E-741

A Search for Double Parton Interactions in 1.8 TeV Proton-Antiproton Collisions

A thesis presented

by

David Nathan Brown

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

June, 1989

FERMILAB
LIBRARY

Copyright © 1989 by David Nathan Brown
All Rights Reserved

Abstract

4-jet events produced in $P\bar{P}$ collisions at 1.8 TeV and observed in the CDF detector at the Fermilab collider are examined for evidence of double parton interactions. Double parton 4-jet events are separated from other types of 4-jet events using topological variables. An upper limit on the double parton cross section is set, expressed as a lower limit on the effective proton cross section; $\sigma_{eff} > 5.14\ mb$ (95% confidence level). This limit is consistent with theoretical predictions.

Contents

List of Figures iii

Acknowledgements vi

The CDF Collaboration vii

1 Introduction 1

2 Double Interaction Theory 6

2.1 The QCD parton model 6

2.2 The Double Parton Cross Section 8

2.2.1 Double Structure Functions 9

2.2.2 The Effective Cross Section 11

2.3 Backgrounds to Double Parton 4-Jet Events 13

2.3.1 Double Collision Background 13

2.3.2 Double Bremsstrahlung Background 14

2.4 4-Jet Rate Calculations 15

2.5 Double Parton Topologies 18

2.6 Jet Fragmentation 20

3 The CDF Detector 23

3.1 Calorimetry 23

3.1.1 Calorimeter Components 24

3.1.2 Readout Segmentation 26
3.1.3 Test Beam Calibration 29
3.1.4 Calorimeter Non-linearity 30
3.1.5 Detector Cracks 31
3.2 Tracking 31
3.3 Trigger 32
3.4 Online Readout and Calibration 33

4 Event Reconstruction **36**
4.1 Energy Reconstruction 36
4.2 Noise Removal 38
4.3 Jet Clustering 42

5 Cluster Correction and Selection **45**
5.1 Underlying Event Contribution 46
5.2 Energy Loss Mechanisms 48
5.3 Correction Algorithm 49
5.4 Cluster Correction Testing 52
5.5 Fake Clusters 54
5.6 Reconstruction Efficiency 56
5.7 Cluster Separation Efficiency 57
5.8 Summary 58

6 4-Jet Analysis **59**
6.1 The CDF 4-Jet Data Sample 59
6.2 The 4-Jet Monte Carlo Data Sample 61
6.3 Double Parton Variables 63
6.3.1 Pairwise Significance 63
6.3.2 Pairwise ϕ Difference 69

6.4 4-jet Data Comparison 70

6.5 Double Parton Cross Section Limits 74

6.5.1 σ_{DP}/σ_{DB} Ratio Fitting 76

6.5.2 σ_{eff} Determination 77

6.5.3 Systematic Errors 78

6.6 Final Results and Conclusions 82

6.7 Future Analysis 83

A The Papageno Monte Carlo **85**

B Jet Fragmentation Tuning **88**

C Gas Spike Removal Algorithm **90**

D Jet Clustering Algorithm **93**

E The Parton Level Jet Simulation **95**

Bibliography **100**

List of Figures

1.1 e^+e^- jet Feynman diagrams 2

1.2 CDF dijet event 3

1.3 $P\bar{P}$ 4 jet Feynman diagrams 5

2.1 Schematic diagram of a $P\bar{P}$ collision 12

2.2 Predicted $d\sigma/dP_t$ for 4-jet events 18

2.3 Predicted $d\sigma/d\Sigma P_t$ for 4-jet events 19

2.4 Stick figure 4-jet topologies 20

3.1 Side view of the CDF detector 25

3.2 Transverse segmentation of the CDF detector 27

3.3 Test beam measured module boundary response 30

4.1 Gas spike energy spectrum 39

4.2 Gas spike removal efficiency 41

5.1 ϕ E_t flow in dijet events 47

5.2 η dependence of underlying E_t density 48

5.3 Jet response function 51

5.4 Fractional dijet P_t imbalance by η 54

5.5 Clustering error and efficiency by P_t 56

5.6 Cluster angular resolution 57

6.1 Offline trigger hardening efficiency 61

6.2 Dijet P_t significance 65

6.3 4-jet P_t significance 66

6.4 A small $\mathcal{S}_{pair}$ 4-jet event 67

6.5 A large $\mathcal{S}_{pair}$ 4-jet event 68

6.6 ΣP_t of 4-jet events 71

6.7 Pairwise significance of 4-jet events 73

6.8 ϕ difference between jet pairs 74

6.9 $\Delta\phi$ versus $\mathcal{S}_{pair}$ 75

6.10 Fit results for $\mathcal{S}_{pair}$ and $\Delta\phi$ 77

B.1 Jet fragmentation measures 89

E.1 Single jet obesity 96

E.2 Idealized dijet event 98

E.3 Dijet K_t balancing 99

Acknowledgements

I'd like to thank Roy for trusting me with responsibility and freedom, allowing me to develop independent critical judgement. Thanks to Marjorie for her counseling, collaboration, and friendship. Both directly and by example she taught me more about science than any other person. thanks to George for teaching me the simplicity of good design, and inspiring me with his calm and humorous intelligence. Thanks to Melissa for the enthusiasm and skill she contributed to our many projects. Thanks to Drew and Tom for taking over my hardware responsibilities, allowing me to finish this work. Thanks to Rob for the hours of blackboard discussion, and the careful thought he brought to our many collaborative efforts. Thanks to Ed for his advice on many technical and aesthetic issues. Thanks to Rick for developing many key ideas used in this analysis. Thanks to William, Johnny, and the other new students for their thought provoking questions. Thanks to Jim and John for the jet physics they taught me, and the role models they provided. Thanks to Jim and Craig and Willy and Henry for their good advice on many subjects. Thanks to the Fermilab staff and all the other CDF members who made this data possible. Finally thanks to Hal and Virginia and Reva and Tera and Kobi and Irene and Alex and Gabe and Brad and Robert, and all the relatives and friends who believed in me, and helped me achieve this goal.

