Recording and reconstructing 10 billion unbiased b hadron decays in CMS

CMS Collaboration

Abstract

The CMS experiment has recorded a sample of 10 billion events containing the unbiased decays of b hadrons. The accumulation, processing, and validation of this data set were delivered without significant impact on the core physics programme of CMS. Events were recorded with peak trigger rates in excess of 50 and 5 kHz at L1 and HLT, respectively, and written (parked) to tape at an average rate of 2 GB/s. New algorithms have been developed to reconstruct and identify electrons with high efficiency at transverse momenta as low as 0.5 GeV. These algorithms were validated using a pilot reconstruction campaign. The entire data set has now been reconstructed.
Recording and reconstructing 10 billion unbiased b hadron decays in CMS

CMS Collaboration

cms-phys-conveners-BPH@cern.ch
Trigger: CMS L1 rates

Rate of the CMS L1 trigger, as a function of time, during the LHC fill 6259. The time interval covers approximately 15 hours in September 2017.

A change in the CMS run number is indicated by a vertical grey dashed line. A change in the CMS prescale column is indicated by a vertical red dashed line.

The L1 trigger rate is initially close to the system design limit of 100 kHz, and reduces gradually over time as the instantaneous luminosity delivered to CMS falls during the LHC fill.
Trigger: CMS L1 rates with B parking

Rate of the CMS L1 trigger, as a function of time, during the LHC fill 7108. The time interval covers approximately 13 hours in August 2018.

A change in the CMS run number is indicated by a vertical grey dashed line. A change in the CMS prescale column is indicated by a vertical red dashed line.

The L1 trigger menu is tuned for the B Parking data stream to deliver close to the system design limit of 100 kHz throughout the fill.
Trigger: CMS HLT rates with B parking

Rate of the CMS High Level Trigger (HLT), as a function of time, during the LHC fill 7108. The time interval covers approximately 13 hours in August 2018.

A change in the CMS run number is indicated by a vertical grey dashed line. A change in the CMS prescale column is indicated by a vertical red dashed line.

The rate for the CMS physics streams (black curve) falls from ~2 kHz during the fill.

The rate for the B Parking stream (blue curve) increases in steps at changes in the prescale column during a CMS run, reaching as high as ~5 kHz.

The L1 and HLT trigger menus are tuned to deliver an average throughput of ~2 GB/s for the B Parking data stream over the timescale of 1 week.
Trigger strategy

During June–Nov 2018, approximately 12 billion events were recorded with a trigger logic that requires the presence of a single, displaced muon. The sample comprises b¯b events with high purity. The muon candidate responsible for the positive trigger decision originates from the "tag-side" b hadron that undergoes a b→µX decay. The "signal-side" b hadron decays naturally as it is not biased by the trigger requirements.

The L1 µ trigger logic requires |η| < 1.5 and is subject to the p_T thresholds summarised in the table below. The HLT trigger logic also requires thresholds to be met on the p_T and IP_{sig} (track impact parameter significance), which improves the trigger purity.

The thresholds evolve during a fill, as the instantaneous luminosity (L_{inst}) falls, to maximise number of signal-side b hadron decays within acceptance. The trigger purity is determined from simulation to be in the range 60–90% depending on the thresholds.

<table>
<thead>
<tr>
<th>Settings</th>
<th>Peak L_{inst} [10^{34} cm^{-2} s^{-1}]</th>
<th>L1 µ p_T threshold [GeV]</th>
<th>HLT µ p_T threshold [GeV]</th>
<th>HLT µ IP_{sig} threshold</th>
<th>Trigger purity [%]</th>
<th>Peak rate [kHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7</td>
<td>12</td>
<td>12</td>
<td>6</td>
<td>92</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>87</td>
<td>2.8</td>
</tr>
<tr>
<td>3</td>
<td>1.3</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>86</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>1.1</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>83</td>
<td>3.7</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>59</td>
<td>5.4</td>
</tr>
</tbody>
</table>
Trigger: \( b \) hadron purity

\[
N(b \to \mu X) = \frac{1}{F_{\text{corr}}} \frac{N(B^0 \to D^{*+} \mu \nu)}{\alpha(D^{*+}) \times \epsilon(D^{*+}) \times \mathcal{B}(D^{*+} \to D^0(\to K\pi)\pi)}
\]

\[
P_b = \frac{N(b \to \mu X)}{N(\mu)}
\]

The purity determination for the B Parking triggers relies on the \( B^0 \to D^{*+} \mu \nu \to (D^0_{\pi \text{soft}}) \mu \nu \to (K\pi_{\pi \text{soft}}) \mu \nu \) decay.

