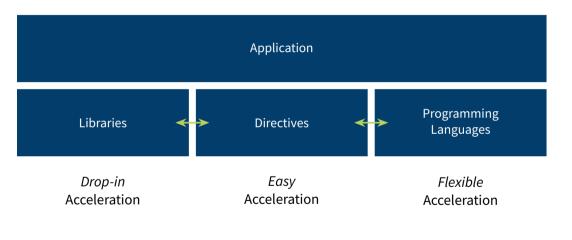
Preface: CPU

A simple CPU program!

```
SAXPY: \vec{y} = a\vec{x} + \vec{y}, with single precision
Part of LAPACK BLAS Level 1
void saxpv(int n. float a. float * x. float * v) {
  for (int i = 0: i < n: i + +)
    v[i] = a * x[i] + v[i]:
}
int a = 42:
int n = 10:
float x[n], v[n];
// fill x, y
saxpy(n, a, x, y);
```

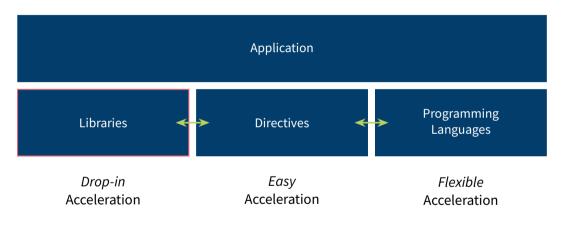


Summary of Acceleration Possibilities





Summary of Acceleration Possibilities







Programming GPUs is easy: Just don't!





Programming GPUs is easy: Just don't!

Use applications & libraries





Programming GPUs is easy: Just don't!

Use applications & libraries





Member of the Helmholtz Association

Libraries

Programming GPUs is easy: Just don't!

Use applications & libraries



ий ІСН

Forschungszentrum

Libraries

Programming GPUs is easy: Just don't!

Use applications & libraries



ий ІСН

Forschungszentrum



Slide 16141



- GPU-parallel BLAS (all 152 routines)
- Single, double, complex data types
- Constant competition with Intel's MKL
- Multi-GPU support
- → https://developer.nvidia.com/cublas http://docs.nvidia.com/cuda/cublas



cuBLAS

Code example

```
int a = 42; int n = 10;
float x[n], y[n];
// fill x. v
cublasHandle t handle;
cublasCreate(&handle):
float * d x, * d y;
cudaMallocManaged(\delta d x, n * sizeof(x[0])):
cudaMallocManaged(\delta d y, n * sizeof(y[0]));
cublasSaxpv(handle. n. a. d x. 1. d v. 1):
cublasGetVector(n. sizeof(v[0]). d v. 1. v. 1):
cudaFree(d x); cudaFree(d y);
cublasDestroy(handle);
```



cuBLAS

Code example

int a = 42; int n = 10;

<pre>float x[n], y[n]; // fill x, y</pre>	
cublasHandle_t handle; cublasCreate(&handle);	Initialize
float * d_x, * d_y; cudaMallocManaged(&d_x, n * sizeof(x[0]));● cudaMallocManaged(&d_y, n * sizeof(y[0]));	Allocate GPU memory
cublasSaxpy(handle, n, a, d_x, 1, d_y, 1);●	Call BLAS routine
<pre>cublasGetVector(n, sizeof(y[0]), d_y, 1, y, 1);</pre>	Copy result to host
cudaFree(d_x);	Finalize



Programming GPUs Directives

GPU Programming with Directives

Keepin' you portable

Annotate serial source code by directives

#pragma acc loop
for (int i = 0; i < 1; i++) {};</pre>



GPU Programming with Directives

Keepin' you portable

Annotate serial source code by directives

```
#pragma acc loop
for (int i = 0; i < 1; i++) {};</pre>
```

- OpenACC: Especially for GPUs; OpenMP: Has GPU support
- Compiler interprets directives, creates according instructions



GPU Programming with Directives

Keepin' you portable

Annotate serial source code by directives

```
#pragma acc loop
for (int i = 0; i < 1; i++) {};</pre>
```

- OpenACC: Especially for GPUs; OpenMP: Has GPU support
- Compiler interprets directives, creates according instructions

Pro

- Portability
 - Other compiler? No problem! To it, it's a serial program
 - Different target architectures from same code
- Easy to program

Con

- Only few compilers
- Not all the raw power available
- A little harder to debug



OpenACC / OpenMP

Code example

```
void saxpy_acc(int n, float a, float * x, float * y) {
    #pragma acc kernels
    for (int i = 0; i < n; i++)
        y[i] = a * x[i] + y[i];
}
float a = 42;
int n = 10;
float x[n], y[n];
// fill x, y
saxpy_acc(n, a, x, y);</pre>
```



OpenACC / OpenMP

Code example

```
void saxpy_acc(int n, float a, float * x, float * y) {
    #pragma omp target map(to:x[0:n]) map(tofrom:y[0:n]) loop
    for (int i = 0; i < n; i++)
        y[i] = a * x[i] + y[i];
}
float a = 42;
int n = 10;
float x[n], y[n];
// fill x, y
saxpy_acc(n, a, x, y);</pre>
```



Programming GPUs CUDA C/C++

Finally...



Finally...

OpenCL Open Computing Language by Khronos Group (Apple, IBM, NVIDIA, ...) 2009

- Platform: Programming language (OpenCL C/C++), API, and compiler
- Targets CPUs, GPUs, FPGAs, and other many-core machines
- Fully open source



Finally...

OpenCL Open Computing Language by Khronos Group (Apple, IBM, NVIDIA, ...) 2009

- Platform: Programming language (OpenCL C/C++), API, and compiler
- Targets CPUs, GPUs, FPGAs, and other many-core machines
- Fully open source

CUDA NVIDIA's GPU platform 2007

- Platform: Drivers, programming language (CUDA C/C++), API, compiler, tools, ...
- Only NVIDIA GPUs
- Compilation with nvcc (free, but not open)
 - clang has CUDA support, but CUDA needed for last step
- Also: CUDA Fortran; and more in NVIDIA HPC SDK



Finally...

OpenCL Open Computing Language by Khronos Group (Apple, IBM, NVIDIA, ...) 2009

- Platform: Programming language (OpenCL C/C++), API, and compiler
- Targets CPUs, GPUs, FPGAs, and other many-core machines
- Fully open source

CUDA NVIDIA's GPU platform 2007

- Platform: Drivers, programming language (CUDA C/C++), API, compiler, tools, ...
- Only NVIDIA GPUs
- Compilation with nvcc (free, but not open)
 - clang has CUDA support, but CUDA needed for last step
- Also: CUDA Fortran; and more in NVIDIA HPC SDK
- HIP AMD's unified programming model for AMD (via ROCm) and NVIDIA GPUs 2016+
- SYCL Intel's unified programming model for CPUs and GPUs (also: DPC++)



Finally...

OpenCL Open Computing Language by Khronos Group (Apple, IBM, NVIDIA, ...) 2009

- Platform: Programming language (OpenCL C/C++), API, and compiler
- Targets CPUs, GPUs, FPGAs, and other many-core machines
- Fully open source

CUDA NVIDIA's GPU platform 2007

- Platform: Drivers, programming language (CUDA C/C++), API, compiler, tools, ...
- Only NVIDIA GPUs
- Compilation with nvcc (free, but not open)
 - clang has CUDA support, but CUDA needed for last step
- Also: CUDA Fortran; and more in NVIDIA HPC SDK
- HIP AMD's unified programming model for AMD (via ROCm) and NVIDIA GPUs 2016+

SYCL Intel's unified programming model for CPUs and GPUs (also: DPC++)

- Choose what flavor you like, what colleagues/collaboration is using
- Hardest: Come up with parallelized algorithm



Finally...

OpenCL Open Computing Language by Khronos Group (Apple, IBM, NVIDIA, ...) 2009

- Platform: Programming language (OpenCL C/C++), API, and compiler
- Targets CPUs, GPUs, FPGAs, and other many-core machines
- Fully open source

CUDA NVIDIA's GPU platform 2007

- Platform: Drivers, programming language (CUDA C/C++), API, compiler, tools, ...
- Only NVIDIA GPUs
- Compilation with nvcc (free, but not open)
 - clang has CUDA support, but CUDA needed for last step
- Also: CUDA Fortran; and more in NVIDIA HPC SDK
- HIP AMD's unified programming model for AMD (via ROCm) and NVIDIA GPUs 2016+

SYCL Intel's unified programming model for CPUs and GPUs (also: DPC++)

- Choose what flavor you like, what colleagues/collaboration is using
- Hardest: Come up with parallelized algorithm



In software: Threads, Blocks

Methods to exploit parallelism:



In software: Threads, Blocks

- Methods to exploit parallelism:
 - Thread



In software: Threads, Blocks

- Methods to exploit parallelism:
 - Threads





In software: Threads, Blocks

Methods to exploit parallelism:







In software: Threads, Blocks

Methods to exploit parallelism:







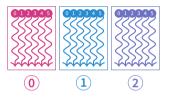


In software: Threads, Blocks

Methods to exploit parallelism:



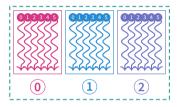
Blocks





In software: Threads, Blocks

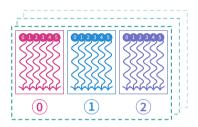
- Methods to exploit parallelism:
 - Threads \rightarrow Block
 - $\blacksquare \quad \mathsf{Blocks} \to \mathsf{Grid}$





In software: Threads, Blocks

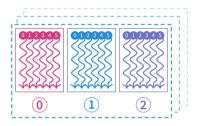
- Methods to exploit parallelism:
 - Threads \rightarrow Block
 - $\blacksquare \quad \mathsf{Blocks} \to \mathsf{Grid}$
 - Threads & blocks in 3D





In software: Threads, Blocks

- Methods to exploit parallelism:
 - Threads \rightarrow Block
 - $\blacksquare \quad \mathsf{Blocks} \to \mathsf{Grid}$
 - Threads & blocks in 3D
- Parallel function: kernel
 - __global__ kernel(int a, float * b) { }
 - Access own ID by global variables threadIdx.x, blockIdx.y,...
- Execution entity: threads
 - Lightweight \rightarrow fast switchting!
 - 1000s threads execute simultaneously \rightarrow order non-deterministic!





