Imperial College London

Recording and reconstructing 10 billion unbiased B hadron decays in CMS

Robert Bainbridge

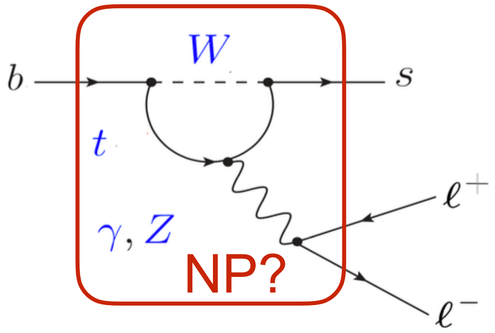
on behalf of the CMS Collaboration

24th International Conference on Computing in High Energy and Nuclear Physics
Adelaide, 4th - 8th November 2019

Physics motivation and challenges

"Flavour anomalies" in measurements from LHCb, BaBar, Belle, ...

Measurements of $R_{\mu\mu}$ and P_5' from CMS using dedicated (μμ) triggers



FCNC transitions such as $b \to s\ell\ell$ are ideal probes for new physics

Anomalies in R_K and R_{K^*} observables that probe lepton flavour universality

Theoretically clean, but experimentally challenging!

$$R_{K^{(*)}} = \frac{\mathcal{B}(B \to K^{(*)}\mu\mu)}{\mathcal{B}(B \to K^{(*)}ee)}$$

e.g. CMS cannot trigger directly on electron channels

Solution:

- Trigger on tag-side μ from $b \to \mu X$ decays
- Use signal-side to access unbiased B decays

$\mathcal{B}[b \to s\ell\ell]$ is $O(10^{-7}) \Longrightarrow$ require sample of 10^{10} B decays

Is this feasible with the CMS trigger?

Tag-side:
$b \to \mu X$

Signal-side:
unbiased
B decays

CMS trigger system (and offline reconstruction)

L1 Trigger

- Hardware-based
- Only muon stations and calorimeters
- Decision time: 3.2 μs

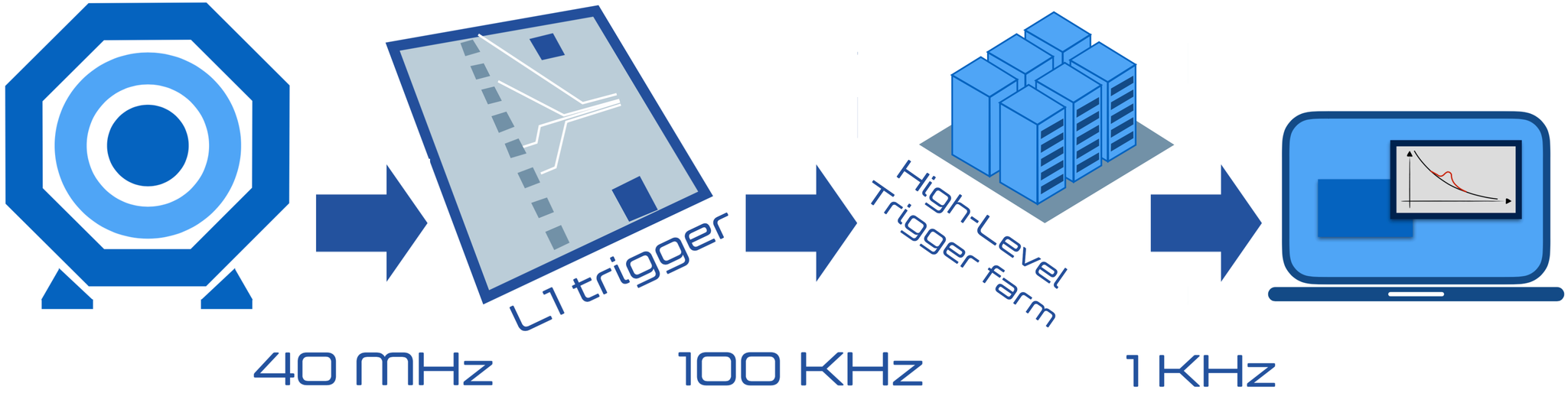
High Level Trigger

- Software-based
- Full detector information
- Simplified reconstruction
- Decision time: ~300 ms

Tier0

- Software-based
- Prompt, full reconstruction
- Processing time: $O(1\ \mathrm{s})$
- At Tier1 within 24−48 hrs

