GPU BASED HOUGH TRACK FINDER

15.06.2021 I PANDA COLLABORATION MEETING | ANNA ALICKE





Mitglied der Helmholtz-Gemeinschaft



Intention

- tracking: most computing intensive part of event reconstruction
- high event rate in panda: 20 mio events/s
- ightarrow tracking algorithm: must be really fast to reduce the amount of data
- \rightarrow test if GPUs can deliver the speed up we need



CPU CPU

Architectural differences: CPU vs GPU

Large cache and controlling

 → fast handling of complex sequential calculations



GPU



Taken from: NVIDIA Corporation (2021), 'NVIDIA CUDA C++ Programming Guide', Version 11.2.1.



Taken from: NVIDIA Corporation (2021), 'NVIDIA CUDA C++ Programming Guide', Version 11.2 .1.

Architectural differences: CPU vs GPU

CPU

Large cache and controlling

 → fast handling of complex sequential calculations

GPU









Does the HoughTrackFinder fulfill the conditions?

Hough transformation + Apollonius problem

X interaction point (IP)

15.06.2021

- 4 -































GPU PROGRAMMING CHALLENGES

- Allocate memory and copy data to GPU is time intensive
- \rightarrow Main aims:
 - Reduce amount of data that have to be copied or allocated
 - Increase computing time of GPU
 - Increase occupancy of GPU

 Inreads (8) 								
▼ ✔ [7697] root.exe 🕞	'I	a selected received	·7					
OS runtime libraries CUDA API	iocti				n		cudaDeviceSynchro)	
 ▶ ▼ [7758] root.exe ▼ 				COP				
🕨 🗹 [7756] cuda-EvtHan 🕶								500000000
5 threads hidden+		-				-		
→ ○ [7721]		4						
• ([//10]	4							
vents View 👻								, P
		/						
Allocate	memo	ory on GP	U		Computing ti	ime on GPU		
							Л Л Л Л Л Л Л Л	
Mitglied der H	Helmholtz-	Gemeinschaft		15.06.202	21 - 7 -		Forschungszentrum	

GPU PROGRAMMING CHALLENGES

- Allocate memory and copy data to GPU is time intensive
- \rightarrow Main aims:
 - Reduce amount of data that have to be copied or allocated
 - Increase computing time of GPU
 - Increase occupancy of GPU

 Many events have to be calculated in parallel

▼ 🗸 [7697] root.exe 🔹	1	a selected restrict	· • • • •					<u> </u>	
OS runtime libraries CUDA API	iii i mi hii hadin iocti	Dili manifi makanan			i mili i i i i i i i i i i i i i i i i i		mmmmd	in finite in the second s	iceSynchro)
Profiler overhead				CUP	,[
▶ 🗸 [7758] root.exe 🕞									
▶ 🔽 [7756] cuda-EvtHan 🗸				000000					
5 threads hidden 🗕 🛨		V					_		
CUDA (GeForce RTX 2060)									
• (7721)									
• (7716]	4	/							
vents View 👻		_/							r
,		/							
Allocate	memo	ory on GF	PU		Сс	omputing ti	me on GF	շՍ	
								🖊 🛦 . IÜL I	СН
Mitglied der H	lelmholtz	-Gemeinschaft		15.06.2	2021	- 7 -		Forschungs	zentrum

COMPARISON CPU VS GPU



• Buffered events: events calculated in parallel





COMPARISON CPU VS GPU



- Buffered events: events calculated in parallel
- Before tracking data have to bee collected.
 - \rightarrow Software specific, different for real detector





COMPARISON CPU VS GPU



- Buffered events: events calculated in parallel
- Before tracking data have to bee collected.
 - \rightarrow Software specific, different for real detector







	CPU time [ms/event]	GPU Time [ms/event]	Speed up
Apollonius calculation	4.25	0.05	85
Hough Space	7.27	2.35	3
Hough transformation (Maximum Finding)	8.37	3.19	2.6
With merging	12.54	4.88	2.5







	CPU time [ms/event]	GPU Time [ms/event]	Speed up
Apollonius calculation	4.25	0.05	85
Hough Space	7.27	2.35	3
Hough transformation (Maximum Finding)	8.37	3.19	2.6
With merging	12.54	4.88	2.5









	CPU time [ms/event]	GPU Time [ms/event]	Speed up					
Apollonius calculation	4.25	0.05	85	Fast thread memory is too small \rightarrow Slow global memory accesses				
Hough Space	7 27	2 35	3	Block 1		Block 2		
nough opace	1.21	2.00	Ū	Thread 1	Thread 2		Thread 1	Thread 2
Hough transformation	8.37	3.19	2.6	Local memory	Local memory		Local memory	Local memory
(Maximum Finding)				Shared memory		Shared memory		
With merging	12.54	4.88	2.5	Global memory				







	CPU time [ms/event]	GPU Time [ms/event]	S u	Speed Ip	A whole method per thread					
Apollonius calculation	4.25	0.05		85	 Fast thread memory is too small → Slow global memory accesses 					
Hough Space	7 27	2.35		3	Block 1			Block 2		
riougii Opuce					Thread 1	Thread 2		Thread 1	Thread 2	
Hough	8.37	3.19		2.6	Local memory	Local memory		Local memory	Local memory	
(Maximum Finding)					Shared	memory		Shared	memory	
With merging	12.54	4.88		2.5	Global memory					
1 Apollonius Circle per thread										
Mitglied der Helmholtz-Gemeinschaft		15.0	06.:	2021	- 11 -			JÜL	ICH Igszentrum	





	CPU time [ms/event]	GPU Time [ms/event]	Speed up	Memory usage [GB]	Buffer Size	Blocks * Threads			
Apollonius calculation	4.25	0.05	85	2	10000	393812 * 256			
Hough Space	7.27	2.35	3	4	4000	125 * 256			
Hough transformation (Maximum Finding)	8.37	3.19	2.6	4	4000	125 * 256			
With merging	12.54	4.88	2.5	4	2000	8 * 256			
		\rightarrow Bad GPU occupancy							
→ Potential for improvement									

15.06.2021

- 12 -



QUALITY + RUNTIME COMPARISON



- GPU version is a bit faster
- Still too slow
- Potential for development



SUMMARY & OUTLOOK



Summary

- GPU version of HoughTrackFinder is implemented
- \rightarrow Large speed up only for one part of the algorithm
- Speed up in total: ~2

Outlook

- Speed: Potential for development
 - \rightarrow memory problem: allocate all Hough spaces for all tracklets
 - \rightarrow better method has to be found
- Secondaries: as new improvement for tracking in PANDA



Thank you for your attention!



Mitglied der Helmholtz-Gemeinschaft