Slide 25141

CUDA SAXPY

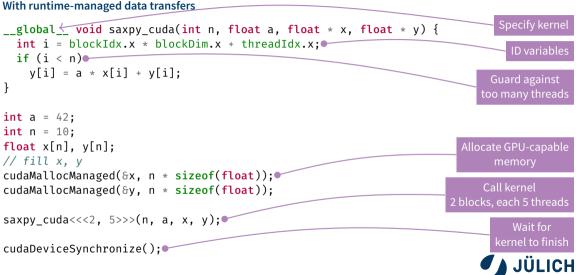
With runtime-managed data transfers

```
global void saxpy cuda(int n, float a, float * x, float * y) {
  int i = blockIdx.x * blockDim.x + threadIdx.x;
  if (i < n)
   v[i] = a * x[i] + v[i]:
}
int a = 42:
int n = 10:
float x[n]. v[n]:
// fill x, y
cudaMallocManaged(&x. n * sizeof(float)):
cudaMallocManaged(&y, n * sizeof(float));
saxpv cuda<<<2. 5>>>(n. a. x. v):
```

```
cudaDeviceSynchronize();
```



CUDA SAXPY



Programming GPUs Performance Analysis

GPU Tools

The helpful helpers helping helpless (and others)

NVIDIA

cuda-gdb GDB-like command line utility for debugging compute-sanitizer Check memory accesses, race conditions, ... Nsight IDE for GPU developing, based on Eclipse (Linux, OS X) or Visual Studio (Windows) or VScode Nsight Systems GPU program profiler with timeline Nsight Compute GPU kernel profiler

AMD

rocProf Profiler for AMD's ROCm stack uProf Analyzer for AMD's CPUs and GPUs



Nsight Systems

•••

\$ nsys pr	ofilestats=tru	e ./poisson	2d 10 # (shor	tened)		
CUDA API	Statistics:					
Time(%)	Total Time (ns)	Num Calls	Average	Minimum	Maximum	Name
90.9	160,407,572	30	5,346,919.1	1,780	25,648,117	cuStreamSynchronize
CUDA Kern	el Statistics:					
Time(%)	Total Time (ns)	Instances	Average	Minimum	Maximum	Name
100 0	150 606 617		15 060 661 7	1/ 525 010	25 652 702	
100.0	158,686,617	10	15,868,661.7		25,652,783	main_106_gpu
0.0	25,120	10	2,512.0	2,304	3,680	main_106_gpured



Nsight Systems



	Tools Help		NVIDIA N	sight Systems 202	20.4.1			
report2.qdre								
Timeline				6	■tx ()=		🕲 1 error.	4 warnings, 16 message
-		18s	19s 2		21s	22s	23s	24s
CPU (256)	6)							
Threads (6)							
	/ (Tesla V100-PCIE-160							
▼ 1.5% Ker						The function of the function o		
▶ 79.6%	main_106_gpu							
▶ 19.0%	main_118_gpu							
3 kern	el groups hidder 🗕 🕂							
▼ 98.5% N	Memory	and inhighteness which	nania ananishin anan	inter distances sta	Annalahi Jentishi	and statistical table	ments another	int consideration californi
<0.1%	Memset				and a feature	period of the second of the second		
51.1%	HtoD memcpy	ملقت المتعدلة لراغد	لمتمر سلما لمتعمر والملاه	and and a design of the second se	an and and, doubled	فلقرار المتقدر أهابه شما	and take a transmission of the	وملقر بداخماها أخدرها
51.1%		ملقت المتعدلة لراغد		and and a design of the second se	an and and, doubled	فلقرار المتقدر أهابه شما	and take a transmission of the	وملقر بداخماها أخدرها
51.1%	HtoD memcpy	and a pharman and an and an and an and an	لمتمر سلما لمتعمر والملاه	and and a design of the second se	an and and, doubled	فلقرار المتقدر أهابه شما	and take a transmission of the	an and a factor of the second se
51.1% 48.9%	HtoD memcpy DtoH memcpy	and a pharman and an and an and an and an	لمتمر سلما لمتعمر والملاه	and and a design of the second se	an and and, doubled	فلقرار المتقدر أهابه شما	and take a transmission of the	معلقات والمعاقل أطلب ها محمد طلبة معاقلا أطلب ها
51.1% 48.9%	HtoD memcpy DtoH memcpy	and a pharman and an and an and an and an	لمتمر سلما لمتعمر والملاه	and and a descent of the second	an and and, doubled	فلقرار المتقدر أهابه شما	and take a transmission of the	معلقات والمعاقل أطلب ها محمد طلبة معاقلا أطلب ها
51.1% I 48.9% Events View	HtoD memcpy DtoH memcpy	and a pharman and an and an and an and an	لمتمر سلما لمتعمر والملاه	and and a descent of the second	an and and, doubled	atat L factor for factor and	which is in the state of the state	le utal juta ani pula si dalam ni shafana nga katana ka uman P
51.1% 48.9% Events View	HtoD memcpy DtoH memcpy v	and a pharman and an and an and an and an	ataly satisfies a stal	a de la constante de la constan La constante de la constante de	الله (الله من , الله (الله عليه الله ع الله (الله عليه الله الله الله الله الله الله الله عليه الله عليه الله الله الله الله الله الله الله ا	lask solak bak tak julijska anajularina alkad utera Name	end table to see to be t	le utal juta ani pula si dalam ni shafana nga katana ka uman P
51.1% 48.9% Events View # 7	HtoD memcpy DtoH memcpy v v Name	und diskundu optim hat almännikan henn	Alaly - stally be - stall adult, makalese, adua	Duration	od tod (tod) daniel obsi bedree best op aan of the GPU	Name Context	end table to see to be t	le utal para la pa arra la para la par
51.1% 48.9% Events View	HtoD memcpy DtoH memcpy v v Name Memset	und a fail and an	 Aladiy askalyste strat Aladi, maketese adea Start 17,9516s 	Duration 1,760 µs	GPU GPU 0 GPU 0 GPU 0 GPU 0	Name Context Stream 14	end table to see to be t	le utal para la pa arra la para la par
51.1% 48.9% Events View # 7 8	HtoD memcpy DtoH memcpy V V V V V V V V V V V V V V V V V V	_ud.i.dusted.ut.i hel.ibated.at.imae	 Start 17,9516s 17,9516s 17,9517s 17,9521s 17,9521s 	рын (1) на на ла мака (1) на на ла рика (1) на на на ла на на на на на 1,760 µs 472,923 µs 3,488 µs 2,080 µs	GPU GPU 0 GPU 0	Inducted John and John and John beschel John Name Context Stream 14 Stream 14	end table to see to be t	le utal para la pa arra la para la par
51.1% 48.9% Events View 7 8 9	HtoD memcpy , DtoH memcpy v v v Name Mamset main_106_gpu main_106_gpu	_ud.i.dusted.ut.i hel.ibated.at.imae	 Start 17,9516s 17,9516s 17,9521s 	ранар (1) (21 - е) (24 - е) (24 - е) (24 - е) Jacobic (2) (20 - е) (24 - е) (24 - е) (24 - е) (24 - e	GPU GPU 0 GPU 0 GPU 0 GPU 0	Industrial Just cal Just and Just cal Just Name Context Stream 14 Stream 14 Stream 14	end table to see to be t	le utal para la pa arra la para la par