B physics allocated ~15% of 1 kHz rate



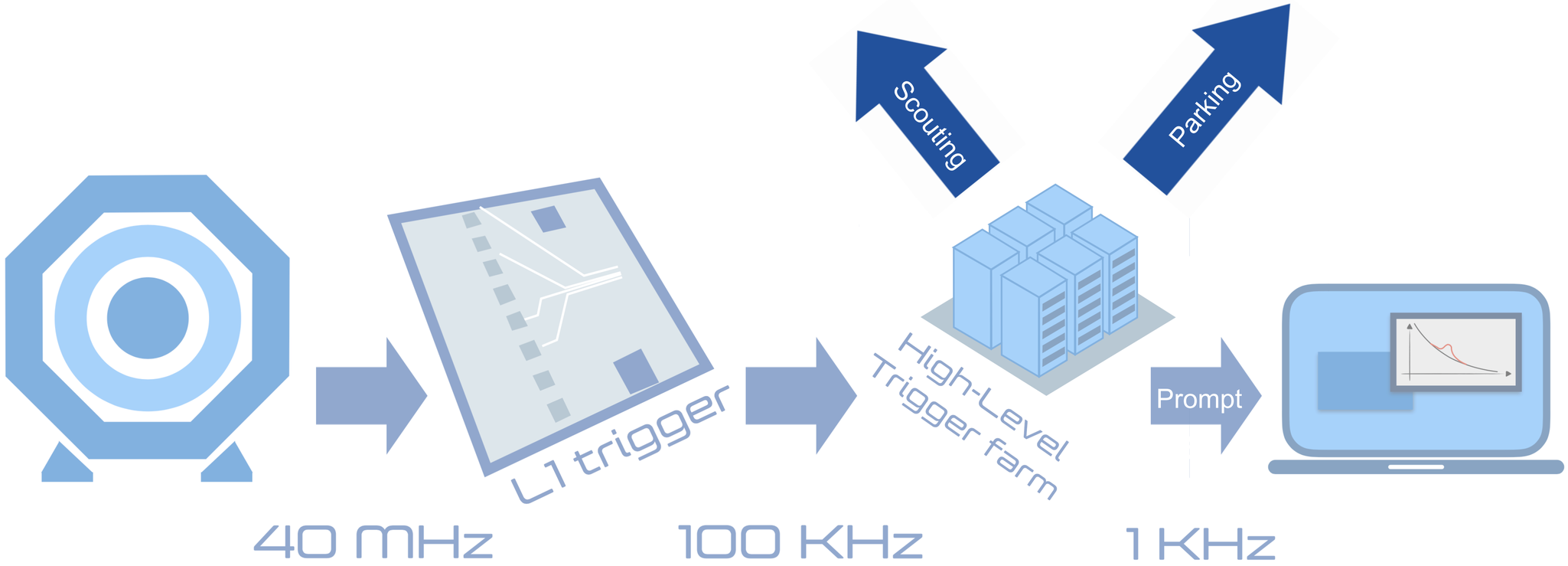
Alternative data streams: scouting and parking

Scouting

- HLT reconstruction
- No reprocessing
- Event size: ~0.01 MB
- ×5−10 more events

Parking

- No prompt reconstruction
- RAW data tier to tape
- Event size: 1 MB/ev
- Process much later (~months)



Trigger strategy and parking

Is it feasible to record 10^{10} B decays with the CMS trigger?

Consider L1 single μ trigger:

$$N(B^0 \to K^* \ell\ell) = f_B \times \mathcal{B}(B^0 \to K^* \ell\ell) \times Rate \times Purity \times t_{LHC}$$

~1000	0.4	$\sim 10^{-7}$	~10 kHz	~0.3	8×10^6 s

Require ~10 kHz at L1 to record ~1000 $B \to K^{(*)}\ell\ell$ decays

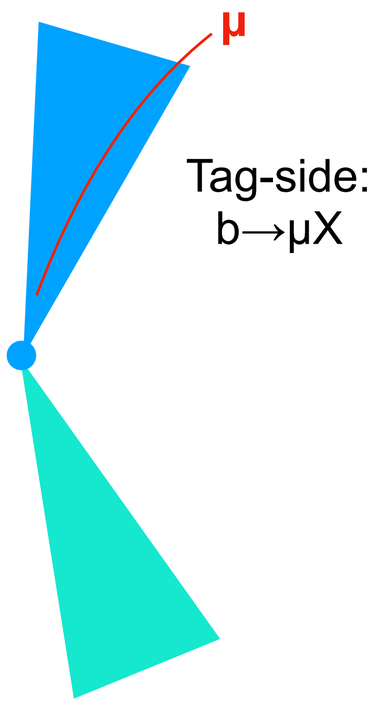
Require HLT to reduce L1 rate and improve L1 purity

