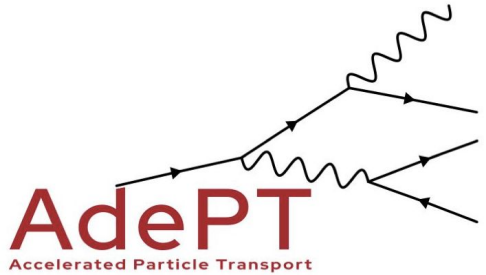
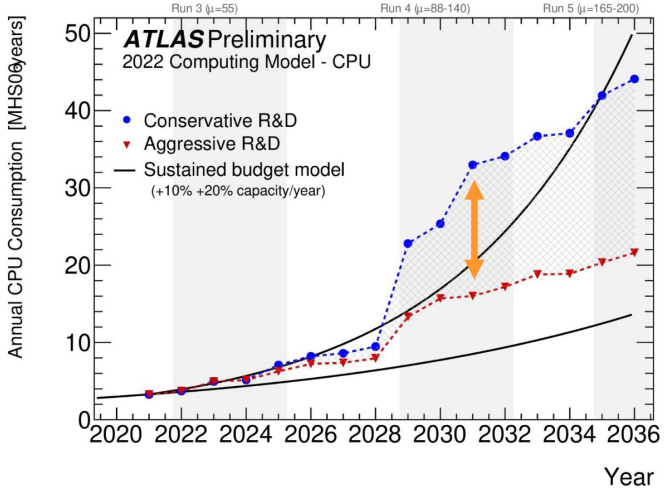
AdePT input to ESPPU

Severin Diederichs, on behalf of the AdePT team
severin.diederichs@cern.ch
CERN, EP-SFT



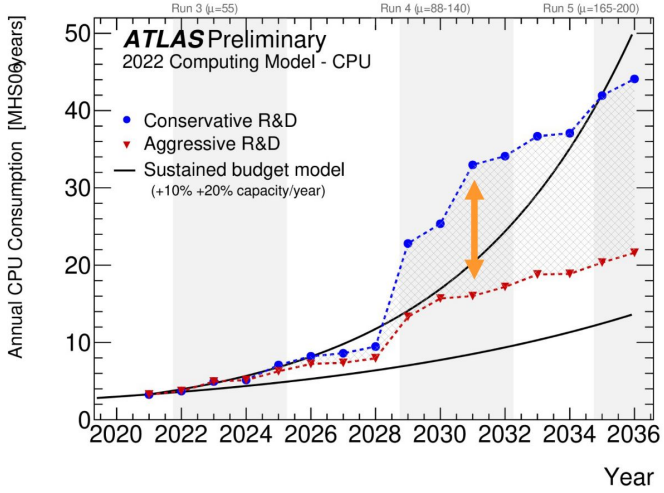
Full Simulation must become more efficient

source



Full Simulation must become more efficient … using GPUs?

source

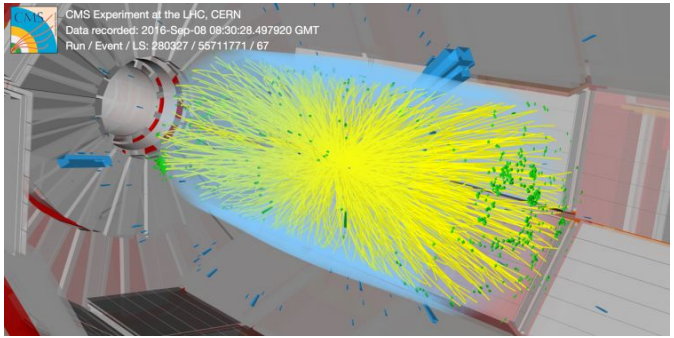


- Top500 dominated by GPU clusters
- Green500 ***completely*** dominated by GPU clusters

  Top 50 of Green500 are **all** GPU-based

Use GPUs to run full simulation?

Porting GEANT4 to GPUs seems easy…

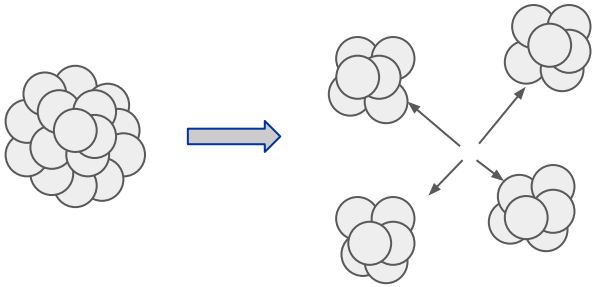


source

- embarrassingly parallel problem
  → **perfect for GPUs!**

Porting GEANT4 to GPUs ~~seems easy~~ is very difficult

Hadronic physics



Fermi-Breakup

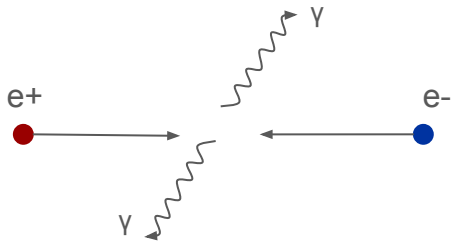
G4: ***991*** reaction channels!