Nsight Compute

GUI

-level overview of the utilization for compute and memory resources of the GPU. For each unit, the Speed Of Light (SoL) reports the achieved percentage of utilization with respect to the refect anximum. High-level overview of the utilization for compute and memory resources of the GPU presented as a nonline chart.	Launch: - m	nc_polymer_ite	eration_352	2_gpu * 🛛 🗑	Add Ba	iseline 💌 Ap	oly <u>R</u> ules							Save as Im	age	
Super Of Light N development of the utilization for compute and memory resources of the OPU, For each unit, the Speed Of Light (SOL) reports the achieved parcentage of utilization with respect to the tetral maximum. High-level conview of the utilization for compute and memory resources of the OPU, For each unit, the Speed Of Light (SOL) reports the achieved parcentage of utilization with respect to the GPU presented as a rooffine chart. Shi {\si} 20. 51 (+12.6%) 22.6.21 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.12 (+12.6%) 22.8.	94 Time: 25.82 n	nsecond Cyc	cles: 28.19	91.300 Reg	s: 144 GF	U: A100-SX	44-40GB	SM Frequer	ncy: 1.09 cy	cle/nsecond	CC: 8.0	Process: [31938] SO	MA 🕀 🛛	Э	
Interformation of the utilization for compute and memory resources of the GPU, for each unit, the Speed Of Light (SQ) reports the software direction with respect to the relation with respect to the utilization for compute and memory resources of the GPU prevented as a rootilic edur. SP (s)	6 Time: 77.17 m	nsecond Cyc	cles: 101.1	23.609 Reg	5: 144 GF	U: Tesla V10	0-SXM2-16GE	3 SM Frequer	ncy: 1.31 cy	cle/nsecond	CC: 7.0	Process: [30412] SOI	AN		
etical maximum. High-level overview of the utilization for compute and memory resources of the GPU presented as a rooflike char. Hemory [k]	l Of Light 🛕												AII	*	۶	
Memory (%) 55.82 (+1.26%) Elapsed Cycles [cycle] 2810308 (-7.2 L2 Cache (%) 61.89 (+4.26%) %friewerky (cycle) (cycle) 227784365.46 (-7.2 L2 Cache (%) 61.89 (+34.96%) %friewerky (cycle) (cycle) 227784365.46 (-7.2 L2 Cache (%) 61.89 (+34.96%) %friewerky (cycle) (cycle) 227784365.46 (-7.2 BM (%) 01.92 (-2.8.2%) DMM Frequency (cycle)rescond) 1.22 (-3.8 00 10,0 20,0 30,0 40,0 50,0 60,0 70,0 80,0 90,0 SOLS M Breakdown SOL Memory Breakdown <td colspa<="" td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>ercentage o</td><td>of utilization</td><td>with respe</td><td>ct to the</td><td></td></td>	<td></td> <td>ercentage o</td> <td>of utilization</td> <td>with respe</td> <td>ct to the</td> <td></td>											ercentage o	of utilization	with respe	ct to the	
LLT/EX 49.19 (+9.49) 94 Active (cycles [cycle] 27784365.46 -77 LLT/EX Cache [N] 61.89 (+9.490) 94 Active (cycles [cycle] 27784365.46 -77 LLT/EX Cache [N] 51.82 (+2.800) BMM Frequency [cycle/nsecond] 1.21 (-98 DRM [N] 51.92 (+2.800) BMM Frequency [cycle/nsecond] 1.22 (-98 M [N] 50.0 50.0 60.0 70.0 80.0 90.0 Solut Breakdown 50.0 50.0 60.0 70.0 80.0 90.0 Solut SM Instance Minite [N] 20.61 (+100.09%) 50.12 The Part Instance Minite [N] 558.2 (+12.400) Solut SM Instance Minite [N] 20.61 (+100.09%) 50.12 The Part Instance Minite [N] 558.2 (+12.400) Solut SM Instance Minite [N] 20.61 (+100.09%) 50.12 The Part Instance Minite [N] 558.2 (+12.400) Solut SM Instance [N] 20.61 (+100.09%) 50.12 The Part Instance [N] 558.2 (+12.400) Solut SM Instance [N] 20.61 (+100.09%) 50.12					20.51	L (+160.099) Duration	[msecond]					25.82	(-66.5	549	
L2 Cache (s) 61.89 (+349.81%) 9f / Frequency (cycle/nsecond) 1.99 (-1.9) DRAH (s) 31.32 (+2.82%) DRAH (s) 1.21 (-38 GPU Utilization GPU Utilization 1.21 (-38 M (s) 50.0 (6.0 70.0 80.0 90.0 90.0 50.0 (6.0 70.0 80.0 90.0 50.0 50.0 60.0 70.0 80.0 90.0 50.0 50.0 50.0 50.0 50.0 50.0 5	[26]				55.82	2 (+1.269) Elapsed	Cycles [cyc]	le]				28191300	(-72.1	12*	
DRAM [%] 31.52 (-42.82%) DRAM Prequency (cycLe/nsecond) 1.21 (-38 SM PG GPU Utilization	Cache [%]				40.19) (+5.499) SM Activ	e Cycles [c	ycle]			277	84365.64	(-72.3	309	
GPU Utilization GPU Utilization M (5) Sols M breakdown Sols M breakdown Sols M Breakdown Sols (+160.0%) Sols M Pip Break Cycles Active [5] 20.51 (+160.0%) 30.12.7 Tbreakdown 55.82 (+124.4) Sols M Pip Break Cycles Active [5] 20.51 (+160.0%) 30.12.7 Tbreakdown 56.82 (+124.4) Sols M Pip Break Cycles Active [5] 13.83 (+165.6%) 30.12.7 Tbreakdown 56.82 (+124.4) Sols M Break Depis Lay Cycles Active [5] 13.83 (+165.6%) 30.12.7 Tbreak Depis Lay View View Fols 38.80 (+71.7) Sols M Break Depis Lay Cycles Active [5] 13.83 (+165.6%) 30.12.7 Tbreak Depis Lay View View Fols 22.46 (+161.1%) Sols M Ind Twater Bips Lay Cycles Active	e [%]				61.89	0 (+149.019) SM Frequ	ency [cycle,	/nsecond]				1.09	(-16.6	57%	
SM (%) Solution <]				31.52	2 (-42.829) DRAM Fre	quency [cyc]	le/nsecon	3)			1.21	(+38.2	21%	
Immory [5] 0.0 10,0 20,0 30,0 40,0 50,0 60,0 70,0 80,0 90,0 Speed Of Light [5] SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Ereakdown SOL SM Ereakdown [54] SOL SM Ereakdown [56] SOL SM Ereakdown [56] <t< td=""><td></td><td></td><td></td><td></td><td></td><td>GPU</td><td>Utilization</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>						GPU	Utilization									
norry [3] 0,0 10,0 20,0 30,0 40,0 50,0 60,0 70,0 80,0 90,0 Speed Of Light (%) SOL SM Breakdown 0.5 M: tost Source (%) 20,5 (+160,00%) 0.5 M: tost Source (%) 20,			_													
0.0 10.0 20.0 30.0 40.0 50.0 60.0 70.0 80.0 90.0 SOL SM Breakdown SOL SM Breakdown OL SM Issue Active [%] SOL Marce State Sta																
0.0 10.0 20.0 20.0 20.0 40.0 50.0 60.0 70.0 80.0 90.0 SOL SM Breakdown SOL Marce Sol SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL Marce Sol SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown SOL SM Breakdown <td col<="" td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td>	<td></td>															
0.0 10.0 20.0 30.0 40.0 50.0 60.0 70.0 80.0 90.0 SOL SM Breakdown SOL SM Breakdown OL SM Issue Active [%] SOL Marce State Sta																
No. Speed of Uight (%) No. No. SOLS M Breakdown SOL SM Breakdown SOL Memory Breakdown 0.5 M: Instructure (%) 20.51 (160.00%) SOL L2: Margins (%) 65.82 (124.41) 0.6 M: Instructure (%) 20.46 (160.00%) SOL L2: Margins (%) 65.82 (124.41) 0.6 M: Instructure (%) 13.83 (166.60%) SOL L2: Tag Requests (%) 65.82 (124.41) 0.6 M: Instructure (%) 13.83 (166.60%) SOL L2: Tag Requests (%) 30.86 (171.11) 0.6 M: Instructure (%) 13.83 (166.60%) SOL L2: Tag Requests (%) 30.86 (171.11) 0.6 M: Instructure (%) 13.83 (166.60%) SOL L2: Tag Requests (%) 30.86 (171.11) 0.6 M: Instructure (%) 13.83 (171.82%) SOL UPU. Dam Trouppingt (%) 33.80 (171.11) 0.6 M: Instructure (%) 13.97 (191.93) SOL L2: D Sectors (%) 24.88 (493.10) 0.6 M: More Read Cycles Active (%) 8.10 (101.99%) SOL L2: D Sectors (%) 12.10 (12.10) 0.6 M: More Read Cycles Active (%) 72.21 (165.69%) SOL L1: Luiw Request (%) 10.89 (172.10) 0.1 SM: More Read Cycles Active (%) 65.85 (172.26%) SOL L2: L1.22.26% Cycles A																