Huge rates ⟹ cannot reconstruct promptly @ Tier0 ⟹ **park in RAW format**

Other considerations

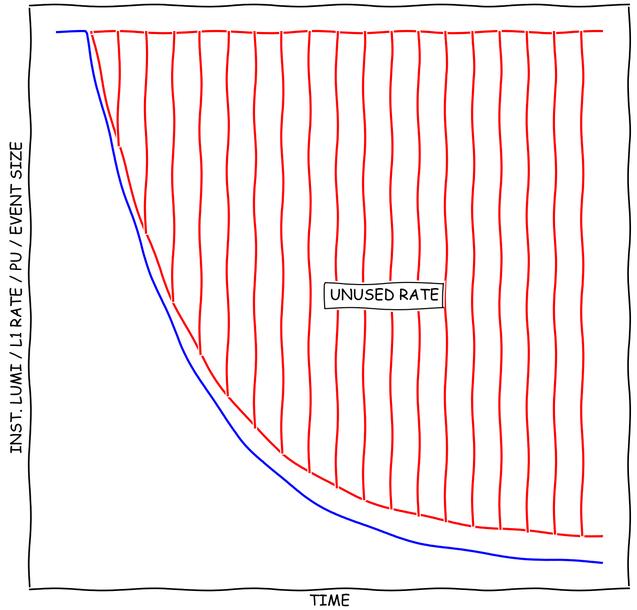
B-enriched trigger

- L1: single μ
- HLT: displaced μ improves purity



Lumi-dependent thresholds

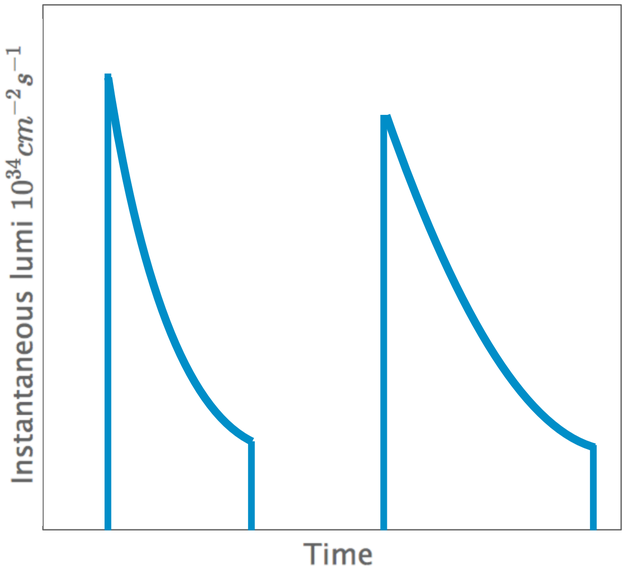
- L1: $\mathcal{L}_{\text{inst}}$ drops ⟹ "free" rate
- HLT: lower pileup ⟹ "free" CPU
- Use resources later in fill
- Tune trigger thresholds vs $\mathcal{L}_{\text{inst}}$ to maximise number of signal-side B decays within acceptance



Bandwidth limitations

- Intra-fill: fill buffers @ P5
- Inter-fill: transfer RAW to tape
- Max. ave. bandwidth: ~2 GB/s

  ⟹ Ave. HLT trigger rate: ~2 kHz



L1 μ trigger

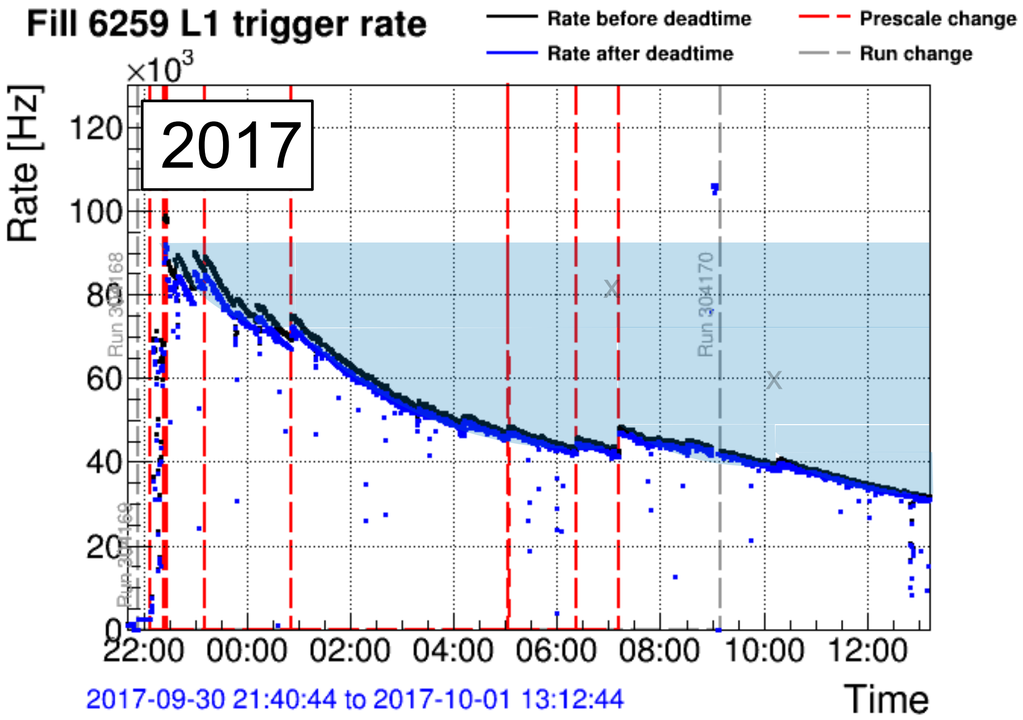
More aggressive evolution of p_T thresholds in 2018

