Tests AMD MI300X

PRODUCT PAGE

Intro

We are interested in throughput tests of MadGraph event generator, that means how many events we can generate per second. An idea would be to run processes such as `g g > t t~ (g) (g) (g)` with different number of events:

- `g g > t t~`;
- `g g > t t~ g`;
- `g g > t t~ g g`;
- `g g > t t~ g g g`.

with varying number of events 1k, 10k, 100k, 1M, 10M.

When running `madevent_<hip/cuda>`, it prints the stats at the end of the run. The timers are C++ functions. Notice that so far only the matrix element computation is offloaded on GPU, so there is still a FORTRAN overhead.

The default vector size used by MadGraph is 16384 (2^{14}), but it can be changed in the `run_card.dat` file. Notice that the GPU is filled entirely independently on the number of events that is selected. This depends on the number of Floating Point double precision cores: they will all work. So, there may be cases in which the generated number of events is smaller than the number of cores, and in this case, most of the generated stuff will be thrown away.

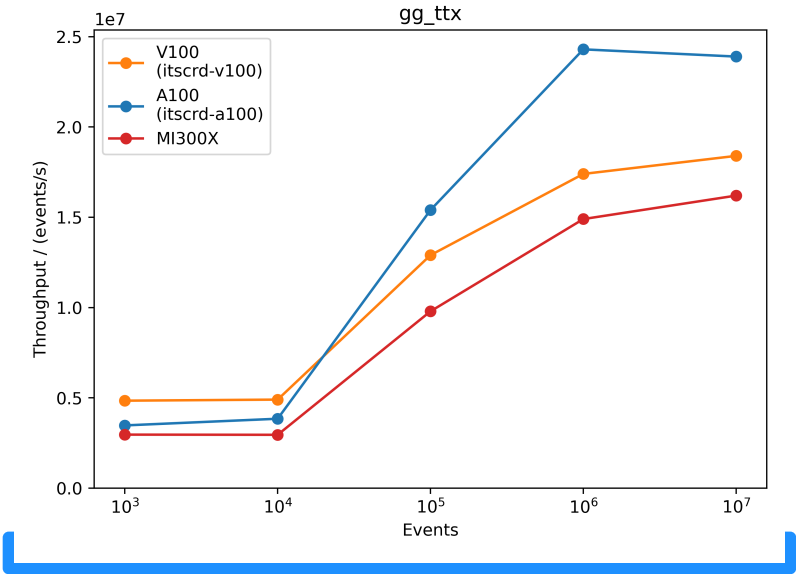
Results

Notice I still need to run the benchmarks on NVIDIA H100.

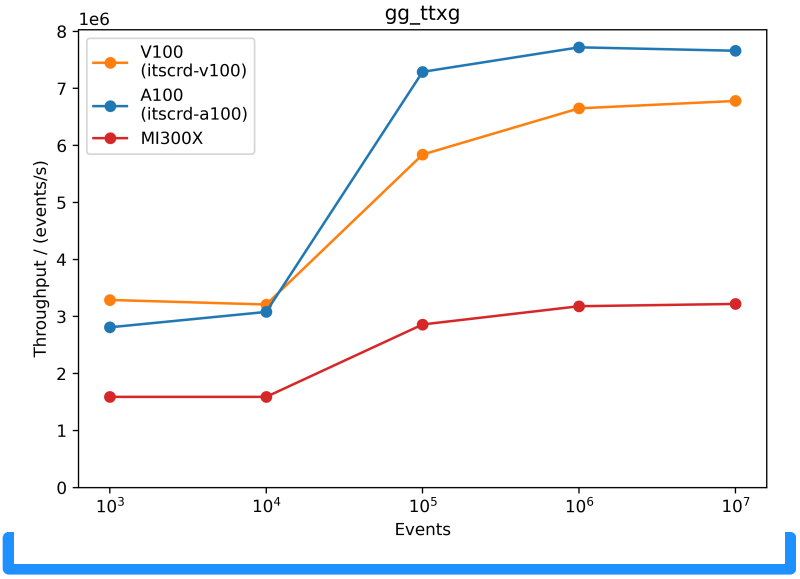
Click on each plot to display the pdf.

