# Production of photons and b jet pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV

Master Thesis by Till Hoffmann*
Under the direction of Prof. Allan Clark

Semestre d'automne 2007

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

Département de Physique Nucléaire et Corpusculaire
Université de Genève

*Till.Hoffmann@cern.ch

**Abstract**

This Master's thesis presents the measurement of the production cross section of the $(b\bar{b} + \gamma)$ final state in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV at the Fermilab Tevatron. The data are collected using a special trigger that uses the Silicon Vertex Trigger (SVT), a device capable of selecting heavy flavor decays online. The efficiency of this selection is calculated directly from data. Two jets are required to be tagged with the SECVTX algorithm, leading to a $b\bar{b}$ purity of the sample of approximately 80 %.

In this data sample, corresponding to an integrated luminosity of approximately $1.1 fb^{-1}$, the cross section is measured to be

$$8.60 \pm 1.07(stat.)^{+1.44}_{-1.56}(syst.) \quad pb$$

This result is approximately 30% higher than a leading-order Monte Carlo calculation. No next-to-leading order calculation was available at the time of writing.

# Table of Contents

# List of Figures

# List of Tables

# 1 Introduction

Measuring the production of photons associated with b-jet pairs is extremely interesting due to the many theoretical handles that can lead to an understanding of the production mechanisms. From the experimental viewpoint, the main difficulties lie in the identification of two b-jets and a photon, in the presence of a much larger background from light- and charm-quark jets on one hand, and from neutral pions decaying into two photons on the other hand.

The ($b\bar{b}$ + photon) signature is produced at lowest order in QCD by two dominant mechanisms, referred to as Bremsstrahlung and gluon splitting (Fig. 1). In the first case, the process consists in the production of a $b\bar{b}$ pair, with one b-quark radiating a photon. This process can be initiated either by gluon-gluon or quark-quark initial states. The second process, always initiated by a same-flavor quark-antiquark pair, has a photon-gluon final state, with the gluon splitting in a $b\bar{b}$ pair.


Figure 1: Dominant Standard Model production mechanisms for $b\bar{b}\gamma$ final states. On the left we have the *Bremsstrahlung* process and on the right gluon splitting.

This final state is useful as a test of perturbative QCD and also in the search for new physics. The decay of supersymmetric (SUSY) particles can result in photons and b-quarks, so the $b\bar{b}\gamma$ signature would be an important QCD background. Technicolor-inspired theories (see [1]) predict the existence of high-mass techni-mesons such as the $\omega_T$, that would produce an excess in this exact final state, following the decay path

$$\omega_T \quad \rightarrow \quad \gamma\pi_T \quad \rightarrow \quad \gamma b\bar{b}$$

An analysis [2] of the signature (b + photon) has already been performed at CDF using an integrated luminosity of 85 $pb^{-1}$. However the statistics at that time were insufficient to perform a meaningful $b\bar{b}\gamma$ measurement. Later, (b+photon) production was again measured using a trigger selecting a high-energy photon[3] using an integrated luminosity of about 300 $pb^{-1}$. A third analysis [4] was performed using a dataset collected using a dedicated trigger path requiring at the trigger level both a photon and a displaced track with large impact parameter. The displaced track requirement ensures a sample with high b-jet content and ensures higher statistics by lowering the trigger threshold on the photon energy.

The current analysis is an extension of the previous measurement of $(b + \gamma)$ using the dedicated trigger path. In the following chapters a first measurement of the $b\bar{b}\gamma$ cross section using a data sample corresponding to approximately 1 $fb^{-1}$ is presented.

# 2 Theoretical Motivation

## 2.1 The Standard Model of particle physics

The Standard Model of particle physics (SM) aims to explain the properties of fundamental particles and their interactions. Matter consists of fermions while the forces are transmitted with the gauge bosons. Fermions are classified in three generations that have the same physical properties except for their masses. For a summary of all the particles discovered up to date, see Tables 1 and 2.

| **Generation** | I | II | III | Q | B | L |
|---|---|---|---|---|---|---|
| Quarks | Up (u) $1.5 - 3MeV$ | Charm (c) $1.25 \pm 0.09GeV$ | Top (t) $170.9 \pm 1.9GeV$ | $+\frac{2}{3}$ | $\frac{1}{3}$ | 0 |
| | Down (d) $3 - 7MeV$ | Strange (s) $95 \pm 25MeV$ | **Bottom (b)** $4.20 \pm 0.07GeV$ | $-\frac{1}{3}$ | $\frac{1}{3}$ | 0 |
| Leptons | Electron (e) $511KeV$ | Muon ($\mu$) $106MeV$ | Tau ($\tau$) $1.78GeV$ | $-1$ | 0 | 1 |
| | $\nu_e$ $< 2eV$ | $\nu_\mu$ $< 0.19MeV$ | $\nu_\tau$ $< 18.2MeV$ | 0 | 0 | 1 |

Table 1: Fundamental spin $\frac{1}{2}$ fermions in the Standard Model with their main quantum numbers. Q is the electrical charge, B the baryon number and L the lepton number. Masses are taken from the Particle Data Group values of 2006 [5], except for the top quark mass which is the most recent TevEWWG world average [6].

| **Particle** | Interaction | Mass | Spin |
|---|---|---|---|
| Gluon (g) | Strong (color) $SU(3)_C$ | 0 | 1 |
| **Photon ($\gamma$)** | Electromagnetic $U(1)_{em}$ | 0 | 1 |
| Z | Weak (neutral) | $91.1876 \pm 0.0021$ GeV | 1 |
| $W^\pm$ | Weak (charged) | $80.400 \pm 0.024$ GeV | 1 |

Table 2: Fundamental gauge bosons in the Standard Model. Masses are the Particle Data Group 2006 values [5], except for the W mass which uses recent results with Run II data [7].

The SM consists of a set of gauge theories describing the strong, weak and electromagnetic forces, while gravity is not included in the SM. In the framework of quantum

field theory, a gauge theory is based on a continuous symmetry group that describes the transformations leaving the Lagrangian invariant. Historically, symmetries were associated to conservation laws by Noether's theorem [8]. However promoting a symmetry to be local[1] lets the force carrier particles emerge in the theory in a natural way. The gauge group of the SM is $SU(3) \otimes SU(2) \otimes U(1)$.

In the SM, leptons only interact through electroweak interactions ($SU(2) \otimes U(1)$, with exchange bosons $W^{\pm}, Z$ and $\gamma$) and are not sensitive to the strong interaction. The quarks carry a $SU(3)$ color charge in addition and interact with gluons.

In order to give mass to the particles in a gauge theory, the gauge invariance must be broken. The Higgs mechanism is invoked to explain this spontaneous symmetry breaking (SSB), in which case the state of minimum energy (vacuum) is not invariant under gauge group rotations but the Lagrangian is invariant. This SSB requires the introduction of a new scalar *Higgs* field. The interaction of this field with the other fundamental particles allows them to become massive while the theory is still renormalizable. In addition a new particle, the Higgs boson appears, which has not yet been seen experimentally.

The SM has been tested experimentally to an impressive level of precision, using in particular data from experiments at the Fermilab Tevatron (proton-antiproton) and the CERN LEP (electron-positron) colliders. However despite these quantitative successes, it cannot be the ultimate theory. One of the fundamental interactions of Nature, gravity, is not described in the SM. Furthermore, the numerous free parameters of the theory (for example the masses of all fermions) cannot be derived from it and must be measured experimentally. In particular it is not understood why the three generations have different mass and why the leptons are so much lighter.

There is also the *hierarchy* problem between the intensities of the electroweak and gravitational forces. The Higgs mass being sensitive to the scale of any new physics beyond the SM, an incredible amount of fine tuning is necessary to keep it at values close to the electroweak scale. This has generated a lot of research in extensions of the Standard Model that would be revealed at higher energies.

## 2.2 Quantum Chromodynamics (QCD)

QCD is the gauge theory of the strong interaction. It describes how quarks, that carry a strong charge interact via the exchange of gluons. Color is the strong charge and is conserved in all interactions. Contrary to the case of weak interactions, the flavor[2] is also conserved in strong interactions. Note that leptons are not sensitive to the strong interaction.

The fact that the symmetry group for color, $SU(3)$, is non-abelian leads to one of the most important properties of QCD, namely the fact that the gluons themselves carry

[1] In a local symmetry, the transformation can even be dependent on the location in (momentum) space.

[2] The 6 quarks are classified according to their weak isospin (up-type and down-type) and the generation they belong to. Flavor is the quantum number assigned to each type.

color charge (actually one color and one anti-color) and can interact with themselves. This clearly distinguishes QCD from another theory with a massless mediator, quantum electrodynamics (QED) which is the theory describing electromagnetic interactions. In that case the photon does not carry electromagnetic charge. The self-coupling of gluons brings a potential between a quark pair of the form

$$V_s(r) = -\frac{4}{3}\frac{\alpha_s}{r} + k \cdot r$$

This property, that the force between two particles grows with their separation leads to the *confinement* of quarks into color-singlet hadrons. Within the context of QCD, quarks must be confined within a short range determined by $V_s(r)$, and experimentally free quarks have been observed. When one attempts to separate two quarks, the energy in the gluons between grows, allowing the creation of a new quark-antiquark pair when a separation of about $10^{-15}m$ is reached. This process is called *fragmentation*.

Another key property of QCD is its coupling constant $\alpha_s$, whose value is close to unity at low values of the momentum transfer ($Q^2$). This makes a perturbative calculation impossible and indeed, very little is known about the inner working of nucleons such as the proton, which will be of interest in the next chapter. Fortunately, this coupling constant evolves to weaker values as the momentum transfer increases between two particles. Because gluons carry a color and an anticolor charge, virtual gluon pairs polarize the vacuum in such a way that the effective charge of a quark augments with growing distance. So when the momentum transfer takes large values and the distance decreases, the strength of the interaction $\alpha_s$ also decreases, as is shown in Fig. 2 This property of *asymptotic freedom* allows a perturbative approach for hard interactions.

## 2.3 Bottom quark production

Heavy quarks must be produced in pairs since they only appear in the proton as virtual particle pairs for a very brief time. The production process of a quark-antiquark pair in a hadronic collider consists roughly of three components, which will be briefly described here:

- The particles in the initial state;
- The hard scattering described by perturbative QCD; and
- The parton shower and fragmentation of the outgoing quarks into observable hadrons

The quark-gluon inner structure of the incoming protons will qualify the particles in the initial state of the hard interaction. The knowledge of this structure is summarized in the Parton Distribution Function (PDF), which is the probability density of finding a given fundamental constituent (parton) as a function of the longitudinal momentum

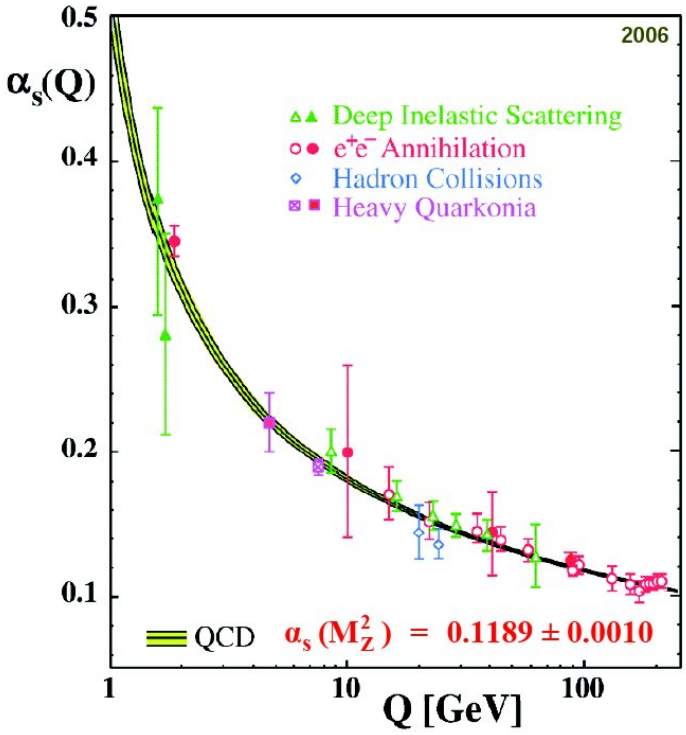


Figure 2: The coupling constant of the strong interaction, $\alpha_s$, as a function of the momentum transfer of the interaction $Q^2$.

fraction $x$. There are no precise calculations available of this since it falls in the non-perturbative regime of QCD, but they are universal functions and can be measured experimentally.

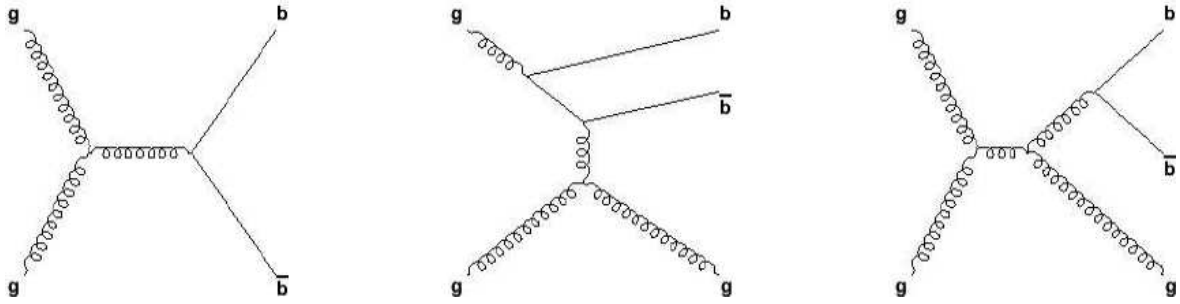


Figure 3: Diagrams for the production of b-quarks. On the left is the process of flavor creation, in the middle flavor excitation and to the right gluon splitting.

The second step can be calculated by summing the contributions of all the different processes that have a b-quark in the final state. For the creation of a $b\bar{b}$ pair, these include three main sources, which are illustrated in Fig. 3.

- In flavor creation, a $b\bar{b}$ pair is produced from gluon-gluon fusion or $q\bar{q}$ annihilation.
- In flavor excitation, a gluon fragments to a $b\bar{b}$ pair in the initial state and one of the b's is scattered by a gluon or quark from the other proton.
- Finally, in gluon splitting the $b\bar{b}$ pair is produced in the final state during the fragmentation of a gluon.

The last step of quark fragmentation and hadronisation will be briefly described in Section 2.6, and mostly relies on phenomenological models.

### 2.3.1 Measurements of bottom production

The first measurement of b-quark production was made at the CERN $p\bar{p}$ collider by the UA1 experiment [REF]. At CDF the inclusive $b$ cross section that was measured in data showed a large disagreement with pure parton-level next-to-leading order (NLO) calculations (see for instance [9]). The correct incorporation of next-to-leading-log resummation of $log(p_T/m_b)$ terms with the NLO hard scatter calculation including massive quarks was then described in the *fixed-order next-to-leading log* (FONLL) framework. This method, in addition to corrections to the heavy-quark fragmentation model has brought better agreement. A recent comparison between data and theory is shown in Fig. 4.

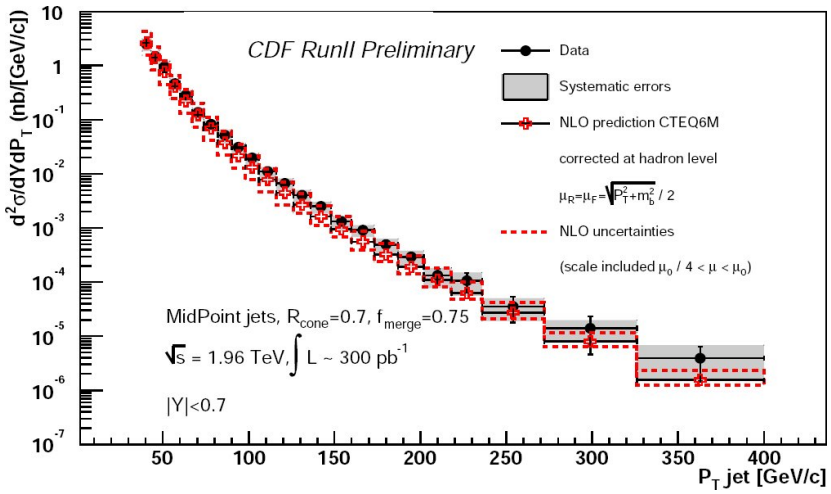


Figure 4: Inclusive b-jet cross section compared to recent next-to-leading order (NLO) QCD calculations. [10]

### 2.3.2 Correlations in $b\bar{b}$ production

The reconstruction of a second b-jet offers more insight in the production mechanisms at hand. The same production mechanisms as for the inclusive b production come into

play, but in addition the angular correlation between the two quarks can be used to isolate the individual contributions (see Fig. 5).

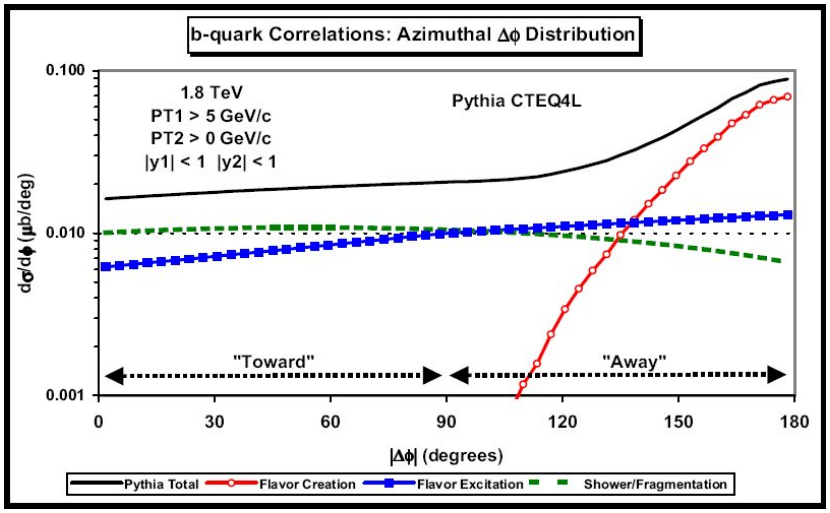


Figure 5: . Contributions from all three production mechanisms to the $b\bar{b}$ cross section as a function of the dijet azimuthal angle. Flavor creation is represented in red circles, flavor excitation in blue squares and shower/fragmentation production with a dashed green line. [11]

A recent analysis by S. Vallecorsa [12] has illustrated another contribution to the correct description of b production. MC simulations were unable to accurately predict the cross section in the case where the two jets are close together. This was thought to be related to NLO effects, but part of this effect has been suggested by Vallecorsa to result at least in part from multiple parton interactions, that is the underlying event.

## 2.4 Photon final states in QCD

The production of prompt photons with high transverse momentum[1] in hadronic collisions is an important testing ground for perturbative QCD due to some unique features of the photon. The production mechanisms consists of mostly $O(\alpha_s\alpha_{em})$ diagrams and electromagnetic coupling ensures the perturbative expansion. Unlike quarks, that cannot be detected directly, photons that participate in the hard interaction can reach the detector without subsequent decays or interactions and can be seen as direct probes of the hard interaction. Additionally the energy measurement (calorimetry) of photons is much more precise than that of jets.

[1]In a hadronic collider, the original momentum (in the direction of the beam) of the two interacting partons is unknown. What is certain is that they carry no momentum in a plane transverse to the beam. Because of this only the transverse momentum of a particle has a significance and also momentum conservation can be used only in the transverse plane.

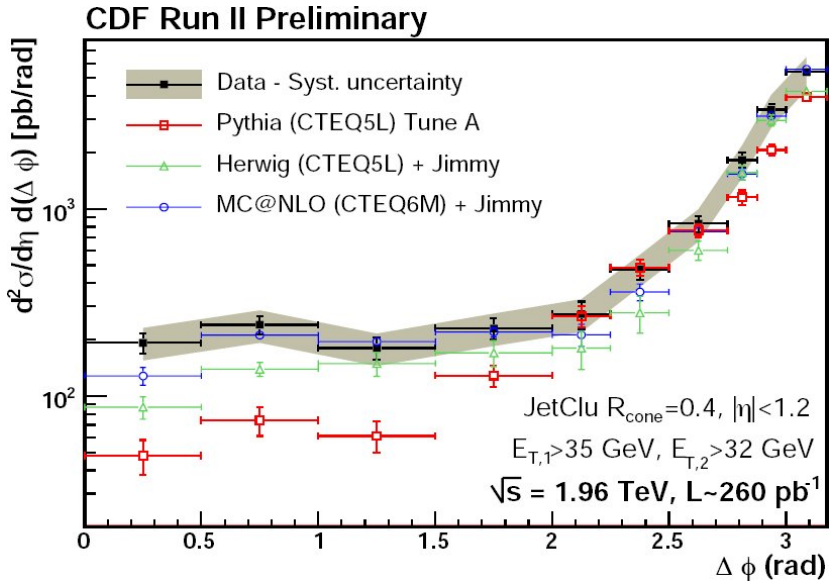


Figure 6: . Differential $b\bar{b}$ cross section as a function of the dijet azimuthal angle [12]. The data are compared to leading-order (LO) and next-to-leading order (NLO) Monte Carlo. The LO MC predictions from PYTHIA [13] (no underlying event) and HERWIG [14] (including a simulation of the underlying event) are compared to the NLO prediction of MCNLO [15], which also includes the underlying event. See [12] for details.

At the tree level, photons are produced via two types of processes (the diagrams are similar to those in Section 2.5).