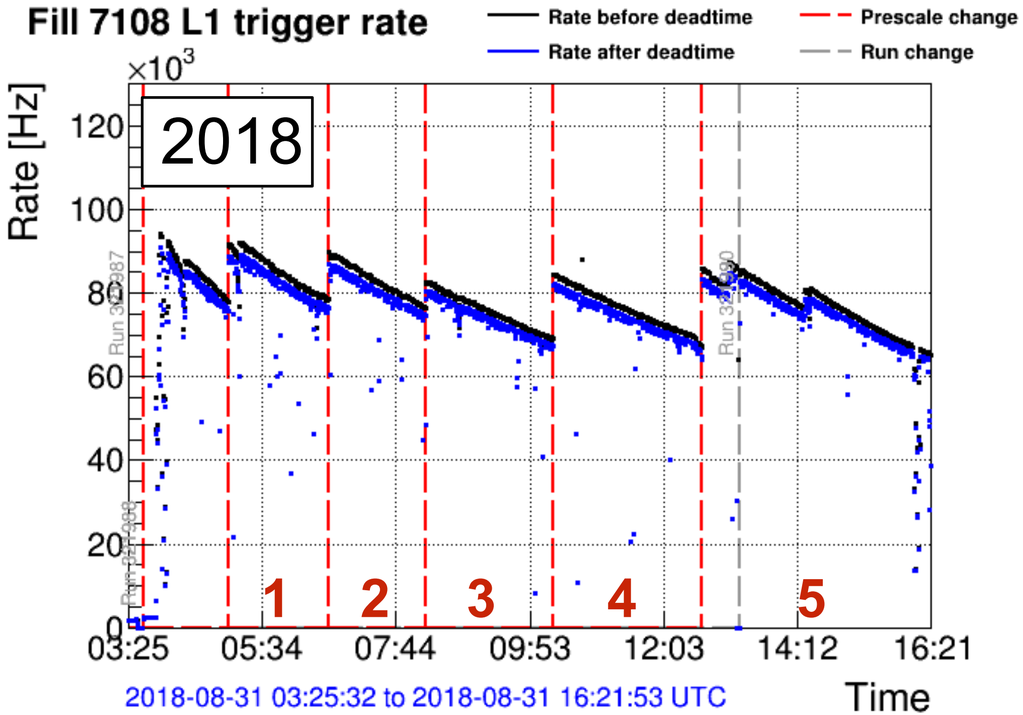
η-restricted trigger:

- Controls rate
- Improves signal-side acceptance

L1 single μ trigger peaks at ~50 kHz (~90 kHz total)

Settings	Peak $\mathcal{L}_{inst}$ [10^{34} cm^{-2} s^{-1}]	L1 seed
1	1.7	Mu12er1p5
2	1.5	Mu10er1p5
3	1.3	Mu9er1p5
4	1.1	Mu8er1p5
5	0.9	Mu7er1p5





High Level Trigger

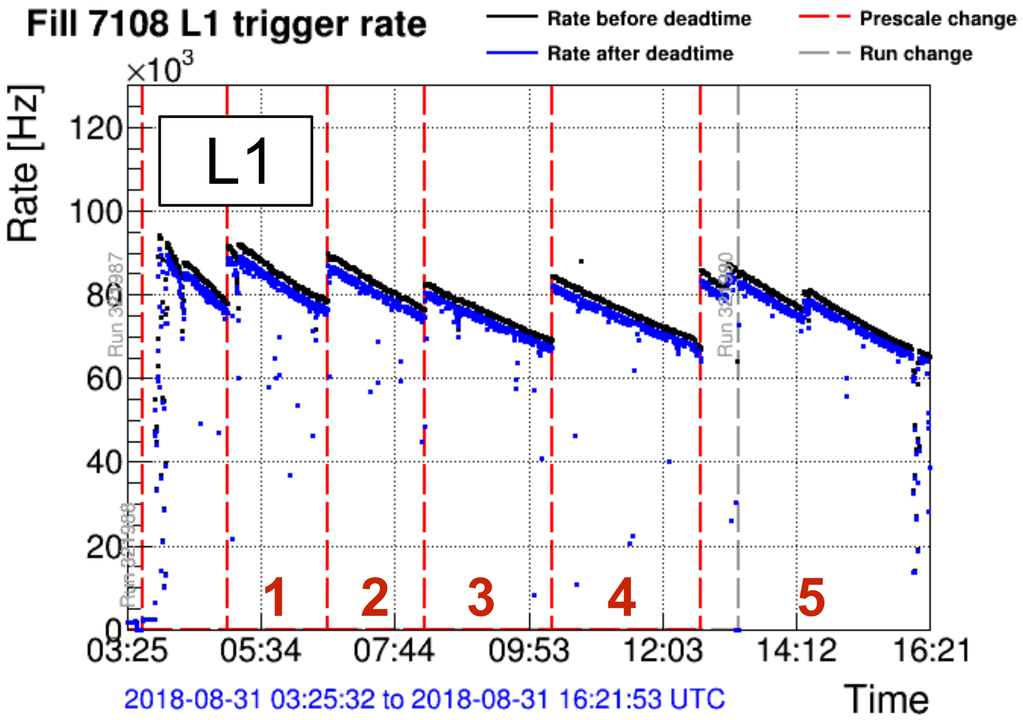
Aggressive reduction in muon p_T threshold