No. Operator Speed of Uight (%) Operator Operator SOLS SM Breakdown SOLL SW breakdown SOLL SW breakdown SOLMemory Breakdown SOLS SM Interactive [%] 20,51 (1680,09%) SOLL 2: Marr21th Cycles Active [%] 6552 (124,41 SOLS SM Interactive [%] 20,46 (1480,10%) SOLL 2: Marr21th Cycles Active [%] 6552 (124,41 SOLS SM: Inter Executed [%] 20,46 (1480,10%) SOLL 2: Tag Requests [%] 6552 (124,41 SOLS SM: Inter Executed [%] 13,86 (-1680,69%) SOLL 2: Tag Requests [%] 3806 (-171,10%) SOLS SM: Inter Executed [%] 13,86 (-1680,69%) SOLL 2: Tag Requests [%] 3806 (-171,10%) SOLS SM: Inter Executed [%] 13,80 (-1680,69%) SOLL 2: Tag Requests [%] 3806 (-171,10%) SOLS SM: Inter Executed [%] 10,80 (-172,96%) SOLL 2: Tag Requests [%] 3806 (-171,10%) SOLS SM: Inter Executed [%] 10,80 (-172,96%) SOLL 2: Das Fore IND (%) 3830 (-171,10%) SOLS SM: Inter Executed [%] 8,10 (-1108,96%) SOLL 2: D Sectors IID Oxice [%] 24,284 (-115,10%) SOLS SM: Inter Executed [%] 8,10 (-105,86%) SOLL 2: D Sectors IID Oxice [%] 12,10 (-1																
SOL SM Breakdown SOL Memory Breakdown SOL, Microsov Active [N] 20.05 (+ 1000/15) SOL 2: Xbar2ns Cycles Active [N] 56.82 (+ 12.42) SOL SM: Inst Executed [%] 20.06 (+ 1000/15) SOL 2: The preparate [N] 56.82 (+ 12.42) SOL SM: Inst Executed [%] 10.83 (+ 1068091) SOL 2: The preparate [N] 30.81 (+ 1000/15) SOL SM: Inst Executed [%] 10.83 (+ 1068091) SOL 1: M LiterAcbar Rec Cycles Active [N] 30.81 (+ 1000/15) SOL SM: Inst Executed [%] 10.30 (+ 107.8991) SOL 1: Lus Windback Active [N] 30.10 (+ 100.8991) SOL SM: Inst Executed Pipe Lus [N] 10.30 (+ 107.8991) SOL 1: Lus Windback Active [N] 31.52 (+ 22.44) SOL SM: Inst Executed Pipe Lus [N] 8.10 (+ 101.8991) SOL 1: Lus Windback Active [N] 22.46 (+ 101.8991) SOL SM: Inst Executed Pipe Lus [N] 2.72 (+ 1068091) SOL L1: Lus Windback Active [N] 22.46 (+ 101.991) SOL SM: Inst Executed Pipe Lus [N] 2.72 (+ 1068091) SOL L1: Lus Windback Active [N] 22.46 (+ 101.991) SOL SM: Inst Executed Pipe Lus [N] 2.72 (+ 1068091) SOL L1: Lus Windback Active [N] 22.46 (+ 101.991) SOL SM: Inst Executed Pipe Lus [N] 2.72 (+ 1068091) SOL	0 100)	20.0	30.0		40.0	50.0	60.0	0	70.0	80.0	2	90.0	1	100	
SOL SM: Instaue Active [M] 20.01 (+160.09%) SOL L2: Xbar2its Cycles Active [M] 65.82 (+124.4) SOL SM: Inst Executed [Pi] 20.46 (+160.05%) SOL L2: Tog Requests [M] 46.67 (+178.4) SOL SM: Inst Executed [Pi] 20.46 (+160.05%) SOL L2: Xbar2its Cycles Active [M] 46.67 (+178.4) SOL SM: Inst Executed [Pi] 13.83 (+165.65%) SOL L2: Tog Requests [M] 36.86 (+170.4) SOL SM: Inst Executed [Pie] 13.83 (+165.65%) SOL L2: Tog Requests [M] 38.80 (+171.7) SOL SM: Inst Executed [Pie] La (M) 13.80 (+165.65%) SOL L2: Tog Requests [M] 38.80 (+171.7) SOL SM: Inst Executed [Pie] La (Pie) 13.03 (+165.65%) SOL L2: La Wavefronts [M] 34.80 (+174.7) SOL SM: Inst Executed [Pie Clu Pie On Any [N] 8.10 (+161.97%) SOL L2: La Wavefronts [M] 22.64 (+154.4) SOL SM: Ind Togenet Cycles Active [N] 8.10 (+161.97%) SOL L2: D Sectors [M] 22.64 (+154.4) SOL SM: Ind Togenet Cycles Active [N] 7.23 (+165.64%) SOL L2: La Subar Cycles Active [N] 10.80 (+172.7) SOL SM: Ind Togenet Cycles Active [N] 7.23 (+165.64%) SOL L2: La Subar Cycles Active [N] 10.80 (+172.7) SOL SM: Ind Togenet Cycles Active [N	,0 10,0)	20,0	30,0				/ -	0	70,0	80,0	D	90,0	1	100	
301. SM: Inst Executed (%) 20.46 (+16010%) SOL 21: Trap Requests (%) 452 (*178) 303. SM: Pipe Struct Cycles Active (%) 18.86 (+16010%) SOL 21: Trap Requests (%) 3061 (+107) 303. SM: Pipe Struct Cycles Active (%) 18.86 (+16050%) SOL L2: T Stertors (%) 3061 (+107) 303. SM: Pipe Struct Cycles Active (%) 18.86 (+16050%) SOL L2: T Stertors (%) 3061 (+107) 303. SM: Pipe Struct Cycles Active (%) 10.80 (+127)/650% SOL L2: T Stertors (%) 38.80 (+71) 303. SM: Pipe Struct Cycles Active (%) 10.37 (+16123%) SOL DUP Dum Throughput (%) 38.25 (-42) 303. SM: And Pipe Struct Pipe Cuc Pied Or Any (%) 8.13 (+71.89%) SOL L1: Law Writeback Active (%) 24.84 (+15) 303. SM: And Pipe Struct Cycles Active (%) 8.10 (+105.86%) SOL L2: D Sectors (%) 22.46 (+15) 303. SM: And Pipe Fina Cycles Active (%) 8.10 (+105.86%) SOL L2: D Sectors (%) 12.10 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (+12.16 (\0 10,0							/ -	-	,-			90,0	1	100	
D3: MP per Shared Cycles Active [N] 10.36 (+166.09%) SOL L1: M. L16x2Abar Reg Cycles Active [N] 30.61 (+107.216.00%) D3: MP per Shared Cycles Active [N] 10.36 (+166.09%) SOL L2: D settors [N] 38.61 (+107.216.00%) D3: MP to End Cycles Active [N] 10.36 (+112.29%) SOL L2: D bas Pipe Lux Wavefronts [N] 38.10 (+71.216.00%) D3: MP to End Cycles Active [N] 10.37 (+112.23%) SOL L2: D bas Pipe Lux Wavefronts [N] 34.25 (+42.26%) D3: MP to End Cycles Active [N] 8.10 (+112.29%) SOL L2: D bas Pipe and Cycles Active [N] 24.83 (+63.26%) D3: MM to End Active [N] 8.10 (+105.29%) SOL L2: D Sectors [N] 22.84 (+115.26%) D3: MM to End Active [N] 7.25 (+165.64%) SOL L2: D Sectors [N] 10.16 (+127.26%) D3: MM to End Active [N] 7.25 (+165.64%) SOL L2: D Sectors [N] 10.64 (+127.26%) D3: MM to End Active [N] 7.25 (+165.64%) SOL L2: D Sectors [N] 10.64 (+167.26%) D3: MM to End Active [N] 7.25 (+165.64%) SOL L2: D Sectors [N] 10.64 (+127.26%) D3: MM to End Active [N] 7.25 (+165.64%) SOL L2: D Sectors [N] 10.66 (+127.26%) D3: MM to End Active [N] 7.25 (+165.66%)<),O 10,C							/ -	-	,-			90,0	1	10	
0LSM: Page Fipt6 Cycles Active [15] 18.36 (+1656.09%) SOL L2: T Sectors [15] 38.00 (+72) 0LSM: Inst Executed Pipe Lau [15] 10.80 (+127.96%) SOL L2: T Sectors [15] 34.00 (+76) 0LSM: Inst Executed Pipe Lau [15] 10.80 (+127.96%) SOL L2: T Sectors [15] 34.10 (+76) 0LSM: Pipe ALV (2564 Active [15] 10.37 (+161.29%) SOL L2: Data Pipe Lau Wavefronts [15] 34.10 (+76) 0LSM: Pipe ALV (2564 Active [15] 0.54 (+161.97%) SOL L2: D Sectors [16] 24.26 (+165) 0LSM: Morg Nether Cycles Active [15] 8.10 (+165.96%) SOL L2: D Sectors [16] 22.64 (+116) 0LSM: Morg Nether Cycles Active [15] 8.10 (+165.96%) SOL L2: D Sectors [16] 22.64 (+116) 0LSM: Morg Nether Cycles Active [15] 7.23 (+165.86%) SOL L2: D Sectors [16] 12.89 (+12.19) 0LSM: Morg Nether Cycles Active [15] 7.23 (+165.86%) SOL L2: US2/bsc Cycles Active [16] 6.83 (+2.19) 0LSM: Morg Nether Cycles Active [16] 7.25 (+165.86%) SOL L2: US2/bsc Cycles Active [16] 6.83 (+2.19) 0LSM: Inst Executed Pipe Xu [16] 4.59 (+165.69%) SOL L1: Data Bank Reade [16] 3.09 (+7.7) 0LSM: Inst Executed Pipe Xu [16] 1.29 (+16					20	5	peed Of Ligh	t [%]	sc	L Memory						
