SMARTHEP

REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

ESR12: Accelerated Anomaly Detection

Pratik Jawahar

Supervisors:
Caterina Doglioni, Jiri Masik, Alex Oh, Maurizio Pierini

Overview:

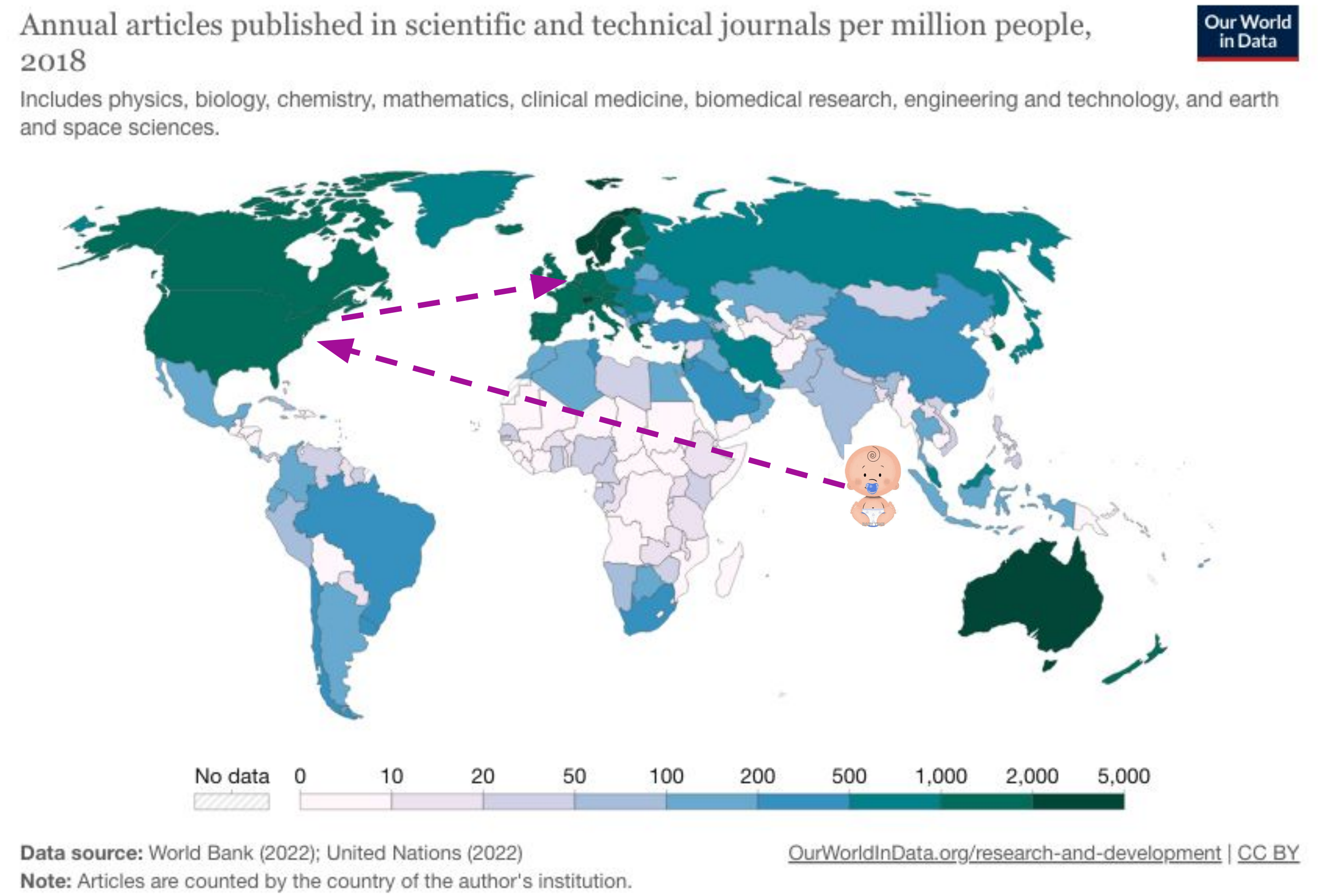
- Qualification Task - Heterogenous Tracking
- ML based Data Compression - Baler
- Misc. Activities

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824 The University of Manchester

About me:

- BS: Mechanical Engineering (2019)
- MS: Robotics Engineering (2022)
- PhD: Robo.. *sike* Particle Physics
- Experience:
    - Summer student (CERN, 2018)
        - Control algorithms for GEM detectors (CMS)
    - Technical student (CERN, 2021), DIANA HEP Fellowship (2020)
        - ML based anomaly detection for new physics searches
- Website: https://www.pratikjawahar.com/



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824
The University of Manchester

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

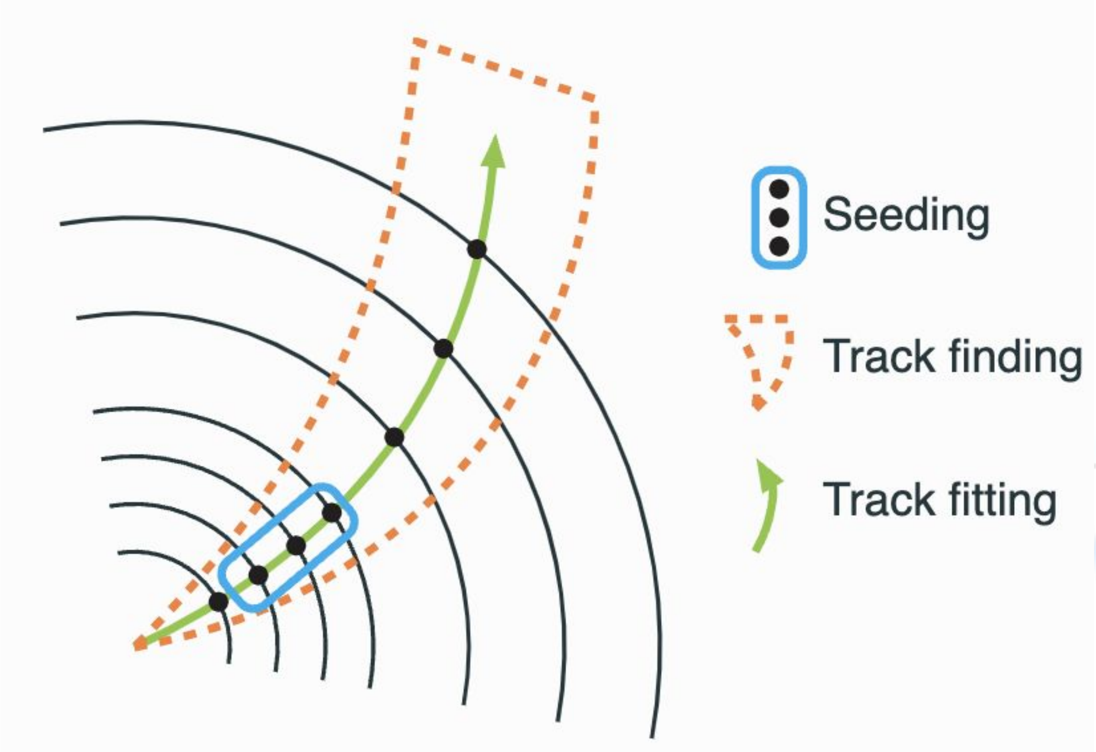
Qualification Task: Heterogenous Track Reconstruction

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824
The University of Manchester

Track Reconstruction:

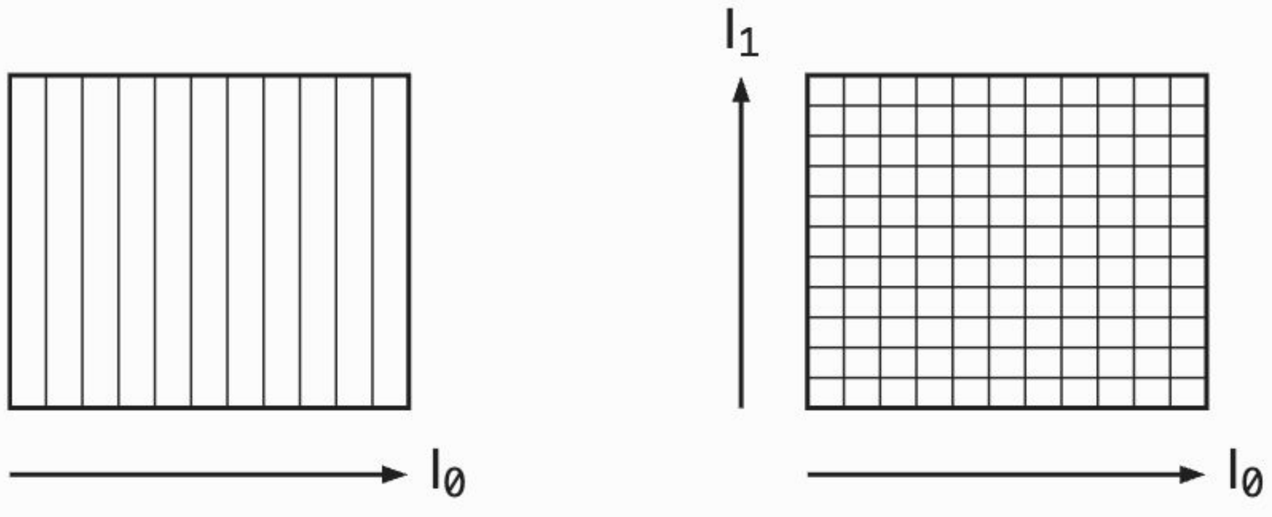
- Reconstruct trajectories of charged particles as they pass through different layers of the detector, subject to a magnetic field



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Track Reconstruction:

- Step 1: Detect a charged particle when it passes through a layer
  - Ionization in semiconductors

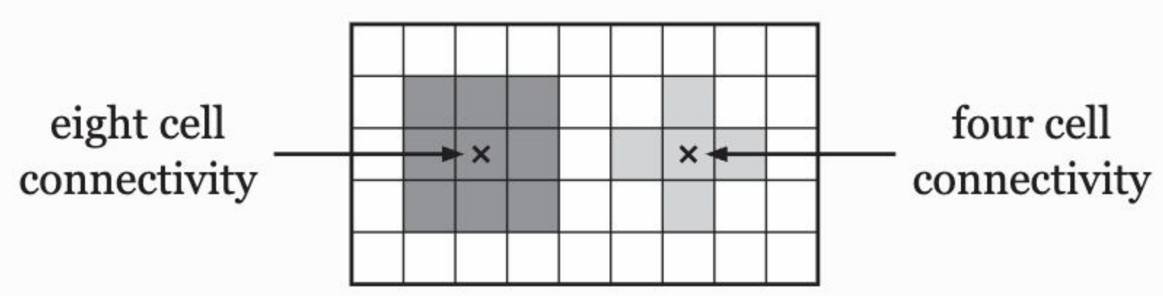


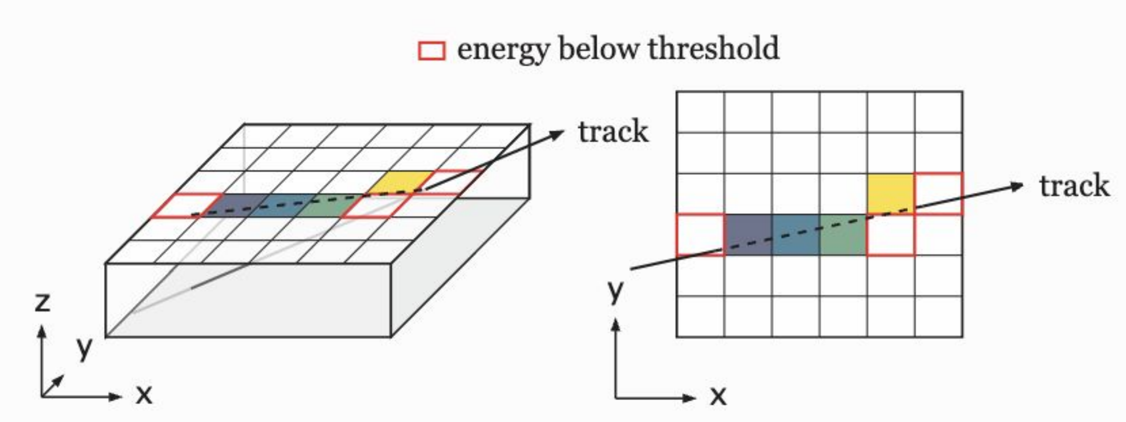
(a) strip (b) pixel

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Track Reconstruction:

- Step 2: Clusterization
  - Depends on how you treat readout, detector design etc.
  - Connected Components Algo (CCA) iterated on edges

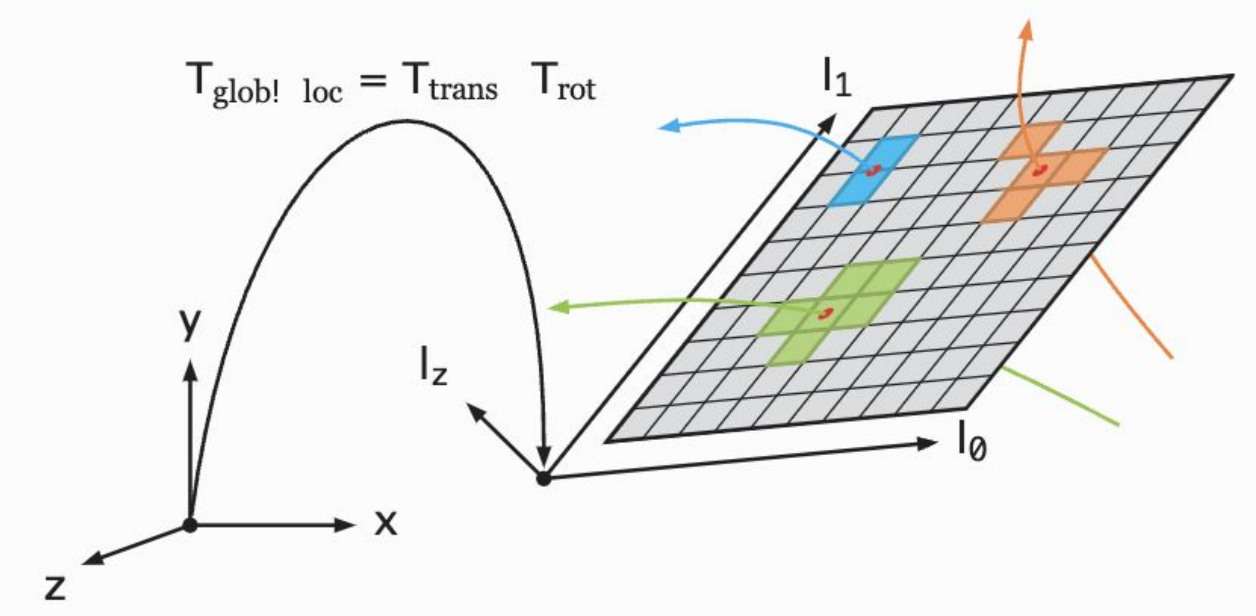




SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Track Reconstruction:

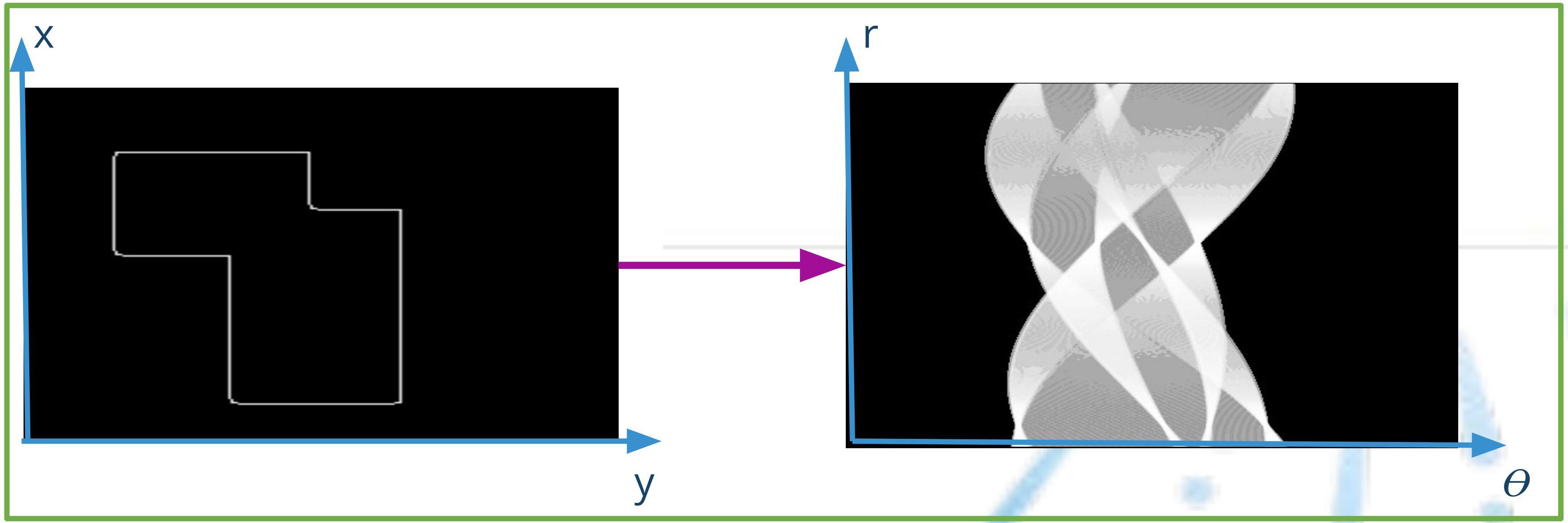
- Step 3: Spacepoint Formation
  - Move from local sub-detector frame to global detector frame