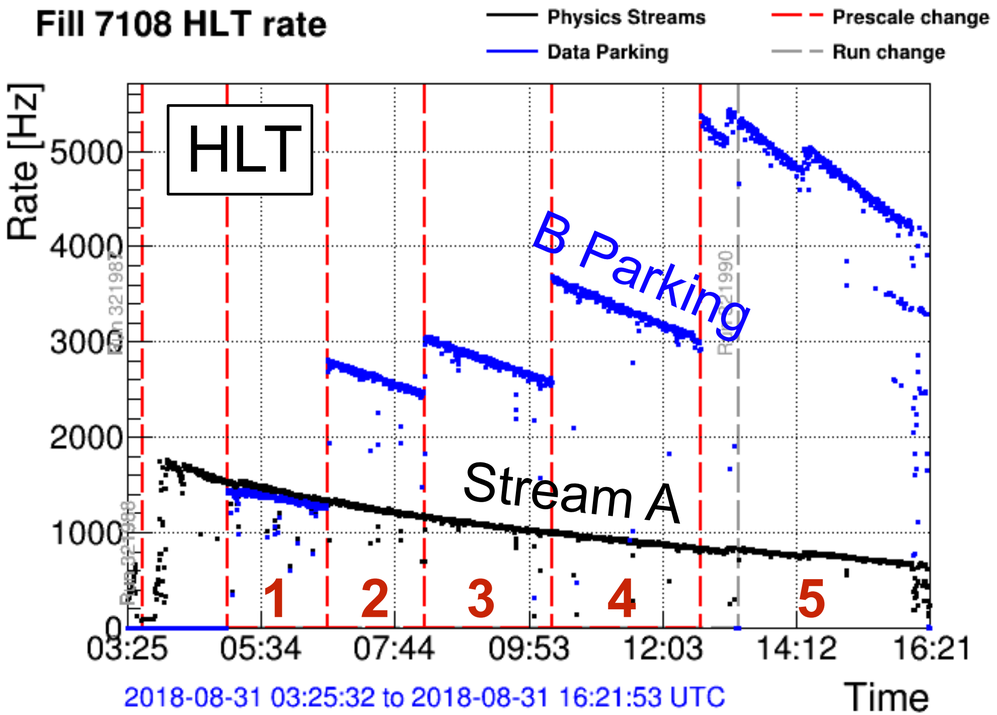
Requirement on muon impact parameter significance

Purities of 60−90% depending on thresholds

Peak rate as high as ~5 kHz

Settings	Peak $\mathcal{L}_{inst}$ [10^{34} cm^{-2} s^{-1}]	L1 seed	HLT path	Peak rate [kHz]
1	**1.7**	Mu12er1p5	Mu12_IP6	1.5
2	**1.5**	Mu10er1p5	Mu9_IP6	2.8
3	**1.3**	Mu9er1p5	Mu9_IP6	3.0
4	**1.1**	Mu8er1p5	Mu9_IP5	3.7
5	**0.9**	Mu7er1p5	Mu7_IP4	5.4





B Parking data set

~12B events recorded during June−Nov 2018 with high purity triggers

$\Longrightarrow$ **~10B unbiased B decays**

Storage: single copy on tape, total size is 7.6 PB

Reconstruction campaign undertaken during LS2 is now nearing completion

Extended "mini" analysis-object data tier (MINIAOD) permanently available for analysis

Signal-enriched ($b \to s\ell\ell$, ℓ=e) skim in RAW and AOD data tiers (temporary shelf life)

