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 ($\mu\mu$) triggers

FCNC transitions such as $b \rightarrow s \ell\ell$ are ideal probes for new physics
Anomalies in $R_K$ and $R_{K^*}$ observables that probe lepton flavour universality
Theoretically clean, but experimentally challenging!

e.g. CMS cannot trigger directly on electron channels

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

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

Is this feasible with the CMS trigger?
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: $O(1 \text{ s})$
- 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 $\mu$ trigger:

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

- $\sim 1000 \quad 0.4 \quad \sim 10^{-7} \quad \sim 10 \text{ kHz} \quad \sim 0.3 \quad 8 \times 10^6 \text{ s}$

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

Require HLT to reduce L1 rate and improve L1 purity

Huge rates $\Rightarrow$ cannot reconstruct promptly @ Tier0 $\Rightarrow$ **park in RAW format**
Other considerations

**B-enriched trigger**
- L1: single $\mu$
- HLT: displaced $\mu$ improves purity

**Lumi-dependent thresholds**
- L1: $\mathcal{L}_{\text{inst}}$ drops $\Rightarrow$ "free" rate
- HLT: lower pileup $\Rightarrow$ "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: $\sim 2$ GB/s
  $\Rightarrow$ Ave. HLT trigger rate: $\sim 2$ kHz

---

Tag-side: $b \rightarrow \mu X$
L1 $\mu$ trigger

More aggressive evolution of $p_T$ thresholds in 2018

$\eta$-restricted trigger:
- Controls rate
- Improves signal-side acceptance

L1 single $\mu$ trigger peaks at ~50 kHz (~90 kHz total)

<table>
<thead>
<tr>
<th>Settings</th>
<th>Peak $\mathcal{L}\text{inst}$ [$10^{34}$ cm$^{-2}$ s$^{-1}$]</th>
<th>L1 seed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7</td>
<td>Mu12er1p5</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>Mu10er1p5</td>
</tr>
<tr>
<td>3</td>
<td>1.3</td>
<td>Mu9er1p5</td>
</tr>
<tr>
<td>4</td>
<td>1.1</td>
<td>Mu8er1p5</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>Mu7er1p5</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Settings</th>
<th>Peak $\mathcal{L}_{\text{inst}}$ [10$^{34}$ cm$^{-2}$ s$^{-1}$]</th>
<th>L1 seed</th>
<th>HLT path</th>
<th>Peak rate [kHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7</td>
<td>Mu12er1p5</td>
<td>Mu12_IP6</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>Mu10er1p5</td>
<td>Mu9_IP6</td>
<td>2.8</td>
</tr>
<tr>
<td>3</td>
<td>1.3</td>
<td>Mu9er1p5</td>
<td>Mu9_IP6</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>1.1</td>
<td>Mu8er1p5</td>
<td>Mu9_IP5</td>
<td>3.0</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>Mu7er1p5</td>
<td>Mu7_IP4</td>
<td>5.4</td>
</tr>
</tbody>
</table>
B Parking data set

~12B events recorded during June−Nov 2018 with high purity triggers
\[ \Rightarrow \sim 10B \text{ unbiased } B \text{ decays} \]

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 ($b \rightarrow s \ell \ell, \ell = e$) 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 ($B^0 \rightarrow D^+\mu\nu, D^+ \rightarrow D^0\pi_{\text{soft}}, D^0 \rightarrow K\pi$ decays)
- Validate data set: reconstruct signal-side B decays
- Commission new low-$p_T$ electron reconstruction
Reconstructed $B \rightarrow K^{(*)} J/\psi (ee)$ 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

---

**Figure 1:**

- **Left:** Distribution of $K^*(K^0\pi^-)e^+e^-$ mass for Run 2018 (13 TeV) data.
- **Right:** Distribution of $K^0\pi^-\pi^+$ mass for Run 2018 (13 TeV) data.

**Legend:**

- **CMS Preliminary**
- **Run 2018 (13 TeV)**

---

Computing in HEP, Adelaide, 4th-8th Nov. 2019

Robert Bainbridge
Soft leptons from (signal-side) B decays

Leptons from $B \rightarrow J/\psi(\rightarrow ee)K$ decays are very soft, typically $p_T \sim 1$ 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 $\Rightarrow$ 10% mistag rate
$\Rightarrow$ 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!

\[ p_{\text{gen}} \text{ [GeV]} \]

Efficiency

CMS Simulation Preliminary 2018 (13 TeV)
Concluding remarks

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

Tag-side $\mu$ 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...