



# Accelerating HEP Software

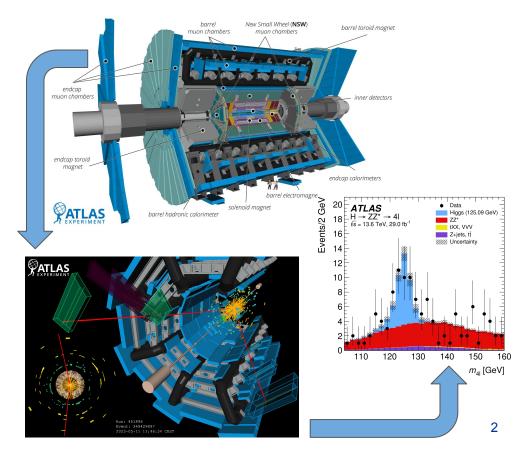
#### Attila Krasznahorkay trying to sample / expose the work of a lot of people...



#### **HEP Software Basics**

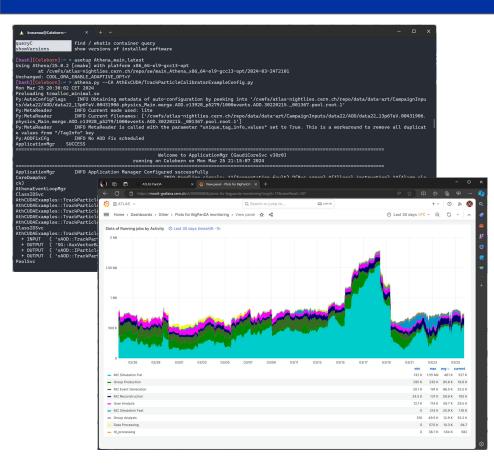


- "Software" is used practically everywhere in HEP
  - But here I'll only focus on software used in "data processing". Software used for controlling the accelerator, detector hardware, etc. is a different beast entirely.
- The goal is to "do physics"
  - Fast and nicely written software is the means, not the actual goal of the endeavour.



### Software ↔ Computing





- Even in data processing, we have 2 distinct areas
  - "Software" covers all the software written to perform the data transformations needed "for physics".
  - "Computing" covers all the software created to allow the first category to execute on up to millions of CPU cores in parallel.

#### Here I will only cover "software"

 Even though using accelerators in job/network optimization, and even just making batch systems "accelerator aware" is a very important thing on our plates.

#### **Development, Build and Deployment**

athenu

h Analyz

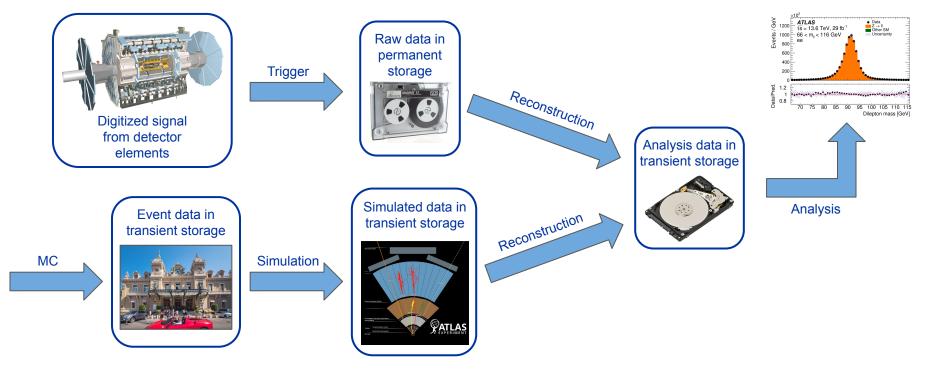
- Interpreting / processing the data
   coming from the LHC detectors takes
   many millions of C++ / Python code
  - We use very customized build systems to allow developers to deal with a small portion of the software at any given time, testing its integration with the rest of the software.
  - This requires highly curated, many GB development environments to be made available around the world.
- The distribution uses <u>CVMFS</u>
  - Which needs to be able to house the full SDKs (making licensing a challenge...)