Early "pilot" reconstruction campaigns, each a small fraction of full data set

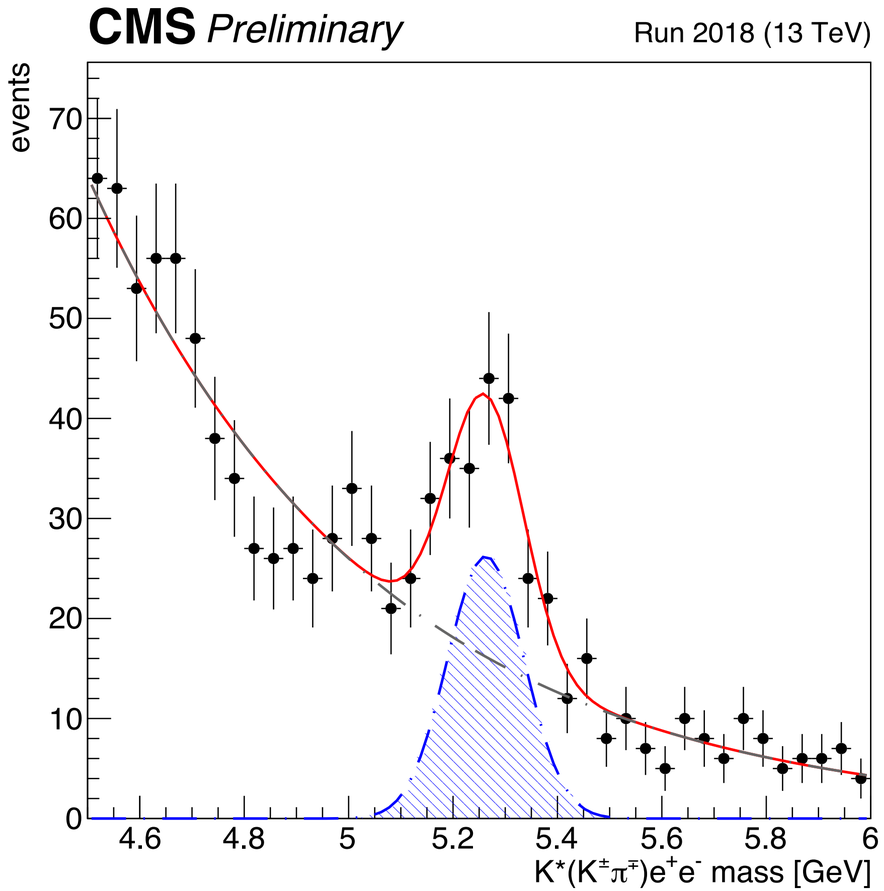
- Trigger purity estimates from MC + checks in data ($B^0 \to D^{*+}\mu\nu$, $D^{*+} \to D^0\pi_{soft}$, $D^0 \to K\pi$ decays)
- Validate data set: reconstruct signal-side B decays
- Commission new low-p_T electron reconstruction

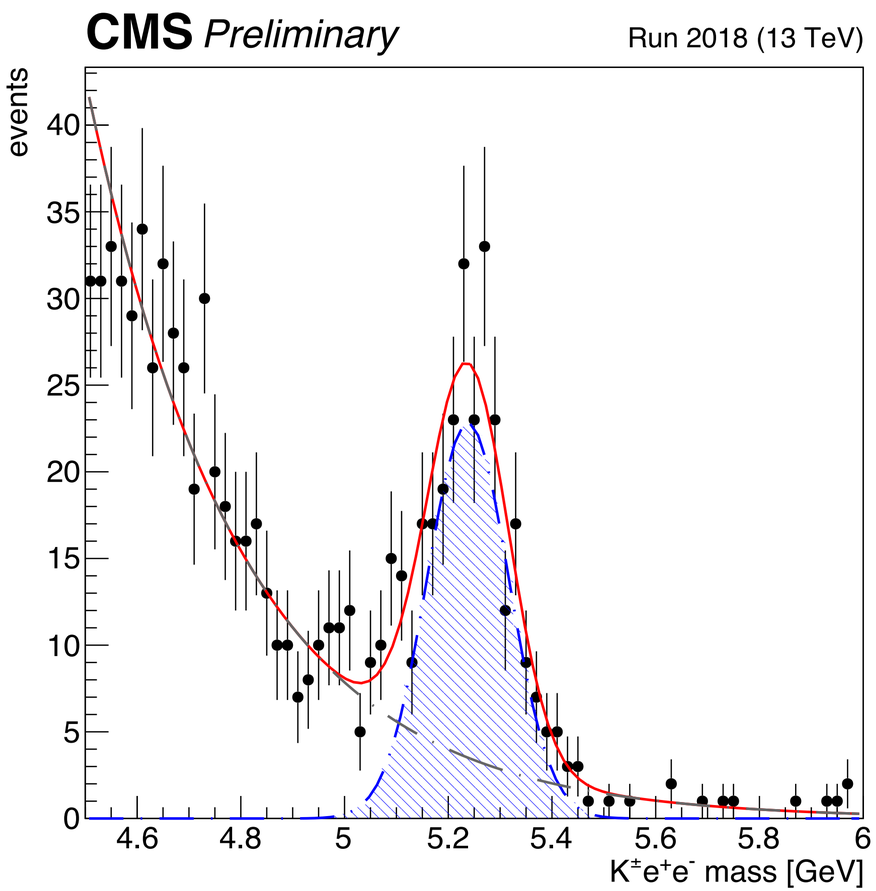
Reconstructed B→K$^{(*)}$J/ψ(ee) decays

First CMS observation of these decays in electron channel!

