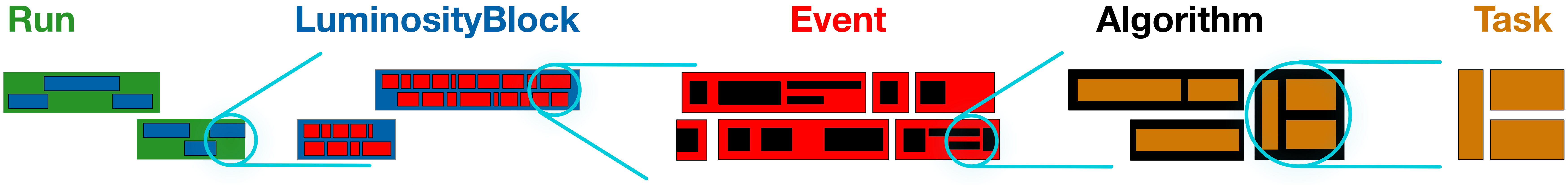
CMSSW Scaling Limits on Many-Core Machines

Dr Christopher Jones[1] and Dr Patrick Gartung[1] on behalf of the CMS collaboration

[1] Fermi National Accelerator Laboratory

FERMILAB-POSTER-23-031-CMS-CSAID

CMS Framework's Levels of Concurrency



Uses Intel Thread Building Blocks library to schedule tasks. *Shared resources (e.g. reading/writing files) are serialized via non-block tasks queues which allows efficient scheduling of other tasks when a task must wait for a resource.*

Measurement Methodology

Hardware

AMD EPYC 7763 CPUs
SSD Storage

2 sockets x 64 cores/socket x 2 hardware threads/core
= **256 threads / node**

Strategy

Completely fill node — *# Jobs = 256 / (threads / job)*
Scale # events as # threads — *Test weak scaling*
Jobs process same events — *Input file contains same 100 events repeated*

Production configurations

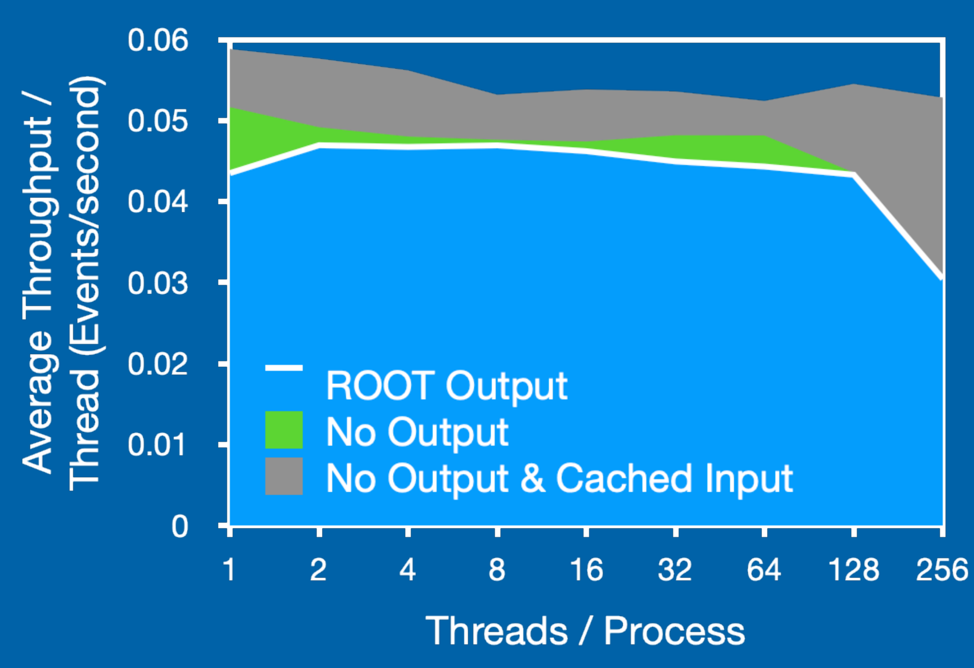
Measurement | Thread Scaling | Memory Usage