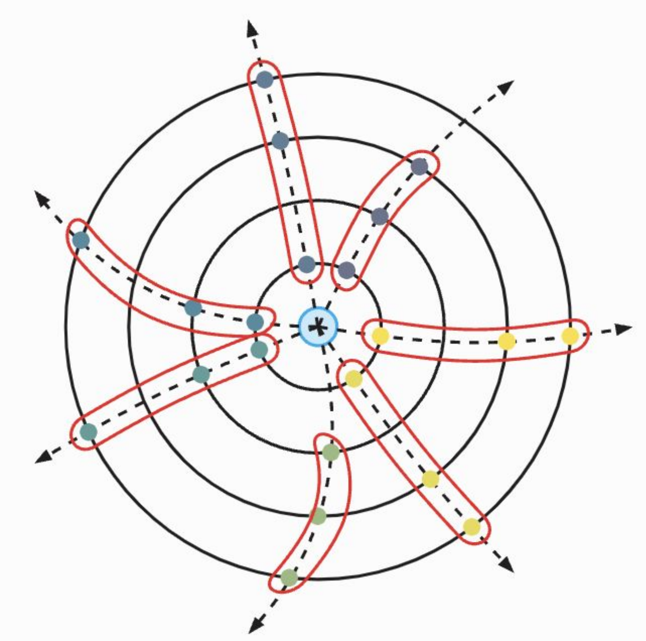


SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Track Reconstruction:

- Step 4: Seeding
  - Find triplets of spacepoints that could potentially belong to the same track
    - Conformal maps => Hough Transform

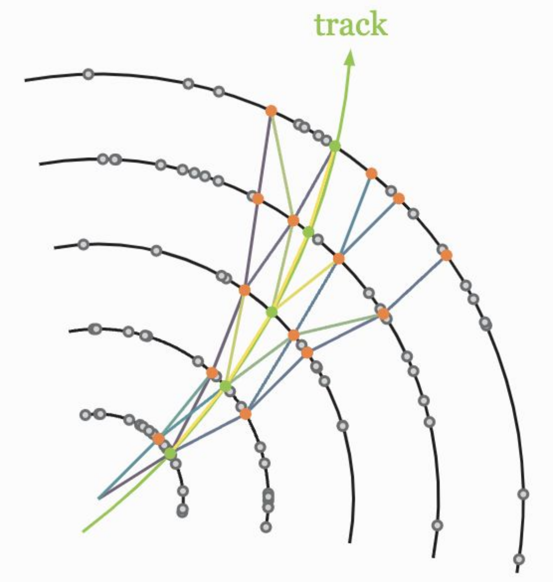




SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Track Reconstruction:

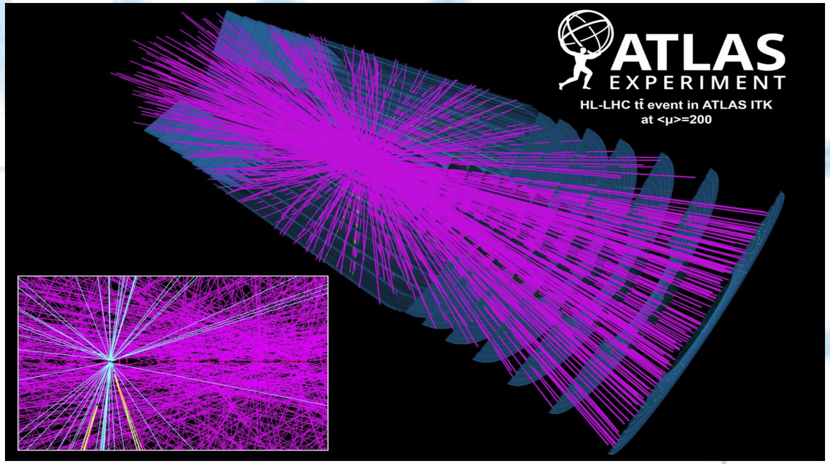
- Step 5: Track finding and fitting
  - Start from a seed
    - Kalman formalism (eg. Combinatorial KF) => Smoothing (eg. global χ^2 fit, or walk back with Kalman fit) => Ambiguity resolution



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

Problem(s)!



- No. of tracks per event for HL-LHC, expected to increase 2.5x
- Current R&D: Use GPUs for speedup via parallel computation
- However, sequential algorithms like CKF do much better on CPUs than GPUs!

acts → **traccc**

Demonstrator tracking chain for accelerators.

Computing term for specific purpose architectures (eg. GPU, TPU, IPU etc.)

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Track Reconstruction:

- Heterogeneous Track Reconstruction:
  - Step 1: Profiling CPU and GPU code to identify speed-up in inference time
    - Ideally without drop in tracking efficiency
  - Step 2: Identify bottlenecks
    - Points where one architecture outperforms the other
  - Step 3: Calculate data-transfer latencies at bottlenecks
    - Data transfer latencies between host (CPU) and device (GPU) eat up speed-up

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

The University of Manchester

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

CUDA Profiling

Step 1: Ensure GPU computations do not decrease physics performance

D2H Calculations - CPU vs D2H CPU

```
Running seeding algo with CUDA spacepoints moved to host
Number of D2H seeds: 28346 (host), 28346 (device)
  Matching rate(s):
    - 88.6898% at 0.01% uncertainty
    - 98.9275% at 0.1% uncertainty
    - 99.2592% at 1% uncertainty
    - 99.2839% at 5% uncertainty
Running track param est with CUDA spacepoints and CUDA seeds moved to host
Number of D2H track parameters: 28346 (host), 28346 (device)
  Matching rate(s):
    - 99.2944% at 0.01% uncertainty
    - 99.7989% at 0.1% uncertainty
    - 99.8024% at 1% uncertainty
    - 99.8024% at 5% uncertainty
```

D2H Calculations - CPU vs CUDA

```
Number of seeds: 28346 (host), 28346 (device)
  Matching rate(s):
    - 87.6208% at 0.01% uncertainty
    - 98.797% at 0.1% uncertainty
    - 99.2345% at 1% uncertainty
    - 99.2733% at 5% uncertainty
Number of track parameters: 28346 (host), 28346 (device)
  Matching rate(s):
    - 98.7617% at 0.01% uncertainty
    - 99.7989% at 0.1% uncertainty
    - 99.8024% at 1% uncertainty
    - 99.8024% at 5% uncertainty
```