- The first one is analogous to Compton scattering in QED, except that the incoming boson is a gluon. This process of $qg \rightarrow \gamma q$ is one of the only ways to directly measure or constrain the gluon distributions in hadrons.
- Annihilation diagrams see a $q\bar{q}$ pair annihilate and radiate a photon and a gluon jet.
- Photons can also be produced during the fragmentation of outgoing partons.

The main motivation for measuring prompt photon production clearly lies in the first process, which can constrain the gluon PDF even at higher fractional momentum ($x$). Indeed, deep inelastic scattering experiments have shed light on the $x < 0.1$ region [16], but at higher values, uncertainties of 50 % or more are compatible with data (see Fig. 7).

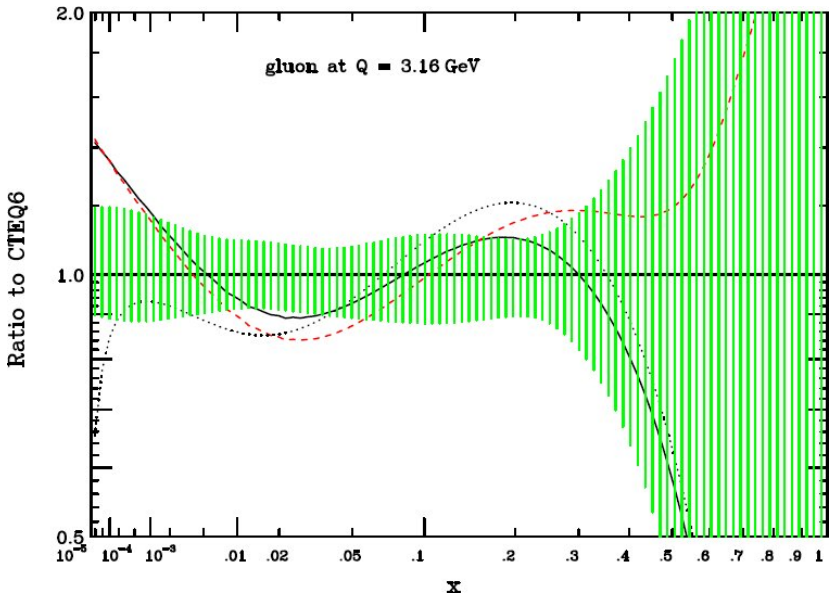


Figure 7: A comparison of several experimental parametrization of the gluon parton distribution function (PDF) within the proton. Several CTEQ5 parametrization (dashed lines) [17] for the gluon PDF are compared to the new fit in CTEQ6 [18]. The vertical green bars represent the uncertainty on this PDF.

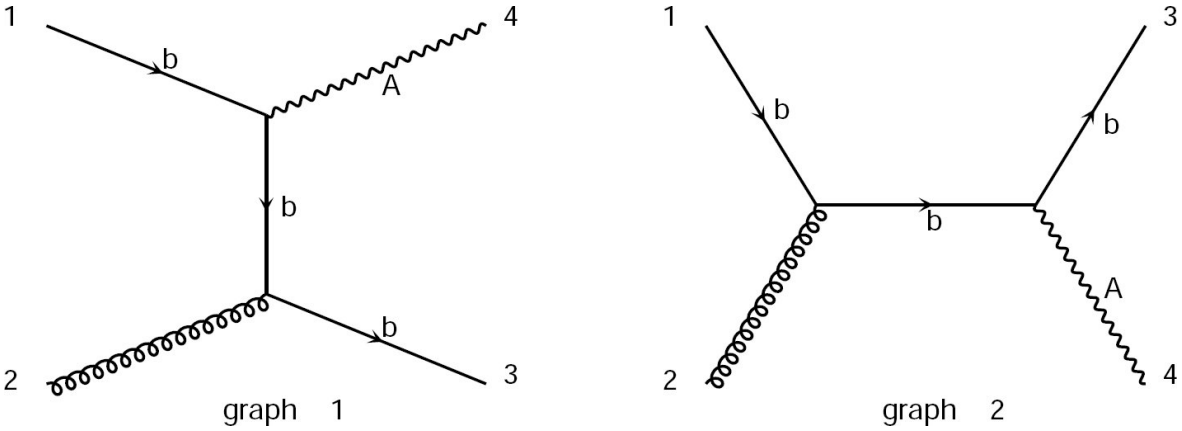


Figure 8: The two leading-order (LO) diagrams for the $(b+\gamma)(b+\gamma)$ signature (not counting antiparticles).

## 2.5 Production of b-quarks and photons

The associate production of a photon and a b-jet is dominated by the Compton scattering of a gluon and a b-quark in the proton sea (see Fig. 8). This means that this channel is sensitive to the b-quark component of the proton. If the gluon parton den-

sity was well known, the cross section would be a direct measurement of the b-quark density. This density can in principle be calculated perturbatively due to the large mass of the b-quark.

## 2.6 Production of $(b\bar{b}+\gamma)$ final states

The production of a b-quark pair and a photon can be seen as a next-to leading order contribution to the inclusive $(b+\gamma)$ production, when only a single b-quark is identified. On the other hand, in extensions to the SM, heavy particles often decay to heavy quarks and photons often appear in their radiative decays. Events with photons and b-quarks can thus be used in searches for new physics. The SM production of $(b\bar{b}+\gamma)$ final states would then be considered as a background and its measurement is therefore important.

As presented in Section 1 (see Fig. 1), there are mainly two types of processes that give a $(b\bar{b}+\gamma)$ final state in QCD. The complete set of diagrams can be found in Appendix A. Their main difference is in the initial state which can be either two gluons or a $q\bar{q}$ pair. In the case of a gluon initial state, the photon is radiated from an outgoing b-quark, while for the other scenario, a photon and a gluon are typically radiated from the incoming quarks and the $b\bar{b}$ pair comes from gluon splitting. However, unlike in the case of $(b+\gamma)$, these b-quarks are not part of the proton sea but rather produced in the hard interaction.

As in the case of $b\bar{b}$ pair production (see Section 2.3.2), angular correlations can give precious information about the production mechanisms at hand. Since we are dealing with a three-body final state, there are now three angular correlations that can be compared. The difference in $\phi$ angle between two objects a and b is defined by $\Delta\Phi(a,b) = \phi(b) - \phi(a)$. The full topology of the event can be determined with only two of these correlations, for example $\Delta\Phi(\gamma, jet1)$ and $\Delta\Phi(jet1, jet2)$ since the sum $\Delta\Phi(\gamma, jet1) + \Delta\Phi(jet1, jet2) + \Delta\Phi(jet2, \gamma)$ is zero by definition[1]. As one can see from Fig.9 and Fig. 10, the quark initial states show a small peak when the two b-quarks are very close.

### 2.6.1 Monte Carlo simulation of $(b\bar{b}+\gamma)$ final states

The theoretical prediction for the $(b\bar{b}+\gamma)$ cross section has been calculated using the MadEvent event generator, which builds on matrix elements computed by MadGraph [19]. That means that the amplitudes for different subprocesses are calculated by MadGraph and the numerical integration over phase space and event generation is performed by MadEvent. The whole program is denoted "MG" in the following. MG can generate events in a specified final state at the parton level. Event information such as particle identities, momenta, spin and color connections are stored in the standard Les Houches event format [20]. The cross section is also calculated at the parton level, and with the generated number of events, an equivalent luminosity of $\mathcal{L}_{eq} = N_{gen}/\sigma_{gen}$

[1] However limited detector resolution and energy mismeasurement can bias this.

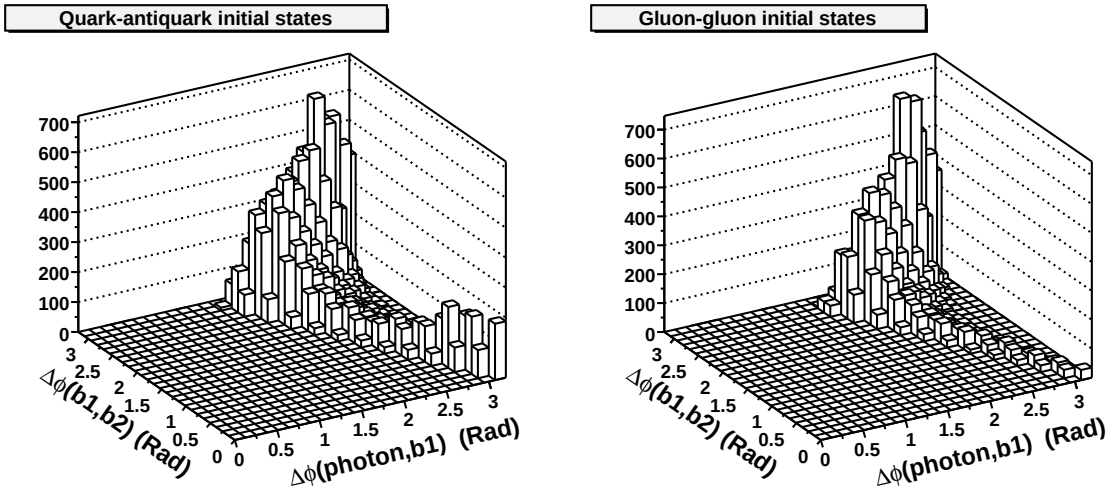


Figure 9: Parton level angular correlations between the photon and the first b-quark, and between the two b-quarks. These two angular correlations define the topology of the event. Events have been separated in the contributions from $q\bar{q}$ (left) and gluon-gluon (right) initial states.

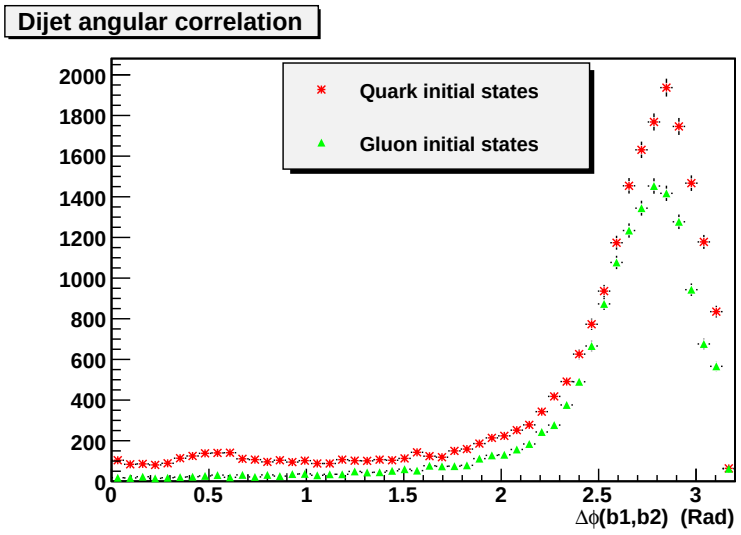


Figure 10: One-dimensional plot of the angular correlation between the generated b-quarks. For small values of $\Delta\phi(b1, b2)$, the $q\bar{q}$ initial states show the stronger contribution.

is calculated. This equivalent luminosity is used to normalize the cross section found after the reconstruction of the MC events.

From this point, the fragmentation of the quarks is made using the version of PYTHIA [13] included in the standard CDF MC production software. In addition to the fragmentation, PYTHIA also adds initial and final state radiation (ISR and FSR) to the partons that are generated. Since this breaks energy and momentum conservation, the energies of all particles are reevaluated at this point. The equivalent luminosity is also reevaluated at this point to account for the generally lower transverse energies of the b-quarks and the photon.

The simulation of events at CDF recreates the actual conditions such as live luminosity and calibrations of the detector. A model of the underlying event (PYTHIA "tune A", see 2.3.2) that is tuned to Run I results is used in addition to account for multiple interactions. The events now contain a large number of particles, and the original partons have fragmented and decayed into more stable hadrons[1]. At this point, the so-called *hadron level* jets are built. The next step of the process is the detector simulation, where the detector response and reconstructed objects are processed. In the end, the event format of the MC sample is exactly the same as the one used for real collision data, and the same analysis is carried out on these simulated events.

This method of generation differs from standard LO MC generators such as PYTHIA and HERWIG [14] which produce inclusive samples of the hard interactions. In that case, to obtain a sample of a rare process such as this one, one would need to generate a huge amount of events[2] and filter out the interesting ones at an early stage. Using the matrix element generator philosophy, one directly obtains a significant event sample at the price of having *only* the contributions from the hard interaction. For the production of a $b\bar{b}gamma$ final state, one could for example imagine to have production of a b-quark pair while the photon could be produced in the underlying event or during the fragmentation.

[1] B hadrons are not decayed with the rather crude PYTHIA decay model but a specific program (EvtGen) is used.

[2] An inclusive photon ($E_T^\gamma > 10$ GeV) MC sample of $6 \times 10^6$ events only yields 24 $b\bar{b}gamma$ events.

# 3 The CDF Experiment

## 3.1 Luminosity

In addition of the center-of-mass collision energy $\sqrt{s}$, the luminosity is one of the most important parameters of an accelerator. It is a measure of how the particles in each beam are interacting with each other at the collision points. It is proportional to the inverse of the physical beam cross section at the collision point and proportional to the revolution frequency. It can be defined as

$$\mathcal{L}_{inst} = \frac{f \times n N_p N_{\bar{p}}}{A} \tag{1}$$

where $N_p$ and $N_{\bar{p}}$ are the number of (anti-)protons in each bunch, $n$ is the number of bunches in each beam and $f$ is the revolution frequency. $A$ is the area in the plane transverse to the beam direction that is occupied by both beams. Luminosity can be measured in units of $[cm^{-2}s^{-1}]$. Physics processes are characterized by a *production cross section*, which is a measure of the probability of the process to happen during a collision. The cross section has the dimension of an inverse area and is usually expressed in milli-, micro-, nano- or even pico*barns*. One *barn* is equal to an area of $10^{-24}cm^2$. The product of production cross section and instantaneous luminosity then gives the production rate of a particular process:

$$\frac{N_{events}}{second} = \mathcal{L}_{inst} \times \sigma_{process} \tag{2}$$

In an analysis where large quantities of data are used, instead of the *instantaneous* luminosity $\mathcal{L}_{inst}$, it is more convenient to use the integrated luminosity, which is noted $\mathcal{L}$. It is defined as the time-integrated instantaneous luminosity and usually expressed in units of inverse femtobarn ($fb^{-1}$) instead of $cm^{-2}$.

## 3.2 The Tevatron

The Tevatron [22] is currently the particle collider in operation achieving the highest collision energy. It is a circular, superconducting proton-antiproton collider with a radius of 1 km and a maximum center-of-mass collision energy of $\sqrt{s} = 1.96$ TeV

**The Accelerator complex** Protons are produced from hydrogen gas that is negatively ionized and accelerated to 750 keV by a Cockroft-Walton electrostatic accelerator. The second stage is a Linac that accelerates the ions to an energy of 400 MeV. Electrons are captured as the hydrogen ions traverse a carbon foil and the remaining naked positively charged protons reach an energy of 8 GeV in the Booster. This Booster synchrotron also packs the protons in bunches.

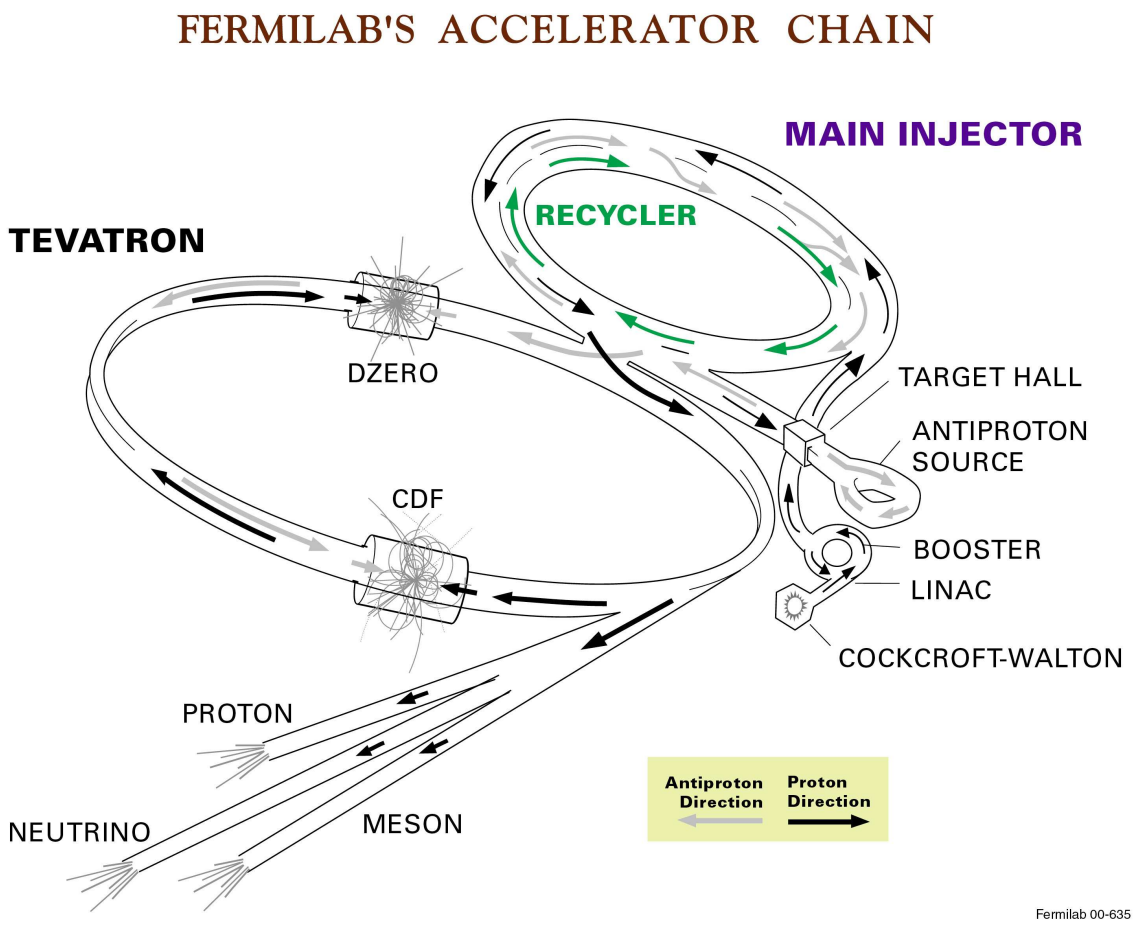


Figure 11: Schematic of the Fermilab Tevatron accelerator complex.

The next stage in the process is the Main Injector[1] (MI) where the protons are prepared for transfer in the actual Tevatron (at 150 GeV).

**Antiproton production** The MI is also used (at 120 GeV) to produce the $\bar{p}$'s that are obtained by bombarding protons on a nickel target. The produced $\bar{p}$'s are collected using a lithium lens and transformed to a continuous beam in the Debuncher. In this beam, however, the $\bar{p}$ momenta are not uniform in both the longitudinal and transverse directions. The process of reducing this spread in momentum so as to obtain a tightly collimated beam is known as *cooling*. After the Debuncher, the $\bar{p}$'s are stored in the Accumulator[2], where stochastic cooling [23] is used. In stochastic cooling, the mean deviation of the beam is measured at one point of the Accumulator ring and this signal

[1]The Main Injector has been added to the accelerator complex in 1999. Before that time, the Fermilab Main Ring, located in the same tunnel as the Tevatron was used.

[2]The Accumulator is a synchrotron with a mean radius of 75 meters.

is transmitted in a direct line to another point where a kicker is activated to correct this deviation, when the $\bar{p}$'s arrive. Over many turns, the momentum spread diminishes and the beam is cooled.

The Recycler Ring (RR) is a fixed energy (8 GeV) storage ring in the Main Injector tunnel that features permanent magnets. It was originally intended to store leftover $\bar{p}$'s but has been used as an intermediate storage ring for antiprotons. This has allowed to take advantage of the higher stacking rate of the Accumulator when it contains few $\bar{p}$'s. Typical stack sizes in the RR, that is the number of $\bar{p}$'s in the beam, reach over $300 \times 10^{10}$ antiprotons. The RR uses electron cooling to reduce the longitudinal emittance of the beam.

The rate at which antiprotons are produced has reached a record of $2.2 \times 10^{11}$ $\bar{p}$/hr to this day. Using the recycler has allowed to prepare large quantities of antiprotons for each Tevatron store, a prerequisite to achieve high luminosities.

The final stage in this scheme is the Tevatron itself, in which protons and antiprotons circulate in opposing directions and are accelerated to the final energy of 980 GeV. Each beam is separated into three trains of 12 bunches each. At a luminosity of $2 \times 10^{32} cm^{-1} s^{-1}$, each proton bunch contains approximately $2.7 \times 10^{11}$ protons, while the number of $\bar{p}$'s in the antiproton bunches is of about $7 \times 10^{10}$. Inside of a train, the bunch spacing is $396 ns$. The empty space between the trains ($2.6 \mu s$) is the "abort gap" necessary to activate the electrostatic kickers. These push the beam into safe dump blocks when a store[1] is aborted either voluntarily or by a safeguard system.

The two beams normally form a complicated helix so as to minimize losses in the Tevatron. Electrostatic separators and large quadrupole magnets allow the two beams to collide inside of the detectors at the two interaction points in B0 (CDF) and DØ(Detector of the same name). In these points the beams are as tightly collimated as possible in order to achieve the highest luminosities.