Reconstruction

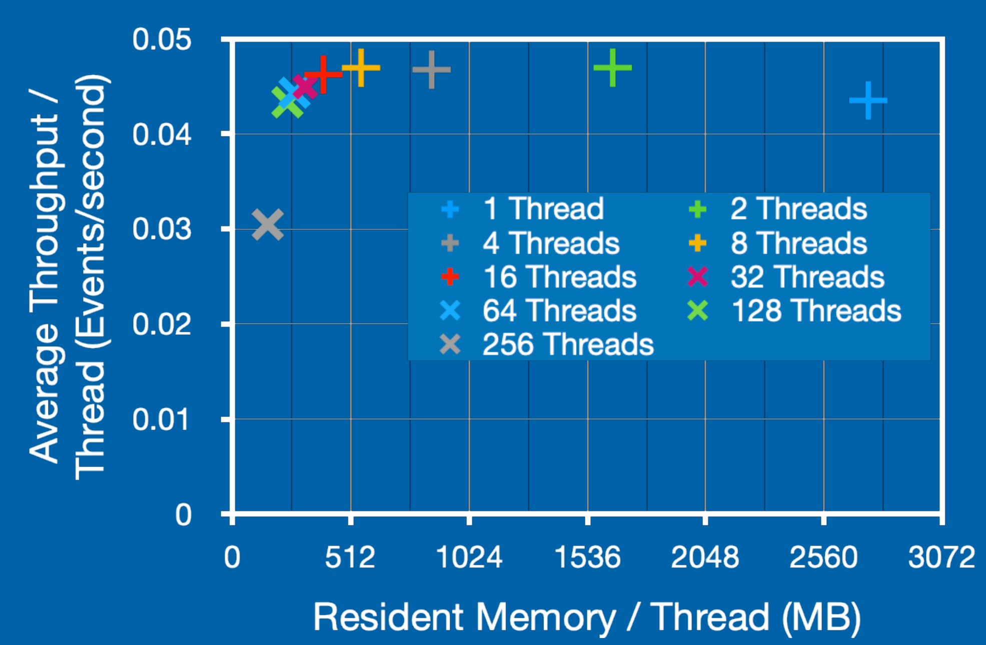
Apply pattern recognition to find physics quantities

Input: simulated HLT output
Cached input: read 100 events into memory and recycle them

Output: physics quantities for analyses



Scaling limited by input
Reading ROOT files must be done serially



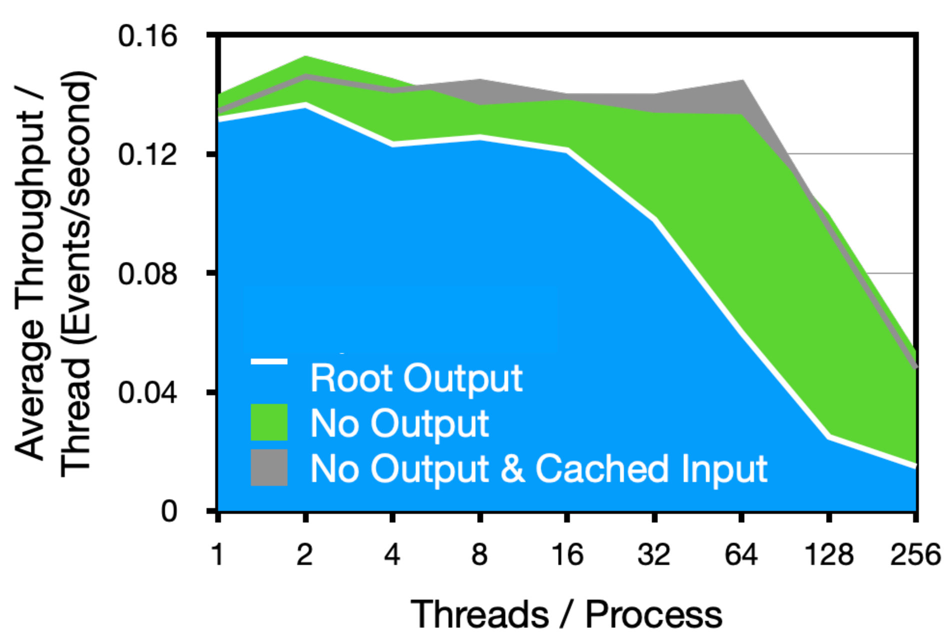
Good throughput even at 256 MB / thread at 128 threads

