High Performance Computing Systems (CMSC714)





Abhinav Bhatele, Department of Computer Science

Summary of last lecture

- MPI trace visualization
- Projections performance analysis tool
- Hatchet: programmable by the user







von Neumann architecture

		Central
		C
Input Device		Arithm
		Me
	http	os://en.wikipedia.o



Abhinav Bhatele (CMSC714)





UMA vs. NUMA



Uniform Memory Access

https://frankdenneman.nl/2016/07/07/numa-deep-dive-part-1-uma-numa/



Non-uniform Memory Access

Abhinav Bhatele (CMSC714)



UMA vs. NUMA



Uniform Memory Access

https://frankdenneman.nl/2016/07/07/numa-deep-dive-part-1-uma-numa/



Non-uniform Memory Access

Abhinav Bhatele (CMSC714)



Fast vs. slow cores

- Intel Core line (Nehalem, Sandy Bridge, Ivy Bridge, Haswell, Broadwell, ...)
- AMD processors (Opteron, Athlon, Zen, ...)
- IBM Power line
- Slower cores: Low frequency, low power

• IBM PowerPC line (440, 450, A2, ...)







Abhinav Bhatele (CMSC714)



Intel Haswell Chip





Abhinav Bhatele (CMSC714)

BQC Chip

- A2 processor core
 - Runs at I.6 GHz
- Shared L2 cache
- Peak performance per core:
 - 12.8 Gflop/s
- Total performance per node: 204.8 Gflop/s





Abhinav Bhatele (CMSC714)

LIVE RECORDING

7



- NVIDIA: Fermi, Kepler, Maxwell, Pascal, Volta, ...
- AMD
- Intel
- Figure on the right shows a single node of Summit @ ORNL







HBM & DRAM speeds are aggregate (Read+Write). All other speeds (X-Bus, NVLink, PCIe, IB) are bi-directional.

Abhinav Bhatele (CMSC714)



0





Abhinav Bhatele (CMSC714)

LIVE RECORDING

9





Volta GV100 SM

Each Volta Streaming Multiprocessor (SM) has:

- 64 FP32 cores
- 64 INT32 cores
- 32 FP64 cores
- 8 Tensor cores

https://images.nvidia.com/content/volta-architecture/pdf/volta-architecture-whitepaper.pdf