## 3.3 Tevatron performance

The first data-taking period that lasted from 1991 until 1996 was labelled "Run I" and allowed to collect data corresponding to an integrated luminosity of approximately 118 $pb^{-1}$ at a center-of-mass energy $\sqrt{s} = 1.8$ TeV. After major upgrades between 1996 and 2001, the ongoing "Run II" has allowed each of the CDF and DØexperiments to collect in excess of 2.5 $fb^{-1}$ of data at an energy of $\sqrt{s} = 1.96$ TeV (see Fig.12 and Fig. 13).

**Outlook** The Tevatron is performing well, with a monthly integrated luminosity reaching 45 $pb^{-1}$. The initial instantaneous luminosity has reached a record of $286 \times 10^{30} cm^{-2} s^{-1}$ in 2007, which is to be compared to the design value of $200 \times 10^{30}$. By the final shutdown of the Tevatron in 2009, a total integrated luminosity of 6 to 7 $fb^{-1}$ is expected.

[1] Period of time when the Tevatron contains colliding beams.

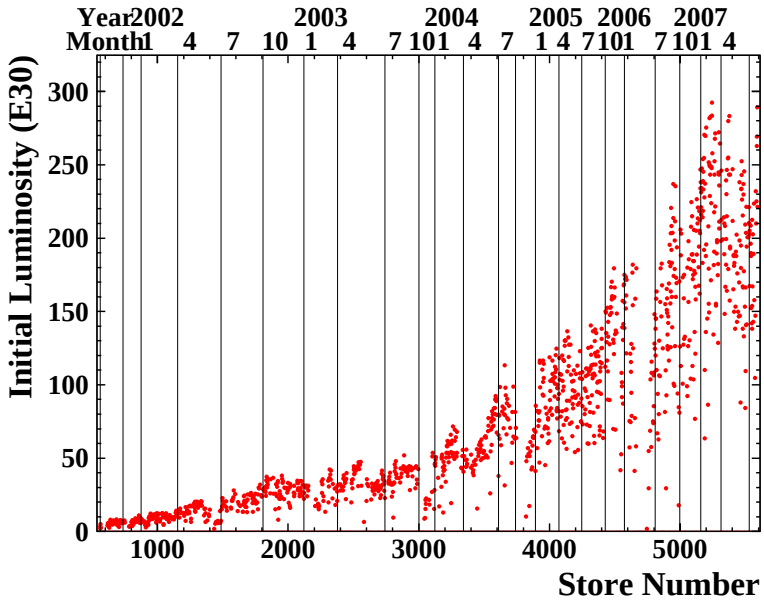


Figure 12: Peak instantaneous luminosities achieved at the Tevatron in units of $10^{30} cm^{-2} s^{-1}$.

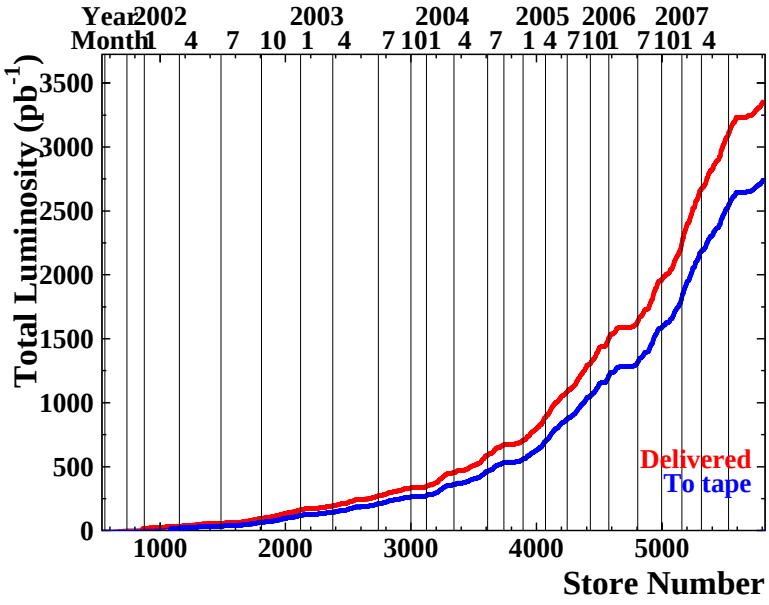


Figure 13: Run II integrated luminosity in $pb^{-1}$ delivered by the Tevatron (red) and collected by CDF (dark blue) until the Summer 2007 shutdown.

## 3.4 The Collider Detector at Fermilab (CDFII)

The CDFII detector is a multi-purpose solenoidal detector with forward-backward and azimuthal symmetry. It features precision tracking of charged particles, fast projective calorimetry and fine grained muon detection. It measures about 27 meters in length, 10 meters in diameter and has a mass of about 5000 tons. The detector has been largely upgraded for the higher luminosities of Run II, including a completely replaced tracking system. An elevation view of the CDFII detector is shown in Fig. 14.

The momenta of charged particles are measured in a homogeneous magnetic field of 1.4 Tesla provided by a superconducting solenoid magnet. The energy of particles is measured with good spatial resolution in the calorimeters.

The detector elements used in this analysis are the tracking and vertexing system, the central electromagnetic and hadronic calorimeters as well as two shower profile detectors which are described in the following paragraphs. A complete description of the detector can be found in [24].

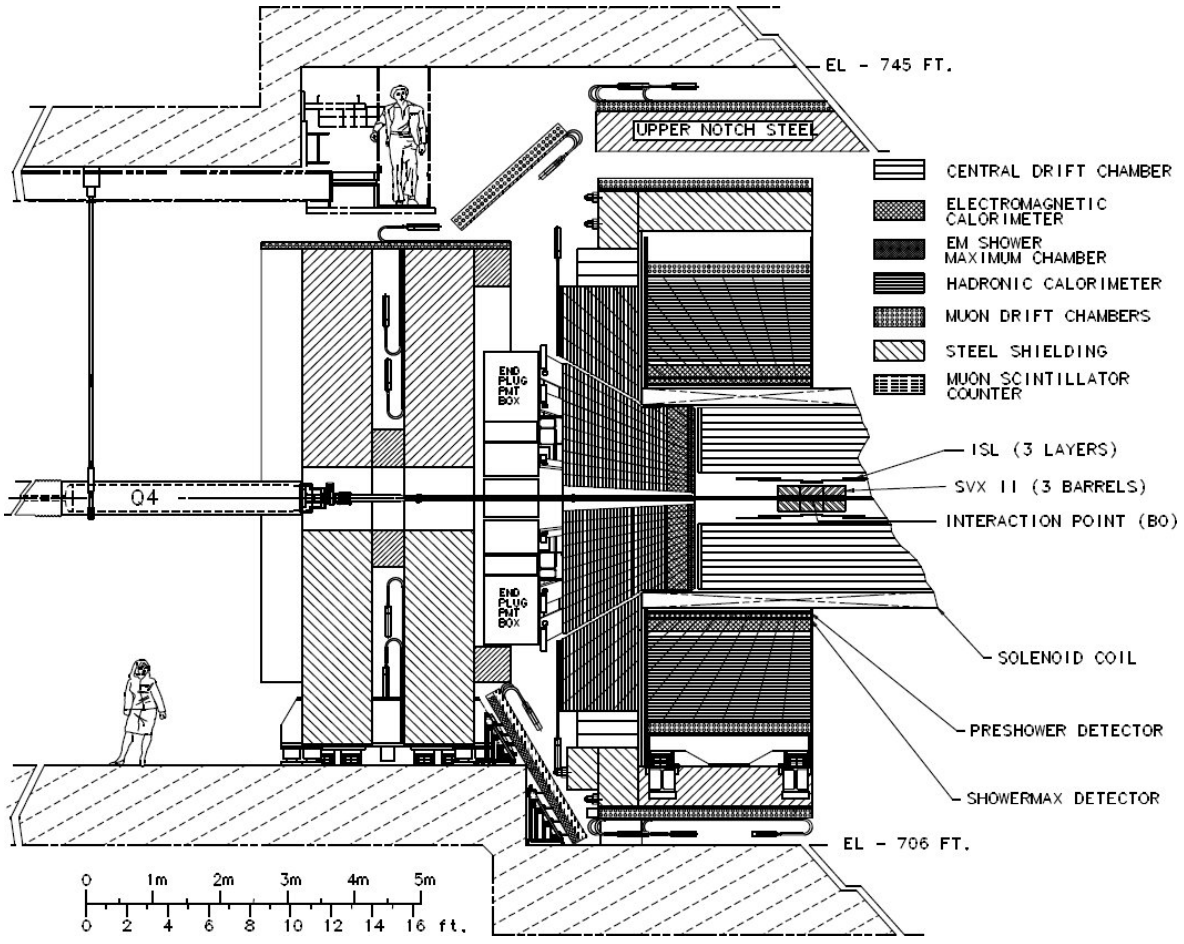


Figure 14: Elevation view of the CDF II detector

### 3.4.1 Geometric definitions

The origin of the coordinate system in the CDF experiment is situated at the center of the detector. Since in hadron colliders the longitudinal momentum of the partons participating in a hard interaction is unknown, quantities measured in the transverse plane to the beam axis are most relevant for considerations of energy or momentum balance. Two variables that are widely used with respect to this are the the transverse energy $E_T$ and the transverse momentum $p_T$, defined by

$$\begin{aligned} E_T &= E \times \sin(\theta) \\ p_T &= p \times \sin(\theta) \end{aligned} \tag{3}$$

Thus the standard coordinate system used is cylindrical, with the z-axis pointing in the direction of the proton beam and the y-axis pointing upwards (see Fig. 15).

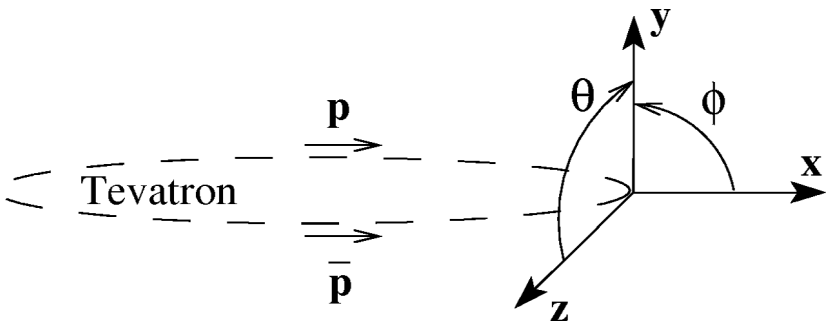


Figure 15: The CDF coordinate system

A variable of interest is the rapidity $y$ of a particle, defined as $y = \frac{1}{2} ln(\frac{E+p_z}{E-p_z})$, which is invariant with respect to boosts in the $z$ direction. In practice, however, what is used is the pseudo-rapidity $\eta = -ln(\tan(\frac{\theta}{2}))$. It has a direct geometric interpretation and is a valid approximation of the rapidity in the massless approximation when $E >> mc^2$.

### 3.4.2 Tracking system

The tracking system has two purposes which are, evidently, the tracking of charged particles but also the reconstruction of decay vertexes by assembling tracks pointing to a common origin. This ability is of crucial importance for heavy flavor identification. A special trigger has been implemented towards that goal and will be described in Section 4.2.

This subgroup of detector components is situated closest to the interaction point and completely inside of the solenoid. The first layer traversed radially is the silicon microstrip detector which extends up to a pseudorapidity of $|\eta| < 2.0$ and to a radius of about 28 cm from the beamline. Located at radii from 40 to 137 cm, the COT is a multiwire cylindrical drift chamber providing full coverage for $|\eta| < 1.0$ (see Fig. 16).

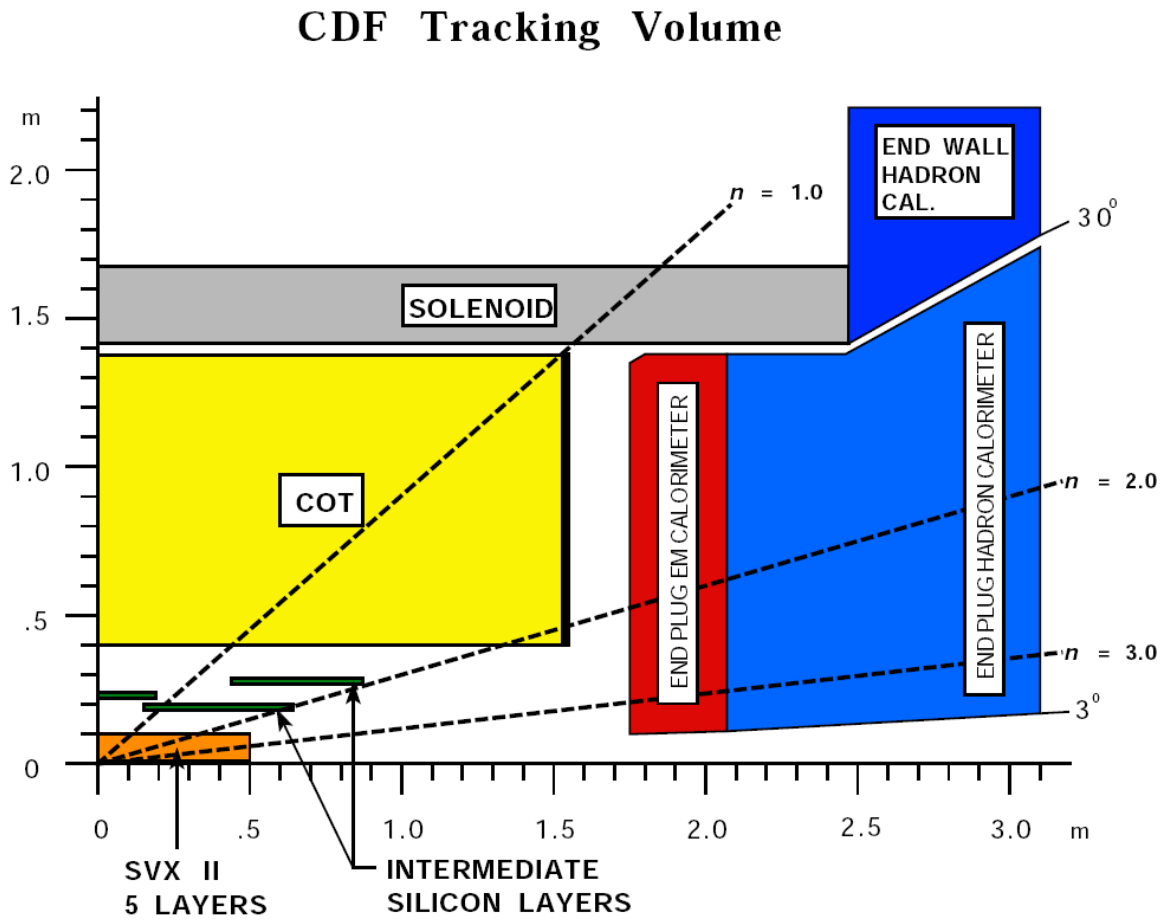


Figure 16: Overview of the tracking system

**Silicon Vertex detector** The silicon tracking system is split into three subsystems: Layer 00, the Silicon Vertex Detector (SVX II) and the Intermediate Silicon Layer (ISL). It consists of up to eight layers of silicon microstrip modules, see Fig. 17. While L00 only records axial hits (the strips are oriented parallel to the beamline), both the SVX and the ISL feature double-sided modules that allow "stereo" recording with the strips on the second side oriented at an angle. The total number of readout channels in the silicon detector is approximately 700'000, about two thirds of the whole CDFII detector.

The silicon system is essential in finding displaced vertexes coming from b hadron decays. The most important parameter measured for this is the transverse distance of closest approach to the beamline of a displaced track, the impact parameter $d_0$. The resolution on the impact parameter measurement in the silicon tracker is approximately 15 $\mu m$ for high-$p_T$ tracks (see Sec. 4.2). When the beam spot spread of 35 $\mu m$ is included, the final resolution on the impact parameter using the full reconstruction

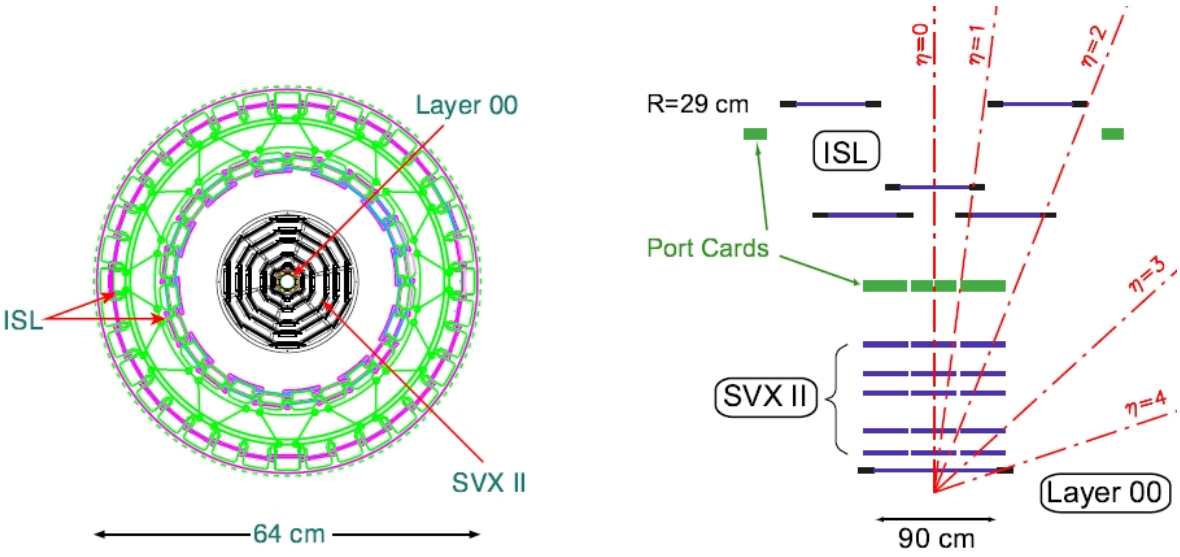


Figure 17: Schematic view of the CDFII silicon detector. Left hand: r-$\phi$ plane. Right hand: r-z plane.

with the silicon tracker is $\sigma(d_0) = 15 \oplus 35 \approx 40 \quad \mu$m. The online recording of events with a displaced track and the reconstruction displaced vertexes is further discussed in Sections 4.2 and 6.6 respectively.

**Central Outer Tracker (COT)** The COT drift chamber is filled with an equal parts mixture of argon and ethane. It consists of eight "superlayers" in the radial direction which are themselves composed of "supercells" in the azimuthal direction. The alternating superlayers provide axial and stereo hit information. The resolution of the COT for a single hit is 140 $\mu m$ and the momentum resolution for charged particles is $\sigma(p_T)/p_T^2 \approx 0.3\% \quad GeV^{-1}$. This resolution reaches less than $0.1\% \quad GeV^{-1}$ when the silicon and COT information are combined. The COT can also provide $\frac{dE}{dx}$ information.

### 3.4.3 Calorimetry

The calorimeters are situated outside of the solenoid and are composed of two parts, the central calorimeter and plug calorimeter. I will briefly describe the former which extends in rapidity up to $|\eta| < 1.5$. The central calorimeter is composed of towers that point back to the interaction point. Each tower approximately spans 15 degrees in the $\phi$ direction and a pseudorapidity range of 0.11. CDF uses sampling calorimeters which alternate passive layers of heavy absorber material (lead or steel) and active layers of scintillator.

**Central Electromagnetic Calorimeter (CEM)** The CEM is composed of alternating layers of lead and scintillator and has a depth of 18 radiation lengths $X_0$. Its

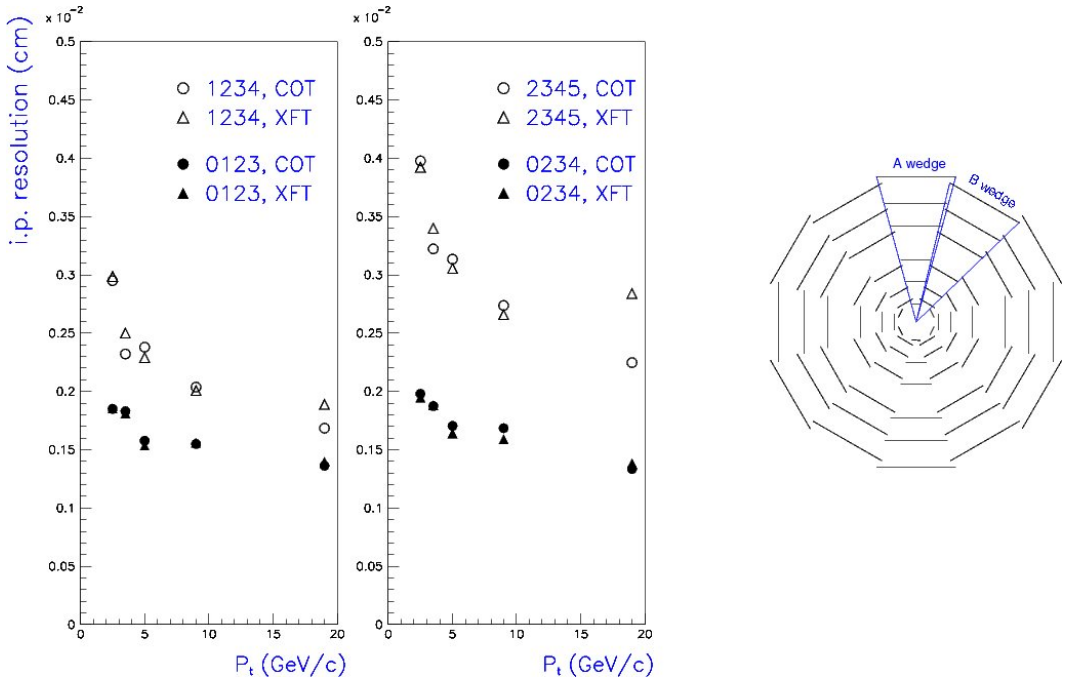


Figure 18: Left: Impact parameter resolution of the silicon tracker as a function of the $p_T$ of the track. The black circles and triangles include L00 in the reconstruction. One can see that the resolution on the impact parameter is dramatically improved by including this detector. Right: Schematic view of the Silicon layers. The two innermost silicon layers represent the L00 detector which sits right on the beam pipe. The difference between the two plots is the configuration of the L00 wedges. In type A wedges (left) the L00 layer is even closer to the beam than in type B wedges (right).

energy resolution reaches $\sigma(E)/E = 13.5\%/\sqrt{E} \oplus 2\%$, where the last term represents the uncertainty on energy calibration.