Overlay pp Collisions (Pile-up) & High Level Trigger

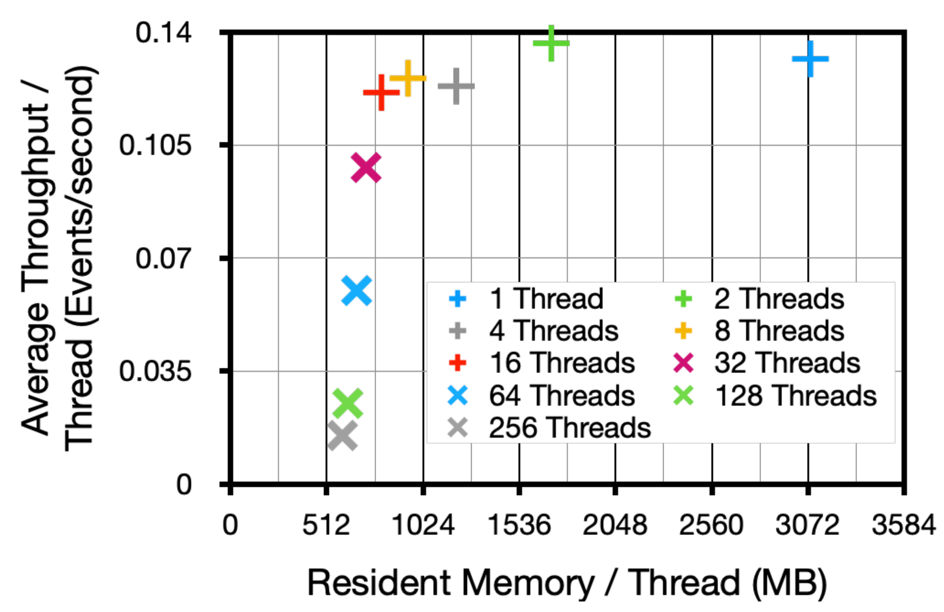
Combine simulated $t\bar{t}$ event with premade *pile-up* events
Each concurrent event reads its own pile-up file
Apply High Level Trigger selections

Input: simulated $t\bar{t}$ events & LHC Run 3 *pile-up* events
Pile-up event: *50 - 75 pp collisions per bunch crossing with 12 bunch crossings per $t\bar{t}$ event*
Cached input: *read 100 $t\bar{t}$ events into memory and recycle them*
Pile-up events not cached

Output: simulated HLT output



Scaling limited by output
Writing ROOT files must be done serially
Further scaling limited by thread-unfriendly algorithm (not input)