							• × •			
https://gitlab.cem							··· 🥠			
						Sign in				
h or go to	🕄 atlas / 🕄 atbena						•			
	🔝 athena		\$	Star 178	a l					
>	->- 127,815 Commits				0					
	The ATLAS Experiment				•					
,	DOI 10.5281/zerodo.26419	2002 10.5251/mmodo.2013997 Donyon main					-			
, ,	Merge branch Edward Moyse			3a84c695 🛱	+					
	main ~ athena	nain v athena								
	🖹 README 🔤 中 LK	README 했 LICENSE								
	Name		Last commit			Last update				
	P: .devcontainer		Fix for .gitconfig in dev containers			9 months ago				
	P: .vscode	VS Code: Update the default Python settings to t				3 months ago				
	P AsgExternal/Asg_T		Update DAOD_PHYSUTE test inputs			2 weeks ago				
		AtlasGeometryCommon Merge remote-tracking branch 'upstream/24.0' i								- 0
	🖹 AtlasTest									
	🗅 Build		https://bigpanda.cem.ch/globalview/							
	En Calorimeter		Dash Tasks Jobs Errors Users Siles H	arvester - My BigPanDA			Job by ID + Enter	P P		- Help - 🛓 Attila
	E Commission	Monitoring for ASCIG or	n BigPanDA				Job by ID + Enter	ρ Pi		<ul> <li>Holp * 🛓 Attila</li> <li>alpanda191 (Rohosh</li> </ul>
	E Commission	Monitoring for ABCIG or ATLAS N	ightlies and CI Global Page				Job by ID · Enter	P Pi		
	E Commission	Monitoring for ASCIG or ATLAS Ni - Nightly and Cl e - For ART GRID 1 - For ART LOCAL	n BigPanDA	e, succeeded, finished (with			s failed), failed are st	own. Follow	result links for c	apanda191 (Refeah letails
	E Commission	Monitoring for ASCIS or ATLAS Ni • Nightly and Cl s • For ART GRID 1 • For ART LOCAL documentation	nBigNarDA ightlies and CI Global Page ystems features and policies test results, the number of tests with statuses active test results, the number of tests with statuses suc	e, succeeded, finished (with			s failed), failed are st	own. Follow	result links for c	apanda191 (Refeah letails
	E Commission	Monitoring for ASCIS or ATLAS Ni • Nightly and Cl s • For ART GRID 1 • For ART LOCAL documentation	n Burhanch ighthies and CI Global Page systems features and policies test results, the number of tests with statuses active test results, the number of tests with statuses active test results, the number of tests with statuses active (tuilt by the tool developed by Oleg Zenin)	e, succeeded, finished (with			s failed), failed are st	own. Follow Follow resu	result links for c	apanda191 (Refeah letails
	E Commission	Monitoring for ASCIG or ATLAS Ni • Nightly and Cl e • For ART GRID • For ART LOCAL documentation Show 100 • e	In BigHeadA	succeeded, finished (with Athe     eeded, finished (with Athe	Build time	Compilation errors	s failed), failed are sh d), failed are shown. CTest (or CI) bast errors	own. Follow Follow resu Search:	result links for c Its links to acces	apanda191 (Polycol) letails s ART local
	E Commission	Morehoming for ASCIG or ATLAS Ni • Nightly and Cli a • For ART GRID • For ART LOCAL documentation Show 100 • • •	Inguinuta ightfiles and CI Global Page ystems factors and polices to reach, for mother of tasks with statuses activ Lear reach, the number of tasks with statuses activ Lear reach, the number of tasks with statuses active Lear reach, the number of tasks with statuses active Lear reach, the number of tasks with statuses active terms	succeeded, finished (with Alhe     Recent     Rotesse     MR-70102-2024-03-25-19	Build time 25-MAR 17:56 23-MAR 05:52	Compilation errors (w/warnings)	s failed), failed are shown. d), failed are shown. CTest (or CI) test errors (wiwarnings)	own. Follow Follow resu Search:	e result links for c Its links to acces	agandat91 (School Intalia & ART local CVMPS (on clience) NIA 23-MAR 09:51
	E Commission	Mentaving for ASCIG of ATLAS Ni • Nightly and Cl a • For ART LOCAL documentation • Show 100 • • • Nightly Group 1 Cl	Inspirate ightfiles and CI Global Page ystems barbare and protes to strast, her number of tasks with statuses and Lear reach, the number of tasks with statuses and Lear reach, the number of tasks with statuses are used to be a statused of the statuses of the statuses of the statuses of the status of the status of the status of the status of the status of the status of the status of the status of	9 successfed, finished (with Athe secessfed, finished (with Athe Recent Recent MR-70102.2024.03-25-19 11	Build time         State           25-MAR         25-MAR           17:56         23-MAR           05:52         24-MAR           23:30         23-30	Compilation errors (wiwarnings) 0 (0)	s failed), failed are sh d), failed are shown.	own. Follow Follow resu Search: ART LOCAL NXA	a result links for c its links to access a ART GRID N(A	etalis s ART local
	E Commission	Menning for ASCIG of ATLAS Ni - Nightly and CI - For ART GRID - For ART LOCAL documentation - Show 100 * e Nightly droup ( CI PRODUCTION	Inguinda ightfiles and CI Global Page systems barbers and prices to reach, for more for task mit shakes activ to reach the number of tasks with attaces are the reaches number of tasks with attaces are used to the system of tasks with attaces are the reaches and the system of tasks with attaces are the reaches and the system of tasks with attaces are the reaches and the system of tasks with attaces are the reaches and the system of tasks with a task of the system of tasks with a task of the system of tasks with a task of the system of tasks of the system of tasks of tasks of the system of tasks of the system of tasks	B	Build time         25-MAR           23-MAR         05:52           24-MAR         23:30           24-MAR         23:30	Compilation errors (wiwarnings) 0 (0) 0 (0)	s failed), failed are sh dy, failed are shown.	own. Follow Follow resu Search: ART LOCAL N/A N/A	result links for costs Its links to access ART GRID NOA NOA	etails s ART local CVMPS (on dirent) NIA 23-MAR 01:11 25-MAR 01:11 25-MAR 22-MAR 23-21