- embarrassingly parallel problem
  → **perfect for GPUs!**
- huge amount of **divergence** in **physics** and **geometry**

  → **terrible for GPUs!**

Porting GEANT4 to GPUs ~~seems easy~~ is very difficult

Electromagnetic physics



Annihilation

- embarrassingly parallel problem
  → **perfect for GPUs!**
- huge amount of **divergence** in **physics** and **geometry**

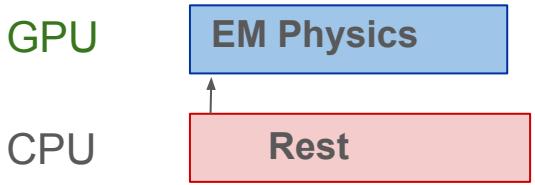
  → **terrible for GPUs!**

→ **Approach**:
offload **electromagnetic** part to GPU

Porting G4 to GPUs

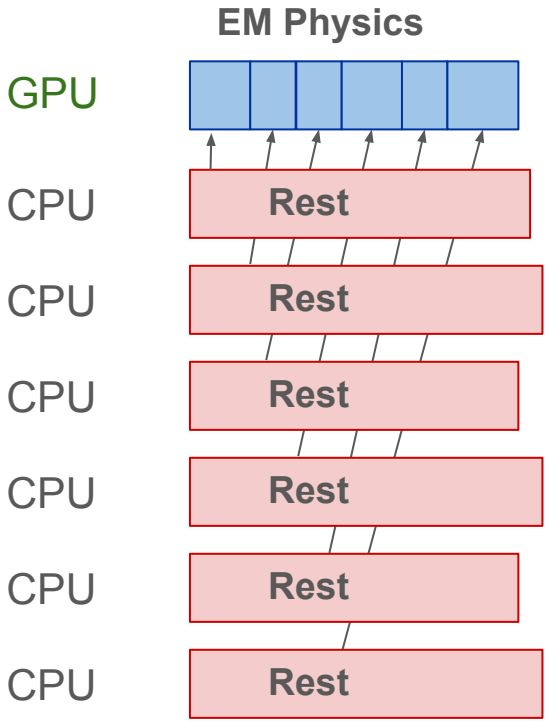
Typical LHC production run

Porting G4 to GPUs



~ **3.3x** theoretical speed-up using GPUs

Porting G4 to GPUs



~ **3.3x** theoretical speed-up using GPUs

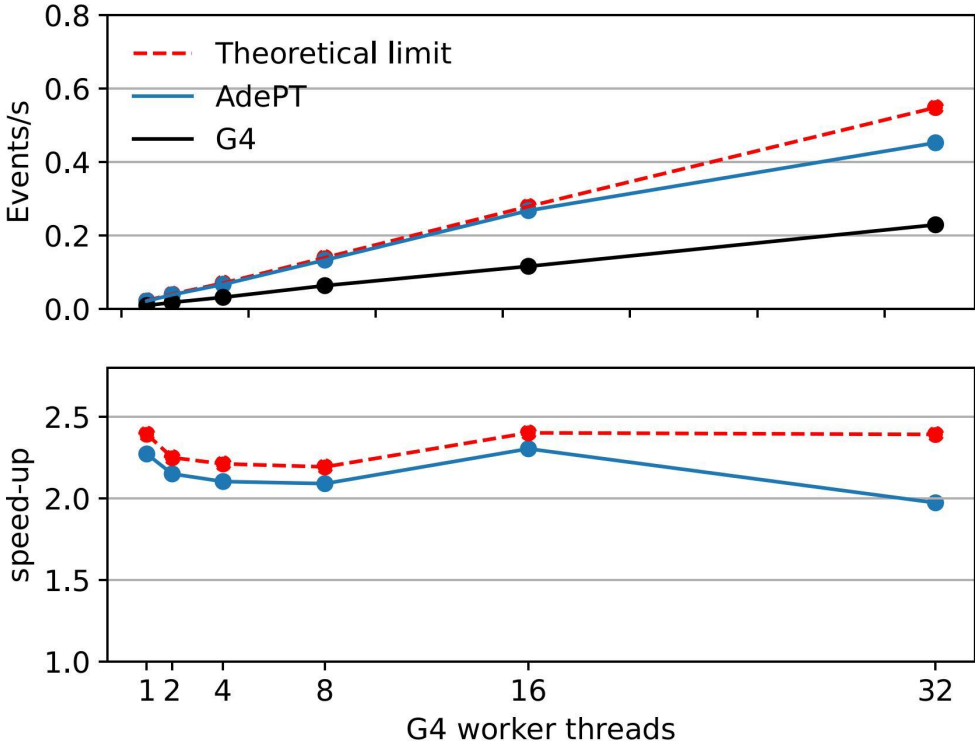
Realistically 16-32 CPU cores per GPU

Ambitious goals:

- close-to-ideal speed-up
- better energy efficiency

AdePT: current performance results

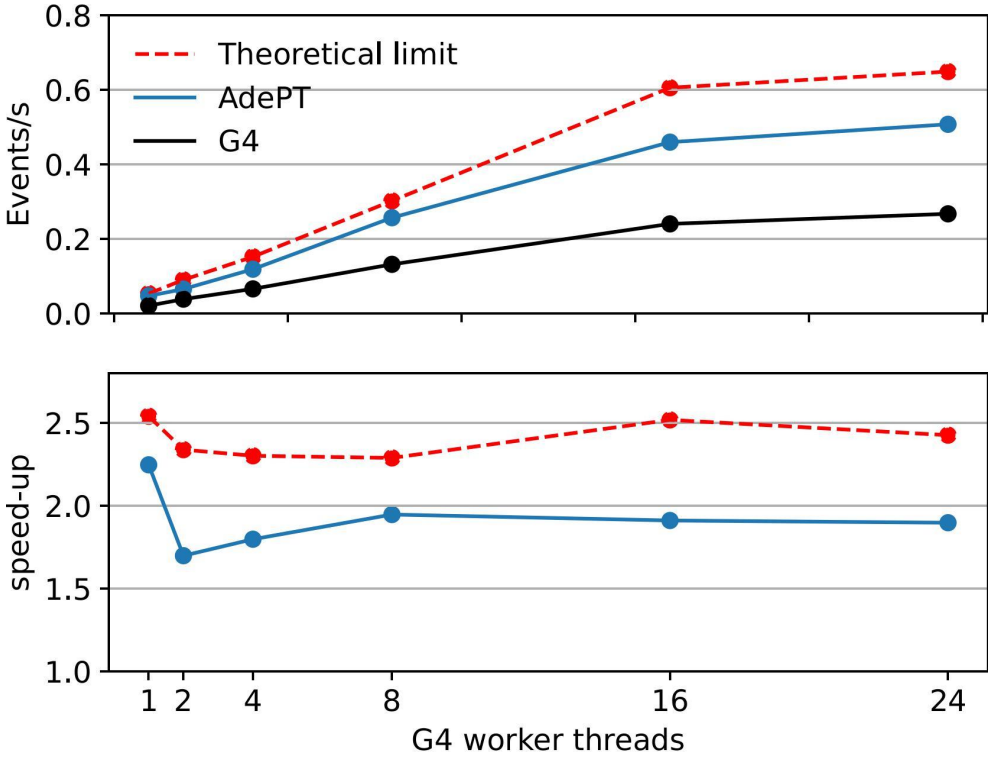
GPU: **Nvidia A100**
CPU: AMD EPYC 7752 96 cores
Input: 4 TTBar per thread
Geometry: CMS2018
No magnetic field
Simple scoring of energy deposition



AdePT: current performance results

GPU: **Nvidia RTX4090**
CPU: AMD Ryzen 9 16 cores
Input: 4 TTBar per thread
Geometry: CMS2018
No magnetic field
Simple scoring of energy deposition

Potential usage of consumer-grade GPUs?



Integration of experimental frameworks crucial

AdePT:

- header-based library for simple integration

  Integration with a **single line**!

- **Generic scoring**: calling the user-defined sensitive detector code on CPU

- Specific experimental needs require additional development and can penalize performance

Gaussino

```
GaussinoSimulation(
    PhysicsConstructors=[
        "GiGaMT_AdePTPhysics",
        "GiGaMT_G4EmStandardPhysics",
        "GiGaMT_G4EmExtraPhysics",
        "GiGaMT_G4DecayPhysics",
        "GiGaMT_G4HadronElasticPhysics",

"GiGaMT_G4HadronPhysicsFTFP_BERT",
        "GiGaMT_G4StoppingPhysics",
        "GiGaMT_G4IonPhysics",
        "GiGaMT_G4NeutronTrackingCut"])
```

Future challenges

- Promising speed-up obtained in complex test simulations - required in **production setups**
- Achieve superior **energy efficiency** on GPU
- **Remove bottlenecks**: faster geometry using a surface model
- Explore usage of GPUs with **FP32 and below**
- **Easy integration** - with a twist! Experiment-specific code needed?
- Maintain GPU code per experiment for a decade ➙ **Portability, stability**?
- Put **more physics** on GPU to increase throughput

Backup

AdePT: current performance results

GPU: Nvidia RTX4090
Input: 1 TTBar per thread
Geometry: CMS2018
uniform magnetic field