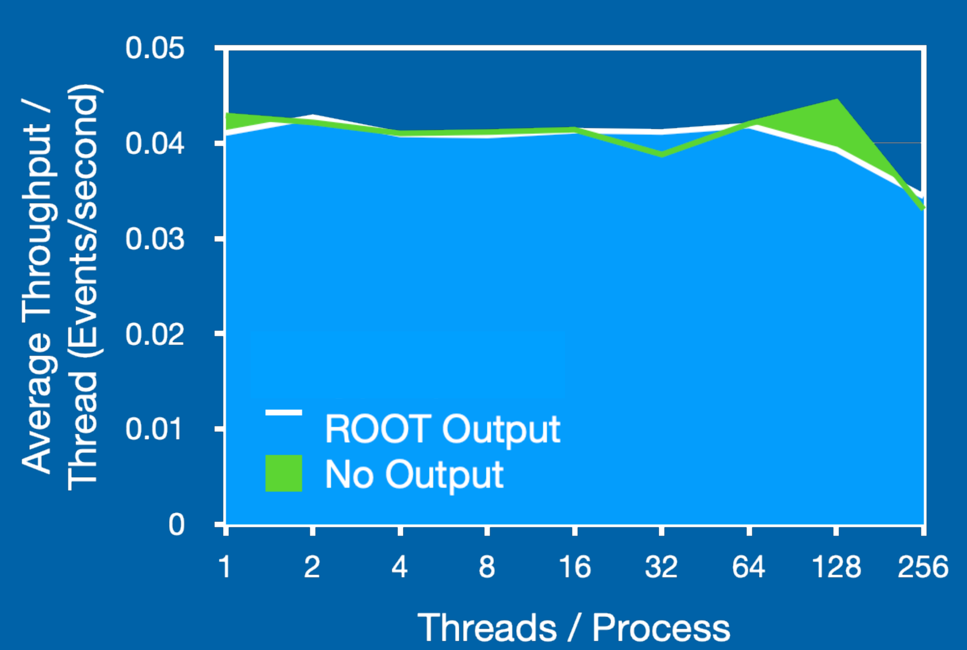
2 threads needed to get below 2 GB / thread

Event Generation & Detector Simulation

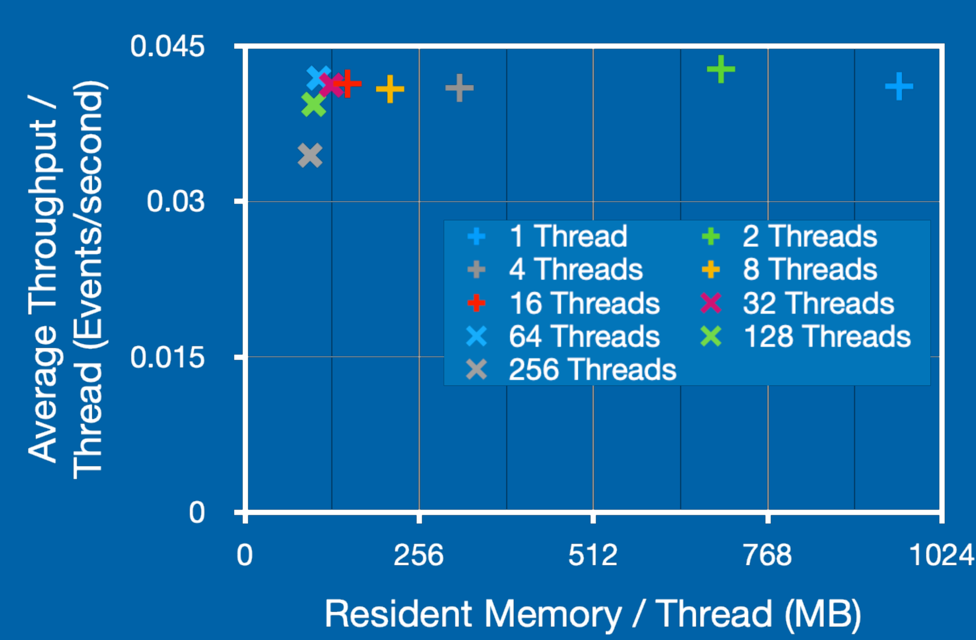
Use Pythia 8 to simulate $t\bar{t}$ events
Use Geant 4 to simulate detector response

Input: none

Output: simulated $t\bar{t}$ events



Scaling limit is under investigation



Good throughput even at 128 MB / thread at 64 threads

This document was prepared by CMS using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.