**Central Hadronic Calorimeter (CHA)** The CHA is also a sampling calorimeter of alternating steel and scintillator tiles. Its energy resolution is about $\sigma(E)/E = 75\%/\sqrt{E} \oplus 3\%$. The tower geometry from the CEM is kept here. The CHA of course is very important in any jet analysis, and the methods used to correct its non-linear energy response are described in Section 6.5.

**Shower profile detectors (CES and CPR)** Embedded in the CEM at the depth where the EM showers reach their maximal spread (about 6 $X_0$) is the Central Electromagnetic Shower maximum detector (CES). It measures the charge deposition on orthogonal strips and wires. Cathode strips running in the azimuthal direction provide z information, while anode wires running in the z direction provide $r - \phi$ information. The CES allows for a position determination of the EM shower and for a measurement

of the shower transverse profile of isolated photons or electrons. The position resolution of the CES is $\pm 2mm$ for photons with $E_T > 30$ GeV.

Just inside of the CEM is the Central Pre-Radiate Chamber (CPR). Its purpose is also to measure electromagnetic shower profiles. The goal is to distinguish the early showers coming from single photons as opposed to multiple photons from pion decays. Since the probability of a conversion from a photon into an $e^+e^-$ pair is constant through the return yoke of the magnet, multiple photons have a higher chance to produce hits in the CPR.

These two detectors (CES and CPR) are essential for the separation of pions and photons, as is explained in Section 6.4.1.

# 4 Data acquisition and triggering

The process of selecting and storing events is described in this section. After an overview of the CDFII data acquisition and trigger system, the method for online selection of heavy-flavor jets is described. The cross section for b-quark production reaches 50 $\mu b$ at the Tevatron, but this is a small fraction of the QCD background of 50 $mb$. The traditional way to trigger B events has been to look for leptons in the decay of the B hadrons. In Run II, a new trigger was developed to select events that have a displaced track with respect to the beamline. This allows also fully hadronic B decays to be recorded. The relevant subsystem is called the Silicon Vertex Tracker (SVT) and is explained in more detail in Section 4.2.

## 4.1 Trigger overview

The trigger plays an important role in hadron collider experiments because the collision rate is much higher than the rate at which events can be stored on tape. The role of the trigger is to extract the interesting events from the huge amount of soft minimum bias collisions. The CDFII trigger has a three level architecture[1] designed to keep the deadtime[2] as short as possible. A schematic view of the data flow and trigger rates at each level is given in Fig. 19.

The first level of the trigger bases its decision on the existence of objects such as calorimeter clusters, hits in the muon chambers or tracks in a time window of $5.5\mu s$.

The data is synchronously processed in three streams. The first one consists of calorimeter objects obtained by applying thresholds on single towers. The total transverse energy and missing $E_T$ are also computed. The second stream extracts muon information from the corresponding detectors. Finally, tracks from the COT are processed in the third stream. This task is assumed by the eXtremely Fast Tracker (XFT) which uses pattern recognition to return the $p_T$ and $\phi$ of recognized tracks. The L1 output rate is about $25kHz$ to be compared to a collision rate of $2.5MHz$.

Having successfully passed the L1 requirements, events are stored in one of four buffers and further analyzed by hardware processors. Events are now treated asynchronously: they remain in the buffer until the are accepted or rejected. The decision of the second level occurs in two steps of $10\mu s$ each. The event is first partially reconstructed and analyzed according to the subdetector used. Muon, electron and jet candidates are now searched for, thus allowing a higher rejection of fakes. The results of this first stage are then examined further during the second step. The accept rate for L2 is $300Hz$. The Silicon Vertex Tracker, described in Section 4.2 is part of the L2 trigger.

After passing the second trigger level, the entire event is read out and analyzed in a Linux PC farm where it is fully reconstructed. The L3 reconstruction code is

[1] The trigger levels are denoted as L1, L2 and L3 in the following.

[2] Deadtime occurs when one part of the DAQ is busy and events cannot be treated even if they pass trigger requirements. These events are then lost, which is why deadtime must be kept to a minimum.

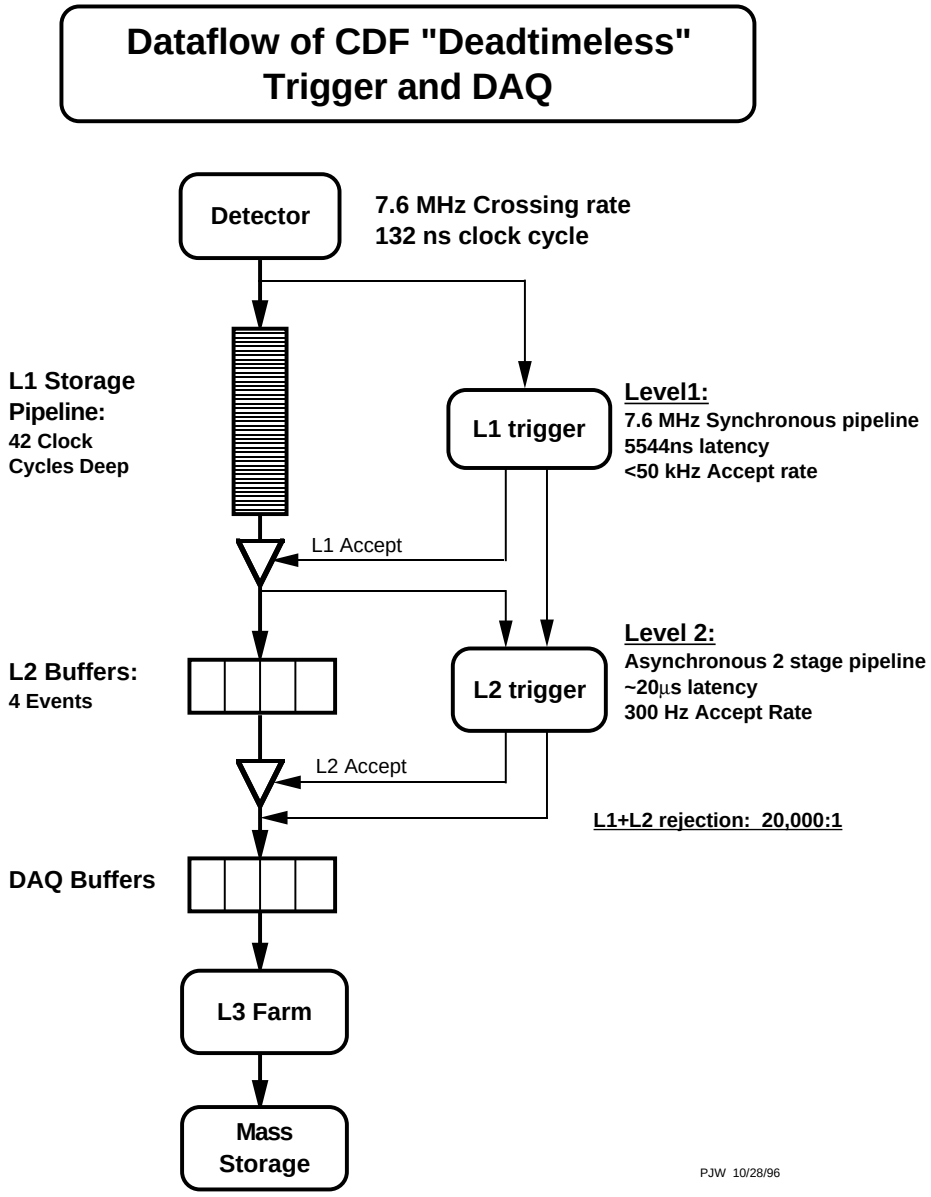


Figure 19: Dataflow of the CDFII Trigger and DAQ systems

identical to the code used in the offline reconstruction of the events. Each event must be processed within about one second and marked with at least one of approximately 140 different trigger paths[1].Events passing this final barrier are sent to permanent storage, at a rate of $75 - 100Hz$.

[1]A trigger path is a collection of successfully passed requirements at each level of the trigger for the event.

## 4.2 Triggering on B-hadron decays

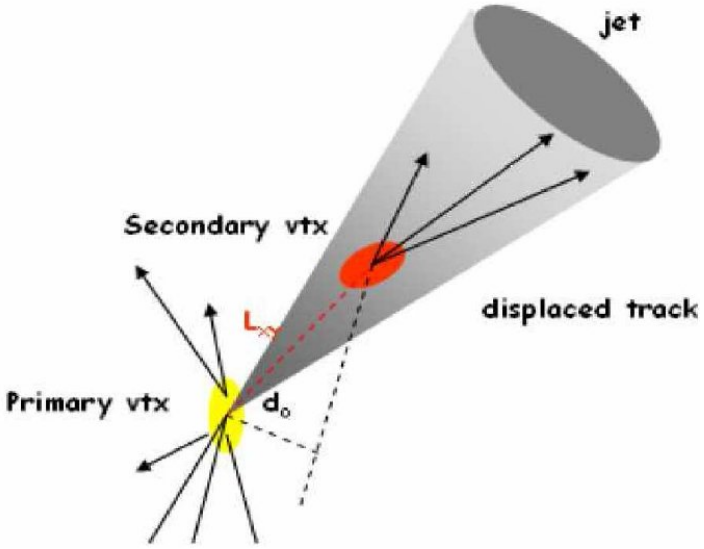


Figure 20: Schematic of a secondary vertex inside of a jet cone. The impact parameter $d_0$ of a track is shown as well as the distance from the primary vertex $L_{xy}$.

To find b-jets, that is jets containing one or more B hadrons, a current technique is to use the long lifetime of B hadrons, which is typically 1.5 ps. This translates to a decay length $c\tau$ of typically 450 $\mu$m, which is detectable. Because of their relativistic momenta after the collision, the actual decay distance of B hadrons can reach several millimeters. Also B hadrons are heavy with masses in excess of 5 GeV, and that ensures that the lighter decay products can escape at a visible angle with respect to the direction of the original hadron. The tracks from these decay particles will point back and intersect at the vertex of the decay. Because of the long lifetime, this "secondary" vertex can be distinguished from the vertex of the hard interaction ("primary vertex").

At the trigger level the required decision time is not sufficient to fully reconstruct this secondary vertex. What can be done however is to trigger on tracks that are not pointing back to the primary vertex, that is tracks with a significant transverse impact parameter $d_0$ (definition in Fig. 20). This ensures a data sample enriched in heavy flavor. It is, however, difficult to implement. The device designed by CDF for this task is described in the following section.

### 4.2.1 The Silicon Vertex Trigger (SVT)

The SVT is part of the L2 trigger. It receives as input a list of COT tracks from the XFT and digitized signals from four silicon layers. Its output is a list of reconstructed 2D tracks, with their direction,momentum and impact parameter. The reconstruction reaches almost offline quality. The resolution on SVT tracks is of $\sigma(\phi) = 1m\text{rad}$,

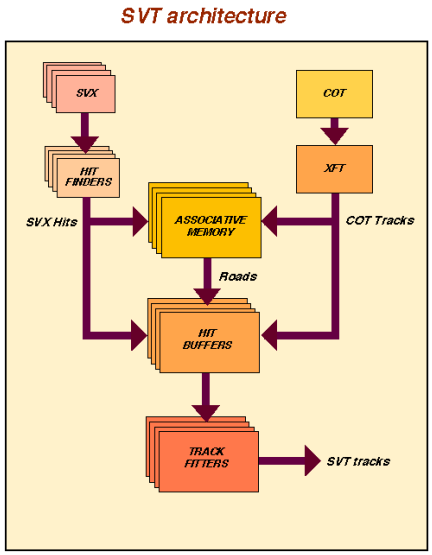


Figure 21: Architecture of the Silicon Vertex Tracker

$\sigma(p_T) = 0.003 \cdot p_T^2 \quad GeV$ and $\sigma(d_0) = 35 \quad \mu m$ where $d_0$ is the impact parameter of the track, that is the closest approach to the beamline (see Fig. 20). The resolution stated for the impact parameter does not take into account the size of the beam spot.

The SVT is divided in several subsystems, which are represented in Fig. 21. It is organized in 12 identical systems, each spanning a wedge in $\phi$. The Hit Finders reconstruct the center of silicon hits and send them to the Hit Buffer for future reference and to the Associative Memory (AM) units. This system performs the pattern recognition, but since the time scale is limited, this is not done with a software algorithm. Instead, the input data is compared to a stored set of precalculated patterns that correspond to legitimate particle trajectories (*roads*). The resolution of these roads is still coarse ($250\mu m$). A list of selected roads is sent to the Hit Buffer which retrieves the full XFT and SVX information and sends them to the Track Fitter. In this last element, the position of the four silicon hits, the curvature and the azimuthal angle of the XFT track give a final track described by the curvature, $\phi$ and $d_0$. The fitting algorithm uses a linear approximation for speed. A $\chi^2$ cut is made on the quality of the fit.

The SVT allows trigger decisions based on the $p_T$, fit quality and most importantly the impact parameter of the track. This means that if the position of the primary interaction vertex is not accurately known during data taking, false trigger decisions would follow. This problem is circumvented by constantly measuring the beam offset (using the SVT and the correlation of $d_0$ and $\phi$ for an offset beam, see Fig. 22) and providing the correction to the impact parameter for trigger decisions.

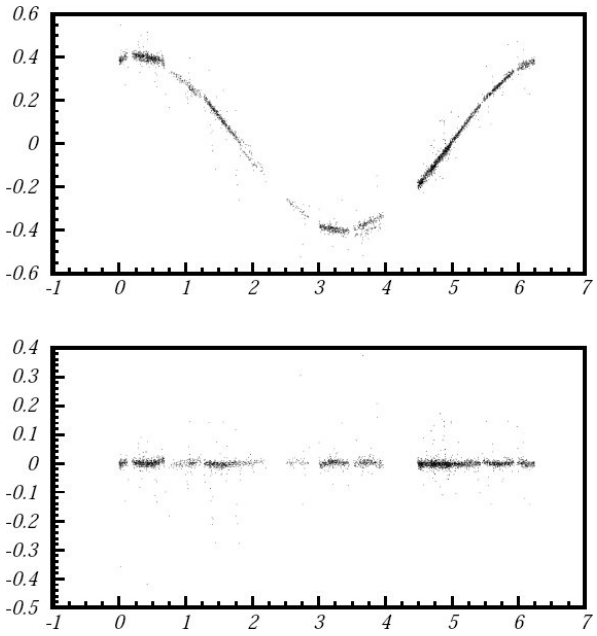


Figure 22: Typical displacement of the beam as visible from the impact parameter distribution as a function of the azimuthal $\phi$ angle. For a position of the beam spot in the x-y plane of $(X_0, Y_0)$, the impact parameter with respect to the origin of tracks pointing back to the beam spot can be expressed by $d = X_0 \times sin(\phi) - Y_0 \times cos(\phi)$. The vertical axis is the measured $d$ in $cm$ and the horizontal is the angle. The upper plot shows the measured values and the lower plot is obtained after using the corrected beam spot position.

The final resolution on the impact parameter, including the beam spot uncertainty, is approximately $\sigma(d_0)_{SVT} = 48\mu m$ which is remarkable for an online reconstruction. The precision is very close to the resolution attainable using the full offline reconstruction ($40\mu m$).

# 5 Day-to-day data collection: detector operations

The scientific success or failure of any particle physics experiment depends on several aspects. The first is the design and overall concept which will determine the projected performance. The second aspect is the execution to specifications or beyond of this design during the construction. Finally, to actually make the best possible use of the previous efforts, comes the careful operation of the apparatus, which represents the daily work of many people. This effort is briefly described here as I had the opportunity to participate in it.

The operation of a huge detector presents many challenges. The first is to keep it functioning and to prevent damage. Secondly, the data that is taken needs to be well understood, which implies up-to-date calibrations of all the subsystems. Thirdly comes the efficiency of data taking, that is the luminosity that is collected and can be used in physics analyzes compared with the luminosity delivered by the accelerator (see Fig. 23).

The detector operations group is in charge of this task, and coordinates the shift crew, which is presented here. The Operations Manager makes the plans for data-taking, accesses to the detector and defines the various studies and calibrations that are to be carried out. The shift crew is composed of four persons that are permanently present in the control room.

- The Scientific Coordinator (SciCo)
- The Consumer Operator (CO)
- Two Aces[1]

The role of the SciCo is to direct the shift crew. He allocates shift resources when many demands are addressed at the crew, communicates with experts or the accelerator control room and leads emergency responses. He also maintains the electronic logbook and ultimately decides whether the collected data is suitable for physics analysis.

The CO is responsible for checking the data quality, using many different plots that represent the detector response. The CO is also in charge of verifying the calibrations and maintains a list of problems to be addressed by the experts.

The Ace has many duties. He is in charge of starting and ending the data acquisition (DAQ) runs, performing the calibrations, turning the high voltage (HV) in subsystems on or off and monitoring the beam conditions. In addition, he must diagnose and where possible fix problems that occur with the DAQ system or the HV.

The CDFII experiment has reached maturity and most of the operational problems already have standard solutions so that they can be directly resolved by the shift crew without needing to call the expert of one particular subsystem. The Aces usually being the crew members which are staying in the Control Room for the longest period, they also have the most experience in dealing with problems. As such, and with very many

[1] "ACE" allegedly stands for *Accelerated CDF Expert*

things to be looked after, they carry quite a large responsibility. More recently, only one ACE is used in the CDF Control Room. This is possible because the detector is mostly functioning very well.

The learning curve at the beginning of the shifts is very steep, since after a short period of one week, the operation of the DAQ system, the control of the high voltages and the response to alarm situations must be integrated. On the other hand, there is a strong support from the group in charge of the detector operations and generally a very good climate in the Control Room that have made these shifts an extraordinary experience for me.

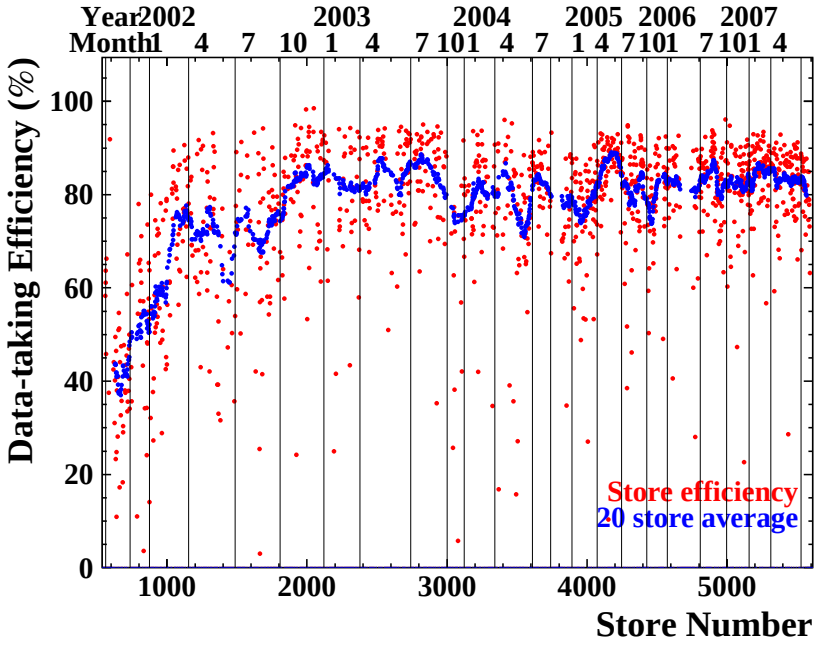


Figure 23: CDF Data taking efficiency in Run II.

# 6 Data Analysis

## 6.1 Outline

The aim of this analysis is to measure the cross section of $b\bar{b}\gamma$ production in proton-antiproton collisions. For that purpose, one needs to obtain a sample with a well-defined photon and two well-defined b-jets. In order to have a reasonable number of candidates, the thresholds on the transverse momentum of the photon or the jets need to be as low as allowed by the trigger bandwidth.

The analysis is performed in parallel using two independent datasets (see Section 6.2), the first used for the physics result, and a second dataset used as a normalization reference to calculate the efficiency of the former (Section 6.3).

We select candidate events with an isolated central ($|\eta| < 1.1$) photon having a transverse energy $E_T^\gamma > 12$ GeV and two central ($|\eta| < 1.5$) jets of $E_T > 20$ GeV. Both jets are required to lie outside of a cone of $\Delta R = \sqrt{d\phi^2 + d\eta^2} = 0.7$ around the photon candidate. We require both jets to have a secondary vertex signaling the decay of a B hadron. The exact selection procedure, efficiencies and background rejection techniques are described in Section 6.4 for the photon and Sections 6.5 to 6.6 for the b-jets.