0.15.M: trait Securited Pape Lux [Vs] 10.80 (+122/96%) SOL L1: bata Pape Lux Wavefronts [Vs] 34.10 (*76. 0.15.M: Pape Autry (Sec Active (N) 10.37 (*1612336) SOL L2: D2: D2: D2: D2: D2: D2: D2: D2: D2: D	sue Active [%]					.51 (+160.09%	() SOL L2:	t [%] Xbar2lts Cycle	SC es Active [%]	L Memory			55.82	(+124.58	%]	
0.13M: Pipe Alu Cycles Active [5] 10.327 (+161.235) SOL OPU: Deam Throughput [6] 31.52 (+22.) 0.13M: Bipe Alu Cycles Active [5] 0.13M: Pipe Alu Cycles Active [5] 24.83 (+31.23) 0.13M: Bipe Alu Cycles Active [5] 8.10 (+161.99%) SOL L1: Use Writeback Active [5] 24.84 (+161.23) 0.13M: Mo2P (Writeback Active [5] 2.81 (+161.99%) SOL L2: D Sectors [10] 22.64 (+116.23) 0.13M: Mo2P (Writeback Active [5] 2.753 (+165.64%) SOL L2: D Sectors [10] 22.64 (+116.23) 0.13M: Mo2P (Writeback Active [5] 7.253 (+165.64%) SOL L2: D Sectors [10] 22.84 (+116.23) 0.13M: Mo2P (Writeback Active [5] 7.252 (+165.64%) SOL L2: L12: Lania Requester [5] 10.80 (+127.23) 0.13M: Mo1P (Writeback Active [5] 7.252 (+165.64%) SOL L2: L12: L12: L12: L12: L12: L12: L12:	sue Active [%]	SOL SM			20.	51 (+160.09% 46 (+160.10%	() SOL L2: SOL L2:	t [%] Xbar2its Cycle T Tag Request	SC es Active [%] ts [%]	L Memory			55.82	(+124.58 (+178.82	%]	
0.15M: https://www.sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sci.uk/sc	isue Active [%] ist Executed [%] ipe Shared Cycles A	SOL SM			20. 18.	51 (+160.09% 46 (+160.10% 36 (+165.69%	(i) SOL L2: (i) SOL L2: (i) SOL L2:	t [%] Xbar2lts Cycle T Tag Request M L1tex2xbar R	es Active [%] ts [%] Req Cycles A	L Memory			55.82 45.27 39.61	(+124.58 (+178.82 (+107.93	%) %) %)	
00. SM: Inst Sexusted Pipe CDe Ped On Any [%] 8.13 (+7.98%) SOLL1: Law Writeback Active [%] 24.83 (+63.53) 0.05. Mic 207 Writes Active [%] 8.10 (+105.94%) SOLL2: D Sectors [#] Onice [%] 22.64 (+115.53) 0.05. Mic Any Private Cycles Active [%] 8.10 (+105.94%) SOLL2: D Sectors [#] Onice [%] 12.10 [= (1.2.12) 0.05. Mic Any Private Cycles Active [%] 8.10 (+105.94%) SOLL2: D Sectors [#] Onice [%] 12.10 [= (1.2.12) 0.05. Mic Any Write Cycles Active [%] 7.22 (+165.86%) SOLL1: Law Requests [%] 0.892 (-21.53) 0.05. Mic Inst Issued [%] 5.55 (+122.64%) SOLL1: Law Requests [%] 6.82 (-21.53) 0.05. Mic Inst Issued [%] 6.55 (+122.64%) SOLL1: Law Requests [%] 6.82 (-21.53) 0.05. Mic Inst Executed Pipe Auf[%] 4.59 (+166.55%) SOLL1: Law Requests [%] 6.89 (-63.64) 0.05. Mic Inst Executed Pipe Auf[%] 1.90 (+77.53) SOLL1: Law Requests [%] 6.90 (-37.54) 0.05. Mic Inst Executed Pipe Auf[%] 1.90 (+17.53) SOLL1: Law Requests [%] 6.90 (-37.54) 0.05. Mic Inst Executed Pipe Auf[%] 1.90 (+17.55, SN) SOLL1: Law Request [%] 6.90 (-75.53) 0.05. Mic Inst Executed Pipe Auf[%]	isue Active [%] ist Executed [%] ipe Shared Cycles A ipe Fp64 Cycles Act	SOL SM ctive [%] ive [%]			20. 18. 18.	51 (+160.09% 46 (+160.10% 36 (+165.69% 36 (+165.69%	() SOL L2: () SOL L2: () SOL L2: () SOL L2: () SOL L2:	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%]	SC es Active [%] ts [%] Req Cycles A	L Memory			55.82 45.27 39.61 38.80	(+124.58' (+178.82' (+107.93') (+71.23'	%) %) %)	
00, SM: Mo27 Writeback Active [%] 8.10 (+101.97%) SOL L2: D Sectors [%] 22.64 (+15) 00, SM: Mo27 Writeback Active [%] 8.10 (+105.96%) SOL L2: D Sectors Fill Device [%] 12.19 (+12.19) 0.5 M/: Mo P4 mice Cycles Active [%] 7.85 (+166.66%) SOL L1: Lisuin Requests [%] 10.80 (+127.37) 0.5 M/: Mo P4 mice Cycles Active [%] 7.23 (+166.66%) SOL L2: LisZback Cycles Active [%] 8.82 (+12.19) 0.5 M/: Mo P4 mice Mice [%] 5.55 (+12.2960) SOL L2: LisZback Cycles Active [%] 6.82 (+12.19) 0.5 M/: Mo P4 mice Mice [%] 5.55 (+12.2960) SOL L2: LisZback Sectors [%] 6.93 (+25.19) 0.5 M/: Inst Executed Pipe Xin [%] 4.59 (+166.69%) SOL L1: Data Bank Writes [%] 3.09 (+7.27) 0.5 M: Inst Executed Pipe Aut [%] 1.96 (+165.59%) SOL L1: Data Bank Writes [%] 1.96 (+23.81) 0.5 M: Inst Executed Pipe Aut [%] 1.96 (+23.83) SOL L1: Texah Bank Writes [%] 0.00 (+7.26.84) 0.5 M: Inst Executed Pipe Aut [%] 1.96 (+35.53%) SOL L1: Texah Bank Writes [%] 0.00 (+23.84)	isue Active [%] ist Executed [%] ipe Shared Cycles A ipe Fp64 Cycles Act ist Executed Pipe Ls	SOL SM ctive [%] cive [%] cu [%]			20. 18. 18. 10.	51 (+160.09% 46 (+160.10% 36 (+165.69% 36 (+165.69% 80 (+127.96%	(i) SOL L2: (i) SOL L2: (i) SOL L2: (i) SOL L1: (i) SOL L2: (i) SOL L1: (i) SOL L2: (i) SOL L1:	t [%] Xbar2lts Cycle T Tag Request M L1tex2xbar R T Sectors [%] Data Pipe Lsu V	SC es Active [%] ts [%] Req Cycles A Wavefronts	L Memory			55.82 45.27 39.61 38.80 34.10	(+124.58 (+178.82 (+107.93) (+71.23) (+76.03	%) %) %) %)	
LO, MK. Moh P, Read Cycles Active (%) 8.10 (-105.89%) SOLL 2: D Sectors FID Drives (%) 12.16 (-12. LO, SM. Moh P, Ward Cycles Active (%) Z55 (-165.66%) SOLL 1: Lisuin Requests (%) 10.80 (-127. LO, SM. Moh P, Ward Cycles Active (%) Z52 (-165.66%) SOLL 1: Lisuin Requests (%) 68.82 (-21. LO, SM. Moh P, Ward Cycles Active (%) Z52 (-165.66%) SOLL 1: Lisuin Requests (%) 68.82 (-21. LO, SM. Moh P, Ward Cycles Active (%) 55.55 (-122.69%) SOLL 1: Lisuin Request Sectors (%) 63.99 (-25. LO, SM. Inst Essued (%) C.S. Min Esseutice (%) A.59 (-166.66%) SOLL 1: Lisuin Request Sectors (%) 63.99 (-25. LO, SM. Inst Essued Pipe Au (%) 1.90 (-167.55%) SOLL 1: Lisuin Request Sectors (%) 63.09 (-72. LO, SM. Inst Essued Pipe Au (%) 1.10 (-165.5%) SOLL 1: Lisuin Reave Rev (%) 0.00 (-72. LO, SM. Inst Esseuted Pipe Au (%) 1.10 (-165.5%) SOLL 1: Lisuin Rev (%) 0.00 (-12.58.	isue Active [%] ist Executed [%] ipe Shared Cycles A ipe Fp64 Cycles Act ist Executed Pipe Ls ipe Alu Cycles Active	SOL SM .ctive [%] .ive [%] .eu [%] e [%]	1 Breakdo		20. 18. 18. 10. 10.	51 (+160.09% 46 (+160.10% 36 (+165.69% 36 (+165.69% 80 (+127.96% 37 (+161.23%	SOL L2: i) SOL L2: ii) SOL L2: iii) SOL L1: iii) SOL L2: iii) SOL L2: iiii) SOL L2: iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	t [%] Xbar2lts Cycle T Tag Request M L1tex2xbar R T Sectors [%] Data Pipe Lsu V U: Dram Throug	SC es Active [%] ts [%] Req Cycles A Wavefronts ighput [%]	L Memory			55.82 45.27 39.61 38.80 34.10 31.52	(+124.58 (+178.82 (+107.93) (+71.23) (+76.03 2 (-42.82	%) %) %) %) %)	