Throughput

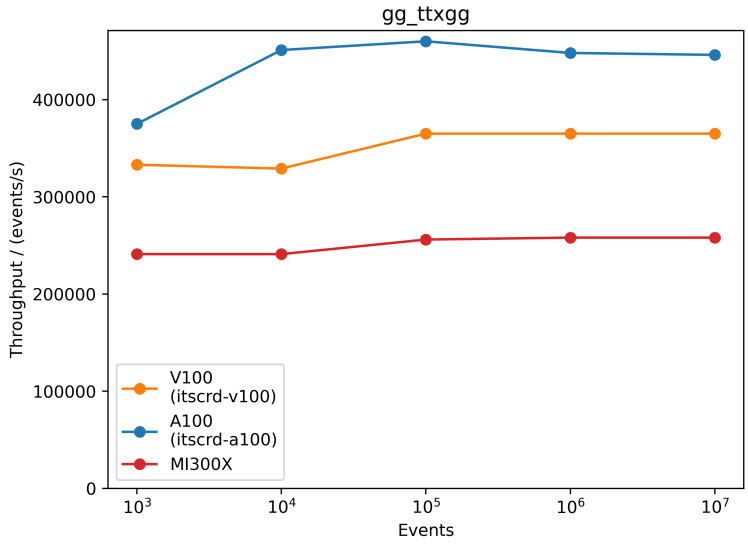
```
g g > t t~
```



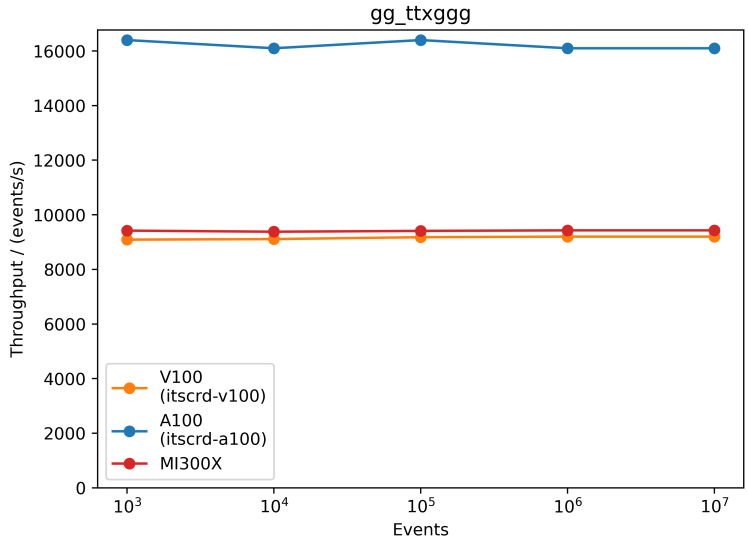

```
g g > t t~ g
```




```
g g > t t~ g g
```




```
g g > t t~ g g g
```



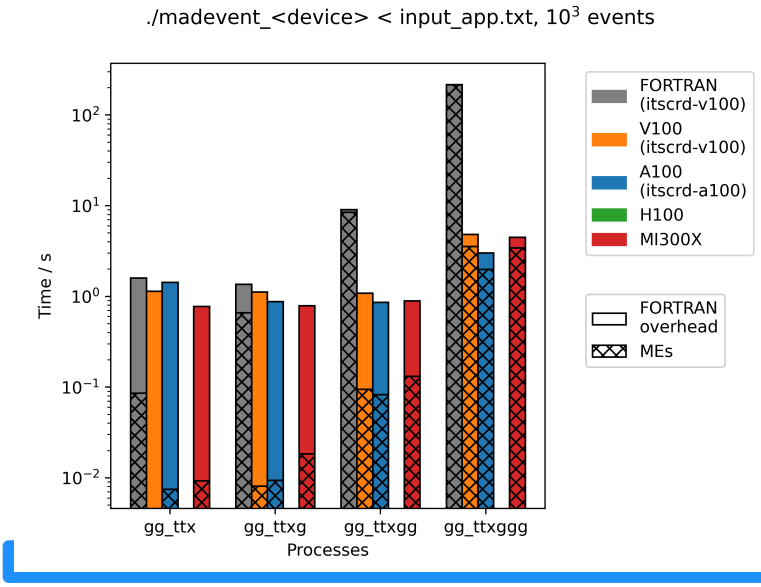
Conclusions

The throughput values shown are referred only to the matrix element computation, so only the part running really on GPU. We see that the AMD performance are comparable to the NVIDIA V100, and they are worse than the NVIDIA A100.