This thesis is dedicated to my father Harold Brown

THE CDF COLLABORATION

F.Abe[p], D.Amidei[c], G.Apollinari[k], G.Ascoli[g], M.Atac[d], P.Auchincloss[n], A.R.Baden[f], A. Barbaro-Galtieri[i], V.E.Barnes[l], F.Bedeschi[k], S.Belforte[k], G.Bellettini[k], J.Bellinger[q], J.Bensinger[b], A.Beretvas[n], P.Berge[d], S.Bertolucci[e], S.Bhadra[g], M.Binkley[d], R.Blair[a], F. Blocker[b], J.Bofill[d], A.W.Booth[d], G.Brandenburg[f], D.Brown[f], A.Byon[l], K. L. Byrum[q], M. Campbell[c], R.Carey[f], W.Carithers[i], D.Carlsmith[q], J.T.Carroll[d], R.Cashmore[d], F.Cervelli[k], K.Chadwick[d,l], T.Chapin[m], G.Chiarelli[k], W.Chinowsky[i], S.Cihangir[o], D.Cline[q], D.Connor[j], M. Contreras[b], J.Cooper[d], M.Cordelli[e], M.Curatolo[e], C.Day[d], R.DelFabbro[k], M.Dell'Orso[k], L.DeMortier[b], T.Devlin[n], D.DiBitonto[o], R. Diebold[a], F.Dittus[d], A.DiVirgilio[k], J.E.Elias[d], R.Ely[i], S.Errede[g], B.Esposito[e], A. Feldman[f], B.Flaugher[n], E.Focardi[k], G.W.Foster[d], M.Franklin[f,g], J.Freeman[d], H.Frisch[c], Y.Fukui[h], A.F.Garfinkel[l], P.Giannetti [k], N.Giokaris[m], P.Giromini[e], L.Gladney[j], M.Gold[i], K.Goulianos[m], C.Grosso-Pilcher[c], C.Haber[i], S.R.Hahn[j], R.Handler[q], R.M.Harris[i], J.Hauser[c], T.Hessing[o], R.Hollebeek[j], L.Holloway[g], P.Hu[n], B.Hubbard[i], P.Hurst[g], J.Huth[d], H.Jensen[d], R.P.Johnson[d], U.Joshi[n], R.W.Kadel[d], T.Kamon[o], S.Kanda[p], D.A.Kardelis[g], I.Karliner[g], E.Kearns[f], R.Kephart[d], P.Kesten[b], H.Keutelian[g], S.Kim[p], L.Kirsch[b], K.Kondo[p], U. Kruse[g], S.E.Kuhlmann[l], A.T.Laasanen[l], W.Li[a], T.Liss[c], N.Lockyer[j], F.Marchetto[o], R.Markeloff[q], L. A. Markosky[q], P.McIntyre[o], A.Menzione[k], T.Meyer[o], S.Mikamo[h], M.Miller[j], T.Mimashi[p], S.Miscetti[e], M.Mishina[h], S.Miyashita[p], N.Mondal[q], S.Mori[p], Y.Morita[p], A.Mukherjee[d], C.Newman-Holmes[d], J. Ng[f], L.Nodulman[a], R.Paoletti[k], A.Para[d], J.Patrick[d], T.J.Phillips[f], H.Piekarz[b], R.Plunkett[m], L.Pondrom[q], J.Proudfoot[a], G.Punzi[k], D.Quarrie[d], K.Ragan[j], G.Redlinger[c], J.Rhoades[q], F.Rimondi[d], L.Ristori[k], T.Rohaly[j], A.Roodman[c], A.Sansoni[e], R.Sard[g], V.Scarpine[g], P.Schlabach[g], E.E.Schmidt[d], P.Schoessow[a], M.H.Schub[l], R.Schwitters[f], A.Scribano[k], S.Segler[d], M.Sekiguchi[p], P.Sestini[k], M.Shapiro[f], M.Sheaff[q], M.Shibata[p], M.Shochet[c], J.Siegrist[i], P.Sinervo[j], J.Skarha[q], D.A.Smith[g], F.D.Snider[c], R.St.Denis[f], A.Stefanini[k], Y.Takaiwa[p], K.Takikawa[p], D.Theriot[d], A.Tollestrup[d], G.Tonelli[k], W. Trischuk[f], Y.Tsay[c], F.Ukegawa[p], D.Underwood[a], R.Vidal[d], R.G.Wagner[a], R.L.Wagner[d], J.Walsh[j], T.Watts[n], R.Webb[o], T.Westhusing[g], S.White[m], A.Wicklund[a], H.H.Williams[j], T.Yamanouchi[d], A.Yamashita[p], K.Yasuoka[p], G.P.Yeh[d], J.Yoh[d], F.Zetti[k]

[a] Argonne National Laboratory- [b] Brandeis University- [c] University of Chicago
[d] Fermi National Accelerator Laboratory- [e] INFN, Laboratori Nazionali di Frascati, Italy
[f] Harvard University- [g] University of Illinois- [h] KEK, Japan
[i] Lawrence Berkeley Laboratory- [j] University of Pennsylvania
[k] INFN, University and Scuola Normale Superiore of Pisa, Italy- [l] Purdue University
[m] Rockefeller University- [n] Rutgers University- [o] Texas A&M University
[p] University of Tsukuba, Japan- [q] University of Wisconsin

Work supported by the U.S. Department of Energy, Division of High Energy Physics, Contract W-31-109-ENG-38, and Harvard University, Contract DE-AC02-76ER03064.

Chapter 1

Introduction

This thesis presents the results of a search for *double parton* interactions in 4-jet events. The double parton mechanism creates 4 jets from hard scatters of 2 pairs of partons within a single proton anti-proton ($P\bar{P}$) collision. The data used comes from the CDF detector, and was taken during the 1987 $P\bar{P}$ collider run at FERMILAB. Besides the final physics result, the main point of the analysis was to explore the use of *topological* jet signatures for partonic processes. The key experimental issues were the systematics of low energy jet reconstruction. Once these have been accounted for, the 4-jet data are shown to be in agreement with QCD models. However, insufficient statistics prevent a proof of the existence of double parton interactions. The upper limit set on the double parton cross section is consistent with current theoretical predictions. CDF data being taken now (1988 run) should allow a determination of the double parton cross section, and possibly a measurement of new aspects of proton structure.