## 6.2 Data samples

In a first stage, the data are collected using a trigger path[1] developed by the University of Geneva that requires a photon with $E_T^\gamma > 12$ GeV, two jets each having $E_T > 10$ GeV and a displaced track with impact parameter of $d_0 > 120\mu$m found by the SVT (see Section 4.2). In the following this dataset will be denominated by "SVT" or simply the biased dataset. Mario Campanelli's $(b + \gamma)$ analysis discussed in Section 1 [4] uses the same dataset, and this analysis is carried out in the same spirit.

The second dataset, used as the control sample, requires an inclusive photon trigger[2] ("ISO" or "unbiased" dataset) which is fully efficient for $E_T(\gamma) > 26$ GeV [3]. However, the transverse energy threshold on the photon must be higher (25 GeV) with this much simpler trigger which only requests an energetic central photon to maintain a reasonable trigger rate. That is why the actual measurement is carried out on the SVT dataset.

Events from the SVT dataset for which $E_T^\gamma > 26$ GeV will be labeled by "ISOSVT" since they should be present in both datasets. This fact is extensively used in the determination of the trigger efficiency of the SVT trigger.

## 6.3 Evaluating the SVT trigger efficiency

To measure the cross section, a measurement of the SVT trigger efficiency is necessary, that is to measure the fraction of events that really pass the trigger with respect to

[1] Named PHOTON_B_JET
[2] PHOTON_25_ISO

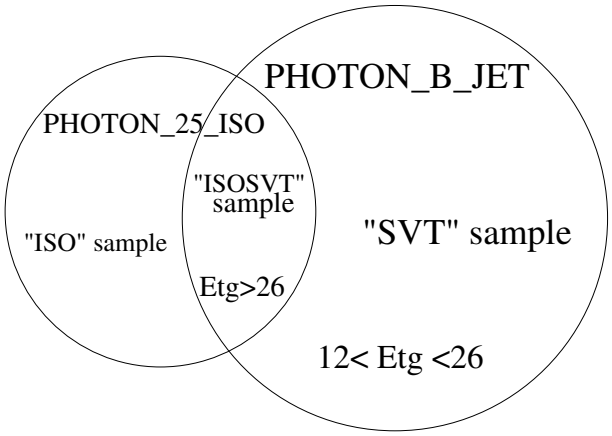


Figure 24: Naming convention for the datasets. The "ISO" sample is obtained with the unbiased PHOTON_25_ISO trigger and has a trigger requirement on the photon $E_T^\gamma > 25$ GeV. The "SVT" dataset is obtained with the dedicated PHOTON_B_JET trigger featuring a displaced track from the SVT and has a photon requirement of $E_T^\gamma > 12$ GeV. The "ISOSVT" sample lies in the overlap of both datasamples when $E_T^\gamma > 26$ GeV and the event has been triggered by PHOTON_B_JET. Since the SVT trigger is not fully efficient, not all the events in the ISO sample are included in the SVT dataset.

those that satisfy the trigger conditions. The trigger efficiency is in this case measured directly from the data using the ISO dataset. The SVT trigger further introduces a bias in the efficiencies necessary to the cross section measurement, but this can be circumvented in a clever way.

While the displaced track requirement of the PHOTON_BJET trigger allows a lower photon threshold and thus gives access to more statistics and phase space, the calculation of the trigger efficiency is complex, and the precise determination of other quantities, for example b-tagging efficiency and purity can only be performed on the basis of simulated events passing a sophisticated trigger simulation.

With some assumptions, the whole issue can be simplified. Since we require at the analysis level that the photon and the jets are well separated, we can assume that efficiency and purity measurements for the photon and jets can be made independently. In that case, the efficiency for finding an SVT track with large impact parameter in a b-tagged jet is the same for photon $E_T$ thresholds of 12 and 25 GeV at the trigger level. It is therefore possible to measure the SVT trigger efficiency using the unbiased dataset. Since the unbiased trigger is fully efficient for $E_T(\gamma) > 26$ GeV, the efficiency of the biased trigger can be computed by counting those events with two positively tagged jets and a photon with $E_T > 26$ GeV in each dataset. The ratio of those events in the ISO sample that have the PHOTON_BJET bit set over the complete number gives the trigger efficiency.

In a more formal way, we can write the cross section for events from the unbiased

sample with $E_T(\gamma) > 26$ GeV (trigger fully efficient) as:

$$\sigma_{ISO} = N_{ISO} \times f^{bb}_{ISO}/(\epsilon^{tag1}_{ISO} \times \epsilon^{tag2}_{ISO} \times \mathcal{L}) \tag{4}$$

where $N_{ISO}$ is the number of candidates in this dataset, $f^{bb}_{ISO}$ is the fraction of true $b\bar{b}$ pairs among them with two tagged jets and $\epsilon^{tag}_{ISO}$ is the tagging efficiency for each jet. The cross section in the SVT dataset is similarly given by

$$\sigma_{SVT} = N_{SVT} \times f^{bb}_{SVT}/(\epsilon^{tag1}_{SVT} \times \epsilon^{tag2}_{SVT} \times \epsilon^{trig}_{SVT} \times \mathcal{L}) \tag{5}$$

The fraction of true $b\bar{b}$ events among candidates $f^{bb}$ and the tagging efficiency $\epsilon^{tag}_{SVT}$ in this case refer to events that have passed the SVT trigger requirements, and extracting them from the MonteCarlo would require a detailed simulation of the SVT behavior. $N_{SVT}$ is the number of candidates in the "SVT" sample and $\epsilon^{trig}_{SVT}$ is the corresponding trigger efficiency. Exploiting the fact that the measured cross section should be equal for both datasets in the overlap region allows us to cancel out these terms. Since the cross section does not depend on the dataset used to measure it, $\sigma_{ISO} = \sigma_{ISOSVT}$, so

$$N^{ISO} \times f^{bb}_{ISO}/(\epsilon^{tag1}_{ISO} \times \epsilon^{tag2}_{ISO}) = N_{ISOSVT} \times f^{bb}_{SVT}/(\epsilon^{tag1}_{SVT} \times \epsilon^{tag2}_{SVT} \times \epsilon^{trig}_{SVT}) \tag{6}$$

Let the number of candidates satisfying $E^{\gamma}_T > 26$ GeV collected with the biased trigger be labeled by $N_{ISOSVT}$. By defining the ratio of events in the overlap region as $\epsilon^{PBJET} = \frac{N_{ISOSVT}}{N_{ISO}}$, we obtain

$$\frac{f^{bb}_{SVT}}{\epsilon^{tag1}_{SVT} \times \epsilon^{tag2}_{SVT} \times \epsilon^{trig}_{SVT}} = \frac{f^{bb}_{ISO}}{\epsilon^{tag1}_{ISO} \times \epsilon^{tag2}_{ISO} \times \epsilon^{PBJET}} \tag{7}$$

Now all the trigger- and jet-related quantities can be factored out in the cross section calculation for the SVT dataset and the cross section for the SVT dataset can be written as

$$\sigma_{SVT} = \frac{N_{SVT}}{\mathcal{L}} \times \frac{f^{bb}_{ISO}}{\epsilon^{tag1}_{ISO} \times \epsilon^{tag2}_{ISO} \times \epsilon^{PBJET}} \tag{8}$$

It is emphasized that this formula does not take into account the necessary corrections for the photon background since only jet-related quantities are considered here. Most importantly, this expression does not involve quantities that would rely on a simulation of the SVT. Instead the factors that would be complicated to calculate in the SVT dataset alone are calculated for the ISO dataset and any remaining difference (which amounts to and will from here be denoted as the *trigger efficiency*) is taken up in the $\epsilon_{PBJET}$ ratio. This simplified formula is only valid for the total signal until the photon background is subtracted. The minor modification necessary to accommodate this is described in Section 6.7.3.

### 6.3.1 Run dependence of the trigger efficiency

The isolation cuts on the photon part of the PHOTON_B_JET trigger were modified for the SVT dataset during the data taking period, resulting in a variation in the $\epsilon^{PBJET}$

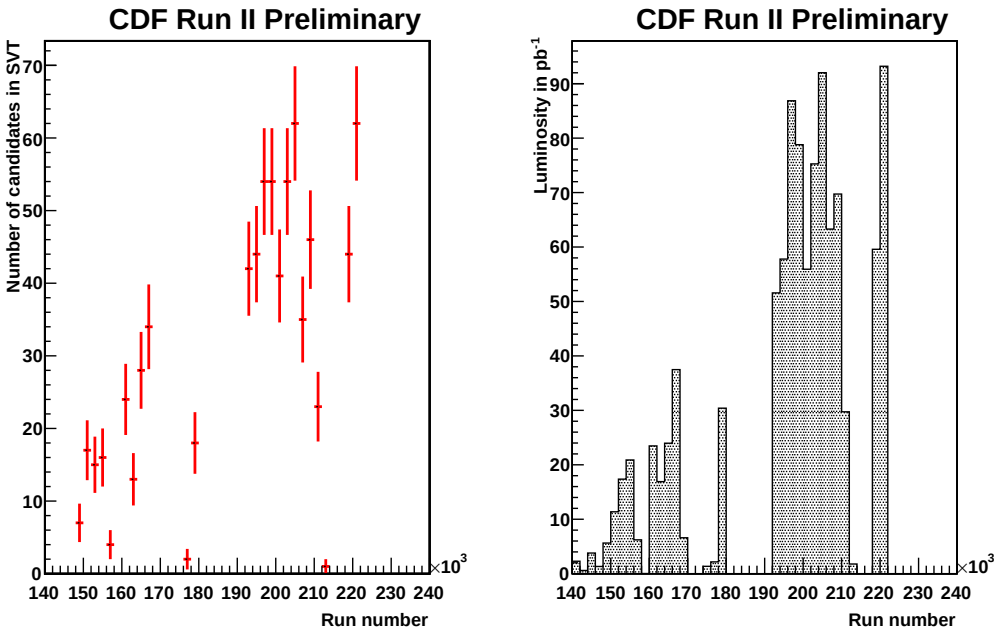


Figure 25: Left plot: number of candidates in the SVT dataset as function of run number. Right plot: collected luminosity as function of the run number.

ratio or trigger efficiency (see also [4]) as a function of time. The detailed requirements at L2 for the biased trigger are shown in Appendix B for versions 1 and 3[1] of this trigger.

Fig. 25 shows the number of candidates with $E_T(\gamma) > 12$ GeV in the biased dataset, $N_{SVT}$, as well as the collected luminosity as a function of the run number. For runs higher than 180'000, the candidates are not proportional to the integrated luminosity. This is due to the changing trigger conditions.

In the overlap region between both datasets, the number of candidates as a function of run number is shown in the left plot of Fig. 26. For the first period the efficiency of the biased trigger is quite high: most of the candidates from the unbiased trigger, which is fully efficient, are also present in the ISOSVT dataset. However, this changes with the L2 biased trigger. In later runs, the biased trigger is less efficient, collecting only every second event. It is nevertheless useful to use this trigger since the extended phase space accessible with the biased trigger makes up for this inefficiency.

The requirement that both jets satisfy $E_T > 20$ GeV is necessary because the b-tagging efficiency (see Section 6.6) falls sharply below that threshold. When determining the trigger efficiency, this requirement is not so crucial and can be dropped to

[1]The second version of the L2 part of the PHOTON_B_JET trigger path has been used in few runs and is not further discussed.

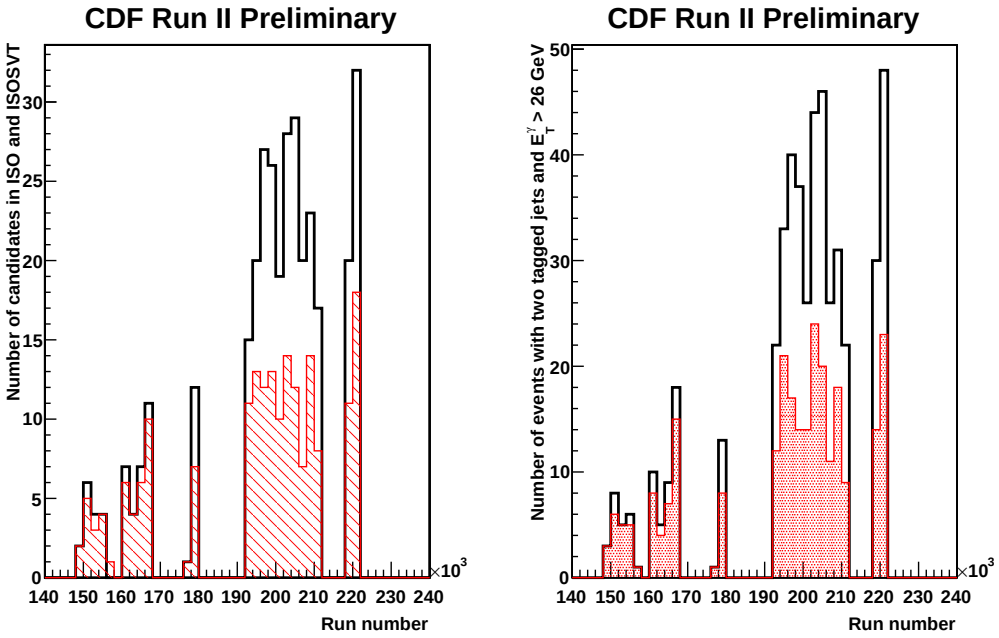


Figure 26: Left plot: candidates from the ISO dataset as function of run number. The shaded histogram shows those that are also selected by the biased trigger (ISOSVT dataset). Right plot: The same plot, but the minimal $E_T$ requirement on the tagged jets has been dropped, resulting in higher statistics.

obtain better statistics. That case is shown on the right side of Fig. 26. Now the actual trigger efficiency is defined as the ratio of events with a 26 GeV photon and two tagged jets found in the in the biased sample compared with the same selection in the unbiased dataset. This efficiency as a function of run number is shown in Fig. 27. The total number of events in the two cases shown in Fig. 26 is displayed in Table 3

| | $N_{ISO}$ | $N_{ISOSVT}$ | Ratio |
|---|---|---|---|
| Jets above 20 GeV | 334 | 189 | $(56 \pm 4)\%$ |
| All jets | 484 | 260 | $(53 \pm 3)\%$ |

Table 3: Trigger efficiency

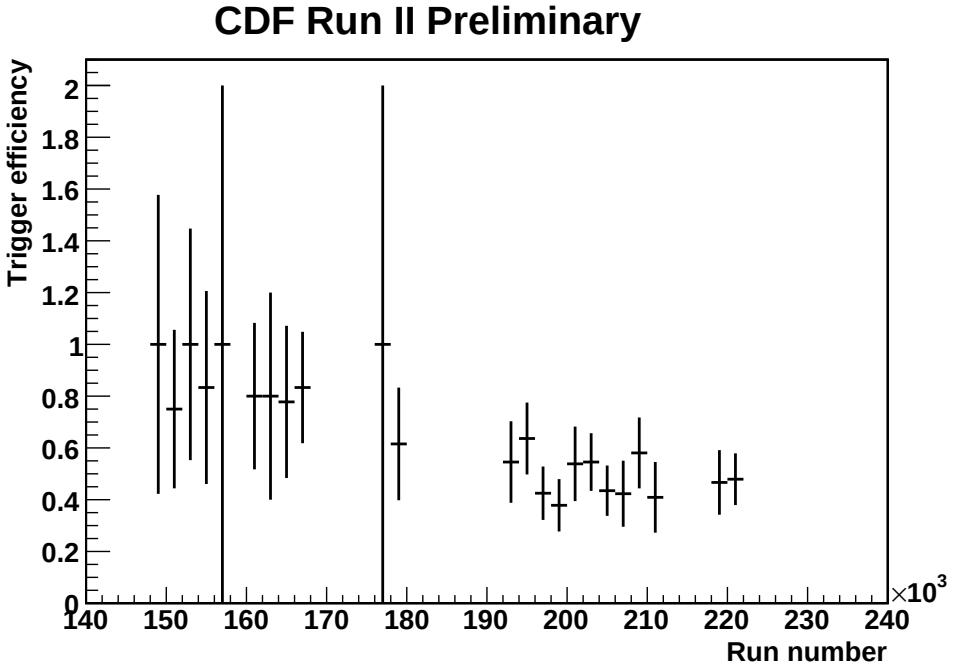


Figure 27: Efficiency of the PHOTON_B_JET trigger with respect to the PHOTON_25_ISO trigger. The efficiency is shown as a function of the run number.

## 6.4 Photon identification

The experimental signature of a photon is an energy cluster in the electromagnetic calorimeter ("em cluster") without any charged particle tracks pointing towards it. The measured energy is corrected for the detector response and denoted $E_T^{corr}$. The exact position of the cluster is determined with the CES shower detector (see Section 3.4.3) and required to lie within 21 cm in $r - \phi$ of the center of the em calorimeter tower (CES $|X| < 21$cm and $< 17$cm if $E_T^{corr} > 35$ GeV). In addition, the position in z (beam direction) must avoid the uninstrumented gap at the center (9 <CES$|Z|$ $< 230$cm).

The photon selection criteria (Photon ID in the following) aim to recognize direct photons, that is photons originating from the hard scattering. However jets usually contain one or more neutral mesons (such as $\pi^0$s) which decay to photons ($\pi^0 \rightarrow \gamma\gamma$). That results in a huge background since the jet cross section is about 1000 times higher than that of prompt photons. The criteria that are briefly described below are used to separate prompt photons from background. The efficiency of this selection was calculated by M. Campanelli in [4] and is shown in Fig. 28.

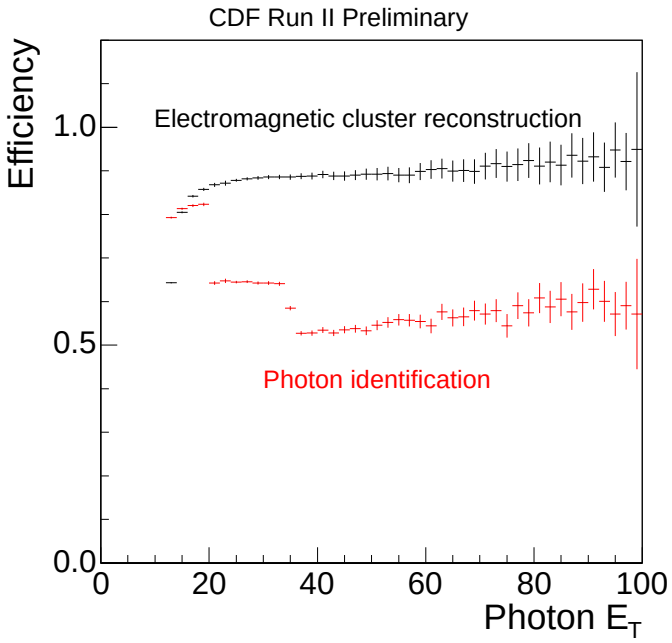


Figure 28: Efficiency for electromagnetic cluster reconstruction and photon identification from [4]. The two steps in the ID efficiency come from the second CES cluster (18 GeV) and the CES$|X|$ cut restriction at 35 GeV.

**Hadronic Leakage** The ratio of energy deposited in the hadronic calorimeter $E_{had}$ directly behind the em cluster is required to be small.

- $\frac{E_{had}}{E_{em}} < 0.125$ or $< 0.055 + 0.00045 \times E_T^{corr}$

**Isolation** The energy deposited in the calorimeters around the em cluster is requested to be less than a defined value. The severity of this isolation cut depends on $E_T^{\gamma}$. Let $E_T^{corr,ISO}$ the difference between the energy deposited in a cone of radius $\Delta R < 0.4$ around the candidate, in both the em and hadronic calorimeters and the energy of the em cluster.

- $E_T^{corr,ISO} < 0.1 \times E_T^{corr}$ for $E_T^{corr} < 20$ GeV
- $E_T^{corr,ISO} < 2.0 + 0.02 \times E_T^{corr}$ for $E_T^{corr} > 20$ GeV

This removes most of the QCD background except for the case when the pion is carrying almost all the momentum of the jet. This represents only one case in 1000 but since the jet cross section is also approximately that much higher, the isolation cut leaves us with a signal-to-background ratio of approximately 1.

**Track rejection and isolation** Photons being uncharged, there should be no charged particle track associated with the em cluster. However, one track pointing to the em cluster may remain acceptable in a photon candidate if the $p_T$ is not characteristic of an electron when compared to the deposited energy. Thus the maximum allowed $p_T$ of this track is limited by a constant term but can grow with the corrected energy of the em cluster $E_T^{corr}$. Since jets contain charged particles that leave tracks, there is also an isolation cut on the $p_T$ sum of all tracks in a cone around the em cluster.

- At most one track associated to the em cluster, provided it satisfies $p_T < (1 + 0.005 \times E_T^{corr})$ GeV
- A requirement on the transverse momenta of tracks within a cone of $\Delta R < 0.4$ around the em cluster that $\Sigma p_T < 2.0 + 0.005 * E_T^{corr}$ GeV.

**Second CES cluster** If a second cluster with consequent energy indicating a close second photon is identified, the event is rejected. This again aims to reject events where an energetic meson has decayed to two photons.

