From: **Kevin Pedro** kpedro88@gmail.com
Subject: Re: Pileup simulation for HL-LHC
Date: January 26, 2025 at 23:41
To: Sridhara Dasu dasu@hep.wisc.edu
Cc: Nick Smith nick.smith@cern.ch

KP

Hi Sridhara,

Indeed, there has been renewed interest in the past ~6 months in new methods to overcome the computing challenges of pileup mixing. My slides from the last O&C week are here: https://indico.cern.ch/event/1460070/contributions/6190298/attachments/2956387/5198603/Pileup%20Mixing%20Alternatives.pdf.

There is an implementation of premixing for the new Phase 2 detectors, but the physics performance has not been fully validated yet. In my opinion, this is mainly the result of recalcitrance by the DPGs; they kept putting off the validation because they kept being allowed to produce classical mixing samples for TDR studies. But premixing still has large resource requirements and implications, so we have started looking at alternatives. (For example, PPD had to delete large fractions of the existing premixed input samples from disk in order to make room for the 2024 sample, which is already quite large because of the increased n_PU in Run 3 and the increased number of cells in the Phase 1 detector. This was a very disruptive action whose consequences are still not resolved yet.)

1. I'm not sure exactly what you mean by "stats" being the same as Run 1 - obviously we produce a number of events proportional to the integrated luminosity for a given year. If you mean the "reuse factor", I don't think this has been re-optimized anytime recently, and it deserves another look. (My personal expectation is that we could reuse a lot more events and nobody would be able to tell...)
2. a. I am currently working with colleagues at SINP to put together a prototype of mixing FastSim MinBias into FullSim. It is almost working now. I will try to provide an update on this during the O&C week next week. Matthew Herndon has also expressed interest in implementing some optimizations from the muon side.
b. I personally think we can greatly reduce the amount of out-of-time pileup we simulate, and probably optimize it on a per-subdetector basis (this may be easier in FastSim because of its simplifications). To the extent that some subdetectors are sensitive to out-of-time pileup, though, I doubt that just adding noise will be sufficient. Pileup is signal, just signal that we don't want...
c. I am also interested in exploring ML-based pileup generation, which affords several interesting possibilities. I will flesh this out a bit more in the upcoming O&C week, since there seems to be some confusion about what was presented in October.
3. Yes, Nick and I have been in communication about the numbers for the current scenario.

Best,
Kevin

On Sat, Jan 25, 2025 at 2:03 AM Sridhara Dasu <dasu@hep.wisc.edu> wrote:

> Dear Kevin,
>
> Pileup simulation for HL-LHC with full simulation is ridiculously expensive. Apparently, it sets the requirements for certain IO paths. It is not clear to me what we are doing for pileup simulation is even appropriate. The last time I had a look at it was decades ago. Apparently, the procedure has not been revised much after pre-mixing ideas of Mike Hildreth.
>
> I think it is important to revisit this. In this context the following questions cropped up. Would you have some comments about these?
>
> 1) Minbias/pileup-premixing stats are the same since Run 1 is the assumption going in to resource needs for IO. Is this fair?
> 2) Do we need to do pileup premixing this way still?
> a) Could we use Fastsim for pileup and mix it in with full sim? Could the data tiers be made compatible?
> b) Could we avoid out-of-time-pileup simulation and simply use salt-and-pepper noise hits?
> c) Other ideas
> 3) Are these numbers sane from your point of view?

Special productions

- Premixed pileup generation
  - In 2023, we used about 1B minBias events to make 200M premixed events with PU60 16 BX (i.e. 960 minBias/premix), and used that for 16B MC events. Each minBias ever (simHits) is 0.2 MB.
  - In 2031 we will simulate PU140 in 7 BX, using 0.4MB/event inputs. Assuming the same re-use factor, 135B MC events needs 1.7B premixed events. Mixing takes 1kHS23s/event, so using all US resources would take 4 days and require **15 Tbps input**

> Nick presented the following at our meeting few weeks ago - the screenshot is one of the slides:
>
> 
>
>
> Best Regards,
> Sridhara
>
> _____
> Professor Sridhara Rao Dasu
> Department of Physics, 4289CH
> University of Wisconsin
> +1-408-829-6625 dasu@hep.wisc.edu
> https://calendly.com/dasu/30min
> http://www.hep.wisc.edu/~dasu