Jets were discovered at SPEAR, an e^+e^- storage ring at SLAC [1]. The particles produced in $e^+e^- \Rightarrow$ hadrons collisions were found to orient preferentially along an axis through the collision point. This linearity formed evidence for an intermediate *quark* stage of the collision, $e^+e^- \Rightarrow q\bar{q} \Rightarrow$ hadrons (see figure 1.1). In this model linear particle alignment results from preferential emission of hadrons along the initial quark directions. The model was confirmed by studying the polar angle distribution of this axis, called the *thrust* axis. The measured distribution was found to agree with the prediction of the quark

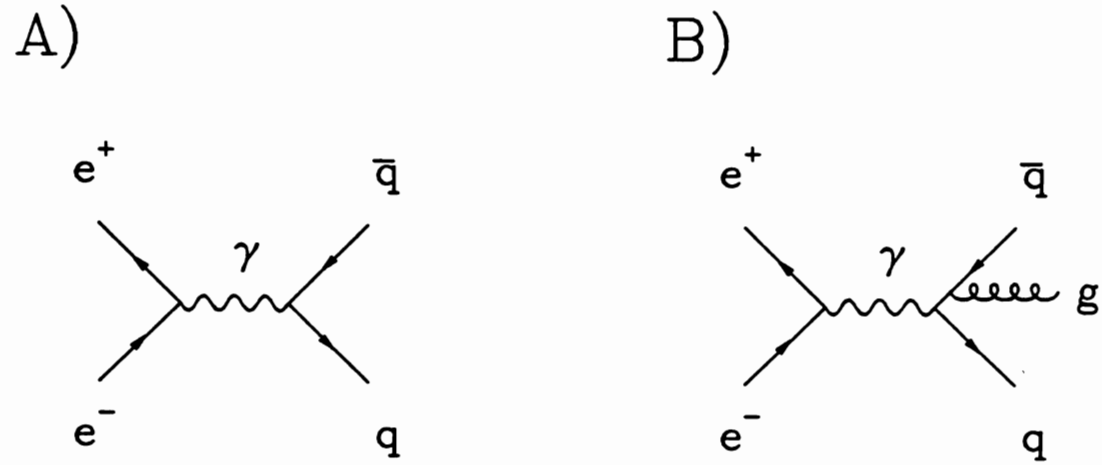


Figure 1.1: *A) The lowest order e^+e^- dijet diagram. B) The lowest order e^+e^- three jet diagram*

direction distribution in $e^+e^- \Rightarrow q\bar{q}$. This demonstrated a direct link between experimental quantities and underlying quark dynamics.

PEP and PETRA, higher energy successors to SPEAR, extended the range of jet topologies accessible to experiment. Experiments there observed events with a *planar* but not linear particle alignment, suggesting a more complicated dynamics [2]. This was interpreted as gluon radiation or *bremsstrahlung* from a quark (see figure 1.1). Such events were called three jet events. Momentum conservation in 3-jet events requires the quark, anti-quark, and gluon to all lie in a plane. Planar particle alignment is then due to the quark-level planar alignment, equivalent to the linear particle alignment in dijet events. The gluon radiation model was confirmed by studies of the particle energy flow within this plane. The average over planar events of the particle energy flow agreed with a QCD-based Monte Carlo calculation.

With their existence clearly proven, new definitions of jets were developed. One such definition reduced jet event complexity without obscuring parton level topology by associating nearby particles together. These *clustering* algorithms attempted to undo the decay or *fragmentation* of quarks and gluons into showers of hadrons. Agreement between cluster angular distributions and parton calculations demonstrated a workable association between