- $E_T$(2nd CES cluster) $< 0.14 * E_T^{corr}$ if $E_T^{corr} < 18$ GeV
- $E_T$(2nd CES cluster) $< 2.4 + 0.01 * E_T^{corr}$ if $E_T^{corr} > 18$ GeV

**CES $\chi^2$(Strip+Wires)** The shower shape of the photon at the maximum lateral extension of the shower is compared with a fit to test beam data and a $\chi^2$ is evaluated. The $\chi^2$ distribution of the CES fit for true photons and for background is shown in Fig. 29.

- $\chi^2$(Strip+Wires)$< 40$

### 6.4.1 Estimation of the fake photon background

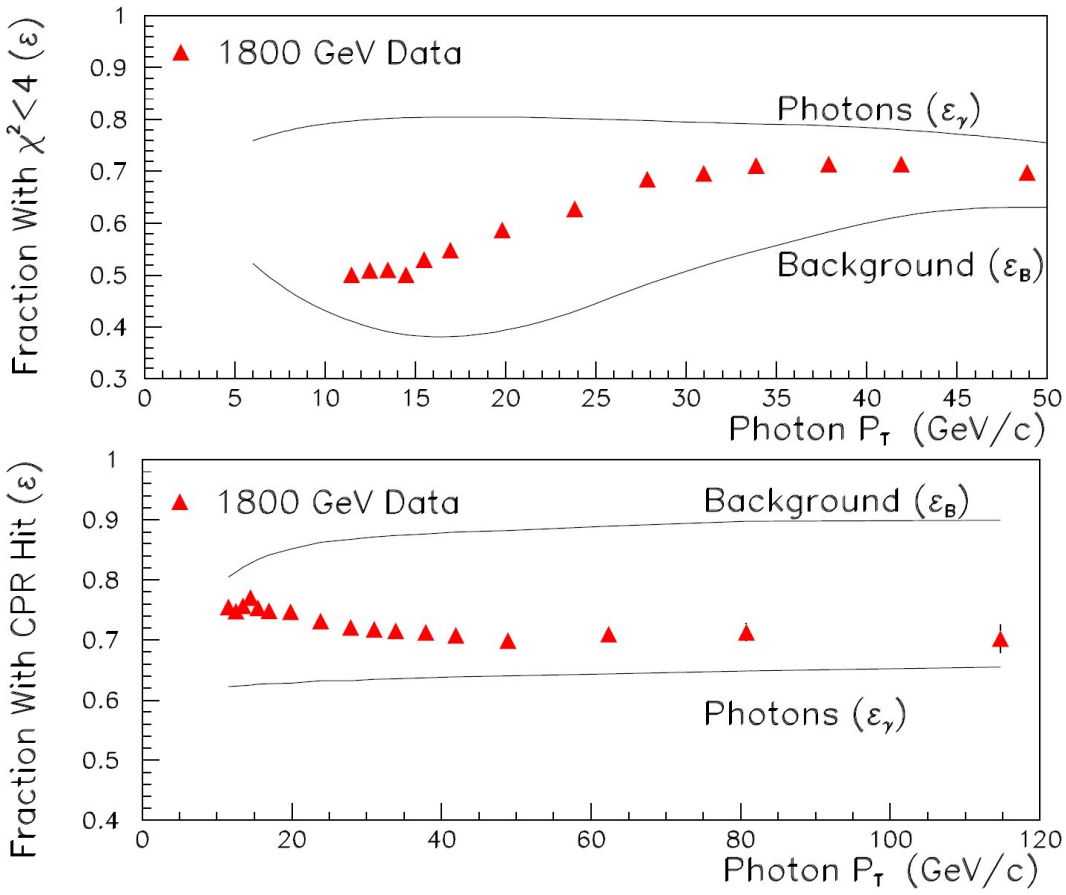


Figure 29: Upper plot: Fraction of events passing the CES $\chi^2$ test in Run I candidates satisfying the photon ID selection (triangles), pure $\gamma$ test beam (upper line) and background jets (lower line). Lower plot: Fraction of events with a CPR hit in Run I collision data for photon candidates (triangles), test beam photons (lower line) and background jets (upper line). See [25].

In candidates passing all the ID cuts from the previous paragraph, only a statistical method allows the separation of true photons from the large hadronic background. There are currently two methods used to separate background from signal, both valid in different transverse energy ranges.

**CES method** For $E_T(\gamma) < 35$ GeV, the transverse profile of the electromagnetic shower measured in the CES is compared to pure photon test beam results. The profile is different for single (prompt) photons and for photon pairs from pion decays. The quantity used is the $\chi^2$ of a fit (see Fig. 29, upper plot), but cannot clearly assign a candidate to signal or background. This is because signal as well as background can

pass the test, but the fraction of events passing it is different in both cases, which allows a statistical separation. Also, as the energy of the candidate rises, the 2-photon shower profile increasingly resembles the single-photon shower profile. Because of this, the CES method is used for candidates with $E_T^\gamma < 35$ GeV.

**CPR method** For higher photon energies ($E_T(\gamma) > 35$ GeV), the conversion probability before the electromagnetic calorimeter is used. Some photons convert in an $e^+e^-$ pair in the coil of the solenoid, before entering the calorimeters. This can be monitored by a charge deposition in the CPR detector. For a single $\gamma$, this probability is of approximately 65%. For two photons coming from a pion decay, the probability that at least one photon converts to an $e^+e^-$ pair in the coil reaches 85 to 90% (see Fig. 29, lower plot). These different conversion probabilities can again be used to estimate the signal to background ratio.

**Statistical separation** With each of these methods, the fraction of real prompt photons in candidate events can be determined. Let there be a quantity $X$ measured for true photons and background. Let the fraction of background events passing this cut be $\epsilon_B$ while the fraction of true prompt photons passing the cut be $\epsilon_\gamma$. These fractions as function of the candidate $E_T$ are known from test beam data (shown in Fig. 29).

For each of the aforementioned methods, one knows the fraction in signal and background events that pass the test on the quantity $X$. The next step is to measure this fraction in the candidate events, in that case a mix of true photons and background. More formally, the information available now is:

- $N_p$ the number of candidates passing the test;
- $N_f$ the number of candidates failing the test;

The numbers of interest are the number of signal and background candidates $N_\gamma$ and $N_B$. Since the fraction of background and signal events that pass the test are known ($\epsilon_B$ and $\epsilon_\gamma$), one can write

$$\begin{aligned} N_p &= \epsilon_B \times N_B + \epsilon_\gamma \times N_\gamma \\ N_f &= (1-\epsilon_B) \times N_B + (1-\epsilon_\gamma) \times N_\gamma \end{aligned}$$

From the total number of candidates $N_{tot} = N_p + N_f$, the number of true photons in candidates is

$$N_\gamma = (N_p - \epsilon_B \times N_{tot})/(\epsilon_\gamma - \epsilon_B) \tag{9}$$

This is evaluated in the following way. A weight is assigned to each candidate depending on wether it passes ($w_p$) or fails ($w_f$) the test:

$$w_p = \frac{1-\epsilon_B}{\epsilon_\gamma - \epsilon_B} \tag{10}$$

$$w_f = \frac{-\epsilon_B}{\epsilon_\gamma - \epsilon_B} \tag{11}$$

Now the sum of weights $\Sigma_p w_p + \Sigma_f w_f$ of all candidates gives the estimated number of true photons as in Equ. 9. In addition, the statistical uncertainty on this quantity is simply given by

$$\sigma(N_\gamma) = \sqrt{\Sigma_i (w_i)^2} \tag{12}$$

Since the weights can take absolute values larger than one, the uncertainty on the true number of photons is significantly larger than that on the number of candidates. This can be quite a handicap if, as in this analysis, the sample has limited statistics. In addition to the number of true photons, a quantity that is used in the calculation of the cross section is the *photon purity*, that is the ratio of the estimated number of true photons over the total number of candidates (see Fig. 30). It is measured for each bin in $E_T^\gamma$.

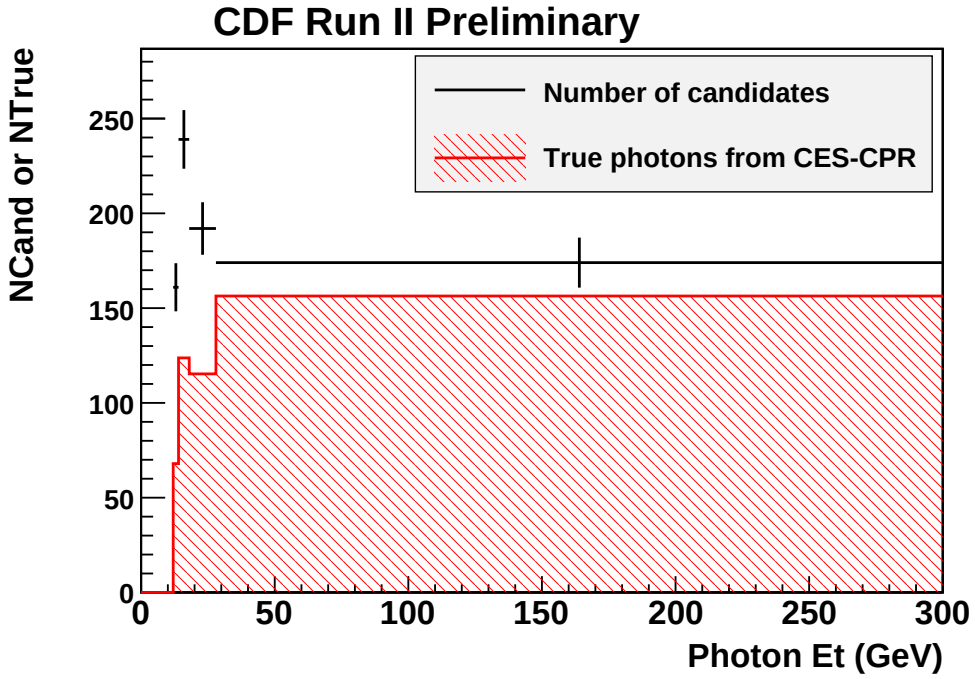


Figure 30: Number of candidates per photon $E_T$ bin for events with two tagged jets. The crosses show the actual number of candidates while the shaded histogram shows the sum of the CES-CPR weights, e.g. the estimated number of true prompt photons.

To test this method, the photon purity has also been measured in different cases. True candidates have a photon candidate and two tagged jets(see Section 6.6). Since the total number of candidates in the biased dataset is of approximately 800, the uncertainty due to the photon weights is large and the purity estimate fluctuates significantly. If one only asks for two jets passing the selection criteria, as described in the following Section 6.5, the statistics are much larger (122'000 events) and a measurement of the

photon purity can be obtained with reduced uncertainty. However, the photon purity does not only depend on the kinematics of the process, so the purity derived from photon plus two-jet events is not applicable to the $b\bar{b}\gamma$ case. To prove that point, the photon purity has also been measured in a sample with two jets and a photon, but with the added requirement that one jet must be positively tagged (13'615 events). Figure 31 shows that the requirement of a tagged jet and a generic one instead of two generic jets significantly changes the photon purity. The additional requirement of a second tagged jet further changes the purity, but this effect is smeared by the much larger uncertainty due to the low statistics.

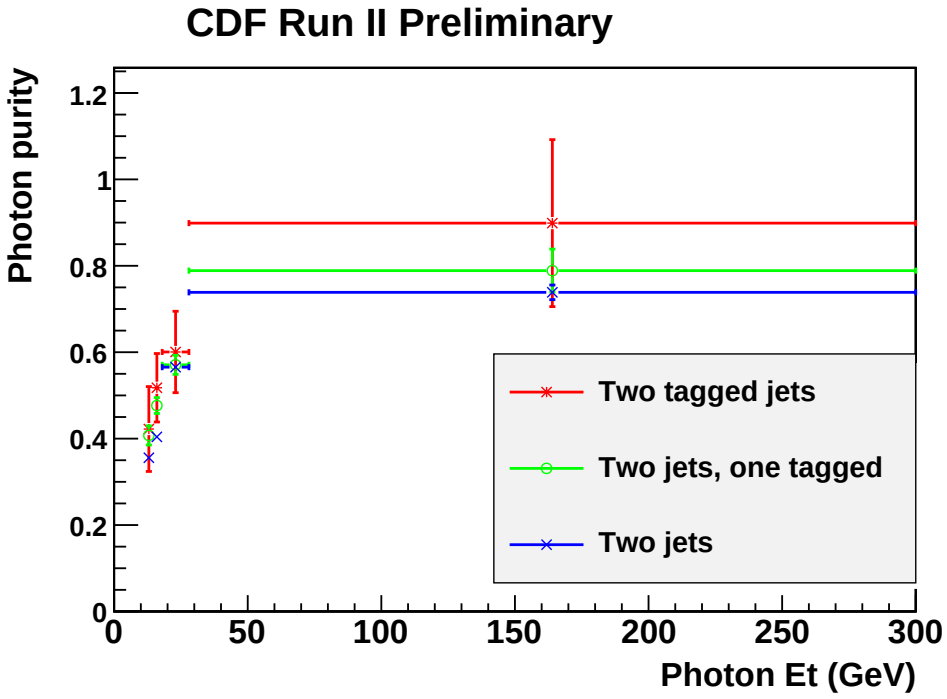


Figure 31: Photon purity as function of photon $E_T$. The purity is the ratio of the CES-CPR estimate of true photons to the total number of candidates. The difference in purity coming from the different signatures requested is clearly visible.

## 6.5 Jets

Unlike photons, the fundamental partons from the hard interaction are not detected directly. Instead they fragment into a large number of more stable hadrons that are roughly collimated in a cone originating at the primary interaction vertex. In the detector, these particles are seen as a number of collimated tracks from charged particles and energy deposition in both the electromagnetic calorimeter (the shower also contains photons and leptons) and the hadronic calorimeter. This is known as a jet, coming from the quark that has participated in the hard interaction.

Several algorithms can be used to identify jets. I briefly explain the jet clustering algorithm (JetClu), as it is the algorithm used in this analysis. The first object from which a jet is constructed is a seed tower in the calorimeter with energy exceeding some given threshold. The surrounding calorimeter towers are then added to the jet, if they lie within a cone of radius $\Delta R < 0.4$[1]. The direction of the jet in this algorithm is the energy-weighted average of the direction of each contributing tower.

$$E_T^{jet} = \Sigma E_T^{tower}$$
$$\eta^{jet} = \frac{\Sigma(\eta^{tower} \times E_T^{tower})}{E_T^{jet}}$$
$$\phi^{jet} = \frac{\Sigma(\phi^{tower} \times E_T^{tower})}{E_T^{jet}}$$

If the resulting jet candidate is stable with respect to small changes in direction, the candidate is retained. Otherwise it is rejected and another seed is used until stability is iteratively achieved.

### 6.5.1 Jet energy corrections

The hadronic calorimeter does not have a uniform or linear response in energy. Additional corrections to the measured energy of the jets are therefore needed. These corrections are described below.

**Relative jet correction** This correction accounts for variations in the detector response as a function of the pseudorapidity $\eta$. The calorimeter response was calibrated with exclusive di-jet events where both jets must physically have the same transverse energy.

**Multiple interactions** The energy from other $p\bar{p}$ interactions that occur during the same bunch crossing can leak into the jet cluster and increase the measured energy. This correction was measured with minimum-bias events as a function of the luminosity and the average contribution is subtracted from the jet energy.

[1] Depending on the analysis, different cone sizes, for example 0.7 or 1.0. may be used.

**Absolute jet energy correction and Energy Scale Corrections** Absolute energy corrections account for the non-linear response of the calorimeter. They were measured during Run I, using jet-photon balancing. Since the response of the detector changes with time, a Jet Energy Scale (JES) has been defined that corrects this time dependence.

In this analysis, jets are reconstructed with the JetClu algorithm using a cone of $\Delta R = 0.4$, and the energy corrections mentioned above are used. For candidate events, two central ($|\eta| < 1.5$) jets are required, both having an offline corrected $E_t > 20$ GeV. Both jets are required to lie outside of a cone of radius $\Delta R = 0.7$ surrounding the photon candidate.

## 6.6 The SecVtx b-jet tagging algorithm

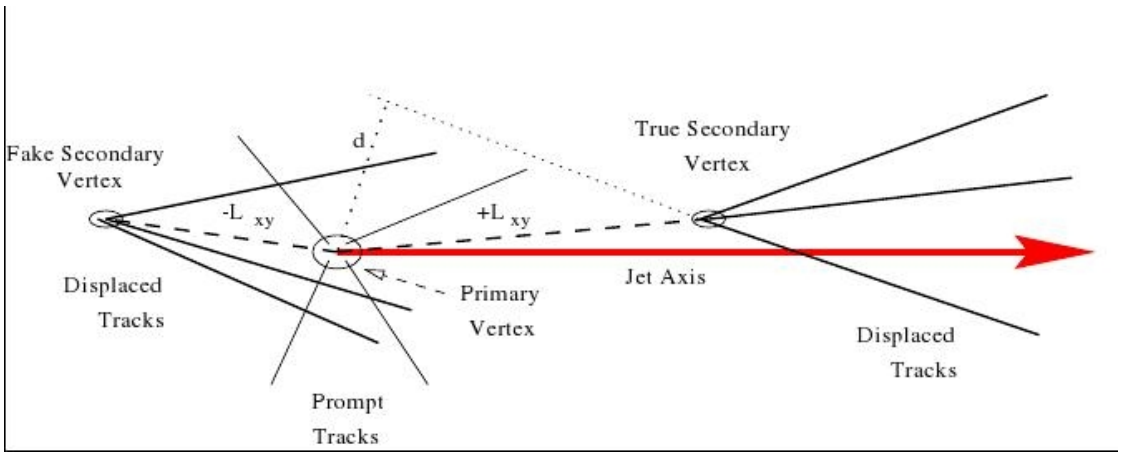


Figure 32: Difference between a positive and a negative tag. The distance $L_{xy}$ from the primary to the secondary vertex is positive for a true secondary vertex.

The SecVtx algorithm used in this analysis aims to reconstruct the secondary vertexes already mentioned in Section 4.2. It first selects good-quality tracks with a significant impact parameter $d_0$ (see Fig. 32 and Fig. 20) with respect to the beamline. The significance of the impact parameter (the quality of the track reconstruction fit) must satisfy $|d_0/\sigma(d_0)| > 2.5$. The algorithm then attempts to reconstruct a secondary vertex using these tracks. If several tracks converge transversely away from the primary vertex, the secondary vertex is reconstructed. The jet is tagged if the distance in $r-\phi$ between the vertexes ($L_{xy}$) has a significance ($|L_{xy}/\sigma(L_{xy})| > 7.5$). Reconstructed secondary vertexes can sometimes appear to lie behind the primary vertex (see Fig. 32), which is mostly due to the limited resolution in reconstructing the primary vertex . A jet is "positively" tagged if the secondary vertex is in the right direction.

The SecVtx algorithm does not successfully tag all jets containing a B hadron, so that an efficiency must be taken into account. This b-tagging efficiency is measured

from a $b\bar{b}\gamma$ Monte Carlo sample. Jets satisfying $|\eta| < 1.5$ and including a B hadron contained in the jet cone ($\Delta R$(B hadron, jet)$< 0.4$) are collected. Within those, the jets that have been positively tagged are selected. The ratio of the tagged jets over all b-jets is the b-tagging efficiency, shown as a function of the jet $E_T$ in Fig. 33. When applied to data, a multiplicative scale factor of $0.94 \pm 0.03$ on the efficiency for each jet is used, since it has been shown that there was a disagreement in the tagging efficiency between data and Monte Carlo (see [26]).

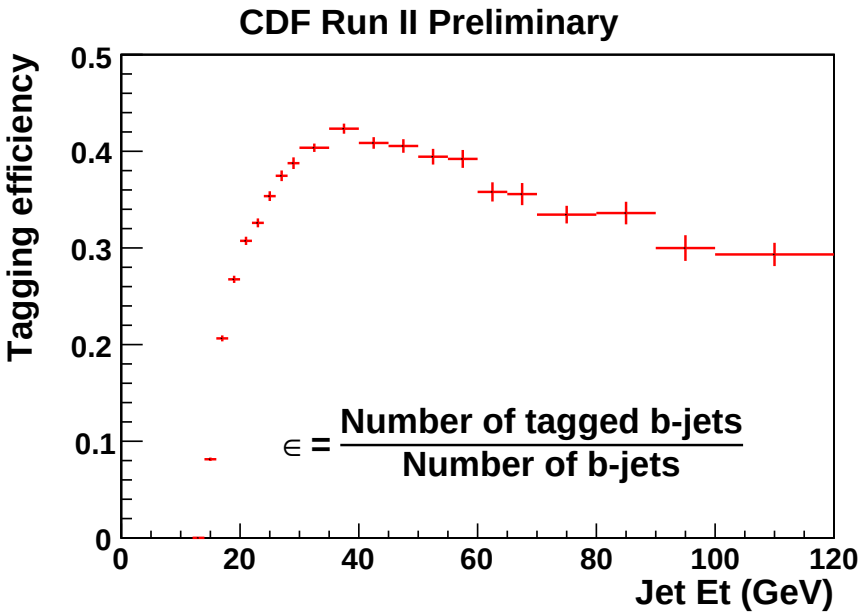


Figure 33: Tagging efficiency for b-jets as a function of the $E_T$ of the jet. At lower transverse energies, the B hadron has only a small relativistic boost so that the secondary vertex is closer to the primary vertex and is more difficult to detect. At high $E_T$ values, the particles from the B hadron decay are more closely collimated which again makes the reconstruction difficult.