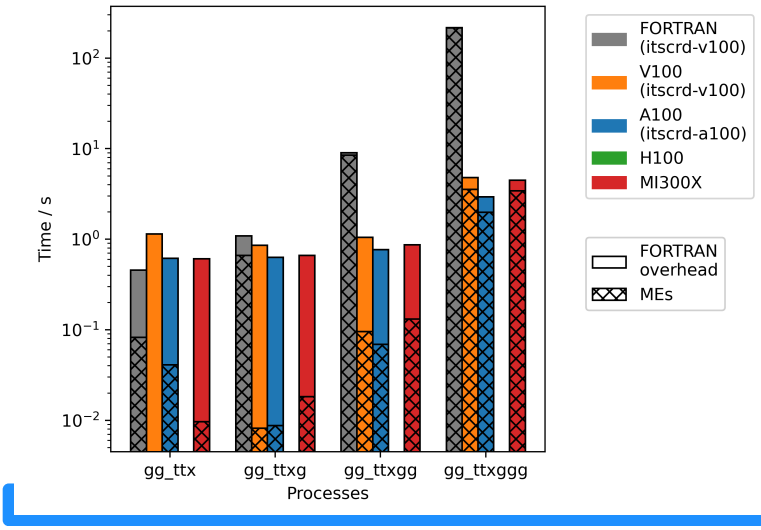
Timings

All together

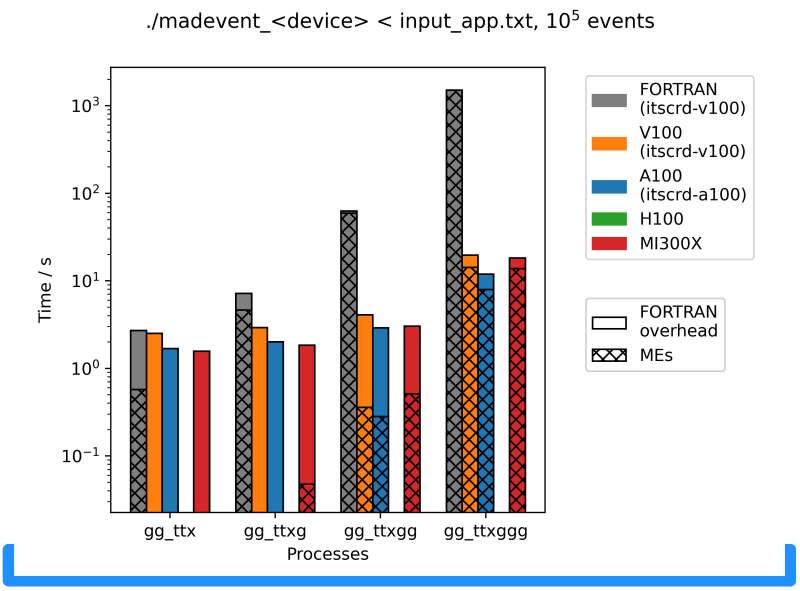
10^3 events



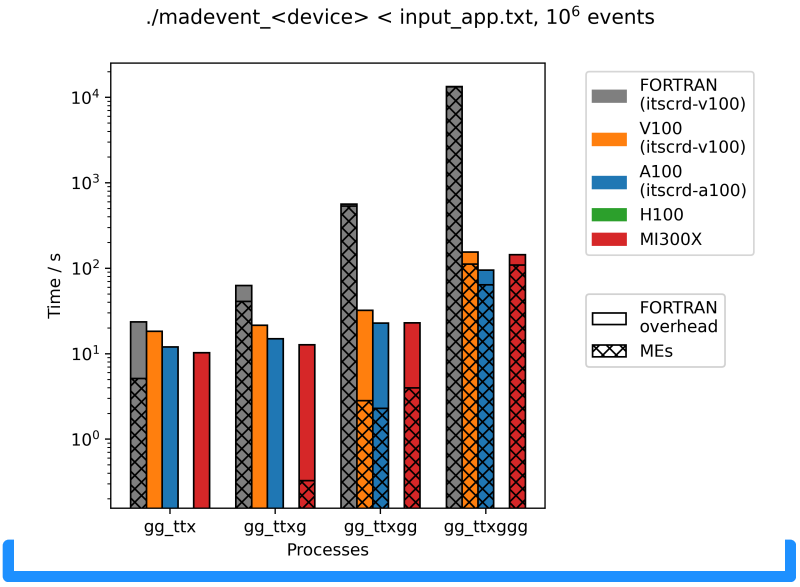
10^4 events



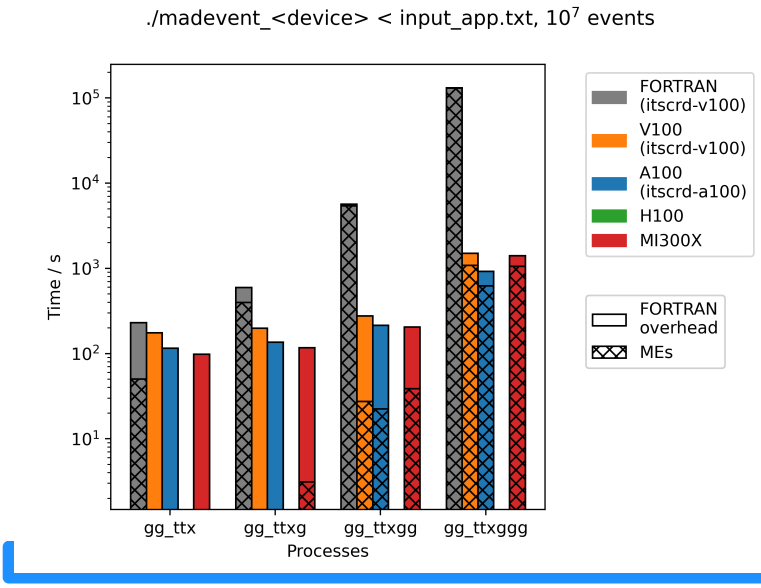