	E Commission	Mentioning for ASCIG of ATLAS NI - Nightly and CI = - For ART ORIO - For ART ORIO - For ART I LOCAL documentation Show 100 * e Nightly droup () 	Ingeneral gottiles and CI Global Page gytems bases and polices to reach, for more for task with statuses and to reach the number of tasks with statuses and the reaches the number of tasks with statuses and the number of tasks with statuses and tasks with statuses and tasks with statuses and tasks with status and tasks with statuses and tasks with status and tasks with statuses and tasks with status and tasks with status and tasks with status an	Recent         Recent           11         2024-02-25-19           2024-02-2370301         2024-02-2370301	Build time         P           2         25-MAR           2         25-MAR           2         25-MAR           2         25-MAR           2         23-MAR           2         22-MAR           2         22-MAR           21:04         21-MAR	Compilation errors (wiwarnings) 0 (0) 0 (0)	s failed), failed are sh dd), failed are shown. Crest (or Ch) set errors (wwwmings) 0 (0) 13 (3) 0 (0)	own. Follow Follow resu Search: LOCAL N/A N/A N/A	ART GRID NVA 0,198.93.9	etals s ART local
	E Commission	Nontransp to AdOD of A	Inspirated ightflies and CI Global Page Systems backes and polices to strank, for muther of tasks that as active the strank, for muther of tasks that as active the strank of the strange of tasks that as active interesting the strange of tasks that as active interesting the strange of tasks that as a strange of tasks MIC studies 24.6 "Advanz, 46.6 staff gas15.deg 24.6 "Advanz, 46.6 staff gas15.deg	Research	Build time         Image: Constraint of the second sec	Compilation errors (er/warnings) 0 (0) 0 (0) 0 (0) 0 (0)	s failed); failed are shown. (c), failed are shown.	own. Follow Follow resu Search: ART LOCAL N/A N/A N/A N/A 3.0.0	RRT GND         ART GND           NXA         NXA           0,198.93.9         0,35.0.1	etals s ART local
	E Commission	Noncomposition of the second s	Inspired	Research         Research           Research         Research           Marca         Research           V         Re	Build time         Build time           2         24-MAR           05-52         22-MAR           22-30         22-MAR           22-40         22-MAR           22-40         22-MAR           21-54         21-54           22-58         22-58	Compliation errors           0 (0)           0 (0)           0 (0)           0 (0)           0 (0)	s falled), failed are stown.           (d), failed	own. Follow Follow resu Search: ART LOCAL N/A N/A N/A N/A	ART ORID ART ORID NIA NIA NIA 0,198,93.9 0,35.0,1 NIA	etails s ART local CVMPS s ART local CVMPS (n client) (n client) (
	E Commission	Norman Service Services Servic	Instruments ightities and Cl Global Page Systems backes and polices ter search, for a more of thesis with statuses and ter search, for a more of thesis with statuses and ter search and the search of the search and the search ter search and the search of the search and the search ter search and the search of the search and the search ter search and the search of the search and the search ter search and the search of the search and the search ter search and the search of the search of the search of the search ter search of the search	Bit         Reserve           a         Reserve           Reserve         Reserve           a         Reserve           b         Reserve           a         Reserve           b         Reserve           a         Reserve           b         Reserve           a         Reserve           b         Reserve           a         Reserve           a         Reserve           a         Reserve           b         Reserve           b         Reserve           b         Reserve           b         Reserve           b         Reserve           b         Reserve           c         Reserve           a         Reserve <td< td=""><td>Build time         Build time           25-MAR         17:56           23-MAR         25-MAR           23-MAR         25-MAR           23-MAR         21-MAR           23-MAR         21-MAR           21-MAR         21-MAR           21-MAR         21-MAR           21-MAR         21-MAR           21-MAR         21-MAR           22-MAR         21-MAR</td><td>Compilation errors (www.errinings) () () () () () () () () () () () () ()</td><td>s failed, tailed are shown (d), tailed are shown (tailed are shown (tearrora)</td><td>own. Follow resu Follow resu search: Iocal Iocal NiA NiA NiA</td><td>ART         GRID           0.198.93.9         0.198.93.9           0.35.0.1         N/A</td><td>apards191         Polessie           etais         aRT local           CvMPS         for diend           N/A         23-MAR           225-MAR         01:11           22-242         23-MAR           01:11         22-24           24-102         24-102           24-102         25-MAR           25-11         25-MAR           25-12         25-MAR           25-14         25-MAR</td></td<>	Build time         Build time           25-MAR         17:56           23-MAR         25-MAR           23-MAR         25-MAR           23-MAR         21-MAR           23-MAR         21-MAR           21-MAR         21-MAR           21-MAR         21-MAR           21-MAR         21-MAR           21-MAR         21-MAR           22-MAR         21-MAR	Compilation errors (www.errinings) () () () () () () () () () () () () ()	s failed, tailed are shown (d), tailed are shown (tailed are shown (tearrora)	own. Follow resu Follow resu search: Iocal Iocal NiA NiA NiA	ART         GRID           0.198.93.9         0.198.93.9           0.35.0.1         N/A	apards191         Polessie           etais         aRT local           CvMPS         for diend           N/A         23-MAR           225-MAR         01:11           22-242         23-MAR           01:11         22-24           24-102         24-102           24-102         25-MAR           25-11         25-MAR           25-12         25-MAR           25-14         25-MAR