## 6.7 $b\bar{b}$ content of the sample

A tagged jet is not necessarily a b-jet. Charmed hadrons also feature a rather long decay length (around 200 $\mu$m) and relatively high masses so that a sample of tagged jets will contain b-jets, but also jets from c-quarks or mistags of light quark or gluon jets. To estimate the fraction of events with two tagged b-jets within the sample of candidates with a photon and two tagged jets, the mass of the secondary vertex in each jet is reconstructed. If the reconstruction was perfect there would be clear peaks on certain masses and one could clearly separate b-jets from others. However neutral particles that are not detected deteriorate the reconstructed mass resolution, so

that no clear distinction remains on an event-by-event basis. Once again a statistical separation can be used to separate the signal from the background. A two-component fit of the distribution of the sum of the masses of both secondary vertexes is performed. Templates are built from MC events for the signal (in this case two b-jets) and for the background (all other combinations of two tagged jets). The fraction $f_{ISO}^{b\bar{b}}$ of real $b\bar{b}$ events in events with two positively SecVtx-tagged jet is measured from the ISO dataset, since the SVT distorts the mass distribution. The restricted statistics do not allow to measure this fraction differentially as a function of the $E_T$ of the jets.

In the cross section calculation of Section 6.3, it is implicitly assumed that the fake photon background has the same $b\bar{b}$ content as the signal, so that efficiency and purity for the photon and jet parts of the analysis can be factorized. Since this is not necessarily true, the cross section needs to be corrected for the different fractions as a function of the photon purity. As already mentioned, the background for prompt photons is mainly due to $\pi^0$ decays in ordinary QCD events where a hadronic jet is mistaken for a photon. A second fit of double-tagged events has been carried out on a sample of 3-jet events[1], where the selection of the first two jets mirrors that used in the photon samples, where the third jet has the same kinematic properties as the photon, but where the photon ID requirements are dropped.

To obtain the number of true $b\bar{b}$ in true photon events $N_{bb\gamma}^{true}$ we have to account for the different b fractions in fake photon events. Let $f_{bb}^{cand}$ be the $b\bar{b}$-purity in candidate events including a photon and $f_{bb}^{bkg}$ that for the jet background. From the number of candidates $N_{tot}$ and the number of real photons $N_\gamma$ estimated using the CES-CPR method, the number of fake photons is $(N_{tot} - N_\gamma)$. The true number of $b\bar{b}\gamma$ events is then given by

$$N_{bb\gamma}^{true} = N_{tot} \times f_{bb}^{cand} - (N_{tot} - N_\gamma) \times f_{bb}^{bkg} \tag{13}$$

### 6.7.1 Templates for the fit

The background template is composed of two tagged jets that are not both b-jets. For example a $(\gamma + c\bar{c})$ event is considered background. Tagged jets are divided in three categories, following the flavor of the quark from which the jet originates: b-jets ("b"), charm jets ("c") or light quark or gluon jets ("l"). All combinations for two tagged jets such as $(c + b)$, $(l + c)$ or $(l + b)$ are considered background.

The templates are obtained from MC events. The $b\bar{b}$ template comes from the dedicated $b\bar{b}\gamma$ MC (see Section 2.6). The non-$b\bar{b}$ template is obtained in a different way. The tagging efficiency for charm or light jets is smaller than for b-jets. The fact that two tagged jets are required implies that these efficiencies are multiplied, resulting in an even lower total efficiency, that would impose a prohibitively large photon plus two jets MC sample. Instead of this, a sample of tagged jets from different (photon + 1 tag) events is collected from an inclusive 10 GeV photon MC dataset. The vertex mass from these tagged jets is stored and summed pairwise to obtain the background

[1]These events are collected with the inclusive jet trigger JET20 that requests a jet with corrected $E_T > 20$ GeV. This trigger is strongly prescaled.

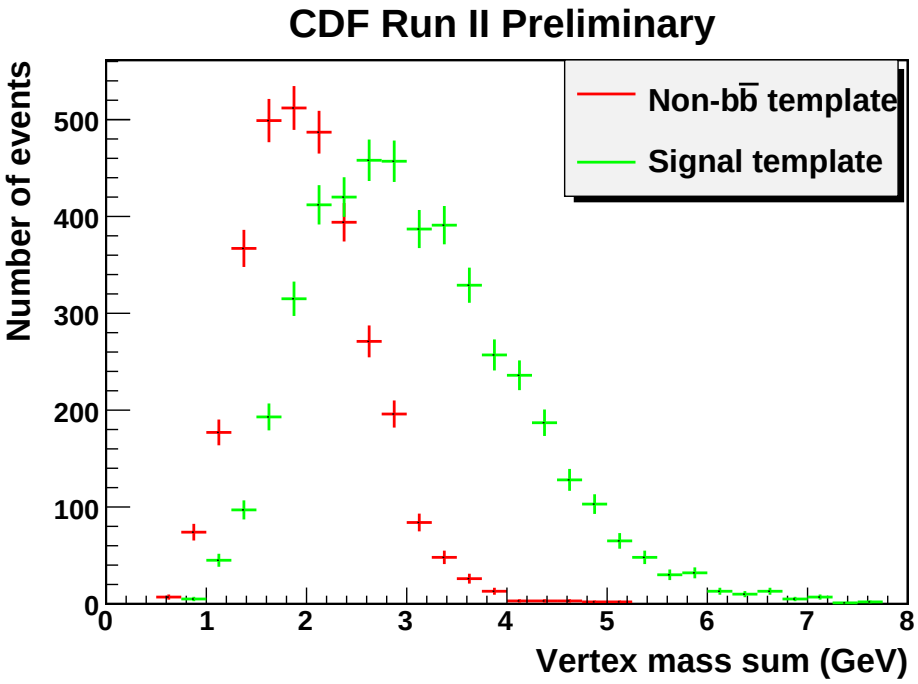


Figure 34: The templates for the $b\bar{b}$ and non-$b\bar{b}$ contributions as they are used in the fit. The templates are not normalized since TFractionFitter (see Sec. 6.7.2) calculates Poisson uncertainties for individual bins.

template. The relative normalization of the different contributions ((l+l), (l+c), (c+b) etc.) is derived from a parton-level photon+dijet MC sample.

As a test of this method, the $b\bar{b}$ signal template is shown in Fig. 35, where it is either directly measured from b-jets belonging to the same event or assembled from jets in different events. No significant difference can be found between the two distributions. This method of obtaining the template has already been used in a previous $b\bar{b}$ analysis [12].

### 6.7.2 Fit to secondary vertex mass

The ROOT routine TFractionFitter [27] is used to fit the data from the unbiased dataset. It employs a binned maximum likelihood technique and Poisson statistics. The originality of this procedure is that uncertainties from the finite MC templates are also taken into account.

Due to the limited statistics available in the ISO sample, we cannot perform the fit in single jet $E_T$ bins, so the $b\bar{b}$ content is determined for the whole sample. The fit result is shown in Fig. 36 and Table 4.

We see that the $b\bar{b}$ purity of this sample is high, in contrast to what happens in the case of a single-b measurement. This is expected since the requirement of a second

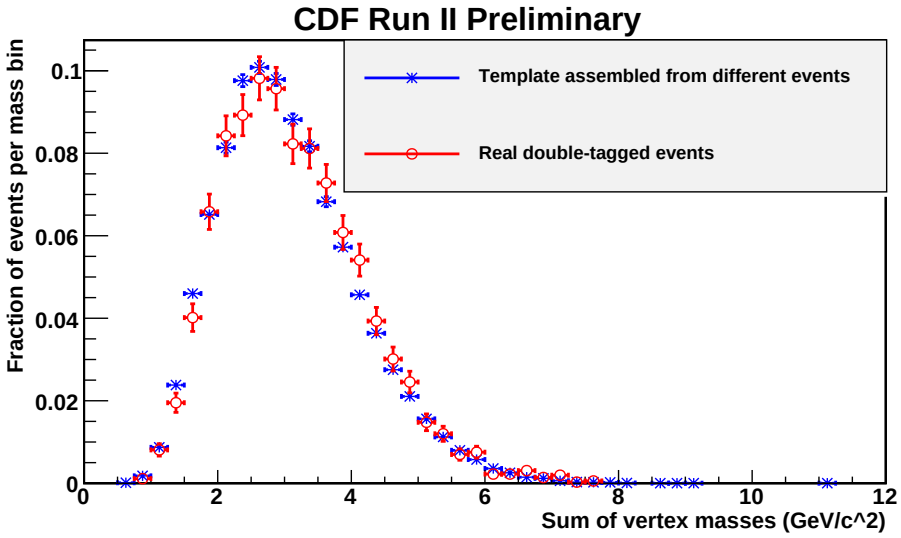


Figure 35: The Vertex mass sum for b-jets has been calculated when the two tagged jets come from the same event (red) and when they come from different events (blue). Both distributions are normalized to unit area.

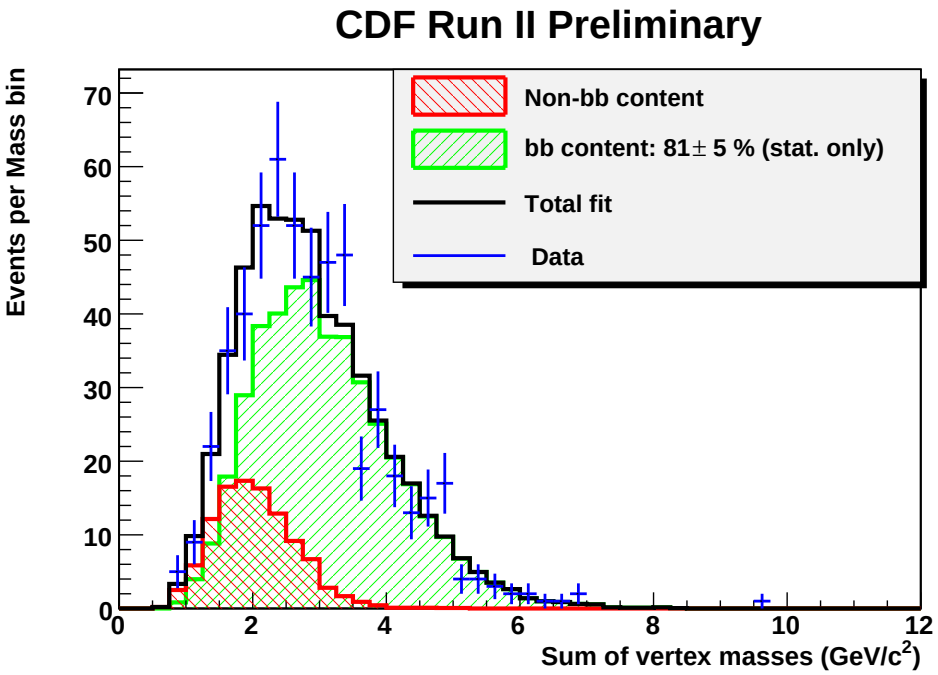


Figure 36: The $b\bar{b}$ purity for candidates.

| Fit result | Fit error | $\chi^2/ndf$ | Fit probability |
|---|---|---|---|
| 80.6% | 5.2% | 22.1/23 | 0.51 |

Table 4: The $b\bar{b}$ purity in the ISO sample as result of the vertex mass fit.

positively-tagged jet significantly enhances the purity, and makes this measurement much more robust with respect to variations in the shape of the background template and relative normalization of its components.

### 6.7.3 Fake photon background composition

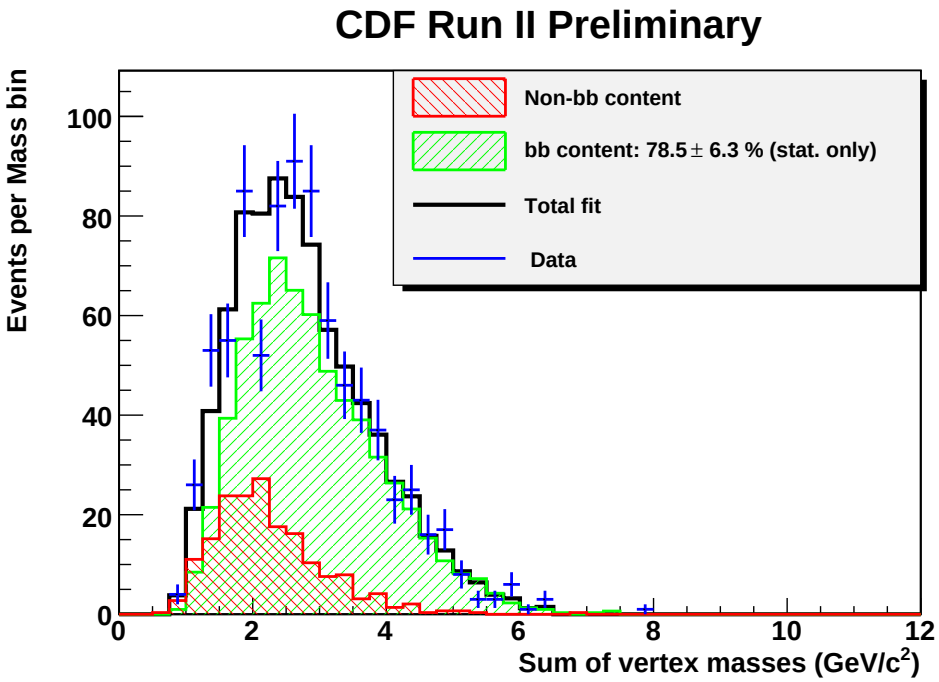


Figure 37: $b\bar{b}$ content of the fake photon background.

For the fake photon background sample, both templates are obtained from a large PYTHIA sample of generic QCD dijet events. The background template is obtained without assembling vertexes from different events. The fit again yields a high purity of the order of 80 % (see Fig. 37 and Table 5). However, the non-$b\bar{b}$ template contains few events with light quarks due to its limited statistics, since only approximately 500 out of 6 million MC events have 2 tagged jets which are not both associated to B hadrons. The influence of the light tagged jets, especially important to account for the low values of the vertex mass sum below 2 GeV appears to be underestimated, resulting in a high $\chi^2$. The low data point just above 2 GeV is also contributing to this.

| Fit result | Fit error | $\chi^2/ndf$ | Fit probability |
|---|---|---|---|
| 78.5% | 6.3% | 36/19 | 0.05 |

Table 5: The $b\bar{b}$ purity in fake photon background as result of the vertex mass fit.

# 7 Results and systematic uncertainties

## 7.1 Measured cross section

The cross section is measured as a function of transverse energy of the photon, the leading jet and the second-to-leading jet. Events are stored in a three-dimensional matrix having as axes these transverse energies. In this way, efficiencies can be applied in different $E_T$ bins for each object.

The number of candidates $N_{tot}$ and the number of real photons $N_p$ from the CES-CPR methods give the number of fake photons as $(N_{tot} - N_p)$ in each photon $E_T$ bin. The $b\bar{b}$ fractions are corrected according to photon purity (Fig. 31). The correction factor to obtain the $b\bar{b}\gamma$ final state in each bin is then

$$F_{bb\gamma}^{true} = \frac{f_{bb}^{cand} - (1 - N_p/N_{tot}) \times f_{bb}^{bkg}}{\epsilon^{tag1} \times \epsilon^{tag2} \times \epsilon^{trigg} \times \mathcal{L}}$$

The differential cross sections are obtained by projecting this three-dimensional matrix on each axis, properly summing the statistical errors. The results are shown in Fig.39, Fig. 40 and Fig. 41. Angular distributions are measured in a similar way, from three-dimensional matrices where the $E_T$ of the second jet is replaced by the angular variable, and a weighted average value for the corrections on the second jet is applied (see Fig. 42 and Fig. 43).

From an integration of these differential cross sections we obtain a total cross section of

$$\sigma = \mathbf{8.60 \pm 1.07(stat.)pb}$$

to be compared to a leading order QCD prediction (from Madgraph) of $6.02 \pm 0.10(stat.)pb^{-1}$ in the reconstructed MC. The 30% discrepancy observed can be expected from the purely leading-order theoretical calculation. To estimate the effect of theoretical uncertainties, the factorization and renormalization scales in the MC generation were varied by a factor of $2^{\pm 1}$. The predicted cross section varies by $^{+30}_{-20}\%$ respectively.

There is no NLO calculation of this final state available at the time of writing.

## 7.2 Systematic uncertainties in the data

Any cross section measurement is directly sensitive to the luminosity measurement, which has an uncertainty of $\pm 6$ %. The scale factor for the b-tagging efficiency has an uncertainty of $\pm 3$ % [26], and since it is applied on both tagged jets, the systematic uncertainty associated with this is $\pm\sqrt{2} \times 3$ %.

The tagging efficiency for jets containing more than one b-hadron is increased by about 20 %. However, since in this study we consider candidates with two tagged jets, we expect the number of such events to be small[1] and neglect this effect.

[1] As we have seen, the $b\bar{b}$ purity of a sample of events with two tagged jets is rather high. If two b-quarks are confined in a single tagged jet, the other tagged jet would have to come from another quark. That means either events with 3 b-quarks or 2 b-quarks in one jet and a mistag. The probability of this occurring is estimated to be small.

The jet energy scale (JES) is varied by $\pm 1\sigma$ to determine its impact on the cross section measurement. The total number of candidates with the nominal value of the JES is 766. When the JES is increased by $1\sigma$ (higher energy per jet, so that some events are selected that were not before), this number increases to 849 (plus 11 %). On the contrary, lowering the JES will give the jets a lower $E_T$ and some events are not selected anymore. The number of candidates decreases to 664 (minus 13 %).

**Systematic error associated with the non-$b\bar{b}$ template** To check the stability of the method of fitting b-quark composition as described in Section 6.7the composition of the non-$b\bar{b}$ template has been varied. To do this, the tagging efficiency for each type of quark (light, charm or beauty) has been modified by a factor of $2^{\pm 1}$. Let $r_{ll}$ be the relative fraction of light-light events in the non-$b\bar{b}$ template, $r_{lc}$ that for light-charm combinations and so on for $lb, cc$ and $cb$, and $f_{bb}$ the $b\bar{b}$ fraction from the fit. The first line of Table 6 shows the composition and fit result for the default values, while the following lines show the results for the modified background template, which as expected from the high purity do not vary much. The fit is most sensitive to contributions from light quarks. A higher content of light quarks (lines 2 or 5 of Table 6) makes the non-$b\bar{b}$ template shift to lower masses and the $b\bar{b}$ fraction must make up the missing contribution in middle values of the mass sum. In that case the estimated purity in the fit is higher.

The deviations in each direction from the default fit result are summed up in quadrature and considered as a systematic error. This asymmetric systematic error is of $\delta\sigma_{sys}/\sigma_{sys} = {}^{+4.2}_{-2.5}$ %.

| $r_{ll}$ | $r_{lc}$ | $r_{lb}$ | $r_{cc}$ | $r_{cb}$ | Fit result | Fit error | Fit $\chi^2/ndf$ | Fit probability |
|---|---|---|---|---|---|---|---|---|
| 15% | 17% | 8% | 54% | 6% | 80.6% | 5.2% | 22.1/23 | 0.51 |
| 36% | 20% | 10% | 31% | 3% | 84.0% | 5.0% | 22.3/23 | 0.50 |
| 5% | 11% | 5% | 71% | 7% | 78.6% | 5.3% | 23.3/23 | 0.44 |
| 5% | 12% | 3% | 75% | 4% | 79.3% | 5.3% | 24.1/23 | 0.40 |
| 31% | 17% | 17% | 28% | 6% | 83.0% | 5.1% | 21.9/23 | 0.53 |
| 14% | 15% | 14% | 47% | 10% | 80.0% | 5.2% | 22.3/23 | 0.50 |
| 17% | 18% | 4% | 57% | 3% | 80.9% | 5.1% | 22.2/23 | 0.50 |

Table 6: $b\bar{b}$ fraction stability

**Tracking efficiency** The final source of systematics is the difference in tracking efficiency in data and MC. This effect can modify the secondary vertex invariant mass distribution used to obtain the $b\bar{b}$ purity. To estimate this systematic error, a parametrization of the efficiency for each track in data and MC is used [28]. It estimates the reconstruction efficiency for a track using variables such as the $p_T$ of the track and the distance to the jet core. The difference in efficiency between data and MC is of the order of 3%.

Tracks are now randomly removed from the secondary vertex according to the ratio of the two efficiencies, and the invariant mass of the sum of all remaining tracks composing the secondary vertex is computed. The SECVTX algorithm requires at least either 3 tracks of intermediate quality or 2 tracks that are very well reconstructed. If the number of remaining tracks after this procedure does not satisfy this condition, the jet is simply not tagged anymore. The total effect is a distortion of the secondary vertex mass template. Figure 38 shows the two templates used in the fit to the data. The fit to the data with this modified template gives a purity of $73.7 \pm 6\%$, to be compared to the nominal purity of approximately 80%. This 8% difference is taken as a systematic error.