SM

LIVE RECORDING

Tex

Tex

							L1 Instructi	ion Cache						
		L0 Ir	nstruct	ion C	ache					L0 Ir	nstruc	tion C	ache	
	Wai	r <mark>p Sc</mark> h	neduler	(32 t	hread/c	clk)			War	<mark>rp Sc</mark> h	nedule	r (32 t	hread/cl	lk)
	Di	spatcl	h Unit (32 th	read/cl	k)			Di	spatcl	h Unit	(32 th	read/clk	;)
	Reg	ister	File (1	6,384	4 x 32-	bit)			Reg	ister	File (′	16,384	4 x 32-b	oit)
FP64	INT	INT	FP32	FP32	\square			FP64	INT	INT	FP32	FP32	\square	H
FP64	INT	INT	FP32	FP32	\square			FP64	INT	INT	FP32	FP32		H
FP64	INT	INT	FP32	FP32	Ħ			FP64	INT	INT	FP32	FP32	\square	Ħ
FP64	INT	INT	FP32	FP32	TENS	SOR	TENSOR	FP64	INT	INT	FP32	FP32	TENS	OR
FP64	INT	INT	FP32	FP32	CO	RE	CORE	FP64	INT	INT	FP32	FP32	COR	E
FP64	INT	INT	FP32	FP32	Ħ			FP64	INT	INT	FP32	FP32	\square	Ħ
FP64	INT	INT	FP32	FP32	H			FP64	INT	INT	FP32	FP32	\square	Ħ
FP64	INT	INT	FP32	FP32	H			FP64	INT	INT	FP32	FP32		Ħ
LD/ LD/ ST ST	LD/ ST	LD/ ST	LD/ ST	LD/ ST	LD/ ST	LD/ ST	SFU	LD/ LD/ ST ST	LD/ ST	LD/ ST	LD/ ST	LD/ ST	LD/ L ST S	LD/ ST
		L0 lr	nstruct	ion C	ache					L0 Ir	nstruc	tion C	ache	
	Wai Di Reg	L0 Ir p Sch spatcl ister	nstruct neduler h Unit (File (1	ion C (32 tl (32 th 6,384	ache hread/c read/cl 4 x 32-	cik) k) bit)			War Dis Reg	L0 Ir rp Sch spatcl ister	nstruc nedule h Unit File ('	tion C r (32 t (32 th 16,384	ache hread/cl read/clk 4 x 32-b	k) ;))it)
FP64	War Di Reg	L0 Ir p Sch spatcl ister	nstruct neduler h Unit (File (1 FP32	ion C (32 th 32 th 6,384 FP32	ache hread/c read/cl 4 x 32-	cik) k) bit)		FP64	War Dis Reg	L0 Ir rp Sch spatcl ister INT	nstruc nedule h Unit File (' FP32	tion C r (32 th (32 th 16,384 FP32	ache hread/cl read/clk 4 x 32-b	ik) ;) pit)
FP64 FP64	War Di Reg	L0 Ir p Sch spatcl ister INT INT	nstruct neduler h Unit (File (1 FP32 FP32	ion C (32 th 32 th 6,384 FP32 FP32	ache hread/c read/cl 4 x 32-	cik) k) bit)		FP64 FP64	War Dis Reg INT	L0 Ir rp Sch spatcl ister INT INT	nstruc nedule h Unit File (' FP32 FP32	tion C r (32 th (32 th 16,384 FP32 FP32	ache hread/cl read/clk 4 x 32-b	ik) ;) pit)
FP64 FP64 FP64	War Dis Reg INT INT	L0 Ir p Sch spatcl ister INT INT	nstruct neduler h Unit (File (1 FP32 FP32 FP32	ion C (32 th 32 th 6,384 FP32 FP32 FP32	ache hread/cl read/cl 4 x 32-	cik) k) bit)		FP64 FP64 FP64	War Dis Reg INT INT	L0 Ir rp Sch spatcl ister INT INT	nstruc nedule h Unit File (' FP32 FP32 FP32	tion C r (32 th (32 th 16,384 FP32 FP32 FP32	ache hread/cl read/clk 4 x 32-b	ik) ;) pit)
FP64 FP64 FP64 FP64	Wai Dis Reg INT INT INT	L0 Ir p Sch spatcl ister INT INT INT	nstruct neduler h Unit (File (1 FP32 FP32 FP32 FP32	ion C (32 th 32 th 6,384 FP32 FP32 FP32 FP32	ache hread/cl 4 x 32-	bit)	TENSOR	FP64 FP64 FP64 FP64	War Dis Reg INT INT INT	L0 Ir rp Sch spatcl ister INT INT INT	nstruc nedule h Unit File (' FP32 FP32 FP32 FP32	tion C r (32 th (32 th 16,384 FP32 FP32 FP32 FP32	ache hread/cl read/clk 4 x 32-b	k) it) OR
FP64 FP64 FP64 FP64 FP64	Wai Dis Reg INT INT INT INT	L0 Ir p Sch spatcl ister INT INT INT INT	nstruct neduler h Unit (File (1 FP32 FP32 FP32 FP32 FP32	ion C (32 th 32 th 6,384 FP32 FP32 FP32 FP32 FP32	ache hread/cl 4 x 32- TENS COF	bit)	TENSOR	FP64 FP64 FP64 FP64 FP64 FP64	War Dis Reg INT INT INT INT	L0 Ir rp Sch spatcl ister INT INT INT INT	nstruc nedule h Unit File (' FP32 FP32 FP32 FP32	tion C r (32 th (32 th 16,384 FP32 FP32 FP32 FP32 FP32	ache hread/cl read/clk 4 x 32-b	k) it) OR E
FP64 FP64 FP64 FP64 FP64 FP64	Wai Dis Reg INT INT INT INT INT	L0 Ir p Sch spatcl ister INT INT INT INT INT	nstruct neduler h Unit (File (1 FP32 FP32 FP32 FP32 FP32 FP32	ion C (32 th 32 th 6,384 FP32 FP32 FP32 FP32 FP32	ache hread/cl 4 x 32- TENS COP	bit)		FP64 FP64 FP64 FP64 FP64 FP64 FP64	War Dis Reg INT INT INT INT INT	L0 Ir rp Sch spatcl ister INT INT INT INT INT	nstruc nedule h Unit File (' FP32 FP32 FP32 FP32 FP32 FP32	tion C r (32 th (32 th 16,384 FP32 FP32 FP32 FP32 FP32 FP32	ache hread/cl read/clk 4 x 32-b	oit)
FP64 FP64 FP64 FP64 FP64 FP64 FP64	Wai Dis Reg INT INT INT INT INT INT	L0 Ir p Sch spatcl ister INT INT INT INT INT INT	nstruct neduler h Unit (File (1 FP32 FP32 FP32 FP32 FP32 FP32 FP32	ion C (32 t) (32 th 6,384 FP32 FP32 FP32 FP32 FP32 FP32 FP32	ache hread/cl 4 x 32- TENS COP	bit)		FP64 FP64 FP64 FP64 FP64 FP64 FP64 FP64	War Dis Reg INT INT INT INT INT INT	L0 Ir p Sch spatcl ister INT INT INT INT INT INT	nstruc nedule h Unit File (* FP32 FP32 FP32 FP32 FP32 FP32 FP32	tion C r (32 th (32 th 16,384 FP32 FP32 FP32 FP32 FP32 FP32 FP32	ache hread/cl read/clk 4 x 32-b	oit)
FP64 FP64 FP64 FP64 FP64 FP64 FP64 FP64	War Dis Reg INT INT INT INT INT INT INT	L0 Ir p Sch spatcl ister INT INT INT INT INT INT INT	nstruct neduler h Unit (File (1 FP32 FP32 FP32 FP32 FP32 FP32 FP32 FP32	ion C (32 t) (32	ache hread/cl 4 x 32- TENS COP	sor		FP64 FP64 FP64 FP64 FP64 FP64 FP64 FP64	War Dis Reg INT INT INT INT INT INT INT	L0 Ir rp Sch spatcl ister INT INT INT INT INT INT INT	nstruc nedule h Unit File (' FP32 FP32 FP32 FP32 FP32 FP32 FP32 FP32	tion C r (32 th (32 th 16,384 FP32 FP32 FP32 FP32 FP32 FP32 FP32 FP32	ache hread/cl read/clk 4 x 32-b	k) bit) OR

Tex

Tex





UNIVERSITY OF MARYLAND

Questions?



Abhinav Bhatele 5218 Brendan Iribe Center (IRB) / College Park, MD 20742 phone: 301.405.4507 / e-mail: bhatele@cs.umd.edu