JOL SM: MIN PQ Write Cycles Active [%] 753 (±165.64%) SOL L1: Lusin Requests [%] 10.80 (±27) JOL SM: Pipe Fma Cycles Active [%] 7.22 (±165.66%) SOL L2: Lts.2bar Requests [%] 8.82 (±21) JOL SM: Min Issi secured [%] 5.55 (±12.56%) SOL L2: Lts.2bar Requests [%] 6.83 (±21) JOL SM: Inst Securited Pipe Xu [%] 4.59 (±165.66%) SOL L1: Total Bark Reads [%] 6.30 (±72) JOL SM: Inst Securited Pipe Xu [%] 4.59 (±165.66%) SOL L1: Total Bark Reads [%] 3.00 (±7) JOL SM: Inst Securited Pipe Xu [%] 1.95 (±12.55%) SOL L1: Total Bark Reads [%] 3.09 (±7) JOL SM: Inst Securited Pipe Aul [%] 1.18 (±165.55%) SOL L1: Total Bark Reads [%] 3.09 (±7) JOL SM: Inst Securited Pipe Aul [%] 1.18 (±165.55%) SOL L1: Total Bark Reads [%] 3.09 (±7)	isue Active [%] ist Executed [%] ipe Shared Cycles Ad ipe Ep64 Cycles Active ist Executed Pipe Ls ipe Alu Cycles Active ist Executed Pipe Ct	SOL SM .ctive [%] .tive [%] .eu [%] .e [%] .bu Pred On An	1 Breakdo		20. 18. 18. 10. 10. 8. 8. 8. 8. 8. 8. 8. 8. 8. 8. 8. 8. 8.	51 (+160.09% 46 (+160.10% 36 (+165.69% 36 (+165.69% 80 (+127.96% 37 (+161.23% 313 (+71.98%	SOL L2: i) SOL L2: ii) SOL L2: iii) SOL L2: iiii) SOL L2: iiiii) SOL L2: iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu V U: Dram Throug Lsu Writeback	SC es Active [%] ts [%] Req Cycles A Wavefronts ughput [%] Active [%]	L Memory			55.82 45.27 39.61 38.8(34.1(31.52 24.83	(+124.58 (+178.82 (+107.93) (+71.23) (+76.03 2 (-42.82 (+63.67	%) %) %) %) %) %)	
JOB, Mr. Dippe Franc Cycles Active [N] 7.22 (+165.80%) SOLL 2: Lts2barb Cycles Active [N] 8.82 (-21. JOB, SM. Mol hant Issued [N] 5.55 (+12.200%) SOLL 1: M Xbar2/Tites Read Sectors [N] 6.39 (-25. JOB, SM. Mol hant Issued [N] 5.55 (+12.200%) SOLL 1: M Xbar2/Tites Read Sectors [N] 6.39 (-25. JOB, Kin Inst Executed Pipe Xu [N] 4.59 (+165.50%) SOLL 1: Data Bank Reads [N] 3.00 (-77. JOB, Kin Inst Executed Pipe Aul [N] 1.19 (+165.55%) SOLL 1: Data Bank Reads [N] 1.39 (-165.5%) JOB, Kin Inst Executed Pipe Aul [N] 1.119 (+165.5%) SOLL 1: Data Bank Reads [N] 1.39 (-17.5%)	isue Active [%] ist Executed [%] ipe Shared Cycles A ipe Fp64 Cycles Activ ist Executed Pipe Lis it Executed Pipe Ct tio2rf Writeback Activ	SOL SM .ctive [%] .ive [%] .u [%] e [%] bu Pred On An ive [%]	1 Breakdo		20. 18. 18. 10. 10 8 8	51 (+160.09% 46 (+160.10% 36 (+165.69% 36 (+165.69% 36 (+165.69% 80 (+127.96% 37 (+161.23% 3.13 (+71.98% 1.10 (+161.91%	SOLL2:) SOLL1:) SOLL2:) SOLL2:) SOLL2:) SOLL2:) SOLL3:) SOLL4:) SOLL4:) SOLL1:) SOLL2:	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu V U: Dram Throug Lsu Writeback D Sectors [%]	SC es Active [%] ts [%] Req Cycles A Wavefronts ighput [%] Active [%]	L Memory			55.82 45.27 39.61 38.80 34.10 31.52 24.83 22.64	(+124.58' (+178.82' (+107.93') (+71.23') (+76.03' 2 (-42.82' (+63.67' 4 (+115.71'	%) %) %) %) %) %)	
XDL SM: Milo inst Issued [%] 5.55 (+123.69%) SOL L1: M Xbar2/Itex Read Sectors [%] 6.39 (-25. OL SM: Inst Executed Pipe Xu [%] 4.59 (+66.69%) SOL L1: M Xbar2/Itex Read Sectors [%] 3.00 (+7. OL SM: Inst Executed Pipe Xu [%] 4.59 (+66.69%) SOL L1: M Xbar2/Itex Read Sectors [%] 3.00 (+7. OL SM: Inst Executed Pipe Inflorm [%] 1.29 SOL L1: To ata Bank Winter [%] 1.95 (+23.5 OL SM: Inst Executed Pipe Adu [%] 1.18 (+165.55%) SOL L1: Texin Sm2tex Req Cycles Active [%] 0.00 (+258.4	isue Active [%] ist Executed [%] ipe Shared Cycles A ipe Fp64 Cycles Active St Executed Pipe Ls ipe Alu Cycles Active st Executed Pipe Cl i02rf Writeback Acti li0 Pq Read Cycles A	SOL SM	1 Breakdo		20. 18. 18. 10. 10. 10 8. 8. 8.	51 (+160.09% 46 (+160.10% 36 (+165.69% 36 (+165.69% 36 (+127.96% 37 (+161.23% 313 (+71.98%) 10 (+161.91% 10 (+105.96%)	SOLL2:) SOLL2:) SOLL2:) SOLL2:) SOLL1:) SOLL2:) SOLL2:	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu V U: Dram Throug Lsu Writeback. D Sectors [%] D Sectors Fill [SC es Active [%] ts [%] Req Cycles A Wavefronts ighput [%] Active [%] Device [%]	L Memory			55.82 45.27 39.61 38.8(34.10 31.52 24.83 22.64 12.11	(+124.58' (+178.82' (+107.93') (+71.23') (+76.03' 2 (-42.82' 1 (+63.67' 4 (+115.71' 9 (-12.29'	%) %) %) %) %) %)	
0L.Sk: trat Securited Pipe Xu [%] 4.59 (+166.69%) SOL L1: bata Bank Reads [%] 3.09 (-77) OL Sk: trat Securited Pipe Xu [%] 1.29 SOL L1: bata Bank Writes [%] 1.95 (+23) OL Sk: trat Securited Pipe Adu [%] 1.18 (+165.55%) SOL L1: Toxin Sm2tz: Ref Cycles Active [%] 0.00 (-258)	isue Active [%] ist Executed [%] ipe Shared Cycles A ipe Fp64 Cycles Activ ist Executed Pipe L iio2rf Writeback Activ iio Pq Read Cycles A iio Pq Write Cycles A	SOL SM .ctive [%] .tv [%] a [%] bu Pred On An ive [%] .ctive [%] .ctive [%]	1 Breakdo		20. 18. 18. 10. 10. 8. 8. 8. 7.	51 (+160.09% 46 (+160.10% 36 (+165.69% 80 (+127.96% 80 (+127.96% 37 (+161.23% 3.13 (+71.98% 1.10 (+161.91% 10 (+105.96% 53 (+165.64%	Append Of Light (a) SOLL2: (b) SOLL2: (c) SOLL1: (c) SOLL2: (c) SOLL1: (c) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL2:	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu V J: Dram Throug Lsu Writeback D Sectors [%] D Sectors Fill [Lsuin Requests	SC es Active [%] ts [%] Req Cycles A Wavefronts ighput [%] Active [%] Device [%] s [%]	L Memory			55.82 45.27 39.61 38.8(34.10 31.52 24.83 22.64 12.11 10.80	(+124.58' (+178.82' (+107.93') (+71.23') (+71.23') (+76.03' 2 (-42.82' (+63.67' 4 (+115.71' 9 (-12.29' (+127.96'	%) %) %) %) %) %) %) %)	
OL SM: Inst Executed Pipe Uniform [½] 1.29 SOL L1: Data Bank Writes [½] 1.95 (+23.3 OL SM: Inst Executed Pipe Adu [½] 1.18 (+165.53%) SOL L1: Texin Sm2tex Req Cycles Active [½] 0.00 (+258.4	isue Active [%] ist Executed [%] ipe Shared Cycles Act ist Executed Pipe Ls ipe Alu Cycles Act ist Executed Pipe Cl lio2rf Writeback Act ito Pa Read Cycles A ipe Fma Cycles Act	SOL SM .ctive [%] .tv [%] a [%] bu Pred On An ive [%] .ctive [%] .ctive [%]	1 Breakdo		20. 18. 10. 10 8 8 8. 7. 7. 7.	51 (+160.09) 46 (+160.10) 36 (+165.69) 36 (+165.69) 80 (+127.96) 37 (+161.23) 313 (+71.98) 10 (+1161.91) 10 (+105.96) 53 (+165.64) 22 (+165.80)	Append Of Light (a) SOLL2: (b) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL1: (c) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL1: (c) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL2: (c) SOLL2:	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu U U: Dram Throug Lsu Writeback D Sectors [%] D Sectors Fill D Sectors Fill Lsuin Requests Lts2xbar Cycle	SC es Active [%] ts [%] Req Cycles A Wavefronts ghput [%] Active [%] Device [%] s [%] es Active [%]	PL Memory			55.82 45.27 39.61 38.80 34.10 31.52 24.83 22.84 12.11 10.80 8.83	(+124.58' (+178.82' (+107.93') (+71.23') (+71.23') (+76.03' 2 (-42.82' (+63.67' 4 (+115.71' 9 (-12.29' (+127.96' 2 (-21.48'	%) %) %) %) %) %) %) %) %)	