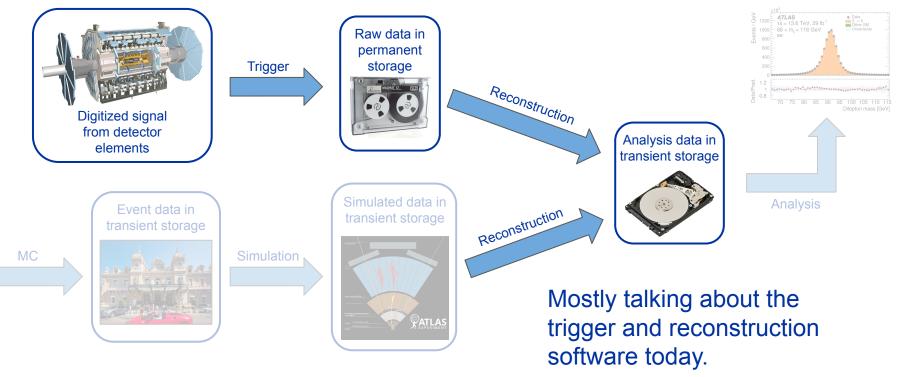
#### HEP Data Processing (1)





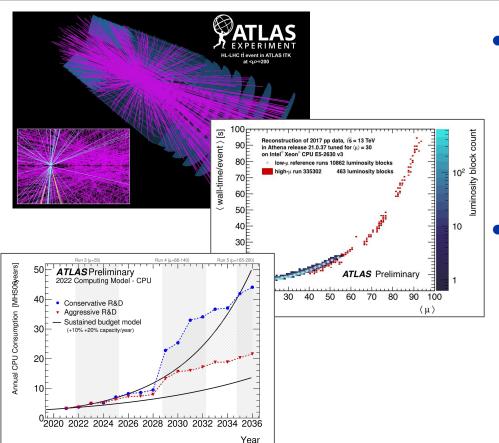
#### HEP Data Processing (2)





### **Trigger / Reconstruction**





- The goal of this software is to figure out what particles, with what exact properties, left signals in the detector
  - A little similar to "finding the cats" in a CCD camera's recorded data 5
  - In trillions of >100 megapixel, "3D camera" pictures
- "Classical" algorithms do this very well in the ongoing LHC data taking. (Though ATLAS is the only one **only** using classical algorithms at this point.)
  - However during the High-Luminosity LHC era these current algorithms would need too much CPU power.