10^5 events



10^6 events

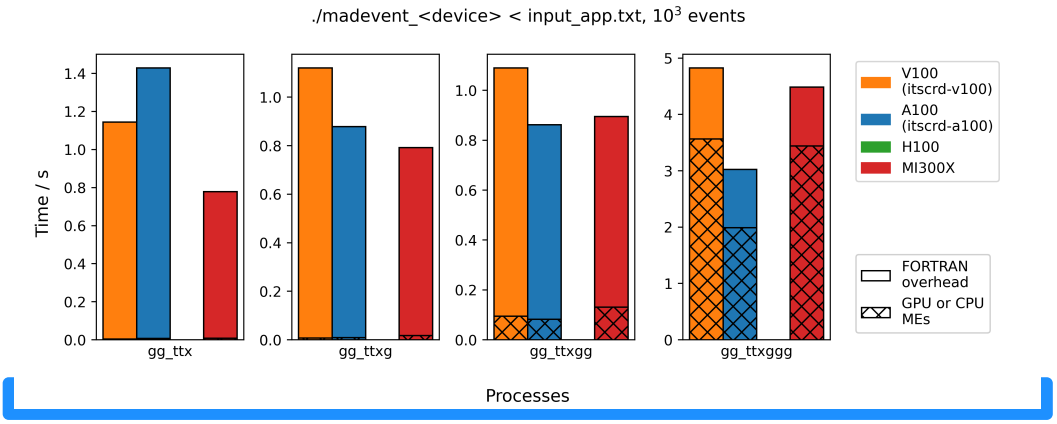


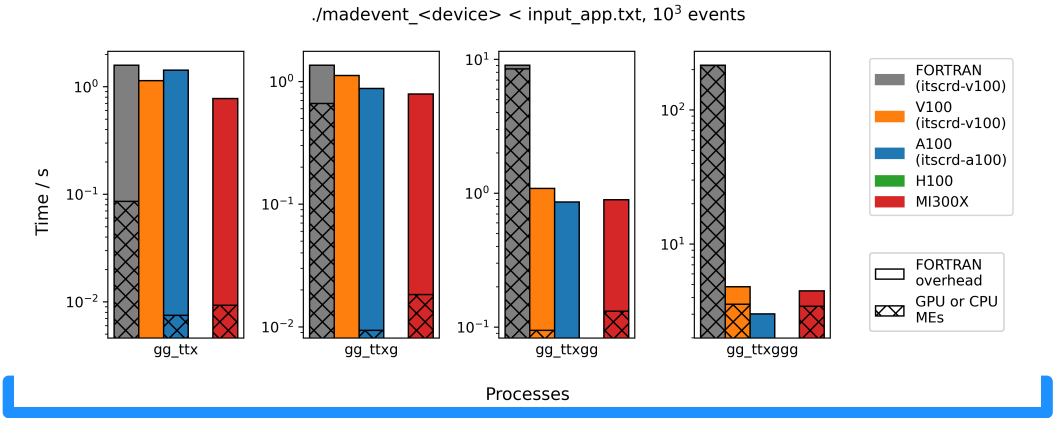
10^7 events



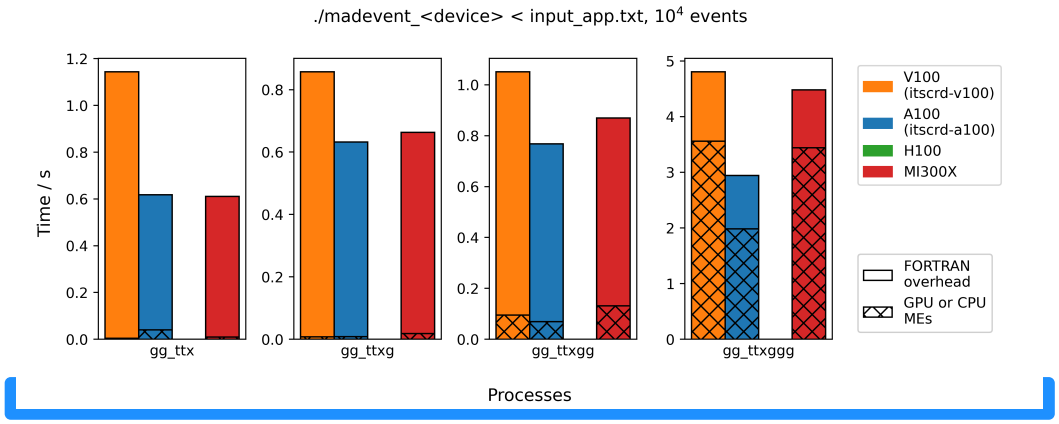
Panel

