GPX: GPU-Accelerated Beam-Beam Simulation for FCC-ee and Future Colliders

Dimitris Ntounis[1,2], Caterina Vernieri[1,2]

[1]Stanford University [2]SLAC National Accelerator Laboratory

A CUDA implementation of GuineaPig, validated statistically against the reference code

Motivation

Beam-beam interactions at the interaction point are the dominant source of beam-induced background at e^+e^- colliders, such as incoherent e^+e^- pairs and beamstrahlung photons. **GuineaPig** is the reference simulation for these processes at FCC-ee, as well as linear colliders (CLIC, ILC, C^3). It is a serial, single-core code.

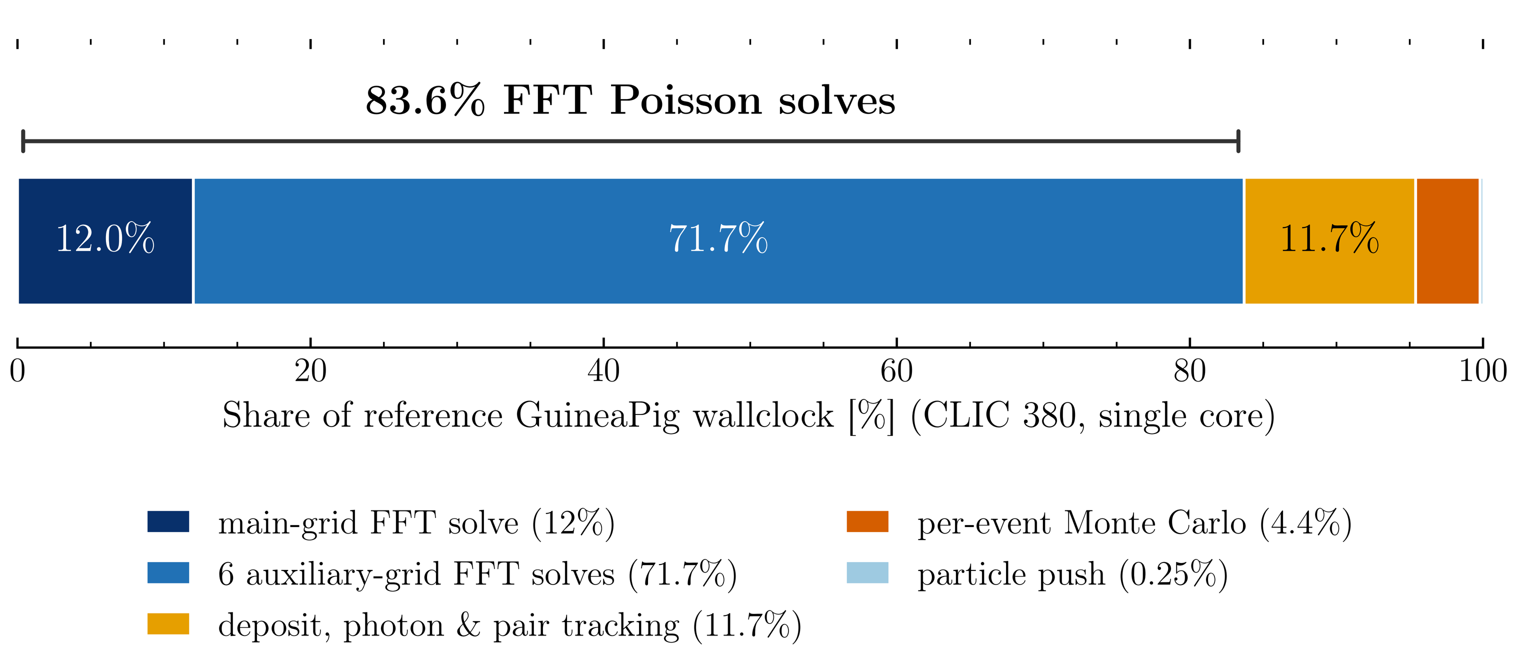
2 h

one FCC-ee Z bunch crossing on one CPU core (256^2 grid)

A background campaign of 10^7 bunch crossings therefore costs about $\mathbf{2 \times 10^7}$ **core-hours** before any detector simulation. The bottleneck is the field solve, which does not parallelise across cores.