KOL SM: Inst Executed Pipe Adu [%] 1.18 (+165.53%) SOL L1: Texin Sm2tex Req Cycles Active [%] 0.00 (+258.63%)	isue Active [%] ist Executed [%] ipe Fp64 Cycles A ipe Fp64 Cycles Active ist Executed Pipe Ls ipe Au Cycles Active ist Executed Pipe Ct lio Pa Read Cycles A tio Pa Write Cycles A tio Pa Write Cycles A ipe Fma Cycles Activ	SOL SM ctive [%] ive [%] au [%] au [%] ou Pred On An ive [%] ictive [%] ictive [%] ictive [%]	1 Breakdo		20. 18. 10. 10 8 8. 7. 7. 7. 5.	51 (+160.09) 46 (+160.09) 36 (+165.69) 36 (+165.69) 80 (+127.96) 81 (+127.96) 81 (+127.96) 81 (+127.96) 81 (+127.96) 81 (+128.99) 53 (+128.69) 55 (+128.69)	Solution a) Solution b) Solution c) Solution	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu U U: Dram Throug Lsu Writeback. D Sectors [%] D Sectors [%] D Sectors Fill (Lsuin Requests Lts2xbar Cycle M Xbar2litex R	SC es Active [%] ts [%] Req Cycles A Wavefronts gghput [%] Active [%] Device [%] s [%] Read Sectors	PL Memory			55.82 45.27 39.61 38.86 34.10 31.52 24.83 22.64 12.11 10.80 8.88 6.39	(+124.58' (+178.82' (+107.93') (+71.23') (+76.03' 2 (-42.82' (+63.67' 4 (+115.71' 9 (-12.29' (+127.96' 2 (-21.48' 9 (-25.58'	%) %) %) %) %) %) %) %) %) %)	
	isue Active [%] inst Executed [%] ipe Shared Cycles At pe Fp64 Cycles At St Executed Pipe Ls ipe Alu Cycles Active inst Executed Pipe Cl ioo2rd Writeback Acti ilio Pa Write Cycles A ipe Fma Cycles Acti- filo Inst Issued [%]	SOL SM	1 Breakdo		20. 18. 10. 10 8 8. 7. 7. 7. 5.	\$ 51 (+160.09) 46 (+166.09) 36 (+165.69) 36 (+165.69) 37 (+161.23) 37 (+161.23) 13 (+71.98) 10 (+161.91) 10 (+165.96) 55 (+165.64) 22 (+165.64) 55 (+165.69) 55 (+165.69) 56 (+165.69)	Soll Soll a) Soll Soll b) Soll Soll	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu \ U: Dram Throug Lsu Writeback. D Sectors [%] D Sectors [%] Lsuin Requests Lts2xbar Cycle M Xbar2ltex R Data Bank Read	SC es Active (%) ts (%) Req Cycles A Wavefronts i gghput (%) Active (%) Device (%) s (%) es Active (%) Read Sectors tds (%)	PL Memory			55.82 45.27 39.61 38.88 34.10 31.52 24.83 22.64 12.11 10.80 8.88 8.6.33 3.0	(+124.58' (+178.82' (+107.93') (+71.23') (+76.03' 2 (-42.82' (+63.67' 4 (+115.71' 9 (-12.29' 2 (-21.48' 9 (-25.58' 9 (+77.11'	%) %) %) %) %) %) %) %) %) %) %) %) %) %	
SUC IDC: Request Cycles Active [76] 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.0 0.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00 (+258.00	isue Active [%] inst Executed [%] ipe Shared Cycles A ipe Fp64 Cycles Active ist Executed Pipe L ipe Alu Cycles Active ist Executed Pipe C ibo27 Writeback Acti ito Pq Read Cycles Active file Pf Read Cycles Active file Pf Read Cycles Active file Fina Cycles Active to Inst Issued [%] ist Executed Pipe U	SOL SM ctive [%] ive [%] tu [%] bu Pred On An ive [%] bu Vre(%] kctive [%] kctive [%] j [%] j [%]	1 Breakdo		20. 18. 10. 10. 10. 5. 8. 8. 7. 7. 7. 5. 4.	\$.51 (+160.09% 46 (+160.10% 36 (+165.69% 36 (+165.69% 37 (+161.23% 1.3 (+71.98% 1.10 (+161.91%) 1.0 (+165.64% 52 (+165.64%) 55 (+165.69%) 55 (+165.69%) 1.2	i) SOL L2: i) SOL L2: i) SOL L1: i) SOL L2: i) SOL L2: i) SOL L1: i) SOL L2: i) SOL L1: i) SOL L1: i) SOL L1: i) SOL L1:	t [%] Xbar2lts Cycle T Tag Request M Litex2xbar R T Sectors [%] Data Pipe Lsu V U: Dram Throug Lsu Writeback. D Sectors Fill D Sectors Fill D Sectors Fill C Lsuin Requests Lts2xbar Cycle M Xbar2litex R Data Bank Read Data Bank Writ	SC es Active [%] ts [%] Req Cycles A Wavefronts ghput [%] Active [%] Device [%] s [%] es Active [%] Read Sectors (%) ftes [%]	Ctive [%]			55.82 45.27 39.61 38.80 34.10 31.55 24.83 22.84 12.11 10.80 8.83 6.39 3.00 1.95	(+124.58' (+178.82' (+107.93') (+716.33') (+716.33' 2 (-42.82' 1 (+63.67' 4 (+115.71' 9 (-12.29' (+127.96' 2 (-21.48' 9 (+77.11' (+23.96'	%] %] %] %] %] %] %] %] %] %] %]	
201 014 (-0.001) 001 (0.000) 001 (0.000) 000 (-0.000)	isue Active [%] ist Executed [%] ipe Shared Cycles Act ist Executed Pipe Ls ist Executed Pipe Ls ist Executed Pipe Cl fio2rf Writeback Act iso Pa Read Cycles A fio Pa Write Cycles A ist Executed Pipe X. ist Executed Pipe X. ist Executed Pipe X.	SOL SM	1 Breakdo		20. 18. 10. 10. 8. 8. 7. 7. 7. 5. 4.	\$ 51 (+160.09% 46 (+160.10% 36 (+165.69% 30 (+127.96% 30 (+127.96% 30 (+127.96% 31 (+71.98%) 10 (+105.96% 53 (+105.96%) 55 (+123.69% 55 (+123.69%) 55 (+123.69%) 12 (+165.53%) 12 (16 (+165.53%) 12 (16 (+165.53%)) 12 (16 (+165.53%))	SOL L2: i) SOL L2: ii) SOL L2: iii) SOL L1: iii) SOL L2: iii) SOL L1:	t (%) Xbar2lts Cycle T Tag Request T Tag Request M LitezZhar R T Sectors (%) Data Pipe Lsu V U: Dram Throupus Lsu Writeback D Sectors (%) D Se	SC es Active [%] ts [%] Req Cycles A Wavefronts gighput [%] Active [%] es Active [%] es Active [%] Read Sectors ids [%] Read Sectors ids [%] Req Cycles A	Ctive [%]			55.82 45.27 39.61 38.80 34.10 31.55 22.64 12.11 10.80 8.83 6.33 3.00 1.95 0.000	(+124.58' (+178.82' (+107.93') (+71.23') (+71.23') (+72.03' 2 (-42.82' 1 (+63.67' 4 (+115.71' 9 (-12.29' (+127.96' 2 (-21.48' 9 (+27.11' i (+23.96' (+258.69'	%) %) %) %) %) %) %) %) %) %) %) %) %) %	
SOL SM: Inst Executed Pipe Tex [%] 0 (+0.00%) SOL L2: D Sectors Fill Sysmem [%] 0.00 (+i SOL SM: Inst Executed Pipe Ipa [%] 0 (+0.00%) SOL L1: D Sectors Fill Sysmem [%] 0.00 (+i	isue Active [%] ist Executed [%] ipe Shared Cycles Active ipe F064 Cycles Active ist Executed Pipe L flio2rf Writeback Active ist Executed Pipe Active ipe Fma Cycles Active flio Fm Cycles Active flio Fma Cycles Active ist Executed Pipe Active ist	SOL SM ctive [%] ive [%] a [%] a [%] uu [%] uu [%] tctive [%] a [%] nform [%] du [%] e [%]	1 Breakdo		20. 18. 10. 10. 8. 8. 7. 7. 7. 5. 4.	\$ 51 (+160.09) 46 (+160.10) 36 (+165.69) 38 (+167.96) 37 (+161.23) 37 (+161.23) 37 (+161.23) 37 (+161.23) 38 (+179.96) 10 (+105.96) 53 (+165.64) 55 (+123.69) 59 (+165.69) 1.2 18 (+165.53)	ppeed of Light) SOL L2:) SOL L3:) SOL L4:) SOL L4:	t (%) Xbar2lts Cycle T Tag Request M Litex2bar7 Data Pipe Lsu u U: Dram Throug Lsu Writeback U: Dram Throug Lsu Writeback D Sectors [%] D Sectors [%] D Sectors [%] D Sectors Cycle M Xbar21tes R Data Bank Rea Tasin Sm2tes R Fuel Sectors (%) Sectors (%) Data Sectors (%) Data Sectors (%) Data Sectors (%) Data Sectors (%) Sectors (%) Sectors (%) Sectors (%) Sectors (%) Data Sectors (%) Data Sectors (%) Sectors (%) Data Sectors (%) Sectors (%) Sectors (%) Sectors (%) Data Sectors (%) D Sect	SC es Active [%] ts [%] Req Cycles A Wavefronts ghput [%] Active [%] es Active [%] Read Sectors kds [%] tes [%] Req Cycles A [%]	ctive [%]			55.82 45.27 39.61 38.86 34.10 31.55 24.83 22.64 12.11 10.80 8.83 6.39 3.0 1.95 0.00 0.000	(+124.58' (+178.82' (+178.82' (+107.93') (+71.23') (+71.23') (+76.03' (+72.82') (+72.84') (+127.96' (+127.96') (+21.48') (+25.58') (+22.96') (+22.96.69') (+258.69')	%) %) %) %) %) %) %) %) %) %) %) %) %) %	