The figure shows the reconstructed mass difference for \( D^{*+} \) and \( D^0 \). Candidate \( D^{*+} \) decays are selected by requiring two tracks from the \( D^0 \to K\pi \) decay that satisfy \( p_T^{\text{min}} > 3 \) and \( p_T^{\text{max}} > 5 \) GeV. The vertex is subject to displacement and quality criteria. The product of the \( K \) and \( \mu \) charges is required to be +1 (right sign) or -1 (wrong sign).

An unbinned likelihood fit is performed with a second-order polynomial and Gaussian PDF for the background and signal, respectively. The number of \( D^{*+} \) candidates are determined from the fitted mass value ±2\( \sigma \).
Trigger: dimuon invariant mass spectrum

Dimuon invariant mass distribution, obtained from a sample of opposite-sign muons originating from a common vertex found within a fraction of the B Parking data set. Both muons are required to be well reconstructed and satisfy a set of identification criteria. The $J/\psi$, $\psi(2S)$, and $Z$ peaks are clearly visible. The inset distributions show the $\rho/\omega$, $\varphi$, and $\Upsilon$ peaks.
Trigger: displaced dimuon invariant mass spectrum

Dimuon invariant mass distribution, obtained from a sample of opposite-sign muons originating from a common vertex found within a fraction of the B Parking data set. Both muons are required to be well reconstructed, satisfy a set of identification criteria, and have an impact parameter significance greater than 3. The J/$\psi$ and $\psi(2S)$ peaks are clearly visible. The large increase in the background counts below $\sim$5 GeV is due to nonprompt tracks from B meson decays.
Commissioning: $B \to J/\psi(\mu\mu)K^*(\pi K)$ decays

Invariant mass distribution obtained from candidate $B \to J/\psi(\mu\mu)K^*(K\pi)$ decays, based on a fraction of the B Parking data set.

Events are subject to requirements on: the $p_T$ of the daughters (K, $\pi$ and $\mu$); the reconstructed $\mu\mu$ and $K\pi$ masses; and properties of the reconstructed vertices. An unbinned likelihood fit is performed with an exponential and Gaussian PDF for the background and signal, respectively.

This distribution was used as part of the commissioning campaign, for which $O(1\%)$ of the B Parking data set was reconstructed with priority early in the 2018 data taking run.
Commissioning: $B \rightarrow J/\psi(\text{ee})K^*(\pi K)$ decays

Invariant mass distribution obtained from candidate $B \rightarrow J/\psi(\text{ee})K^*(K\pi)$ decays, based on a fraction of the B Parking data set.

Events are subject to requirements on:
- the $p_T$ of the daughters (K, π and e);
- the reconstructed ee and Kπ masses;
- and properties of the reconstructed vertices. An unbinned likelihood fit is performed with an exponential and Gaussian PDF for the background and signal, respectively.

This distribution was used as part of the commissioning campaign, for which $O(1\%)$ of the B Parking data set was reconstructed with priority early in the 2018 data taking run.

This distribution provides a first observation of the resonant $B \rightarrow J/\psi(\text{ee})K^*(K\pi)$ decay in CMS.
Commissioning: $B \to J/\psi(\text{ee})K$ decays

Invariant mass distribution obtained from candidate $B \to J/\psi(\text{ee})K$ decays, based on a fraction of the B Parking data set.

Events are subject to requirements on: the $p_T$ of the daughters (K and e); the reconstructed ee mass; and properties of the reconstructed vertex. An unbinned likelihood fit is performed with an exponential and Gaussian PDF for the background and signal, respectively.

This distribution was used as part of the commissioning campaign, for which $O(1\%)$ of the B Parking data set was reconstructed with priority early in the 2018 data taking run.

This distribution provides a first observation of the resonant $B \to J/\psi(\text{ee})K$ decay in CMS
Reconstruction: $p_T$ spectra of $B \to K \ell\ell$ final state

Generator-level $p_T$ distributions of the daughter particles from probe-side $B \to J/\psi(\to \ell\ell)K$ decays.

The lepton $p_T$ distributions are very soft, with most probable values of 2 and 1 GeV for the leading and sub-leading leptons, respectively.

The lepton $p_T$ distributions are shown following requirements on the $(p_T, \eta)$ acceptance for the muon from the tag-side $b$ hadron decay to mimic the trigger conditions during data taking.