Where the runtime goes

Profiling of the reference code attributes **83.6%** of the wallclock to FFT-based Poisson solves: one on the main transverse grid and six on auxiliary longitudinal grids, executed every slice step.

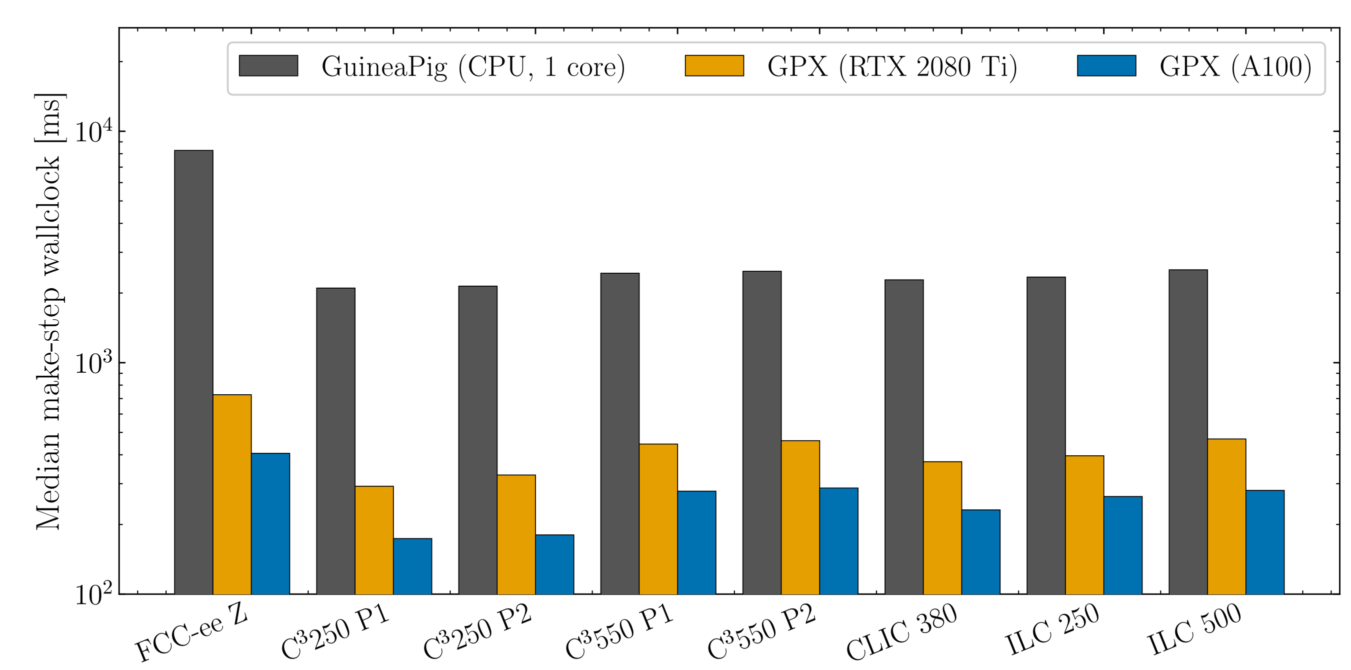


Because the per-step solve is a serial FFT, the parallelisable fraction of a step is only $p \approx 0.10$–0.45. An ideal shared-memory parallelisation of GuineaPig would reach ~ 1.1–$1.8\times$ and saturate by 32 cores. A GPU port instead **replaces** the FFT solve with cuFFT and moves the surrounding particle work to the device.