Programming GPUs Advanced Topics

Advanced Topics

So much more interesting things to show!

- Optimize memory transfers to reduce overhead
- Optimize applications for GPU architecture
- Drop-in BLAS acceleration with NVBLAS (\$LD_PRELOAD)
- Tensor Cores for Deep Learning
- Libraries, Abstractions: Kokkos, Alpaka, Futhark, HIP, SYCL, C++AMP, C++ pSTL, ...
- Use multiple GPUs
 - On one node
 - Across many nodes \rightarrow MPI
-
- Some of that: Addressed at dedicated training courses



Using GPUs on JSC Systems

Compiling

- CUDA Module: module load CUDA/12
 - Compile: nvcc file.cu
 - Example cuBLAS:g++ file.cpp -I\$CUDA_HOME/include -L\$CUDA_HOME/lib64
 -lcublas -lcudart
- OpenACC Module: module load NVHPC/23.7-CUDA-12
 - Compile: nvc++ -acc=gpu file.cpp

MPI CUDA-aware MPIs (with direct Device-Device transfers) ParaStationMPI module load ParaStationMPI/5.9.2-1 MPI-settings/CUDA OpenMPI module load OpenMPI/4.1.5 MPI-settings/CUDA



Running

 Dedicated GPU partitions JUWELS

```
\begin{array}{l} --\text{partition=gpus} \quad 46 \text{ nodes (Job limits: } \leq 1 \text{ d}) \\ --\text{partition=develgpus} \quad 10 \text{ nodes (Job limits: } \leq 2 \text{ h}, \leq 2 \text{ nodes}) \\ \text{JUWELS Booster} \\ \quad --\text{partition=booster} \quad 926 \text{ nodes} \\ --\text{partition=develbooster} \quad 10 \text{ nodes (Job limits: } \leq 1 \text{ d}, \leq 2 \text{ nodes}) \\ \text{JURECA DC} \end{array}
```

--partition=dc-gpu 192 nodes

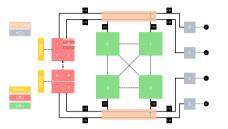
Needed: Resource configuration with --gres=gpu:4

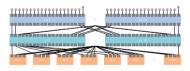
 \rightarrow See online documentation

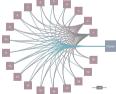


Running JUWELS Booster Topology

- JUWELS Booster: NPS-4 (in total: 8 NUMA Domains)
- Not all have GPU or HCA affinity!
- Network is structured into two levels: In-Cell and Inter-Cell (DragonFly+ network)







JÜLICH Forschungszentrum

→ Documentation: apps.fzjuelich.de/jsc/hps/juwels/

Example

- 16 tasks in total, running on 4 nodes
- Per node: 4 GPUs

```
#!/bin/bash -x
#SBATCH --nodes=4
#SBATCH --ntasks=16
#SBATCH --ntasks-per-node=4
#SBATCH --output=gpu-out.%j
#SBATCH --error=gpu-err.%j
#SBATCH --time=00:15:00
#SBATCH --partition=gpus
#SBATCH --partition=gpus
```

```
#SBATCH --gres=gpu:4
```

```
srun ./gpu-prog
```



Conclusion

- GPUs provide highly-parallel computing power
- We have many devices installed at JSC, ready to be used!



- GPUs provide highly-parallel computing power
- We have many devices installed at JSC, ready to be used!
- Training courses by JSC next year
- See online documentation and sc@fz-juelich.de



- GPUs provide highly-parallel computing power
- We have many devices installed at JSC, ready to be used!
- Training courses by JSC next year
- See online documentation and sc@fz-juelich.de
- Further consultation via our lab: NVIDIA Application Lab in Jülich; contact me!



- GPUs provide highly-parallel computing power
- We have many devices installed at JSC, ready to be used!
- Training courses by JSC next year
- See online documentation and sc@fz-juelich.de
- Further consultation via our lab: NVIDIA Application Lab in Jülich; contact mel-





Appendix

Appendix Glossary References



Glossary I

- AMD Manufacturer of CPUs and GPUs. 52, 53, 54, 55, 56, 57, 88, 90 Ampere GPU architecture from NVIDIA (announced 2019). 4, 5, 6
 - API A programmatic interface to software by well-defined functions. Short for application programming interface. 52, 53, 54, 55, 56, 57
 - CUDA Computing platform for GPUs from NVIDIA. Provides, among others, CUDA C/C++. 2, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 77, 90
 - HIP GPU programming model by AMD to target their own and NVIDIA GPUs with one combined language. Short for Heterogeneous-compute Interface for Portability. 52, 53, 54, 55, 56, 57



Glossary II

- JSC Jülich Supercomputing Centre, the supercomputing institute of Forschungszentrum Jülich, Germany. 2, 82, 83, 84, 85, 89
- JURECA A multi-purpose supercomputer at JSC. 6
- JUWELS Jülich's new supercomputer, the successor of JUQUEEN. 3, 4, 5, 78
 - MPI The Message Passing Interface, a API definition for multi-node computing. 75, 77
 - NVIDIA US technology company creating GPUs. 3, 4, 5, 6, 26, 27, 28, 52, 53, 54, 55, 56, 57, 70, 82, 83, 84, 85, 88, 90
- OpenACC Directive-based programming, primarily for many-core machines. 46, 47, 48, 49, 50, 77



Glossary III

- OpenCL The Open Computing Language. Framework for writing code for heterogeneous architectures (CPU, GPU, DSP, FPGA). The alternative to CUDA. 52, 53, 54, 55, 56, 57
- OpenMP Directive-based programming, primarily for multi-threaded machines. 46, 47, 48, 49, 50
 - ROCm AMD software stack and platform to program AMD GPUs. Short for Radeon Open Compute (*Radeon* is the GPU product line of AMD). 52, 53, 54, 55, 56, 57
 - SAXPY Single-precision $A \times X + Y$. A simple code example of scaling a vector and adding an offset. 34, 67, 68
 - Tesla The GPU product line for general purpose computing computing of NVIDIA. 3



Glossary IV

- CPU Central Processing Unit. 3, 6, 10, 11, 12, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 34, 52, 53, 54, 55, 56, 57, 88, 90
- GPU Graphics Processing Unit. 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 33, 37, 38, 39, 40, 41, 42, 45, 46, 47, 48, 51, 52, 53, 54, 55, 56, 57, 68, 69, 70, 74, 75, 76, 78, 79, 80, 82, 83, 84, 85, 88, 89, 90
- SIMD Single Instruction, Multiple Data. 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
- SIMT Single Instruction, Multiple Threads. 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
 - SM Streaming Multiprocessor. 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
- SMT Simultaneous Multithreading. 19, 20, 21, 22, 23, 24, 25, 26, 27, 28



References I

- [2] Karl Rupp. Pictures: CPU/GPU Performance Comparison. URL: https://www.karlrupp.net/2013/06/cpu-gpu-and-mic-hardwarecharacteristics-over-time/ (pages 8, 9).
- [6] Wes Breazell. Picture: Wizard. URL: https://thenounproject.com/wes13/collection/its-a-wizards-world/ (pages 37-41).



References: Images, Graphics I

- [1] Forschungszentrum Jülich GmbH (Ralf-Uwe Limbach). JUWELS Booster.
- [3] Mark Lee. Picture: kawasaki ninja. URL: https://www.flickr.com/photos/pochacco20/39030210/ (pages 10, 11).
- [4] Shearings Holidays. Picture: Shearings coach 636. URL: https://www.flickr.com/photos/shearings/13583388025/ (pages 10, 11).
- [5] Nvidia Corporation. Pictures: Volta GPU. Volta Architecture Whitepaper. URL: https://images.nvidia.com/content/volta-architecture/pdf/Volta-Architecture-Whitepaper-v1.0.pdf.