The lepton $p_T$ spectra imply challenges related to muon acceptance and electron reconstruction efficiency.
Reconstruction: low-$p_T$ electron candidates

A custom low-$p_T$ electron reconstruction has been developed for the B Parking data set.

The reconstruction of electrons relies heavily on GSF tracking, which is computationally expensive. The GSF tracking is seeded by a more computationally efficient logic that identifies electron candidates.

A new seeding logic implements two independent boosted decision trees (BDT) that provide discrimination based on a "kinematically agnostic" BDT (green curve) that exploits tracking and calorimeter information; and a model-dependent "kinematically aware" BDT (blue curve) that also utilises the $p_T$, $\eta$, and the track impact parameter of an electron candidate.

A loose "seeding working point" is defined for each BDT that yields a 10% mistag rate while providing a factor $\sim 2$ gain in efficiency over that obtained from the "baseline seeding" logic (red circle). A logical OR of the loose working points from the two BDTs is used to identify electron candidates and seed the reconstruction of low-$p_T$ electrons.
Reconstruction: efficiencies for low-$p_T$ electrons

The figure shows the reconstruction efficiency for PF electrons (blue squares) as a function of the generator-level electron $p_T$. No identification criteria are applied to the PF electrons.

The figure also shows the efficiencies obtained for low-$p_T$ GSF tracks (red circles) and electrons (green triangles) that are reconstructed from electron candidates of the seeding logic described in the previous slide, which uses a logical OR of the loose seeding working points (10% mistag rate) for the two BDTs. No identification criteria are applied to the low-$p_T$ electrons.
Reconstruction: low-p_T electrons from photon conversions

Photons converting to an $e^\pm e^\mp$ pair provide a powerful control sample in data to study the performance of the electron identification algorithm.

All low-p_T GSF tracks are considered by a dedicated algorithm that identifies electron pairs consistent with a photon conversion.

The figure shows the vertex position of conversion candidates in the transverse plane for the region $|\eta| < 1$. The material of the CMS tracker and pixel barrel subdetectors can be clearly seen.

The control sample can be used to train and/or validate the performance of machined-learned algorithms used to identify electrons.
Reconstruction: low-$p_T$ electrons from photon conversions

A zoomed version of the figure shown in the previous slide. The structure of the inner layer of the CMS pixel barrel subdetector is clearly visible. The converted photons originate from the beam spot (red diamond), which is offset from the geometrical centre of the CMS detector.

Electrons from the photon conversion candidates, with a transverse displacement of 2-4 cm, provide a suitable control sample for electrons produced from the decays of $b$ hadrons.
Modes of unbiased B hadron decays on tape

The table indicates the number of unbiased decays of different types of B hadrons recorded to tape in 2018 ($N_{2018}$). The fractions of B hadron type that are produced ($f_B$) and their branching fraction ($\mathcal{B}$) are also indicated.

<table>
<thead>
<tr>
<th>Mode</th>
<th>$N_{2018}$</th>
<th>$f_B$</th>
<th>$\mathcal{B}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic b hadrons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_d^0$</td>
<td>$4.0 \times 10^9$</td>
<td>0.4</td>
<td>1.0</td>
</tr>
<tr>
<td>$B^\pm$</td>
<td>$4.0 \times 10^9$</td>
<td>0.4</td>
<td>1.0</td>
</tr>
<tr>
<td>$B_s$</td>
<td>$1.2 \times 10^9$</td>
<td>0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>b baryons</td>
<td>$1.2 \times 10^9$</td>
<td>0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>$B_c$</td>
<td>$1.0 \times 10^7$</td>
<td>0.001</td>
<td>1.0</td>
</tr>
<tr>
<td>Total</td>
<td>$1.0 \times 10^{10}$</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Events for $R_K$ and $R_{K^*}$ analyses**

<table>
<thead>
<tr>
<th>Decay</th>
<th>Events</th>
<th>$f_B$</th>
<th>$\mathcal{B}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0 \rightarrow K^*\ell^+\ell^-$</td>
<td>2600</td>
<td>0.4</td>
<td>$6.6 \times 10^{-7}$</td>
</tr>
<tr>
<td>$B^\pm \rightarrow K^{\pm}\ell^+\ell^-$</td>
<td>1800</td>
<td>0.4</td>
<td>$4.5 \times 10^{-7}$</td>
</tr>
</tbody>
</table>