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

CUDA Profiling - NSight Compute

Step 2: Profile code in as close-to-deployment conditions as possible

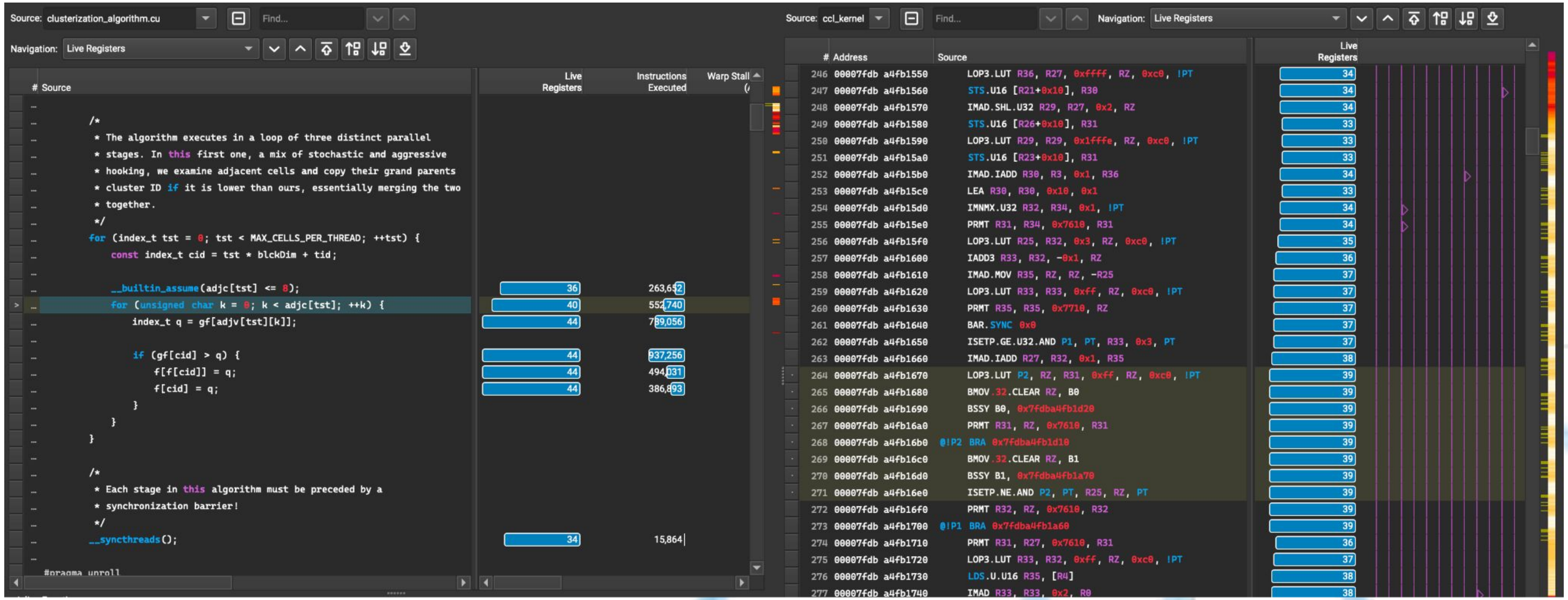
CUDA tools: NSight Compute, NSight Systems

ID	Issues Detected	Function Name	Demangled Name	Process	Device Narr	Grid Size	Block Size	Cycles [cycle]	Duration [msecond]	Compute Throughp	Memory Throughput [%]	# Registers [register/thread]
0	20	ccl_kernel	traccc::cuda::kernels::ccl_kernel(v...	[1138765] tracc...	Tesla T4	474, 1, 1	128, 1, 1	472435	0.81	43.75	43.72	49
1	14	form_spacepoints	traccc::cuda::kernels::form_space...	[1138765] tracc...	Tesla T4	137, 1, 1	1024, 1, 1	95281	0.16	5.95	36.43	36
2	11	count_grid_capacities	traccc::cuda::kernels::count_grid_...	[1138765] tracc...	Tesla T4	545, 1, 1	256, 1, 1	18692	0.04	22.31	50.49	20
3	14	populate_grid	traccc::cuda::kernels::populate_gri...	[1138765] tracc...	Tesla T4	545, 1, 1	256, 1, 1	35039	0.06	11.99	54.14	22
4	15	fill_prefix_sum	traccc::cuda::kernels::fill_prefix_su...	[1138765] tracc...	Tesla T4	18, 1, 1	32, 1, 1	23989	0.04	7.45	26.34	28
5	10	count_doublets	traccc::cuda::kernels::count_doubl...	[1138765] tracc...	Tesla T4	1131, 1, 1	64, 1, 1	848561	1.45	71.08	16.92	42
6	13	find_doublets	traccc::cuda::kernels::find_doublet...	[1138765] tracc...	Tesla T4	479, 1, 1	64, 1, 1	1211907	2.07	33.78	35.93	38
7	9	count_triplets	traccc::cuda::kernels::count_triplet...	[1138765] tracc...	Tesla T4	24496, 1, 1	64, 1, 1	1957486	3.33	68.04	23.78	45
8	15	reduce_triplet_counts	traccc::cuda::kernels::reduce_tripl...	[1138765] tracc...	Tesla T4	479, 1, 1	64, 1, 1	19298	0.03	2.52	11.65	16
9	15	find_triplets	traccc::cuda::kernels::find_triplets(...	[1138765] tracc...	Tesla T4	720, 1, 1	64, 1, 1	294522	0.51	29.96	27.57	45
10	18	update_triplet_weights	traccc::cuda::kernels::update_tripl...	[1138765] tracc...	Tesla T4	930, 1, 1	64, 1, 1	18140	0.04	13.81	48.89	23
11	19	select_seeds	traccc::cuda::kernels::select_seed...	[1138765] tracc...	Tesla T4	479, 1, 1	64, 1, 1	102384	0.18	26.16	26.16	39
12	14	estimate_track_params	traccc::cuda::kernels::estimate_tra...	[1138765] tracc...	Tesla T4	443, 1, 1	64, 1, 1	78502	0.14	5.62	44.97	52

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

CUDA Profiling - NSight Compute

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824
The University of Manchester

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

CUDA Profiling

Data File	Detector Geometry	No. of Events
tml_full/ttbar_mu300	tml_detector/trackml-detector	1

Old Version

CPU				GPU		
	Parent process	**Duration [mu-sec]**			**Parent process**	**Duration [mu-sec]**
	Container Instantiation	**8**			Container Instantiation	**6**
	File reading	**1,720,207**			File reading	NA
	Clusterization	**40,197**			Clusterization	NA
	Spacepoint Formation	**6,078**			Spacepoint Formation	NA
	Clusterization + Spacepoints	**46,275**			Clusterization + Spacepoints	**40,197**
	Seeding	**1,075,583**			Seeding	**80,072**
	Track param est	**27,909**			Track param est	**2,651**

New Version - with FastSV

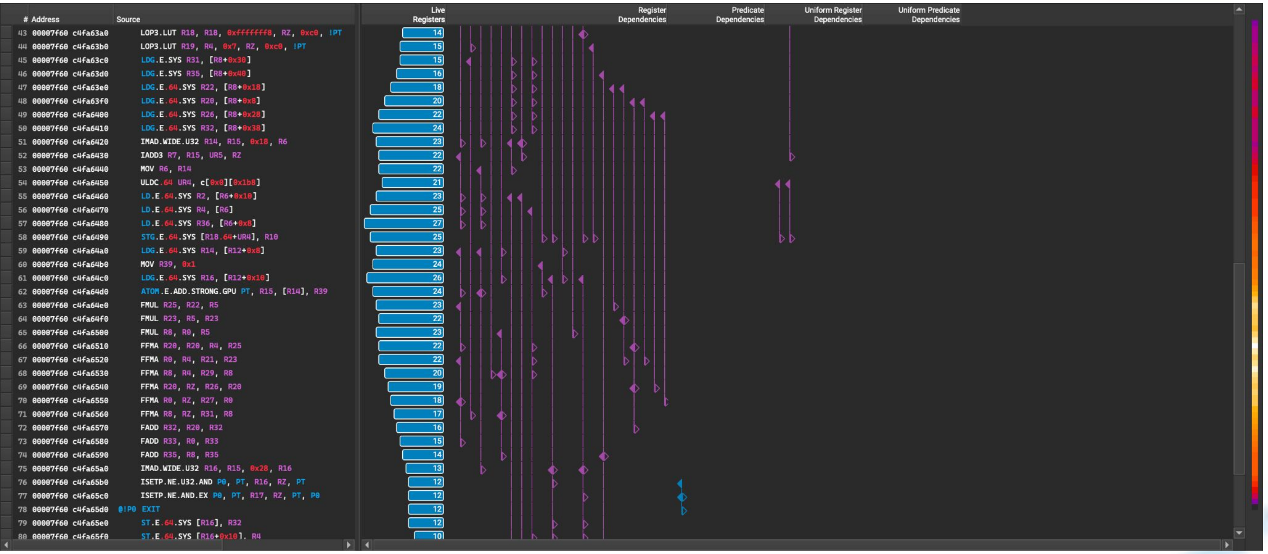
CPU				GPU		
	Parent process	**Duration [mu-sec]**			**Parent process**	**Duration [mu-sec]**
	Container Instantiation	**8**			Container Instantiation	**6**
	File reading	**1,288,893**			File reading	NA
	Clusterization	**36,070**			Clusterization	NA
	Spacepoint Formation	**3,131**			Spacepoint Formation	NA
	Clusterization + Spacepoints	**39,201**			Clusterization + Spacepoints	**2,433**
	Seeding	**792,729**			Seeding	**11,686**
	Track param est	**9,316**			Track param est	**411**