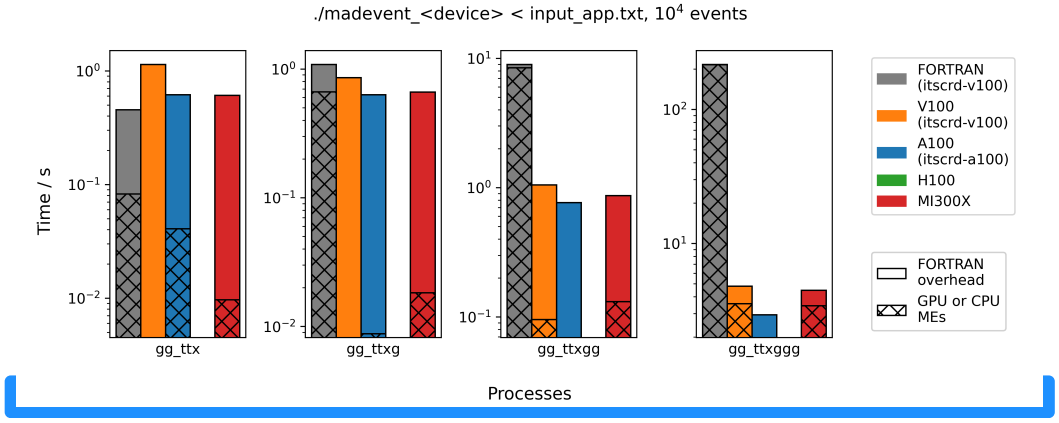
10^3 events



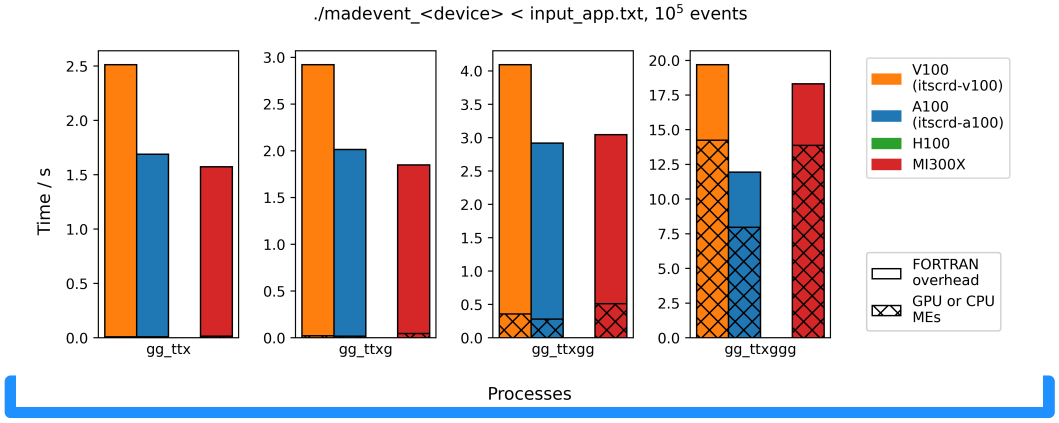


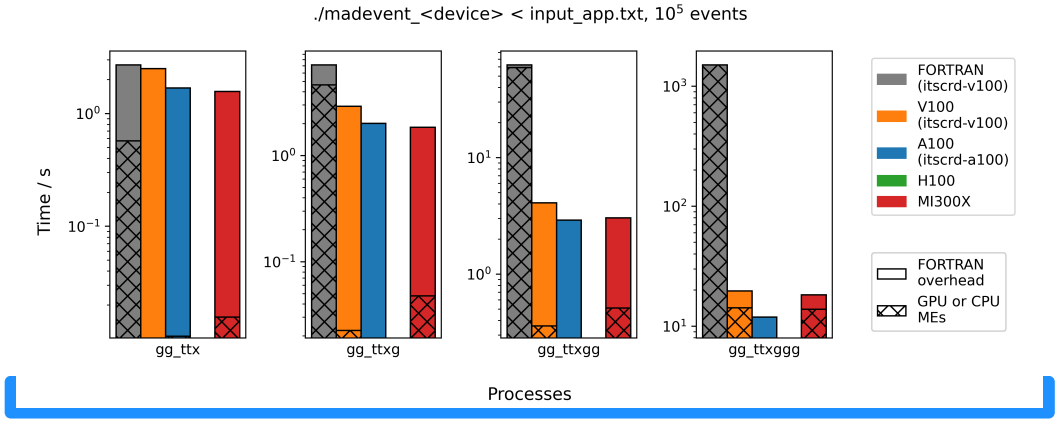
10^4 events



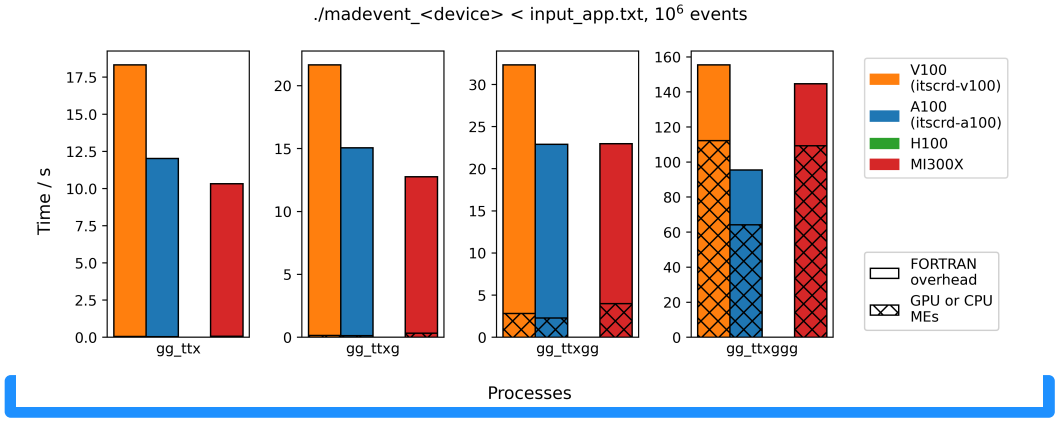