Electrons are too soft to trigger on; only possible with B Parked data set

Here, using standard electrons from CMS Particle Flow algorithm





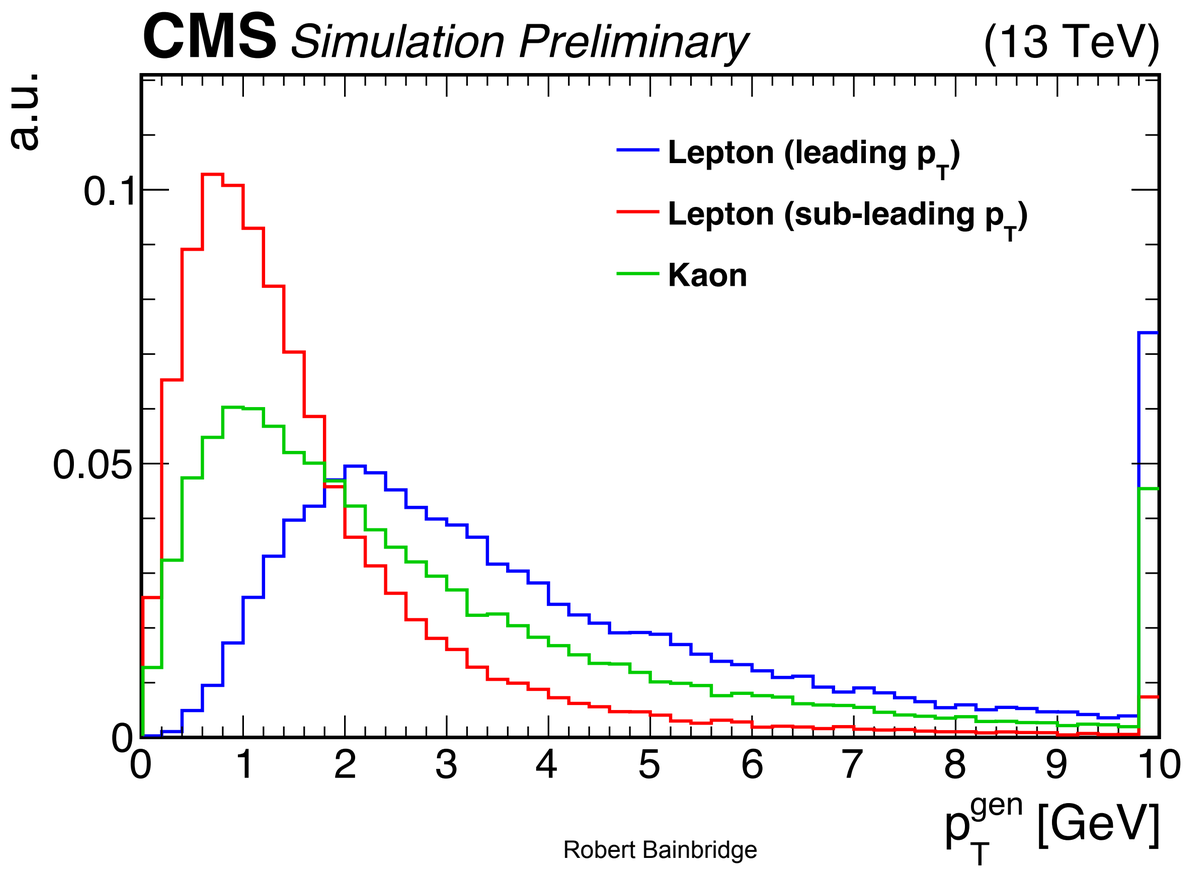
Soft leptons from (signal-side) B decays

Leptons from B→J/ψ(→ee)K decays are very soft, typically p_T ~ 1 GeV

Muons have high reconstruction efficiency but are acceptance limited

Electrons reach ECAL barrel with p_T > 0.7 GeV but ...

... zero efficiency for p_T < 2 GeV from CMS Particle Flow algorithm (tuned for high p_T)



New "low-p_T" electron reconstruction

Tracker-driven approach, independent of Particle Flow algorithm

Bremsstrahlung energy loss of electron modelled by Gaussian mixture

Tracking uses Gaussian Sum Filter (GSF) algorithm

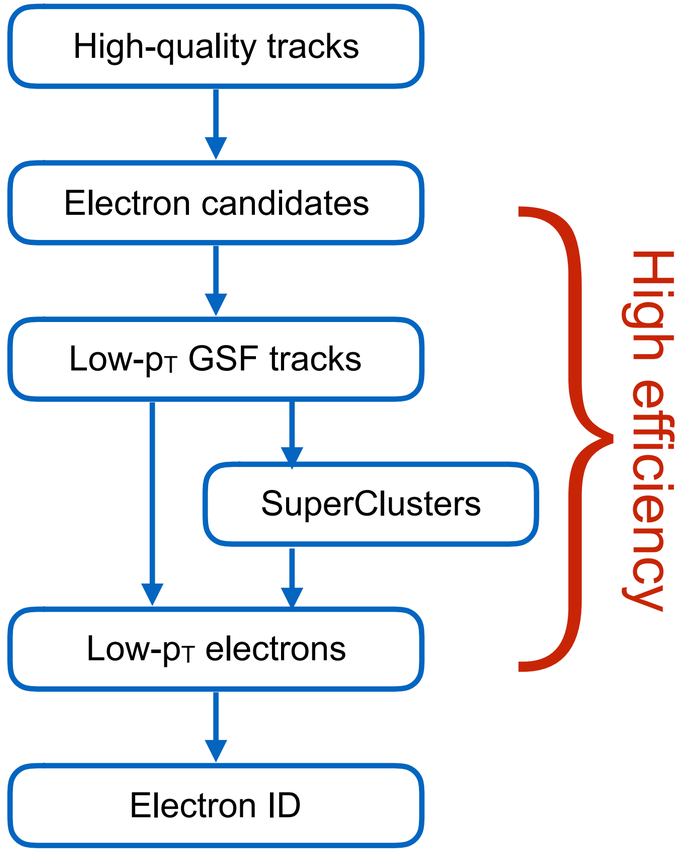
Computationally expensive!