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

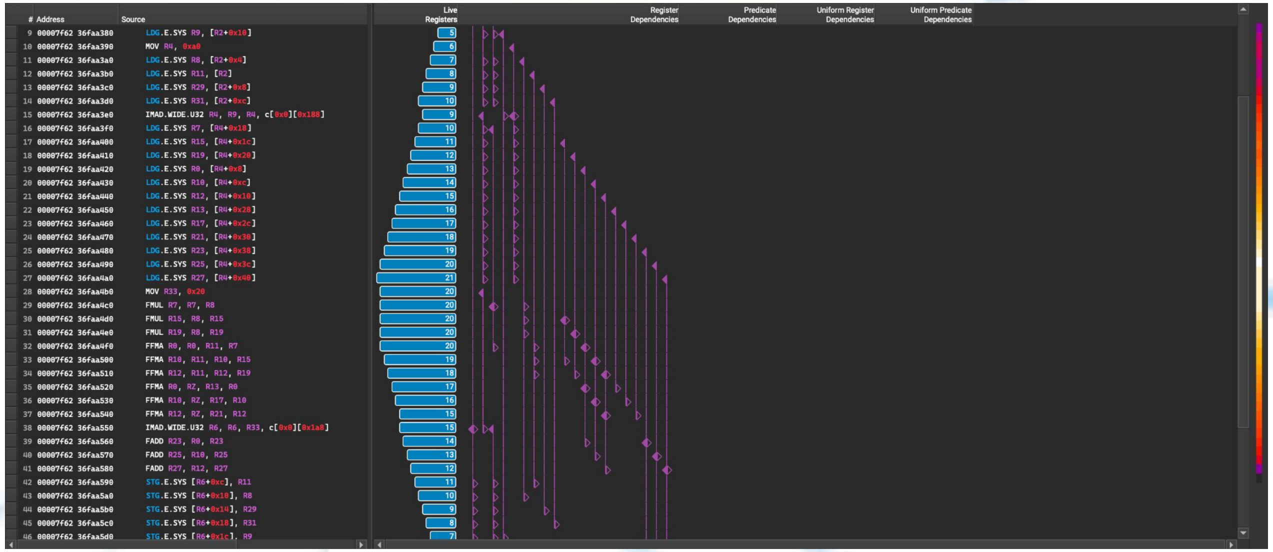
CUDA Profiling - NSight Compute

Data File	Detector Geometry	No. of Events
tml_full/ttbar_mu300	tml_detector/trackml-detector	1

Old Version

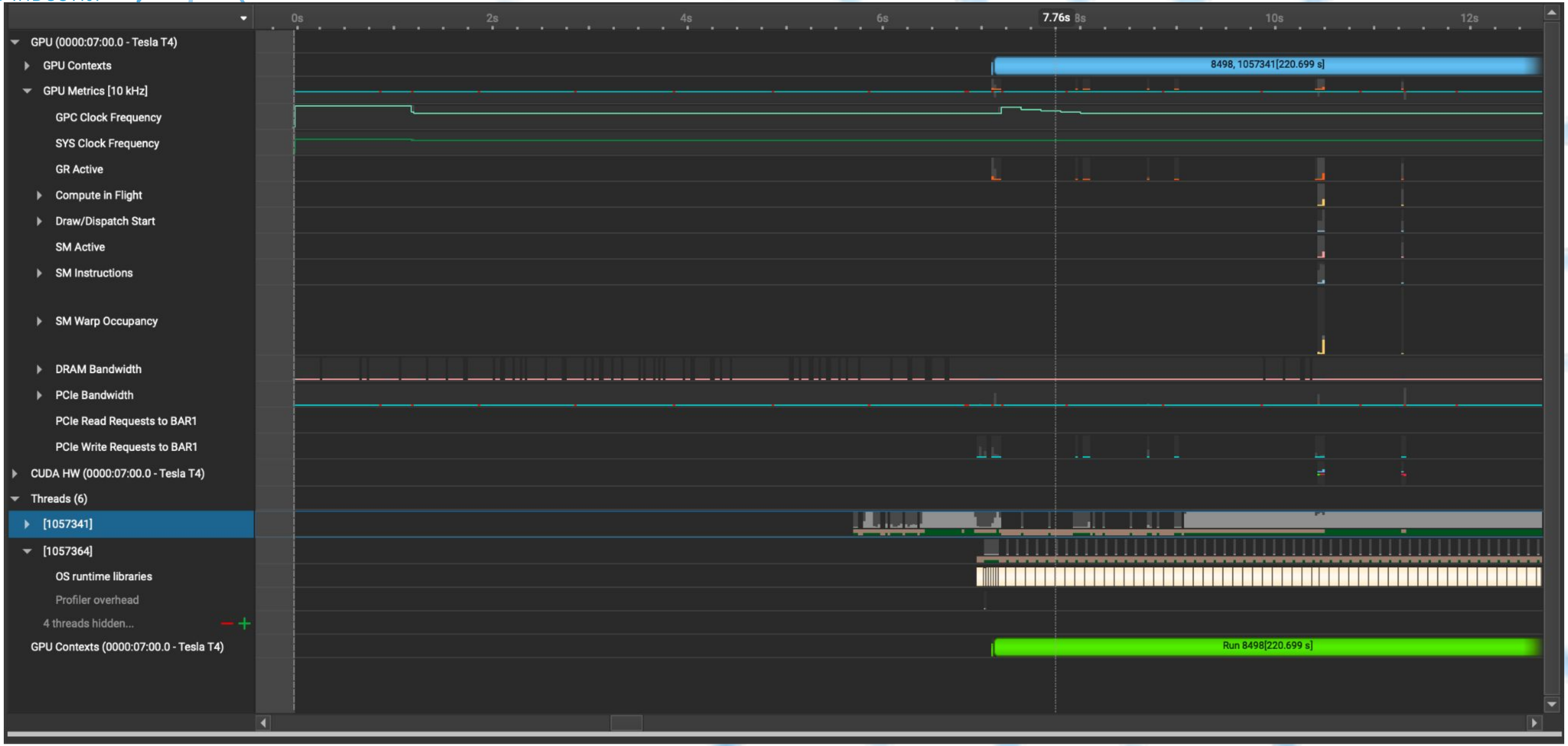


New Version - with FastSV



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

CUDA Profiling - NSight Systems



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

CUDA Profiling

- POC feasibility example:
  - Clusterization, Spacepoint formation, Seeding are significantly faster on GPU
  - Considering Host-Device and Device-Host wall-time overheads,
    - there is still a speedup of ~O(800 msec)
- *Note* This is only an example for one event

Data File	Detector Geometry	No. of Events
tml_full/ttbar_mu300	tml_detector/trackml-detector	1

CPU				GPU		
	Parent process	Duration [mu-sec]			Parent process	Duration [mu-sec]
	Container Instantiation	15			Container Instantiation	5
	File reading	1,341,172			File reading	NA
	Clusterization	36,070			Clusterization	NA
	Spacepoint Formation	3,131			Spacepoint Formation	NA
	Clusterization + Spacepoints	39,201			Clusterization + Spacepoints	2,433
	Seeding	814,789			Seeding	11,686
	Track param est	9,316			Track param est	411

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824
The University of Manchester

CUDA Profiling

- POC feasibility example:
  - Clusterization, Spacepoint formation, Seeding are significantly faster on GPU
  - Considering Host-Device and Device-Host wall-time overheads,
    - there is still a speedup of ~O(800 msec)
- If CKF on GPU is considerably slower, it may still be feasible to run the first part of the chain on Device and move data to Host for CKF
  - Potential solution for net speedup in the tracking chain
- *Note* This is only an example for one event

Data File	Detector Geometry	No. of Events
tml_full/ttbar_mu300	tml_detector/trackml-detector	1

CPU			GPU		
Parent process	**Duration [mu-sec]**			**Parent process**	**Duration [mu-sec]**
Host to Device [Cells]	**3035**			Device to Host [Cells]	NA
Host to Device [Spacepoints]	**2,703**			Device to Host [Spacepoints]	**1087**
Host to Device [Seeds]	**781**			Device to Host [Seeds]	**349**
Host to Device [Track params]	**1,655**			Device to Host [Track params]	**1,161**

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Next Steps

- Extend studies to larger number of events with more realistic data (eg. mu=60, 140, 200 etc)
- Use seeds generated by traccc in ACTS CKF function to understand Device-Host costs
- Repeat overhead studies wrt other parameters such as Memory, GPU Occupancy etc.