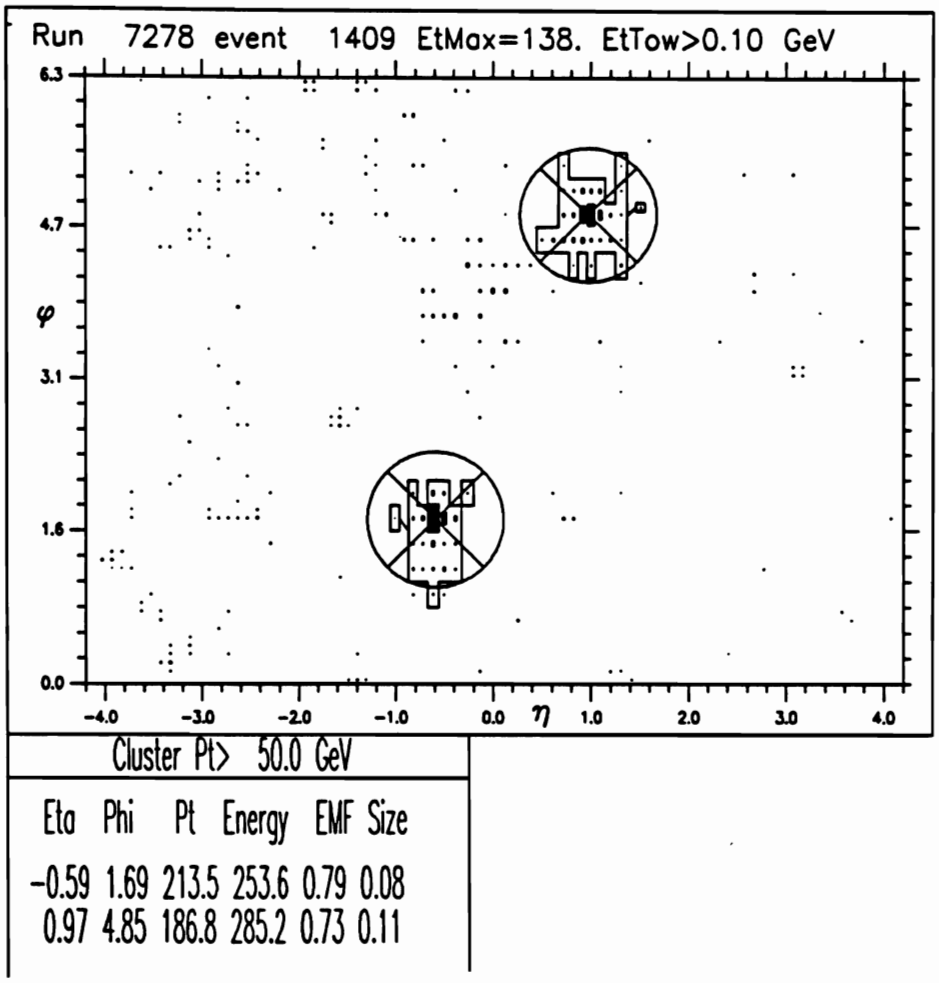


Figure 1.2: *A high transverse energy ($E_t \equiv E \sin\theta$) dijet event from in the CDF detector from a 1.8 TeV $P\bar{P}$ collision. The plot shows transverse energy flow across the solid angle of the CDF detector. The vertical axis plots azimuthal angle, the horizontal value is related to the polar angle. The rectangle sizes represent E_t density, divided according to the true CDF segmentation. The intense clumping of E_t in two regions opposite in azimuth evidences hard parton scattering. The square and circular outlines show the association of towers given by the CDF clustering algorithm.*

clusters and partons. By ascribing particles in a cluster to a single parton the process of fragmentation itself was studied.

Jets have also been observed at hadron colliders [3]. Unlike electrons, protons are *composite* particles, containing a variety of *partonic constituents.* Jets at $P\bar{P}$ colliders arise from hard collisions between pointlike partons. Because these constituents carry only part of the proton momentum, the jets produced contain only a fraction of the available center of mass energy. The typical $P\bar{P}$ beam energy is very large (.9 TeV for the FNAL collider), so that the observed jets can be very energetic, as shown in figure 1.2. The simple *dijet* structure of most jet events corresponds to the linear particle alignment of the early SPEAR jets. Agreement of $P\bar{P}$ jet angular and energy distributions with partonic calculations have given strong substantiation of both QCD and the parton model [5, 6].

Multiple jet production in $P\bar{P}$ collisions is more complicated than in e^+e^-. For instance, because the initial partons carry color charge, 3-jet events can be produced by gluon bremsstrahlung off either the initial or final interaction legs. As with dijets, the initial interacting partons can be any combination of quarks or gluons. In spite of this complexity, angular measures in $P\bar{P}$ 3-jet events have been analyzed and found to agree well with QCD Monte Carlo predictions [7].

4-jet events in $P\bar{P}$ can be produced by *double bremsstrahlung*, the radiation of 2 gluons from any of the primary interaction legs (see figure 1.3). Additionally, 4 jets can be produced by 2 pairs of partons within the proton interacting (see figure 1.3). This *double parton* scattering isn't a new prediction about proton structure, but a consequence of the parton model.

While double parton 4-jet events may be difficult to separate from double brem 4-jets, there are reasons to want to do so. Simultaneously probing of pairs of partons within a proton would provide information about parton *correlations.* Parton correlation measurements would give unique information about non-perturbative QCD and hadron structure. Double parton events also form a background to any processes that results in double pairs of jets,

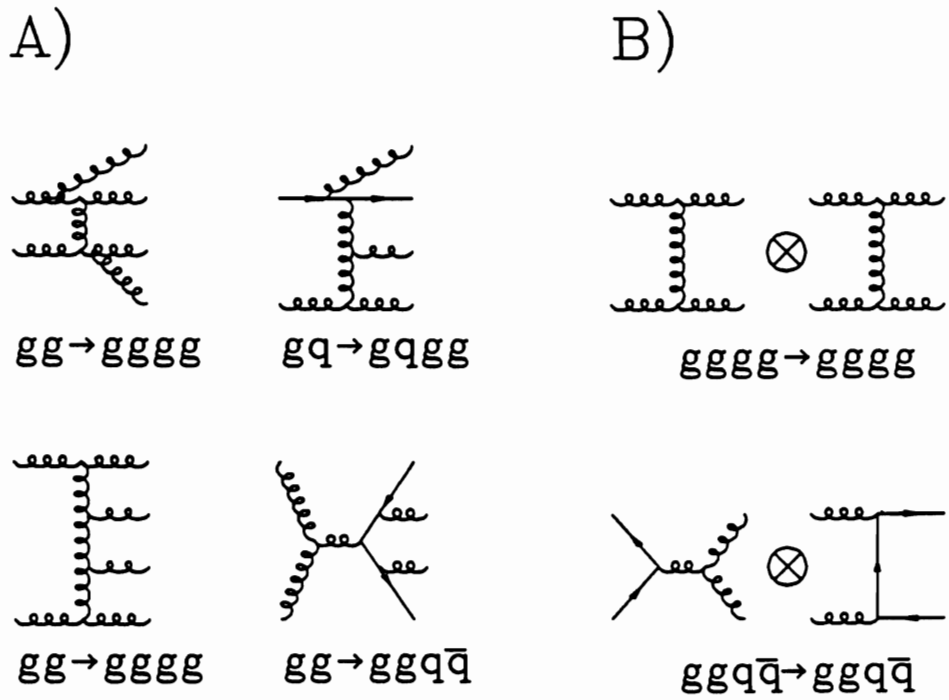


Figure 1.3: *A) Some of the over 300 lowest order $P\bar{P}$ QCD bremsstrahlung 4-jet diagrams. B) Some of the lowest order double parton scattering 4-jet diagrams. These are the same order in α_s as the bremsstrahlung diagrams.*

such as double boson production or Higgs decay. A good understanding of double parton scattering is necessary to search for such rare processes. Details of the current theoretical and experimental understanding of double parton interactions are presented in the following chapters.