What runs on the GPU

- **Field solve** — all seven Poisson solves as cuFFT convolutions with the Green's function, whose transform is kept in device memory.
- **Particle kinematics** — cloud-in-cell charge deposition, field interpolation and the particle push. Particles and photons stay resident on the GPU in structure-of-arrays layout for the whole crossing, so no per-slice host–device transfer is needed.
- **Monte Carlo** — $\gamma\gamma$ collision enumeration and the Breit–Wheeler pair production run as batched kernels with per-thread counters seeded from the CPU host random stream.
- Double precision is maintained throughout, whereas the physics models, input decks, and output formats of GuineaPig are unchanged.

Speedup over single-core GuineaPig



Median wallclock of the per-step field-solve and collision loop (`make_step`), 3 seeds per point, GPU and CPU measured on the same node. CPU: AMD EPYC 7542 (Zen 2, 2.1 GHz), single thread, `-O3`, transparent huge pages disabled. GPUs: NVIDIA RTX 2080 Ti and A100-SXM4-40GB.

	RTX 2080 Ti	A100
FCC-ee Z	**11.4×**	**20.4×**
Linear colliders (7 configurations)	5.4–7.2×	8.6–12.1×

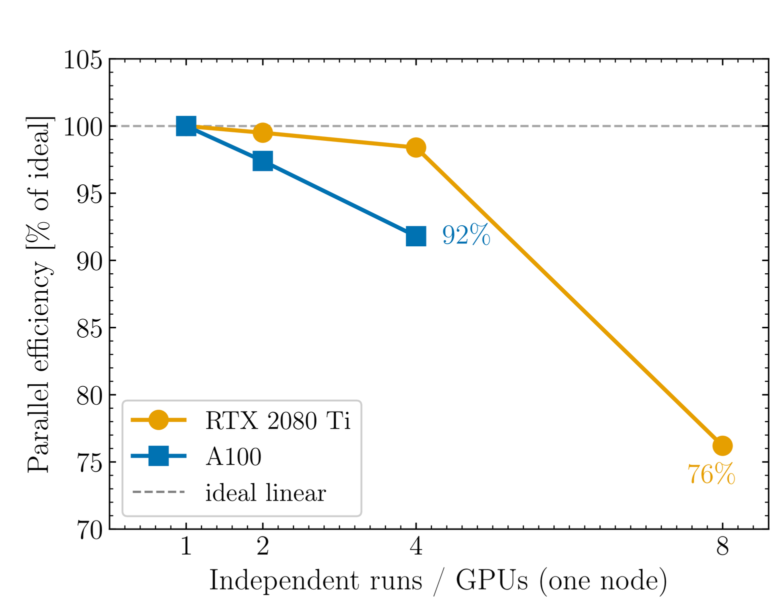
FCC-ee Z gains the most because it is the most field-solve-dominated configuration in the set, and the field solve is exactly the part the GPU takes over. The linear-collider configurations spend a larger share of their time in the per-event Monte Carlo, which is intrinsically less parallel.

Time per bunch crossing, end to end

	CPU core	RTX 2080 Ti	A100
FCC-ee Z	2.2 h	12.0 min	**6.9 min**
CLIC 380	56 min	5.9 min	3.5 min
ILC 500	45 min	7.9 min	5.1 min
C^3 250 PS1	47 min	5.3 min	2.9 min

Complete executable wallclock including start-up and output, median of 3 seeds, same node as above. A 10^7-crossing FCC-ee Z campaign moves from 2×10^7 core-hours to 1.1×10^6 A100-hours.

Many crossings in parallel on one node



GPX runs **one bunch crossing per GPU**. On a node with N GPUs, N independent crossings run concurrently, each driven by its own CPU core; the runs share only the host memory bandwidth, PCIe and disk.