Data File	Detector Geometry	No. of Events
tml_full/ttbar_mu300	tml_detector/trackml-detector	1

CPU			**GPU**		
Parent process	**Duration [mu-sec]**			**Parent process**	**Duration [mu-sec]**
Host to Device [Cells]	**3035**			Device to Host [Cells]	NA
Host to Device [Spacepoints]	**2,703**			Device to Host [Spacepoints]	**1087**
Host to Device [Seeds]	**781**			Device to Host [Seeds]	**349**
Host to Device [Track params]	**1,655**			Device to Host [Track params]	**1,161**

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

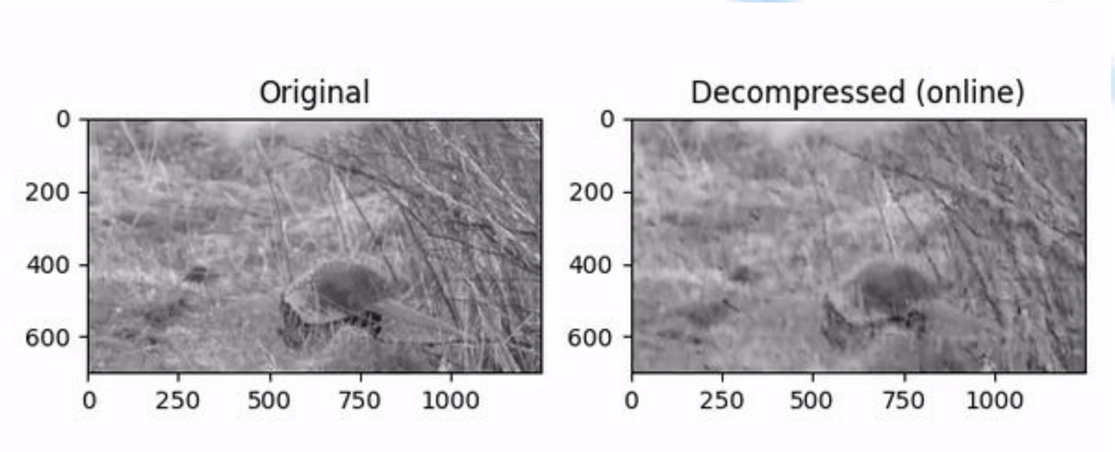
SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

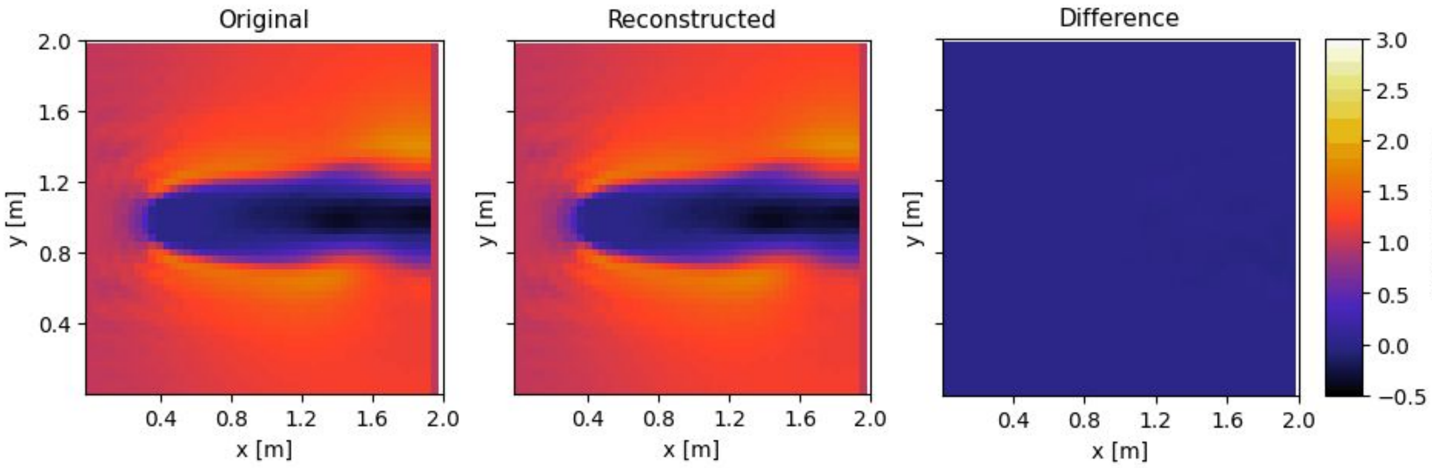
Baler: ML based Data Compression

SMARTHEP is funded by the European Union’s Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824
The University of Manchester

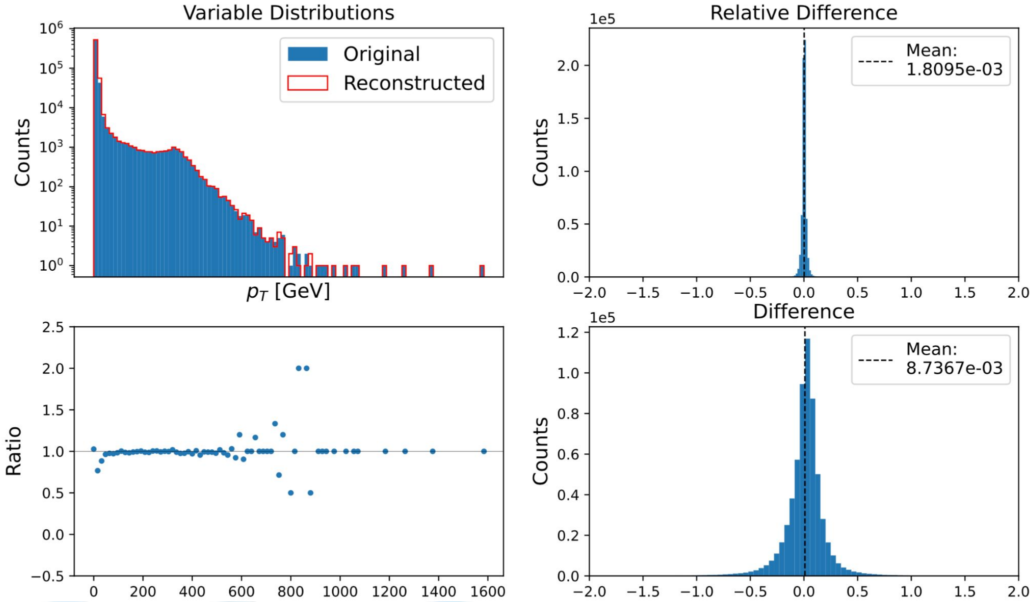
Baler: ML based Data Compression





Paper: https://arxiv.org/abs/2305.02283

Code: https://github.com/baler-collaboration/baler



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Misc Activities

Side Quests:

- Helped lift new wire bonder into the Lund Uni clean room
- Working on DDMTD corrections for mu-CTP trigger chip for temperature variance with Kalman Filters

Hackathons:

- DL in HEP - MIT
- Baler (Lund)

Workshops/Conferences:

- IOP APP&HEPP UK National - Talk
- FastML - Talk
- Hammers and Nails - Poster

Schools:

- tCSC Split
- HASCO particle physics
- STFC UK theory summer school

Teaching/Outreach:

- Jupyter Notebooks Intro - Lund
- Advanced C++ - Manchester

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Conclusions & Future Work

- Track reconstruction may not scale to HL-LHC
  - R&D required to speed up track reconstruction
    - without loss in tracking efficiency
    - GPUs could be a viable option
      - GPUs may not be able to speed up the full chain due to sequential algo (eg. CKF)
  - Heterogeneous computing solutions may be feasible
    - POC studies show:
      - First half of the chain on GPUs, followed by offloading to CPUs for CKF, could show order of 100msec speedup per event
- Work on novel anomaly detection solutions?
- Contribute to the HLS4ML project?
- Analysis: Semi-visible jet signatures for dark sector searches - ML based signal-background discrimination?
- TBD!

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Thank you!

Questions?