Chapter 2

Double Interaction Theory

The QCD parton model of proton structure accurately describes experimental results which pick out single partons from protons. This model can be extrapolated to predict the basic form of the double parton interaction cross section. The extrapolation is incomplete because it does not predict the absolute double parton rate or the effects of parton *correlations* within the proton. The total rate of double parton events can be used to extract information about the effective size of the proton. Correlations can be used to study parton interactions within the proton. The most substantial background to double parton 4-jet events comes from gluon bremsstrahlung (*double brem*). Because the *topology* of double parton 4-jet events is substantially different from double brem, these can in principle be separated.

2.1 The QCD parton model

As has been proven in many experiments, the proton is a composite particle. One type of experiment measures the inelastic cross section for pion-proton interactions. The *resonances* in these cross sections indicate excited proton states, which require the proton to be composite. Another type of experiment uses leptons to probe deep within the proton. These deep inelastic scattering experiments (DIS) study the proton constituents from the momentum spectra of the scattered leptons [8].

The current theory of the proton describes it as a bound state of 3 fractionally charged fermionic *quarks*. These quarks interact through the *strong force*, which provides the bind-

ing energy of the proton. The strong force is described by a non-abelian gauge theory called quantum chromo-dynamics (QCD). The carrier of this force, called a *gluon*, is also necessarily a proton constituent. QCD has successfully described many aspects of hadron physics. In principle the structure of the proton can also be predicted using QCD. In practice, the equations involved are non-perturbative, and thus are not solvable using standard techniques. While some progress has been made describing hadron structure using computer models of QCD, most descriptions still rely on phenomenological models, which in turn are based on experimental data.

The distributions of quarks and gluons in the proton are described quantitatively by *structure functions*. These model the probability of finding a given type of parton, as a function of the kinematics of that parton. Structure functions are expressed in terms of 2 kinematic variables, called x and Q^2. x measures the fraction of proton 3-momentum carried by the parton, defined formally in the infinite momentum frame of the proton. Q^2 measures the 4-momentum transfer between the parton and whatever probe is reacting against it. Probes in DIS experiments are either photons or weak bosons, while the probes of a hadron collision are simply the partons of the other hadron. Conceptually, Q^2 measures how deeply one must probe into the proton to find a parton of a given type and x value.

The Q^2 dependence of structure functions arrises from higher order QCD interactions between partons. Accordingly, the dependence is weak, resulting in the approximate *scaling* of their x dependence [9]. Scaling is expected for protons made of pointlike, indivisible, and weakly interacting partons.

Scaling violation can be thought of as due to the creation of virtual partons within the proton. From the uncertainty relation, we know that the momenta of virtual partons are limited by the time and distance scale at which they are sampled. Thus, their abundance will depend on the Q^2 of the probing. The general effect of scaling violation is to increase the number of partons at small x, thereby decreasing the number at large x, as Q^2 increases.

Quantitative estimates of structure function *evolution* with Q^2 have been made using

perturbative QCD. The leading contributions come from successive gluon radiation from quark and gluon lines. Summing these diagrams gives the *leading log* approximation to structure function evolution [10]. The neglected terms involve interactions between quarks and/or gluons which start on separate lines.

QCD evolution predictions agree well with DIS data for quark structure functions in the experimentally accessible regions of x and Q^2 [8]. Gluon structure functions cannot be directly measured in DIS because gluons are electro-weak charge neutral. Gluon structure functions can be infered in a very limited range of x and Q^2 from anti-quark distribution functions [11]. Outside of this range, the only information on gluon structure functions comes from requiring the momentum of all the constituents to sum to the proton momentum. Gluon structure functions are therefore more uncertain than quark structure functions.

Many regions of x and Q^2 important to $P\bar{P}$ scattering are inaccessible to DIS. Various functional forms have been suggested to extrapolate DIS measurements into these regions [12]. These forms typically agree to within a factor of two for the low x values important in the double parton analysis ($x \sim .02$). The dominant parton species in these regions are gluons. The discrepancy between models is taken as part of the systematic error later in this analysis.

2.2 The Double Parton Cross Section

The cross section for a single dijet event in a $P\bar{P}$ collision can be described conceptually as;

$$\frac{d\sigma^{P\bar{P}}_{dijet}(\vec{\chi})}{d\vec{\chi}} = \sum_{i,j} \int \frac{d\sigma^{part}_{ij}}{dx_i dx_j} F_i(x_i, Q^2) F_j(x_j, Q^2) \delta(\vec{\chi}'(x_i, x_j) - \vec{\chi}) dx_i dx_j. \tag{2.1}$$

i and j sum over specific parton types (gluons, quarks, and anti-quarks of all colors and flavors) within the proton and anti-proton. The sum incoherently adds the *sub processes* that, while unique at the parton level, give rise to indistinguishable jets. $F_i(x_i, Q^2)$ are the proton structure functions expressed as longitudinal momentum probability functions. $\frac{d\sigma^{part}_{ij}}{dx_i dx_j}$ is the parton level differential cross-section for a specific sub process. $\vec{\chi}$ represents an arbitrary choice of final state kinematic parameters. The δ function transforms from parton

momentum fractions into the final kinematic basis ($\vec{\chi}$).