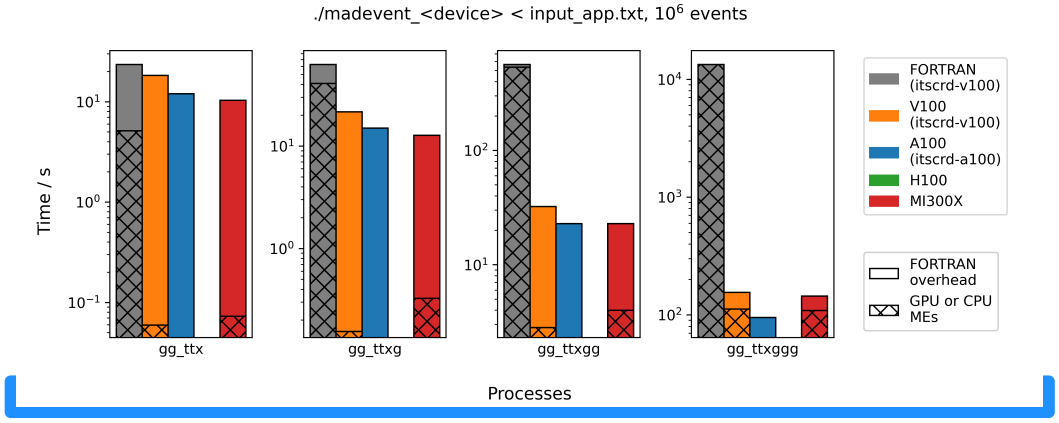
10^5 events



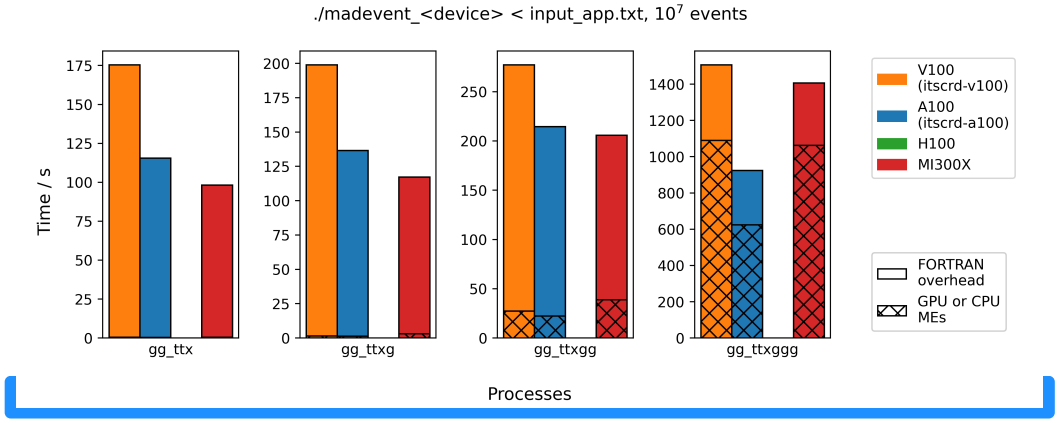


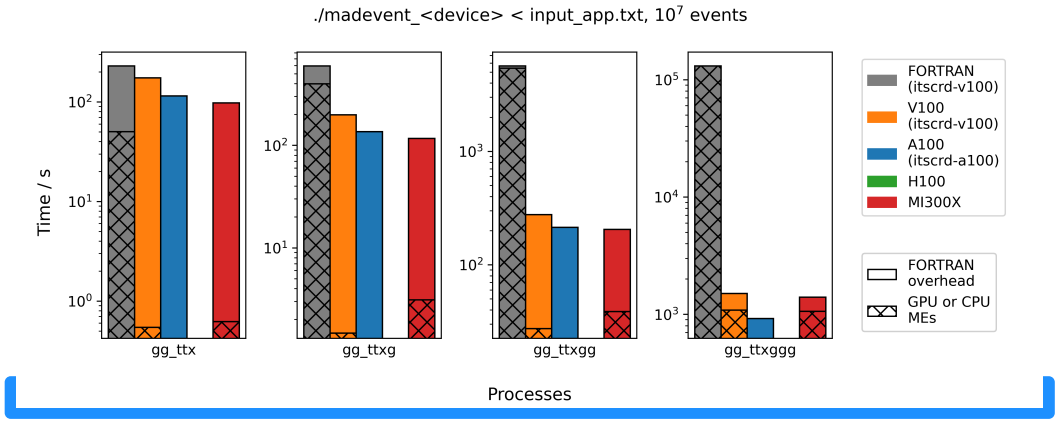
10^6 events





10^7 events





Conclusions

The GPU part is hatched, and we can clearly see that the AMD GPU is characterized by a higher GPU time with respect to the NVIDIA GPUs. This is even more evident when the process is very complicated, like `g g > t t~ g g g`.