People are hungry, just putting it out there :)

Feel free to chat later as well! :)

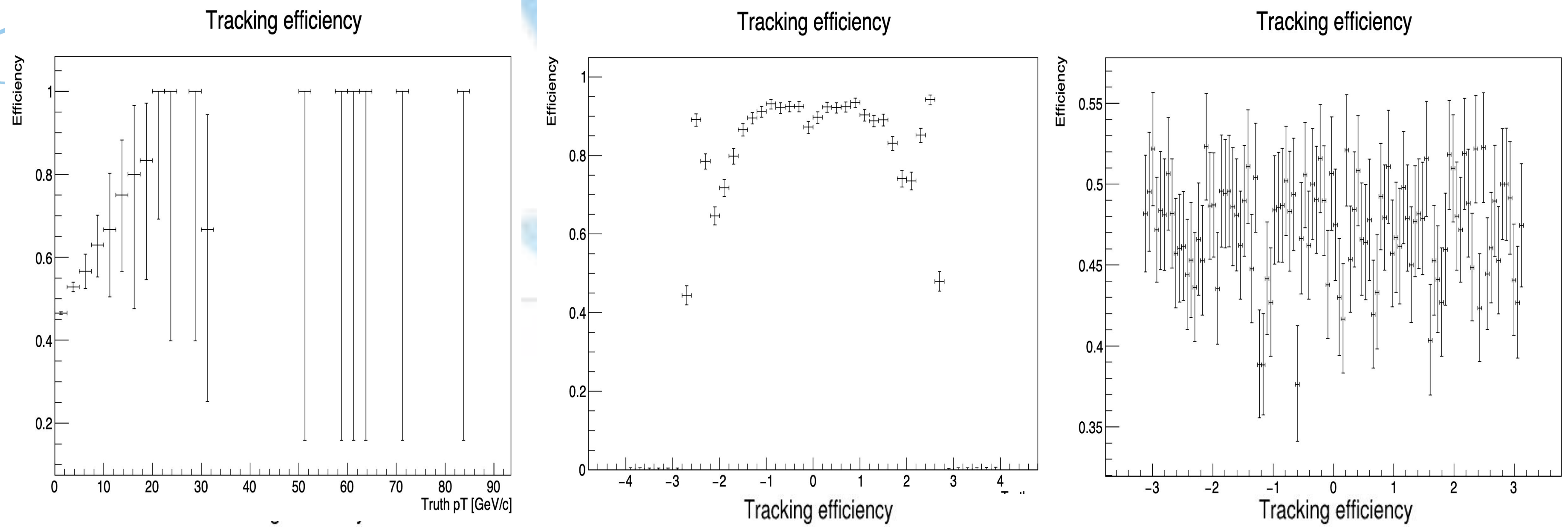
SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

BACKUP

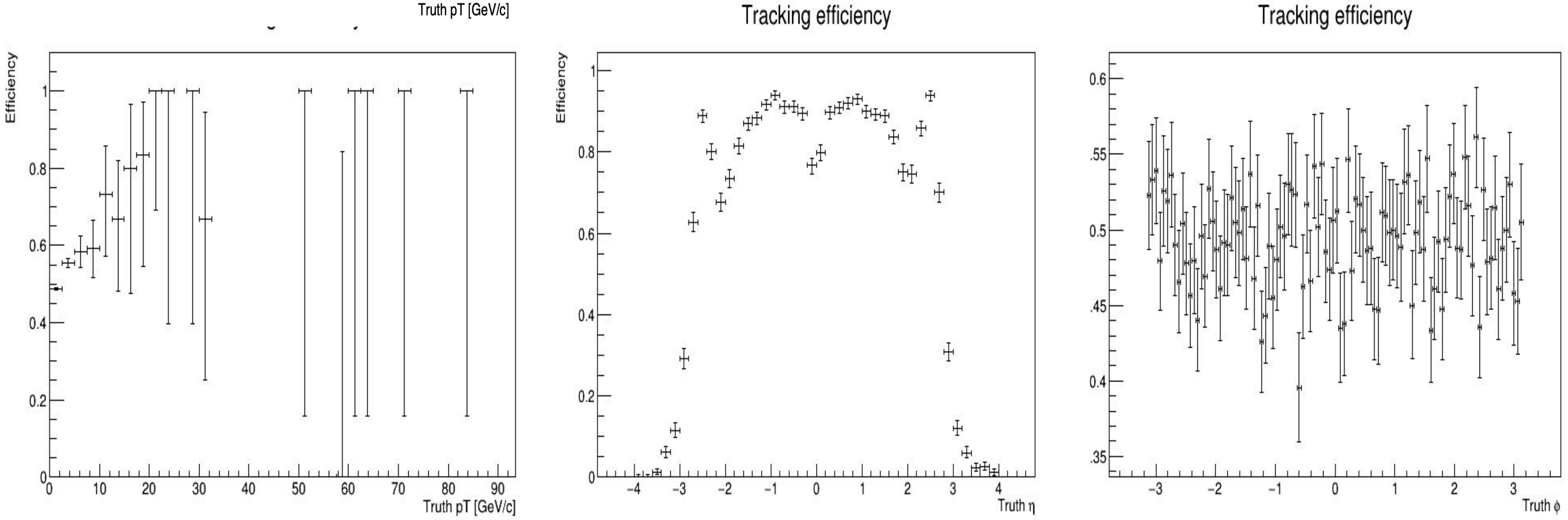
SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

CPU



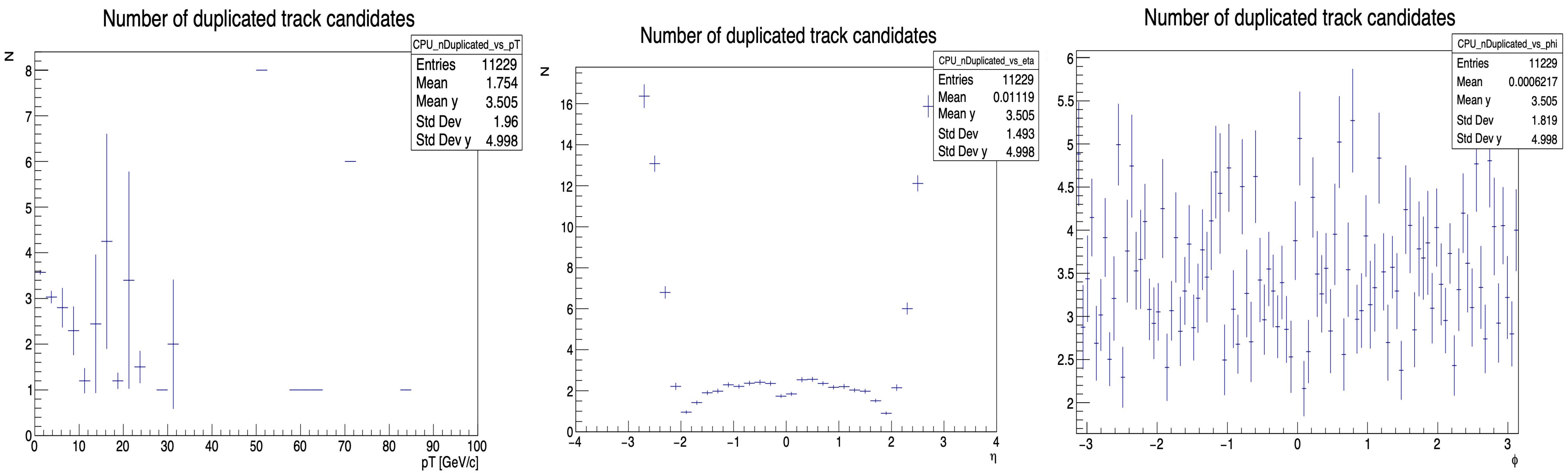
CUDA



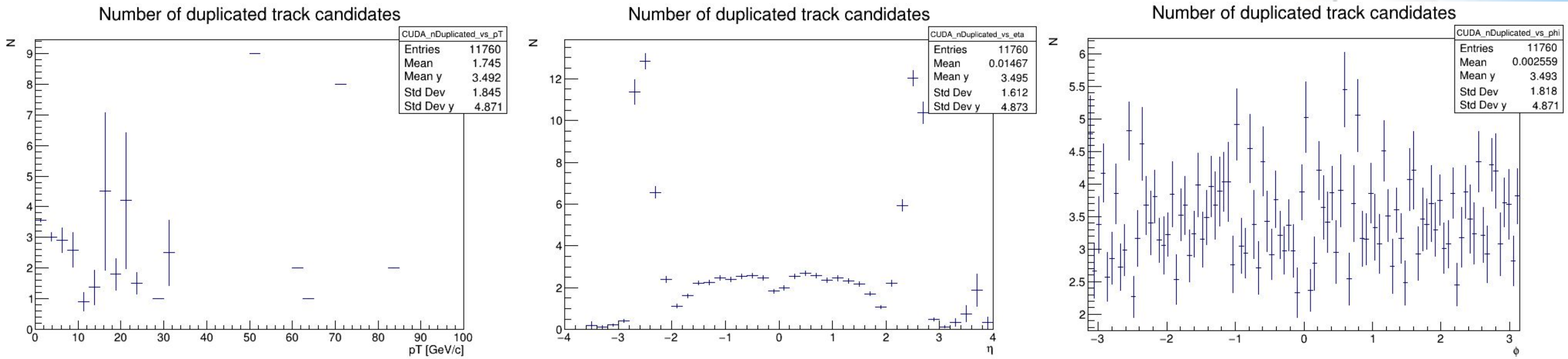
SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824
The University of Manchester