Analogous to equation 2.1, we can write down the double parton 4-jet scattering cross section as;

$$\frac{d\sigma^{PP}_{double}(\vec{\chi_1},\vec{\chi_2})}{d\vec{\chi_1}d\vec{\chi_2}} = \frac{1}{\sigma_{eff}}\sum_{i,j,k,l}\int\frac{d\sigma^{part}_{ij}}{dx_i dx_j}\frac{d\sigma^{part}_{kl}}{dx_k dx_l}G_{ik}(x_i,x_k,Q_1^2,d)G_{jl}(x_j,x_l,Q_2^2,d)\times$$
$$\delta(\vec{\chi_1'}(x_i,x_j)-\vec{\chi_1})\delta(\vec{\chi_2'}(x_i,x_j)-\vec{\chi_2})dx_i dx_j dx_k dx_l. \quad (2.2)$$

This is essentially the square of the dijet cross section. Again the sum incoherently adds the possible sub processes, which now include all the combinations of 2 separate scatterings. Single structure functions are replaced by *double structure functions* representing the *simultaneous* probability of picking out 2 partons of given type and momentum fraction. Correlations between partons are explicit in the d dependence of the double structure functions, where d measures the *transverse separation* of the partons. The sub process cross sections are unchanged from the single dijet case. The σ_{eff} term explicitly preserves the dimensionality.

The physics content of double parton scattering is contained in the double structure functions and the σ_{eff} terms. The former of these measures non-trivial correlations between partons. The latter measures the effective transverse size of the proton and determines the absolute double parton rate.

2.2.1 Double Structure Functions

The double structure functions G describe possible non-trivial parton correlations. These could arise from the QCD bremsstrahlung responsible for parton evolution as described in section 2.1. Correlations could also arise from diquark states within the proton.

The importance of correlation terms in double structure functions has been estimated using a simple proton model [13]. In this model, interactions are considered important when partons are forced to *overlap*. Parton overlap is defined from the number of partons in the proton (N), the size of the proton (R), and the effective size of the partons (S). Only

overlap in the transverse direction to the collision axis will affect scattering, which gives;

$$\text{overlap} \equiv \frac{\alpha_s^2 S^2}{R^2} N.$$

The factor of α_s^2 assumes a lowest order QCD interatcion between the partons. An effective parton size comes from their momentum (q) via the uncertainty principle; $S \sim 1/q$. The number of partons is estimated by a low-x extrapolation of structure function parametrizations; $N(x) \sim x^{-\frac{1}{2}}$. x is chosen as the minimum value possible for the machine energy and the interaction energy; $x = 2q/\sqrt{S}$. Taking $\sqrt{S} = 1800$ GeV for the Tevatron, and requiring $q > 10$ GeV for the resulting jets to be experimentally resolvable gives an overlap of;

$$\text{overlap} \sim 10^{-5}.$$

Since this is much less one, the conclusion is that parton correlations are unimportant for Tevatron double parton events. The unimportance of parton correlations has also been derived from first-order perturbative QCD calculations [14].

Assuming parton correlations are negligible, we can factorize the double structure functions as;

$$G(x_1, x_2, Q^2, d) \approx F(x_1, Q^2) F(x_2, Q^2)$$

Plugging this expression back into equation 2.2 we find that the entire double parton cross section now factorizes;

$$\begin{aligned}
\frac{d\sigma^{P\bar{P}}_{double}(\vec{\chi_1}, \vec{\chi_2})}{d\vec{\chi_1} d\vec{\chi_2}} &\approx \frac{1}{\sigma_{eff}} \sum_{i,j,k,l} \int \frac{d\sigma^{part}_{ij}}{dx_i dx_j} F_i(x_i, Q_1^2) F_j(x_j, Q_1^2) \delta(\vec{\chi_1'}(x_i, x_j) - \vec{\chi_1}) dx_i dx_j \times \\
&\quad \int \frac{d\sigma^{part}_{kl}}{dx_k dx_l} F_k(x_k, Q_2^2) F_l(x_l, Q_2^2) \delta(\vec{\chi_2'}(x_k, x_l) - \vec{\chi_2}) dx_k dx_l \\
&\approx \frac{1}{\sigma_{eff}} \frac{d\sigma^{P\bar{P}}_{dijet}(\vec{\chi_1})}{d\vec{\chi_1}} \frac{d\sigma^{P\bar{P}}_{dijet}(\vec{\chi_2})}{d\vec{\chi_2}}.
\end{aligned} \tag{2.3}$$

Thus the double parton cross section breaks down into the product of 2 independent $P\bar{P}$ dijet cross sections, normalized by $1/\sigma_{eff}$. The factorized expression is only $\approx$ to show that we have omitted parton correlations.

The factorization of the double structure functions does not imply that parton correlations are unmeasurably small. Instead, it says that the dominant features of the double parton cross section can be simply modeled as the product of 2 dijet cross sections. Before parton correlations can be measured, methods must first be found to reliably identify double parton events. This analysis focuses on double parton event identification. The extraction of parton correlations from these events is left for the future.

2.2.2 The Effective Cross Section

A simple example demonstrates the physical significance of the σ_{eff} term in the above cross section. Imagine a single $P\bar{P}$ beam crossing of integrated luminosity L. Using equation 2.3 we can write the *probability* of a double interaction as;

$$\mathcal{P}_{double} = L \cdot \sigma_{double} = (L \cdot \sigma_{dijet})(\frac{1}{\sigma_{eff}}) \cdot \sigma_{dijet},$$

where the terms are grouped to suggest that the 2 interactions occurred sequentially. The first term is simply the probability of getting a single dijet interaction. The second term must then represent the probability of getting an *additional* dijet interaction. The $1/\sigma_{eff}$ term is therefore an *effective luminosity* for the second interaction. For the first interaction to have occurred at all, there must have been a $P\bar{P}$ collision. Naively, we can imagine that the remaining partons interact as in a normal inelastic $P\bar{P}$ collision. We can therefore loosely associate the effective cross section with the total inelastic $P\bar{P}$ cross section, $\sigma_{eff} \sim \sigma_{tot}$.

A geometric argument shows that σ_{eff} must actually be somewhat smaller than the total inelastic cross section. Consider a $P\bar{P}$ collision as a function of *impact parameter* Δ, as shown in figure 2.1. The proton is modeled naively as a sphere of uniform parton density and a fixed size. In this model, collisions at all impact parameters smaller than the proton size contribute equally to the total cross section. However, because the sphere is thicker at its middle, The number of partons passing through each other is greater when the impact parameter is smaller. Thus parton interactions (*ie* dijet events) are more likely at low impact parameters than at high.