New: electron candidate "seeding" logic

New: SuperCluster algorithm

New: ID, based on MINIAOD quantities

Highly efficient chain, requires sophisticated ID



Reconstruction campaign is a 1-time effort

Seeding logic: use loosest possible working point (bounded by computing)

Low-p_T electrons available in extended MINIAOD data tier

Electron ID is under development

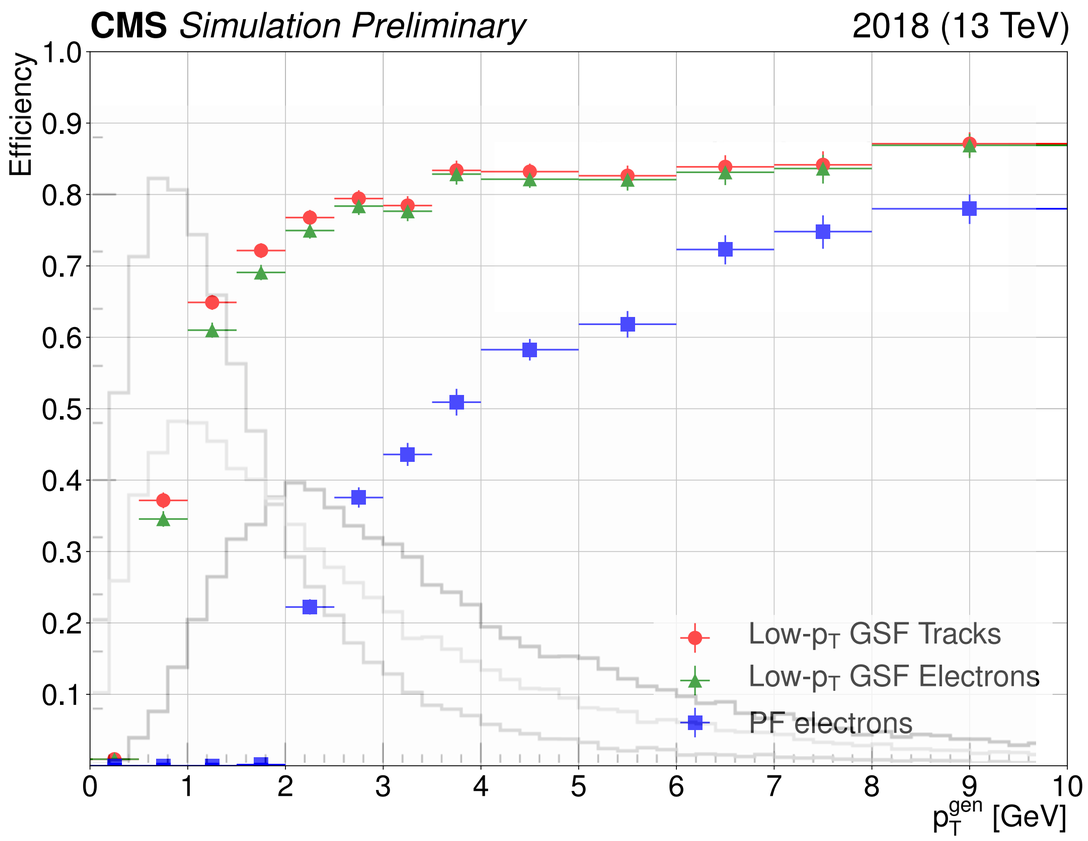
New low-p_T electron efficiencies

Reconstruction uses loosest possible working point in seeding logic ⟹ 10% mistag rate

⟹ 25% increase in CPU and 130% increase in MINIAOD size

Factor 2 gain in electron efficiency w.r.t PF baseline

However, we require a sophisticated ID to control purity - very difficult!



Concluding remarks

CMS now has a sample of ~10B unbiased B decays

Tag-side μ trigger strategy coupled with parking of a 2 GB/s RAW data stream

Reconstruction of the B Parking data set is almost complete

Analysis-level data tier available, including new low p_T electrons

Data set commissioning with "pilot" reconstruction campaigns

Huge effort: PC, RunCoord, Trigger, DAQ, O&C, T0 Ops, PPD, EGamma, ...

Parked data set is a huge asset to CMS, with rich possibilities

Delivered without significant impact on the CMS core physics programme

We are beginning to study it in earnest ...