# Things To Do Right: Event Data Model

#ifndef DataFormats PortableTestObjects interface TestSoA b



// Global "average" of something (non-const)

VECMEM HOST AND DEVICE

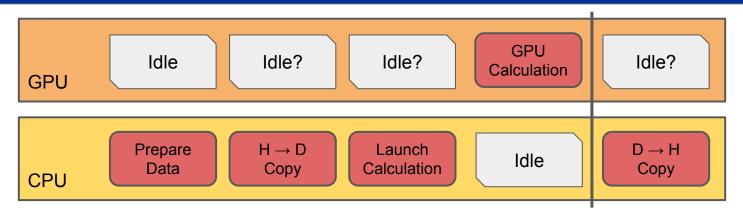
#### • Our software uses "rich" data

- Many, well defined properties, in very jagged arrays
- Our 20+ year old existing software tends to manage these in ways not suited for accelerators (AoS vs. SoA)
- We need to move this data with a high frequency between H↔D
- Must be manageable to implement/maintain the code that describes all our data types
  - Note that a single "type" usually requires multiple C++ classes to describe it

	#include <array></array>	47	auto& average() { return BASE::template get<3>(); }
5	wincibue von ay	48	<pre>/// Global "average" of something (const)</pre>
6	#include <eigen core=""></eigen>	49	VECMEM_HOST_AND_DEVICE
7	#include <eigen dense=""></eigen>	50	<pre>const auto&amp; average() const { return BASE::template get&lt;3&gt;(); }</pre>
8		51	
9	<pre>#include "DataFormats/Common/interface/StdArray.h"</pre>	52	/// "Indices" of something (non-const)
10	<pre>#include "DataFormats/SoATemplate/interface/SoACommon.h"</pre>	53	VECMEM_HOST_AND_DEVICE
11	<pre>#include "DataFormats/SoATemplate/interface/SoALayout.h"</pre>	54	<pre>auto&amp; indices() { return BASE::template get&lt;4&gt;(); }</pre>
12	<pre>#include "DataFormats/SoATemplate/interface/SoAView.h"</pre>	55	/// "Indices" of something (const)
13		56	VECMEM HOST AND DEVICE
14	namespace portabletest {	57	<pre>const auto&amp; indices() const { return BASE::template get&lt;4&gt;(); }</pre>
15		58	
16	<pre>// the type aliases are needed because commas confuse macros</pre>	59	/// "Index" of something (non-const)
17	using Array = edm::StdArray <short, 4="">;</short,>	60	VECMEM HOST AND DEVICE
18	<pre>//using Array = std::array<short, 4="">;</short,></pre>	61	<pre>auto&amp; index() { return BASE::template get&lt;5&gt;(); }</pre>
19	using Matrix - Eigen::Matrix <double, 3,="" 6="">;</double,>	62	/// "Index" of something (const)
20		63	VECMEM HOST AND DEVICE
21	// SoA layout with x, y, z, id fields	64	<pre>const auto&amp; index() const { return BASE::template get&lt;5&gt;(); }</pre>
22	GENERATE_SOA_LAYOUT(TestSoALayout,	65	const dated intenty const ( recard order temptate Berts ()) }
23	<pre>// columns: one value per element</pre>	66	}; // class jagged interface
24	SOA_COLUMN(double, x),	67	J, // Class Jaggeu_incernace
25	SOA_COLUMN(double, y), SOA COLUMN(double, z),	68	/// "Jagged" container for the tests
26	SOA_COLUMM(double, 2), SOA COLUMM(int32 t, id),	69	using jagged soa container =
-20	// scalars: one value for the whole structure	70	edm::container <jagged_soa_interface, edm::type::scalar<int="">,</jagged_soa_interface,>
	SOA_SCALAR(double, r),		0.00 = =
30	// column of arrays: one fixed-size array per element	71	edm::type::vector <float>, edm::type::jagged_vector<double>,</double></float>
31	SOA COLUMN(Array, flags),	72	<pre>edm::type::scalar<float>, edm::type::jagged_vector<int>,</int></float></pre>
32	<pre>// Eigen columns: each matrix element is stored in a separate column</pre>	73	<pre>edm::type::vector<int> &gt;;</int></pre>
33	SOA EIGEN COLUMN(Matrix, m))	74	
34		75	} // namespace testing
35	using TestSoA = TestSoALayout<>;	76	} // namespace vecmem
		_	