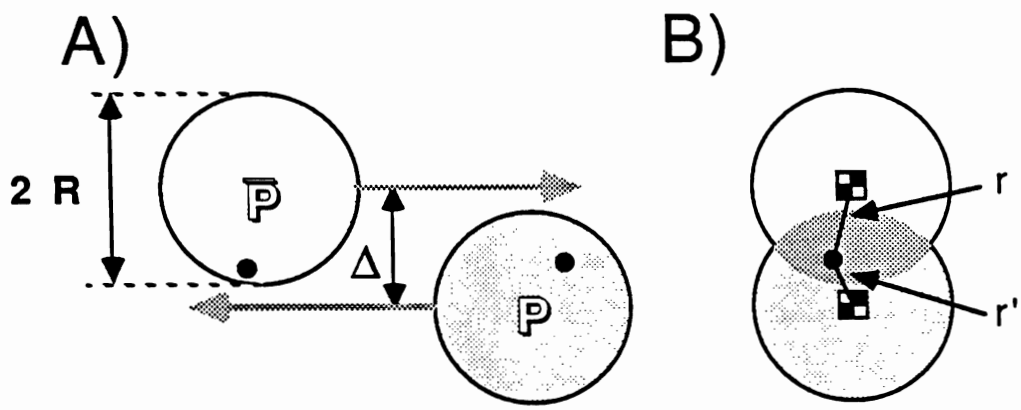


Figure 2.1: *Shown is a highly stylized view of a $P\bar{P}$ collision in (A) a side view and (B) a head-on view. The protons are separated transversely by* impact parameter Δ. *The chunks of parton density indicated by the vectors r and r' will collide, giving a contribution to the* parton luminosity

Consider now double parton interactions as two sequential dijet interactions. By the above logic, the first interaction preferentially selects low impact parameter collisions. Thus the effective luminosity for the second interaction is larger than that from a random $P\bar{P}$ collision, making $\sigma_{eff} < \sigma_{tot}$.

This simple proton model can be used to quantitatively estimate the size of σ_{eff}. We define a uniform sphere parton density function as;

$$\mathcal{D}(r) \equiv \begin{cases} C & \text{if } r < R \\ 0 & \text{otherwise} \end{cases} .$$

In this model the $P\bar{P}$ total inelastic cross section is just $4\pi R^2$. For a $P\bar{P}$ collision at impact parameter Δ the *parton luminosity* is then

$$L_{eff}(\Delta) = \int_{overlap} \mathcal{D}(r)\mathcal{D}(r')dV,$$

where the integral is taken over the overlap volume defined by figure 2.1. The dependence of L_{eff} on Δ is implicit in the definitions of r and r' shown in the figure.

We can then solve for the $P\bar{P}$ dijet cross section as;

$$\sigma_{dijet} = \sigma_{dijet}^{parton} \cdot \int L_{eff}(\Delta) 2\pi\Delta d\Delta,$$

where σ_{dijet}^{parton} is the parton level dijet cross section. The integral averages over all impact parameters, assuming azimuthal symmetry. Similarly, we can derive the cross section for a double dijet event, and use equation 2.3 to relate it to the single dijet cross section;

$$\sigma_{double} = (\sigma_{dijet}^{parton})^2 \cdot \int L_{eff}^2(\Delta) 2\pi\Delta d\Delta = \frac{\sigma_{dijet}^2}{\sigma_{eff}}.$$

Solving for σ_{eff} gives;

$$\sigma_{eff} = \frac{(\int L_{eff}(\Delta) 2\pi\Delta d\Delta)^2}{\int L_{eff}^2(\Delta) 2\pi\Delta d\Delta}.$$

This can be numerically integrated, subject to the constraint that $\sigma_{tot} = \int_0^R 2\pi\Delta d\Delta$. This gives $\sigma_{eff} \cong \sigma_{tot}/2.3$, independent of the size R or the value of the parton density C. We can interpret the *enhancement factor* of 2.3 as due to the trivial correlation between partons owing simply to their confinement within the proton.

2.3 Backgrounds to Double Parton 4-Jet Events

Double parton scattering is not the only source of 4 jet events. In order to measure σ_{eff} and parton correlations, the double parton events must be separated from the other types of 4-jet events. As will be shown, the background is large and the experimental resolution is poor, so the separation is difficult. In this section we examine the 4-jet background from the most likely sources.

2.3.1 Double Collision Background

From equation 2.3 we know that the double parton cross section approximately factorizes into the product of two single interaction cross sections. This implies an obvious background to double parton events from *double collisions*, where two $P\bar{P}$ pairs interact in the same beam crossing. This background is indistinguishable from double parton scattering in terms of the jets produced. Fortunately, the rate of double collisions was very small in the 1987 run, and they form a negligible background. This is demonstrated by a simple exercise.

In our double parton analysis, we will be looking at 4-jet final states. Consequently for this background estimate, we are only interested in double collisions where each collision

independently produces a dijet event. We can therefore write the number of double collision background events as the number of single dijet events times the probability of getting a second dijet event in the same crossing;

$$N_{double\ collision} = (L_{tot} \cdot \sigma_{dijet}) \cdot (L_{1\ crossing} \cdot \sigma_{dijet}).$$

L_{tot} and $L_{1\ crossing}$ are the integrated luminosities of the entire 1987 run and of a single crossing, respectively. The integrated luminosity of a single crossing is determined from the instantaneous luminosity and the time between crossings,

$$L_{1\ crossing} = \mathcal{L} \cdot \tau_{crossing}.$$

Most of the 1987 data was taken at a luminosity of $\sim 10^{28} cm^{-2} sec^{-1}$. The total luminosity of the 1987 run was $\sim 25 nb^{-1}$. Three bunches in the Tevatron gave a time between beam crossings of $\sim 7 \mu sec$ so that;

$$N_{double\ collision} \sim 2 \times 10^{-8} \sigma^2_{dijet}(nb).$$

Lets compare this to the estimated double parton rate. Taking $\sigma_{eff} = 40mb/2.3 \cong 17mb$ in equation 2.3 gives;

$$N_{double\ interaction} \sim 2 \times 10^{-6} \sigma^2_{dijet}(nb) \sim 100 \times N_{double\ collision}.$$

Therefore double collisions are a negligible background to double parton interactions.