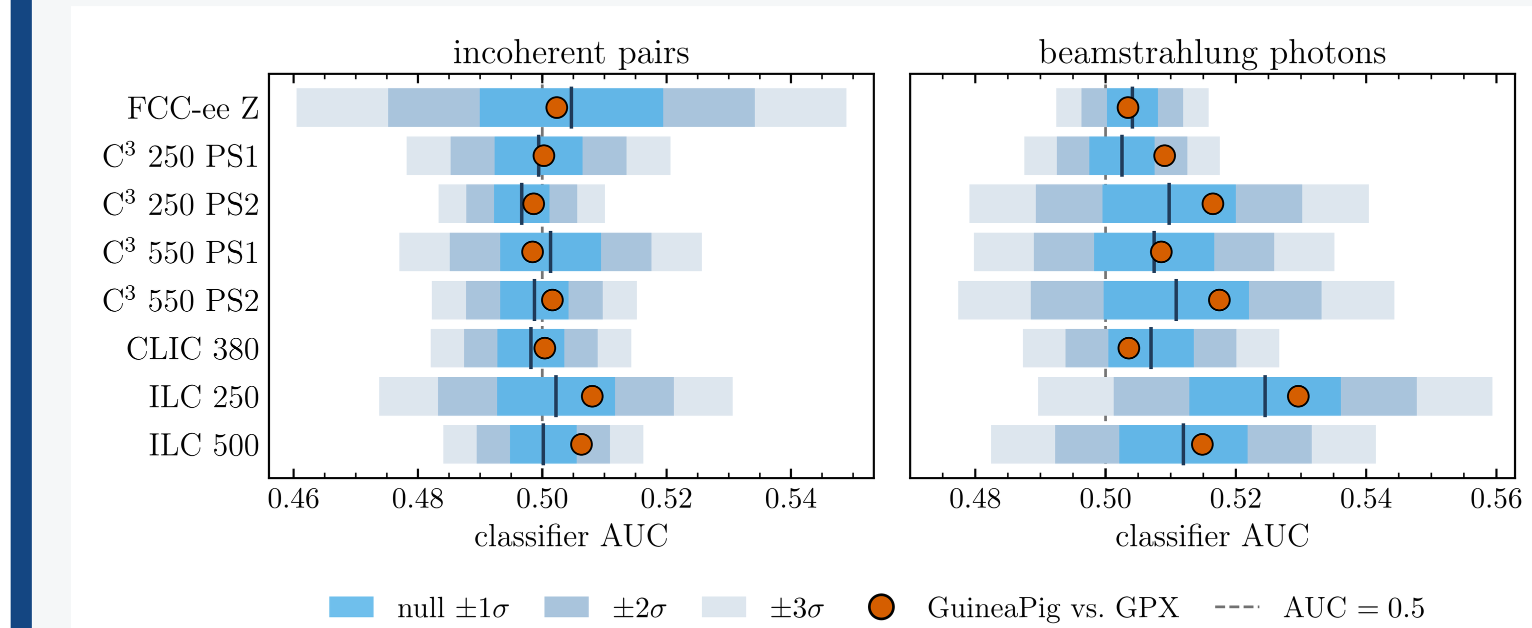
- **98%** of ideal scaling up to 4 GPUs.
- 76% at 8 GPUs, from contention on the shared host resources (not GPU limit).

Parallel efficiency: time of a single run divided by the time per run with N concurrent runs. CLIC 380.