# Things To Do Right: Task Scheduling



- Not all aspects of our data processing are suited for accelerators
  - Any calculation not using an accelerator must be using the available CPU cores efficiently in parallel with whatever the GPU is doing, maximising the usage of all components
- A number of different solutions are used by the experiments currently
  - Built around <u>TBB</u> (for ATLAS and CMS) and <u>ZeroMQ</u> (for ALICE and LHCb)
- But new, hopefully better solutions seem to be on the horizon
  - Using <u>Boost::fiber</u> and/or <u>C++ coroutines</u>

## Things To Do Right: Algorithms



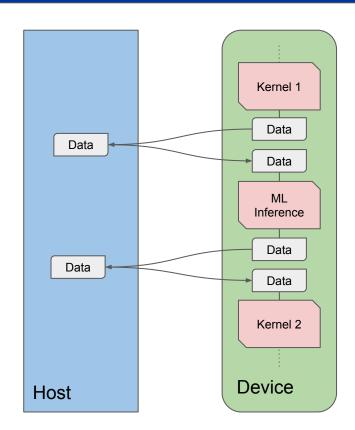
#### namespace traccc::device {

	/// Implementation of a FastSV algorithm with the following s	steps:
	/// 1) mix of stochastic and aggressive hooking	
	/// 2) shortcutting	
	/// The implementation corresponds to an adapted versiion of	Algorithm 3 of
	/// the following paper:	
	/// https://www.csioncodinect.com/ssionce/onticle/pii/507433	39global
		40 // _launch_bounds_(256,4) 41 void
	/// This array only gets updated at the	
		42 v finaclus(uinti6_t const*restrictid, // module id of each pixel 43 uinti6_t const*restrict x, // local coordinates of each pixel
		44 uinti6_t constrestricty, // local conditates of each pixel
	With Sharon [Tul] and so the number of and accure sector	44 units_c const _restrict_y, // index of the first pixel of each module 45 uint32 t const* restrict moduleStart, // index of the first pixel of each modul
	/// uparamiting augv vector of augacent cerrs	46 uint32 t* _ restrict _ nClustersInModule, // output: number of clusters found in each
	/// wparam[in] tid The thread index	47 uint32_t* _restrict _ moduleId, // output: module id of each module
	///@param[in]blckDim The block size	48 int32_t*restrict clusterId, // output: cluster id of each pixel
		49 int numElements) {
	/// iteration.	<pre>50 if (blockIdx.x &gt;= moduleStart[0])</pre>
	/// @param[inout] gf array holding grandparent cell ID fr	
	/// iteration.	
	/// @param[in] barrier A generic object for block-wide sync	53 auto firstPixel = moduleStart[1 + blockIdx.x];
		54 auto thisModuleId = id[firstPixel];
	cemprate Cypename Darrier_C2	<pre>55 assert(thisModuleId &lt; MaxNumModules);</pre>
	<pre>/ TRACCC_DEVICE void fast_sv_1(index_t* f, index_t* gf,</pre>	
	unsigned char adjc[MAX CELLS PE	57 #ifdef GPU_DEBUG
	index_t adjv[MAX_CELLS_PER_THRE	58 if (thisModuleId % 100 == 1)
	const index_t tid, const index_	
	* The algorithm finishes if an iteration leaves the arr	
	* This varible will be set if a change is made, and dic	
		65 // find the index of the first pixel not belonging to this module (or invalid)
		66 shared int msize;
		67 msize = numElements;
	<pre>bool gf_changed;</pre>	68syncthreads();
		70 // skip threads not associated to an existing pixel
		<pre>71 for (int i = first; i &lt; numElements; i += blockDim.x) {</pre>
		<pre>72 if (id[i] InvId) // skip invalid pixels</pre>
	nuce a change co che gi ai rayi	
		74 if (id[i] != thisModuleId) { // find the first pixel in a different module
	gt_cnanged = talse;	75 atomicMin(&msize, i);
	* stages. In this first one, a mix of stochastic and	
	* hooking, we examine adjacent cells and copy their	
	* cluster ID if it is lower than ours, essentially m	
	* together.	and the state of t
	<pre>for (index_t tst = 0; tst &lt; MAX_CELLS_PER_THREAD; ++t</pre>	
63	<pre>const index_t cid = tst * blckDim + tid;</pre>	