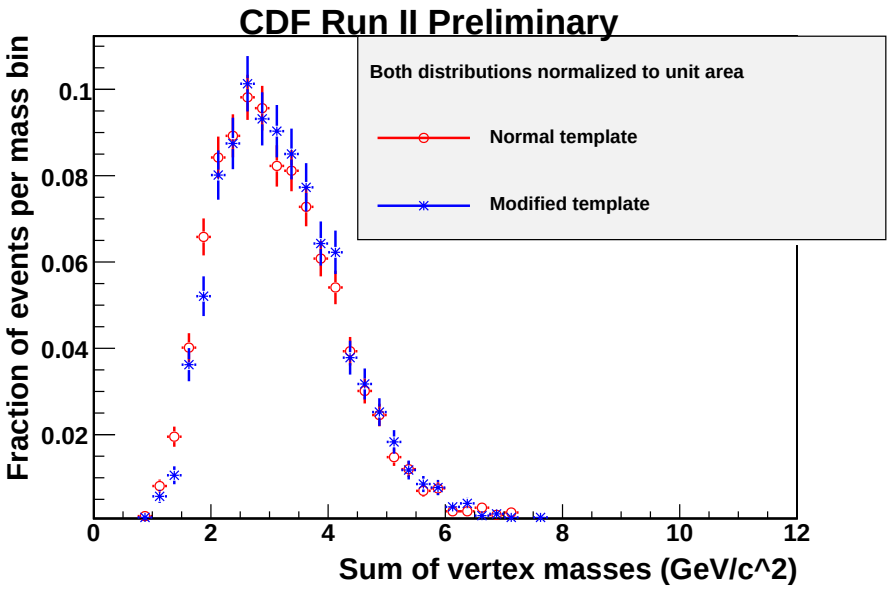


Figure 38: The two templates for the $b\bar{b}$ component of the fit to the mass sum of both secondary vertexes. In red is the original template and in blue the one in which tracks are randomly killed according to their efficiency. Both distributions are normalized to unit area.

| Source of systematics | (%) |
|---|---|
| Luminosity | $\pm 6$ |
| b-tagging scale factor (2 jets) | $\pm 4.2$ |
| non-$b\bar{b}$ template | $^{+4.2}_{-2.5}$ |
| Jet energy scale | $^{+10.8}_{-13.3}$ |
| Tracking efficiency | $\pm 8$ |
| Photon efficiency | $\pm 1$ |
| Photon background | $\pm 5$ |
| Total | $+16.7 - 18.1$ |

Table 7: Systematic errors for the total cross section

## 7.3 Comparison of the measured cross section with the theoretical prediction

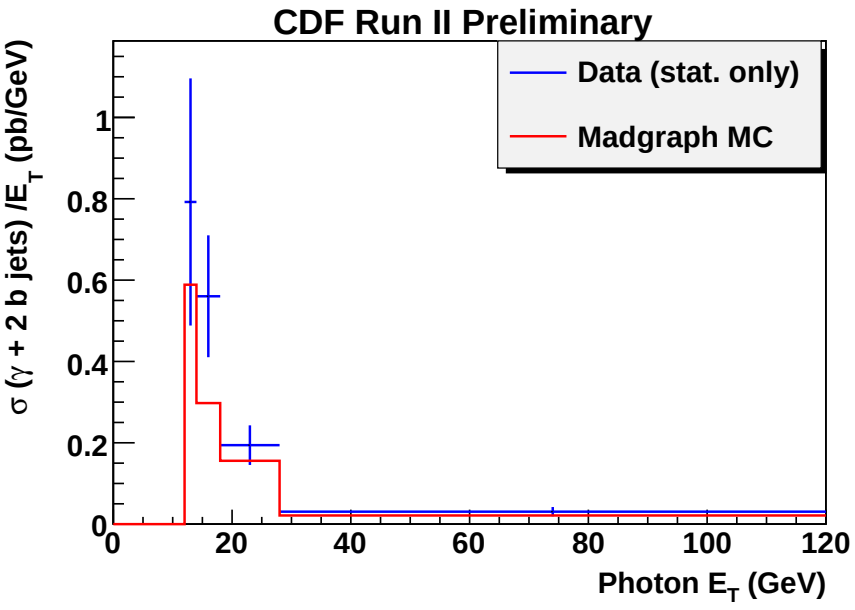


Figure 39: Differential cross section as function of photon $E_T$.

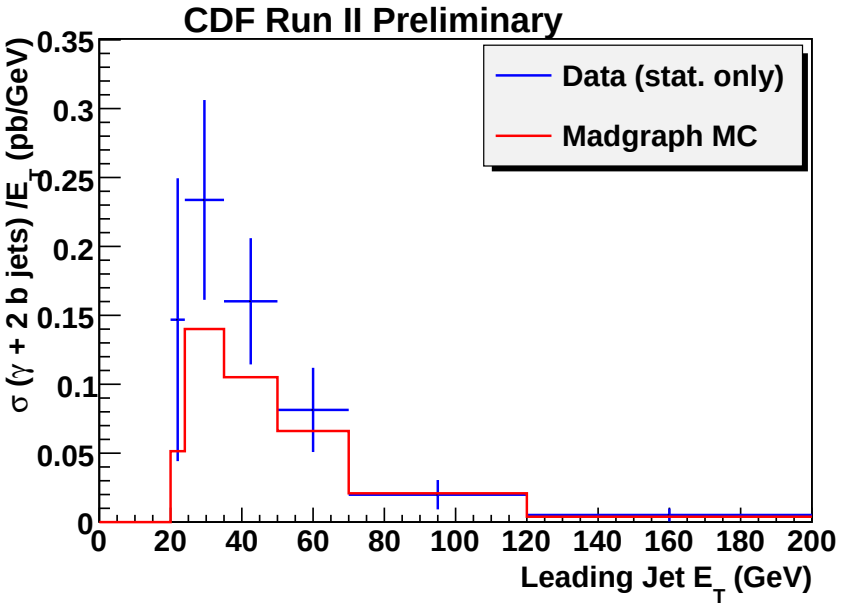


Figure 40: Differential cross section as function of leading jet $E_T$.

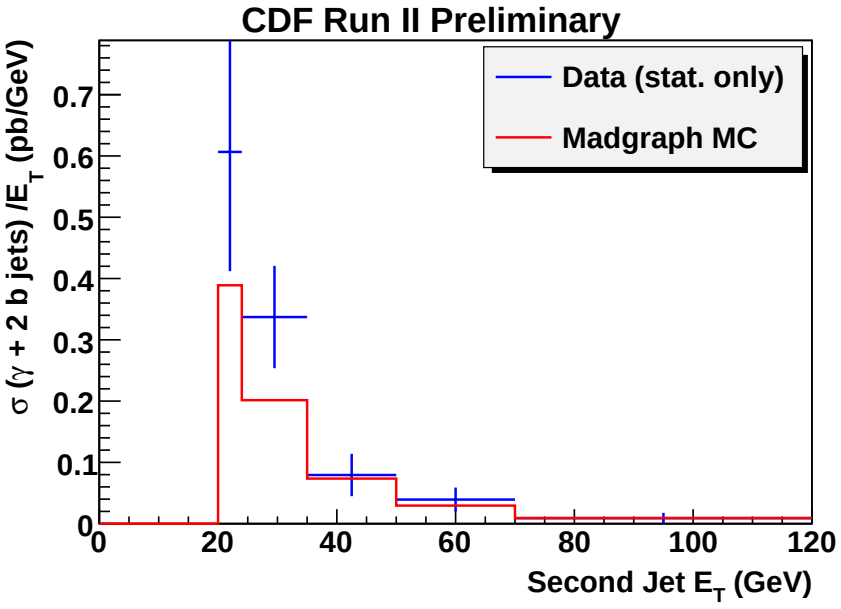


Figure 41: Differential cross section as function of the second jet $E_T$

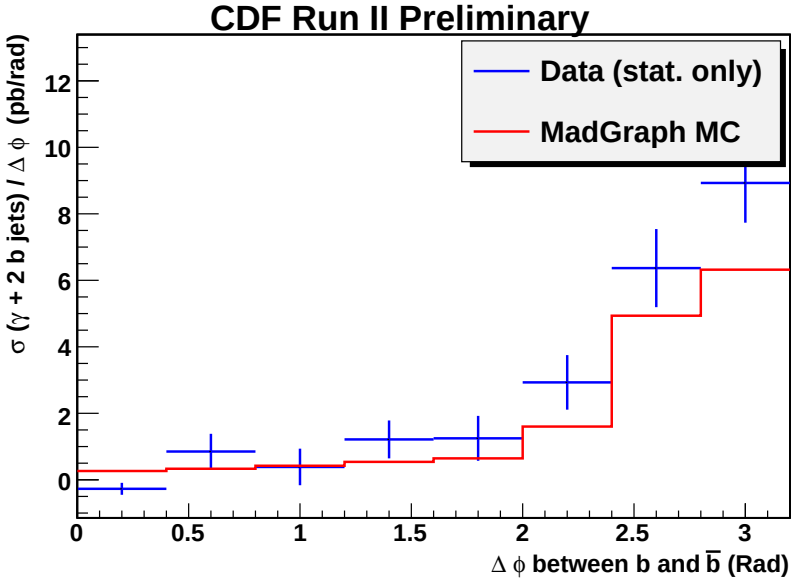


Figure 42: Angular correlation of the two jets.

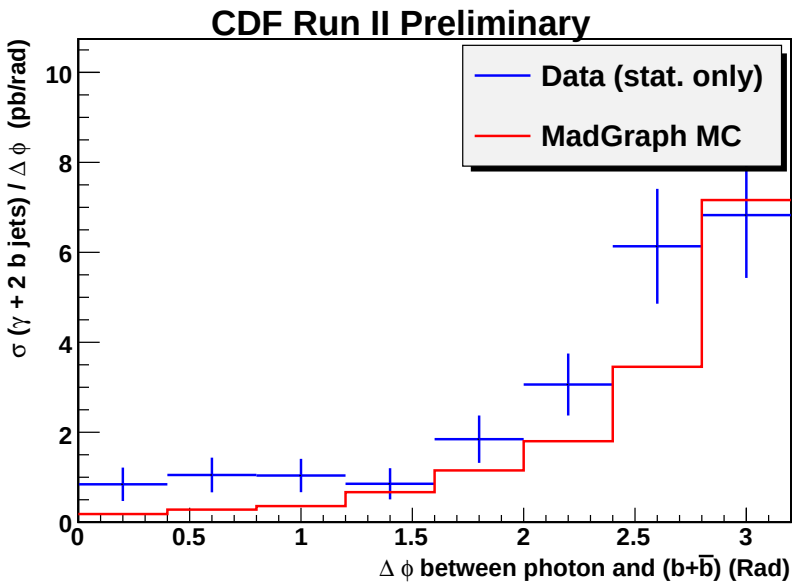


Figure 43: Phi angle between the vector sum of the two jets and the photon

# 8 Conclusions

We have presented the first measurement of the $p\bar{p} \to b\bar{b} + \gamma + X$ cross section at the Tevatron, using the dedicated dataset PHOTON_BJET which makes use of the CDF Silicon Vertex Tracker (SVT). The trigger efficiency has been computed directly from the data, using the fraction of events with a track passing the SVT-trigger requirement in an unbiased dataset where only a high-$E_T$ photon is required. Offline, we request two jets with positive b-tagging, leading to a $b\bar{b}$ purity of the tagged sample (also measured from data) of approximately 81%. We measure a total cross section of

$$8.60 \pm 1.07 pb(stat.)^{+1.44}_{-1.56}(syst.)pb.$$

This result is approximately 30% higher than a leading-order QCD calculation. This ratio is expected from important non-leading-order effects applied to similar channels. The angular information is expected to give important information on the production mechanism. However, an analysis of the angular correlations in the fake photon contribution to the data is complex. A careful and delicate background analysis to study angular correlations is beyond the scope of this first measurement, but should certainly be the goal of any follow-up of this work.

# References


[1] F. Abe *et al.*, Search for a technicolor $\omega_T$ particle in events with a photon and a b quark jet at CDF, *Phys. Rev. Lett.* **83** (1999) 3124

[2] A. Affolder *et al.*, Searches for new physics in events with a photon and b-quark jet at CDF, *Phys. Rev.* **D 65** (2002) 052006

[3] A. Gajjar, R. McNulty, T. Shears, Photon + heavy flavor production at CDF, CDF/PHYS/CDF/PUBLIC/7072

[4] M. Campanelli, Production of photons and b-jets using the PHOTON_B_JET dataset, CDF/PUB/JET/GROUP/8623

[5] W.-M. Yao *et al.*, Review of particle physics, *J. Phys.* **G 33** (2006) 1

[6] The Tevatron Electroweak Working Group, For the CDF and D0 Collaborations, `hep-ex/0703034`

[7] T. Aaltonen *et al.*, First Measurement of the W-Boson Mass in Run II of the Tevatron, *Phys. Rev. Lett.* **99** (2007) 151801

[8] Emmy Noether, Invariant Variation Problems, `physics/0503066v1`

[9] F. Abe *et al.*, Measurement of the B meson differential cross-section, $d\sigma/dp_T$, in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV, *Phys. Rev. Lett.* **75** (1995) 1451

[10] Monica D'Onofrio, Measurement of the Inclusive b-jet Cross Section in p-pbar Collisions at $\sqrt{s} = 1.96$ TeV, CDF/PUB/JET/PUBLIC/8418

[11] R. Field, The Sources of b-Quarks at the Tevatron and their Correlations, CDF/ANAL/BOTTOM/PUBLIC/5813

[12] S. Vallecorsa, Measurement of the $b\bar{b}$ jet cross section using a dedicated trigger at CDF, CDF/PHYS/JET/PUBLIC/8939

[13] T. Sjöstrand *et al.*, High-Energy-Physics Event Generation with PYTHIA 6.1 ,*Comput. Phys. Commun.* **135** (2001) 238, `hep-ph/0010017`

[14] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour and B.R. Webber, *J. High Energy Phys.* **0101** (2001) 010

[15] S. Frixione and B.R. Webber, Matching NLO QCD computations and parton shower simulations,*J. High Energy Phys.* **0206** (2002) 029;`hep-ph/0204244`;

[16] S. Aid *et al.*, A Measurement and QCD analysis of the proton structure function f2 (x, $q^2$) at HERA *Nucl. Phys.* **B 470** (1996) 3

[17] H.L. Lai *et al.*:Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distribution, `hep-ph/9903282`

[18] J.Pumplin *et al.*: New Generation of Parton Distributions with Uncertainties from Global QCD Analysis, `hep-ph/0201195v3`

[19] F. Maltoni and T. Stelzer, MadEvent: automatic event generation with MadGraph, *J. High Energy Phys.* **02** (2003) 027,`hep-ph/0208156`

[20] J. Alwall *et al.*, A Standard format for Les Houches event files, *Comput. Phys. Commun.* **176** (2007) 300

[21] S. Tsuno, LesHouches package – Generic Interface for Les Houches Accord at CDF offline software, http://nkek15.fnal.gov/ tsuno/leshouches.html

[22] The Tevatron Rookie Book, http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v1.pdf

[23] S. van der Meer, Stochastic cooling and the accumulation of antiprotons, *Rev. Mod. Phys.* **57** (1985) 3

[24] CDF II Collaboration, The CDF II Detector: Technical Design Report, FERMILAB-PUB-96/390-E (1996)

[25] F. Abe *et al.*, Precision Measurement of the Prompt Photon Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV, *Phys. Rev. Lett.* **73** (1994) 2662. Erratum in *Phys. Rev. Lett.* **74** (1995) 1891

[26] A. Abulencia *et al.*, Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV, *Phys. Rev. Lett.* **97** (2006) 082004

[27] R. Barlow and C. Beeston, Fitting using finite Monte Carlo samples, *Comput. Phys. Commun.* **77** (1993) 219

[28] Simon Sabik and Pierre Savard, Track reconstruction efficiency in jets, CDF/ANAL/JET/CDFR/6894

[29] F.Abe *et al.*, Properties of photon plus two jet events in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV, *Phys. Rev.* **D 57** (1998) 67

# A Feynman diagrams for $p\bar{p} \to b\bar{b}\gamma$

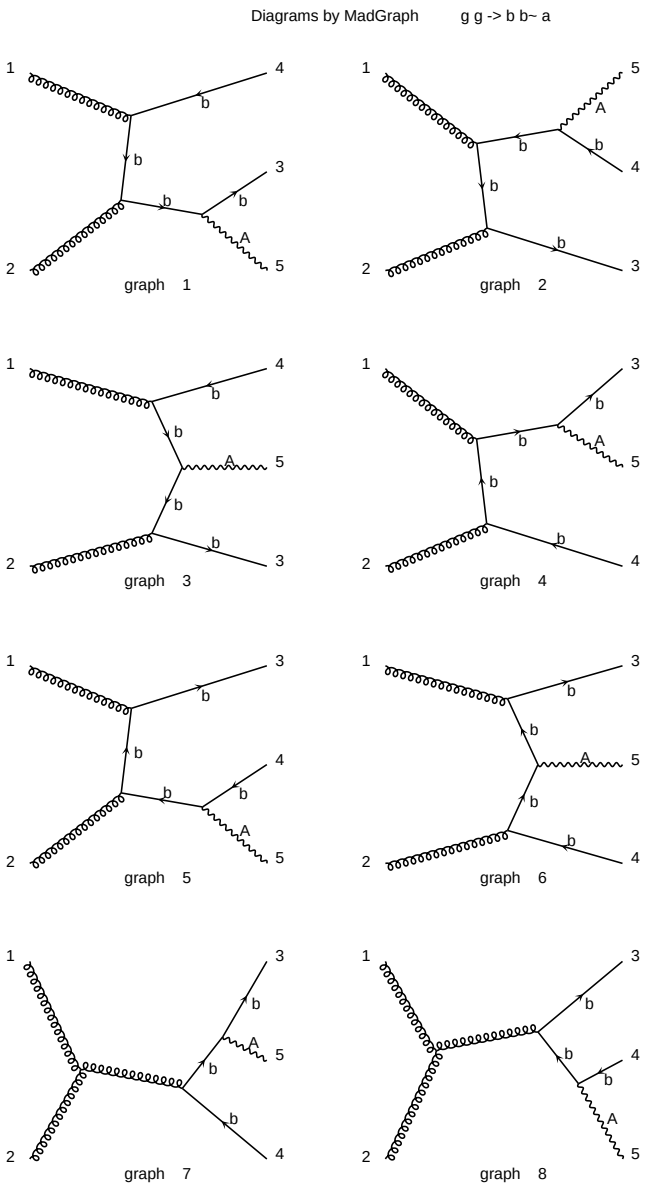


Figure 44: Diagrams with gluon initial states

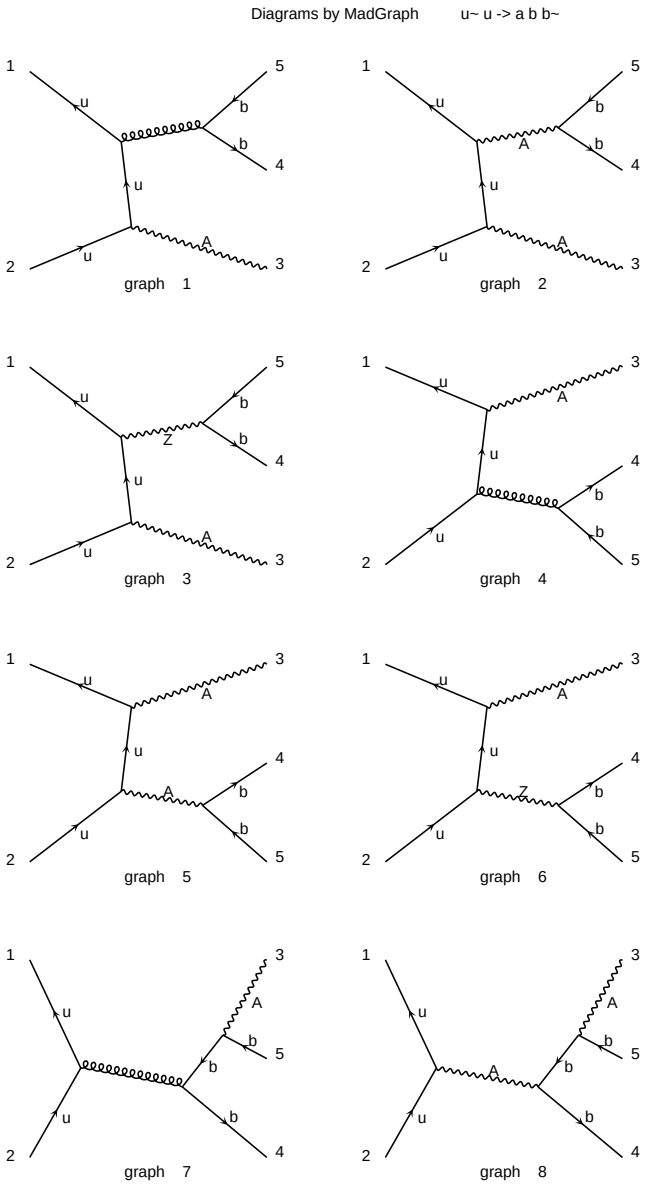


Figure 45: Diagrams with $q\bar{q}$ initial states

# B The PHOTON_B_JET trigger at Level 2

The L2 photon part of the trigger has the following requirements :
**Trigger version 1**

- Minimum $E_T$:
  12 GeV
- Isolation energy:
  3 GeV
- Energy-dependent additional energy allowed for the isolation cut:
  $0.15 \times E_T$
- Ratio of hadronic over em energy deposition:
  0.125
- Maximum rapidity:
  1.1

**Trigger version 3**

- Minimum $E_T$:
  12 GeV
- Isolation energy:
  **1 GeV**
- **No loosening of the isolation with higher energy**.
- Maximum allowed number of tracks pointing to the EM cluster:
  1 track
- Ratio of hadronic over em energy deposition:
  0.125
- Maximum rapidity:
  1.1