Statistical agreement with GuineaPig

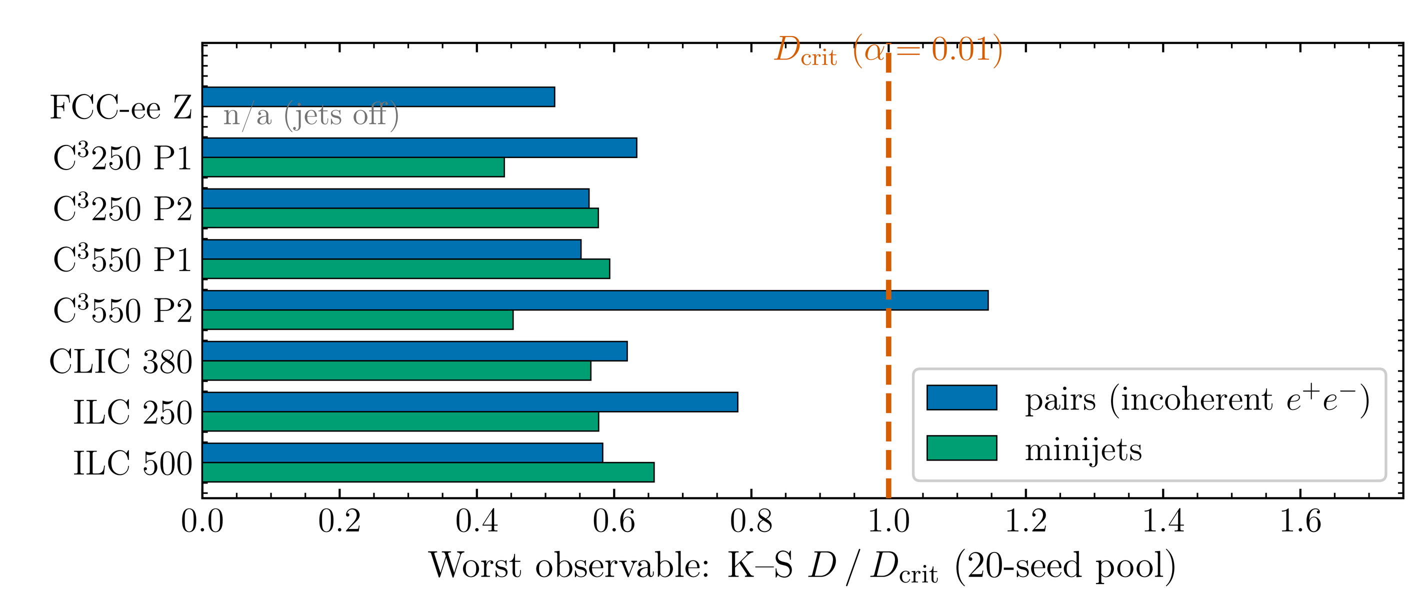
GPX is not bit-reproducible with respect to the legacy GuineaPig code, because charge deposition uses atomic floating-point accumulation. Agreement is therefore assessed on the physics output: for each of the eight collider configurations tested, 20 seeds per code are pooled and compared with **two-sample Kolmogorov–Smirnov tests** on the full luminosity and beam-induced background distributions at $\alpha = 0.01$, and with a **per-record classifier** trained to separate GuineaPig from GPX records, calibrated against the null from random pairings of GuineaPig seeds.

Classifier: records are indistinguishable



Classifier AUC for GuineaPig vs. GPX against the null band from GuineaPig-only pairings. All 16 cases lie within the null; the largest deviation is 1.3σ. AUC = 0.5 is random guessing.

Kolmogorov–Smirnov: worst observable



Largest K–S statistic over all output columns, relative to the critical value; for pairs it is always z or v_z, for minijets a jet energy. The single exceedance (C^3 550 PS2 pair z, 1.15 $D_{\rm crit}$) is a 0.15% difference in the cumulative distribution at $N = 3.1 \times 10^6$, at the resolution limit of the test.

Conclusions and outlook

- GPX statistically reproduces the luminosity and background distributions of GuineaPig on all tested collider configurations.
- One GPU replaces a small CPU allocation: 11–20× faster on FCC-ee Z, one bunch crossing in 7–12 minutes instead of 2.2 hours.
- The remaining cost is the FFT compute itself; further gains need a different Poisson solver rather than more porting.

dntounis@stanford.edu