- We have a wealth of experience with writing efficient data processing code in a single thread
  - Even our multi-threaded jobs are built from single threaded tasks that do independent, well defined operations
- We have been developing novel algorithms for many years now, that could take advantage of hundreds of thousands of threads
  - But I believe we are still only at the beginning of doing this correctly <sup>(5)</sup>

### Things To Do Right: ML Inference

- HEP software has used machine learning since a long-long time
  - At the same time, the innovations in industry do present very new opportunities for us
- ML/AI techniques already exist for doing very complicated operations in the trigger / reconstruction software of the experiments
- What is missing are efficient algorithm chains that would "stay on a device", interleaving ML/AI inference and hand-written algorithmic steps





#### **Next Generation Triggers**





**Next Generation HEP Triggers Proposal** 

CERN - European Organization for Nuclear Research

V 1.2\_dist (20231212)

#### **Executive Summary**

The High Energy Physics (HEP) program at CERN has achieved major breakthroughs in particle physics, technology, and algorithms, including the discovery of the Higgs boson in 2012. This allowed the validation of compatibility of the theoretical construction behind the Standard Model (SM) of particle physics with the data, but the existing uncertainties leave room for models beyond the SM. With the experimental collider framework in place, scientific exploration continues to answer questions around the origin of dark matter, the disproportionately low abundance of antimatter and the nature of the discovered Higgs boson. Hard physics problems aside, much can be gained from improvements to the data acquisition pipeline allowing for capturing a richer set of collision events, furthering scientific understanding.

The Large Hadron Collider (LHC) consists of a 27 km tunnel where superconducting magnets guide bunches of protons, circulating in opposite directions, which are then caused to collide at experimental sites (e.g. ATLAS and CMS) at a rate of 40 million times per second. The collision events emit various particles, which are tracked through a multitude of radiation-hardened detectors and fed into the L1 trigger system, which needs to reject >99% of the events within 10 microseconds due to detector cache constraints and available network capacity.

This data is further reduced by >99% in the High-Level Trigger (HLT) to conform to the current event analysis and simulation capacity. HEP experimentation is fundamentally stochastic, so without changing other factors, an increase in data collection throughput would allow for higher confidence in current results while increasing the likelihood of detecting novel particles in the current LHC setup. Furthermore, this capacity increase is absolutely needed for future LHC upgrades where each collision will have many more interesting events.

The interpretation of the LHC data relies on theoretical simulations of particle interactions in the Standard Model (SM) and in scenarios of new physics beyond the SM (BSM). The full exploitation of the immense HL-LHC datasets, and in perspective of the data from Future Colliders, will require radical improvements in the computing strategies of theory calculations, to increase their accuracy while keeping affordable computing times. A multitude of theoretical tools must be addressed, in a coordinated effort, to preserve their interoperability and harmonize the overall precision. In addition to the several ingredients needed to describe the final states of proton collisions, the infrastructure developed for the triggers, e.g. the GPU cluster, also supports the advancement of software and algorithms for lattice Quantum Field Theory (LQFT) calculations, as a unique approach to control relevant non-perturbative ingredients. The engagement of LQFT experts would also bring to the triggers  All of the previously highlighted areas are pursued by the <u>"Next Generation</u> <u>Triggers" project</u>

- With external funding, CERN is starting a 5-year R&D project that will hire a large number of people, and get a large amount of accelerator hardware
- The goal is to find new ways for using the latest types of hardware in processing data from the HL-LHC
- If you're a graduate/post-doc, and interested in such developments, consider applying!

# Summary



- HEP experiments worldwide (and especially at CERN) are implementing support for accelerators in their software
  - With industry moving to providing computing power primarily in such a form in the future, this is a must
- CERN (experiments) have a rich development program for the following years to put accelerator usage into production
  - Both on our custom-built clusters, and the Worldwide LHC Computing Grid
- Some of this development is difficult to get help with from non-physicists
  - But with most of it we can certainly take advantage of industry support! Partnerships through CERN
     OpenLab can be of great help for the HL-LHC data taking / analysis.



http://home.cern