CPU



CUDA



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

CUDA Profiling

Old Version

- Seeding - Synchronous kernel launches
  - fill_prefix_sum used for synchroniszation

ID	Issues Detected	Function Name	Demangled Name
0	2	find_clusters	traccc::cuda::kernels::find_...
1	2	fill_prefix_sum	traccc::cuda::kernels::fill_p...
2	2	count_cluster_c...	traccc::cuda::kernels::coun...
3	2	connect_compo...	traccc::cuda::kernels::conn...
4	3	create_measure...	traccc::cuda::kernels::creat...
5	2	fill_prefix_sum	traccc::cuda::kernels::fill_p...
6	2	form_spacepoints	traccc::cuda::kernels::form...
7	2	fill_prefix_sum	traccc::cuda::kernels::fill_p...
8	2	count_grid_capa...	traccc::cuda::kernels::coun...
9	2	populate_grid	traccc::cuda::kernels::popu...
10	3	fill_prefix_sum	traccc::cuda::kernels::fill_p...
11	2	count_doublets	traccc::cuda::kernels::coun...
12	3	fill_prefix_sum	traccc::cuda::kernels::fill_p...
13	2	find_doublets	traccc::cuda::kernels::find_...
14	3	fill_prefix_sum	traccc::cuda::kernels::fill_p...
15	2	count_triplets	traccc::cuda::kernels::coun...
16	3	fill_prefix_sum	traccc::cuda::kernels::fill_p...
17	3	find_triplets	traccc::cuda::kernels::find_...
18	3	fill_prefix_sum	traccc::cuda::kernels::fill_p...
19	3	update_triplet_w...	traccc::cuda::kernels::upda...
20	2	select_seeds	traccc::cuda::kernels::sele...
21	2	estimate_track_...	traccc::cuda::kernels::esti...

New Version - with FastSV

- Seeding - Asynchronous kernel launches
  - Increases paralellization

ID	Issues Detected	Function Name	Demangled Name
0	3	ccl_kernel	traccc::cuda::kernels::ccl_kernel(v...
1	3	form_spacepoints	traccc::cuda::kernels::form_space...
2	2	count_grid_capacities	traccc::cuda::kernels::count_grid_...
3	2	populate_grid	traccc::cuda::kernels::populate_gri...
4	3	fill_prefix_sum	traccc::cuda::kernels::fill_prefix_su...
5	2	count_doublets	traccc::cuda::kernels::count_doubl...
6	2	find_doublets	traccc::cuda::kernels::find_doublet...
7	2	count_triplets	traccc::cuda::kernels::count_triplet...
8	2	reduce_triplet_counts	traccc::cuda::kernels::reduce_tripl...
9	3	find_triplets	traccc::cuda::kernels::find_triplets(...
10	3	update_triplet_weights	traccc::cuda::kernels::update_tripl...
11	3	select_seeds	traccc::cuda::kernels::select_seed...
12	2	estimate_track_params	traccc::cuda::kernels::estimate_tra...

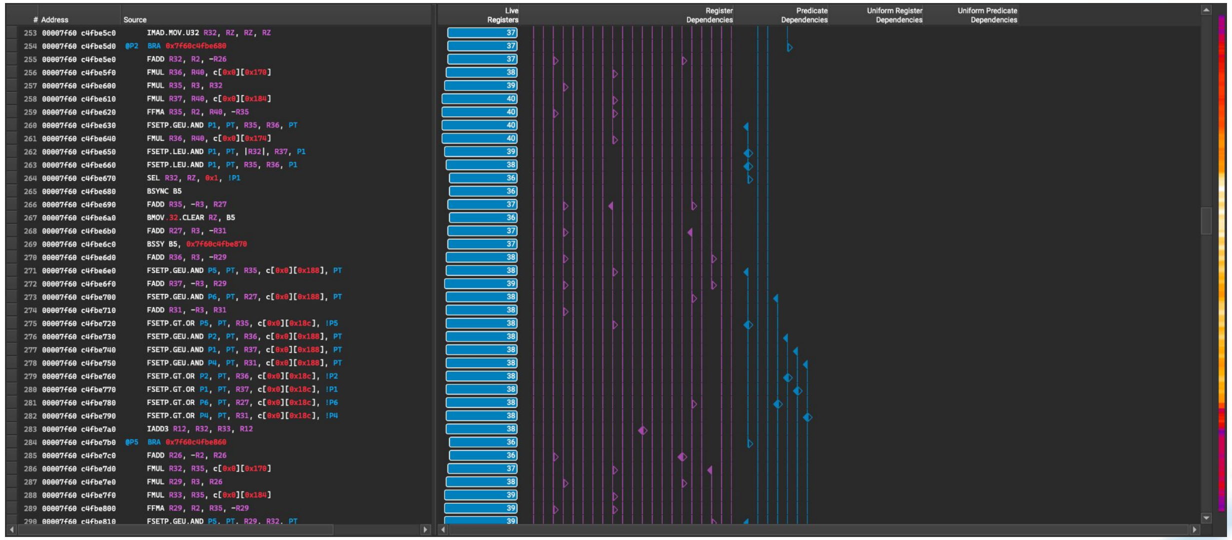
SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

MANCHESTER 1824
The University of Manchester

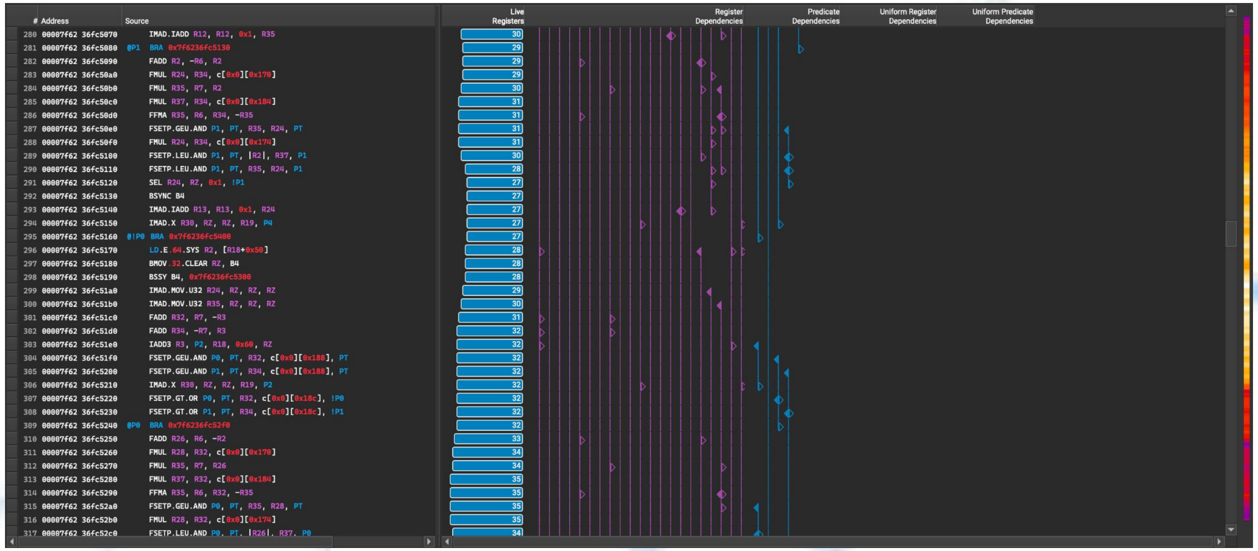
CUDA Profiling- NEW

Data File	Detector Geometry	No. of Events
tml_full/ttbar_mu300	tml_detector/trackml-detector	1

Old Version
count_doublets()



New Version - with FastSV
count_doublets()



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086