Double collision events can also be eliminated directly using vertex tracking information. Vertex tracking allows the resolution of separate $P\bar{P}$ collision points to within about 5 cm. Given that the bunch length in the Tevatron is about 60 cm, this implies a double collision rejection factor of about 10. Vertex rejection of double collisions was not done in this analysis, but will be necessary for 1988 CDF data, taken at 20 times higher instantaneous luminosity.

2.3.2 Double Bremsstrahlung Background

Another source of background to double parton scattering comes from *double bremsstrahlung*. These are events where a single pair of partons collide. Normally, such collisions give

rise to a simple dijet event, as shown in figure 1.2. Occasionally, extra jets are produced through QCD bremsstrahlung off either initial or final parton legs. Double brem is when two such extra jets are produced, giving a 4-jet final state. Examples of a few double brem diagrams are shown in figure 1.3. In total over 300 diagrams are needed to describe this process to lowest order (α_s^4).

Fortunately, the formidable task of evaluating these first order QCD diagrams has been done. This required the work of many physicists, and took several years [15]. Unfortunately, the final result is so complicated that it is computationally impractical. For instance, the term which describes the sub-process $qg \rightarrow qggg$ involves over 1000 separate subterms of parton momenta dot products. Evaluating the full matrix element for a single event requires an average of 10 cpu seconds (VAX 8600 equivalent). The full matrix element is therefore not practical for use in Monte Carlo event generators. Instead, approximations are used.

The approximation used in this analysis includes the exact matrix element on only the most common double brem sub-process, namely $gg \rightarrow gggg$. This term is computationally simple due to its symmetry. Sub-processes involving quarks (such as $gq \rightarrow gqgg$) are approximated using a technique called *gluon reduction* [16]. This models $q\bar{q}$ pairs as a single gluon, giving a simple, symmetric matrix element term with a reduced number of partons. This is then multiplied by the first order QCD probability for a gluon to split into the $q\bar{q}$ pair. This approximation has been shown to be very similar to the exact matrix element.

2.4 4-Jet Rate Calculations

To compare quantitatively double brem production with double parton production, we use a Monte Carlo program called Papageno [17]. This uses first-order perturbative QCD matrix elements to describe many parton-level processes, including 2 and 4 jet production. These matrix elements are combined with structure functions and phase-space density functions to calculate $P\bar{P}$ collision rates for various final states. Papageno works by sampling the kinematic space of possible final states, and generating *events* corresponding to the events

measured in an experiment. For each event it returns 4-vectors of the partons plus an overall *weight* which describes the event probability. Event weights are scaled such that their sum gives the total cross section for the process. A histogram of an event parameter (such as total energy) weighted by this value is effectively a calculation of the differential cross section in terms of that parameter. The details of how Papageno works are described in appendix A.

Because the matrix elements inside Papageno are calculated only to first order, we must choose a Q^2 scale at which to evaluate α_s. The proper choice of Q^2 scale can account for the effects of higher order processes. For complicated final states (such as 4-jet final states), there is no obvious value of Q^2 to use. For what follows in this section, we set Q to be half the average P_t of the final 4 partons of the event, $Q =< P_t > /2$. Because the same value is used for both double parton and double brem 4-jet calculations, comparisons between them are relatively insensitive to how Q^2 is calculated. Estimates of the systematic errors for double parton and double brem rates due to Q^2 scale uncertainty are made in section 6.5.3.

To model $P\bar{P}$ collisions we must also decide on a structure function parametrization, and a Q^2 scale at which to evaluate it. We choose the structure function Q^2 scale to be the same as that of the matrix element. The choice of structure function affects the relative double parton to double brem rate, as the two processes sample different Feynman x ranges. In double brem a single scatter generates all the observed energy while in double parton the two scatters combine to produce the energy. Consequently, we expect double parton to sample lower x values than double brem for similar final states. Structure function uncertainty therefore results in a systematic error when comparing double brem with double parton cross sections. This error is estimated in section 6.5.3, but is ignored for this simple comparison.

To calculate the double parton cross section we start with Papageno dijet events. These are generated according to the lowest order QCD $P\bar{P}$ dijet cross section. Double dijets

events are created by combining two consecutively generated dijet events. A combined weight is calculated as the product of the individual weights, scaled by $2N_{gen}/\sigma_{eff}$, where N_{gen} is the total number of dijet events generated. The $1/\sigma_{eff}$ term comes from equation 2.3. The $2N_{gen}$ term comes from combinatorics. Combining consecutive events samples only $N_{gen}/2$ of the N_{gen}^2 possible combinations of two dijet events. The $2N_{gen}$ term scales the double dijet events to compensate for the unused combinations. In our calculation, we take $\sigma_{eff} = 40mb/2.3 \cong 17mb$.

To calculate the double brem cross section, we use the approximate 4-jet matrix element described in section 2.3.2. To eliminate the colinear divergence in the cross section, we require an angular separation between all partons of at least 1 radian. We also require an angular separation between the partons and the beam directions of .06 radians. To avoid infra-red divergence all partons were required to have a minimum *transverse momentum* relative to the beam direction of 10 GeV. These cuts were also made on the double parton events.

Figure 2.2 plots the double brem and double parton differential cross sections as a function of the transverse momentum (P_t) of the lowest P_t parton. Though smaller than the double brem, the double parton rate falls at approximately the same rate as the double brem with P_t. Experimentally, detector and physics effects force jet analysis to require a minimum P_t of around 15 GeV (see section 5.5).

Figure 2.3 plots the differential cross sections as a function of the sum P_t (ΣP_t) of all 4 partons. This plot includes only events with all partons above 15 GeV P_t. Here the double parton rate drops much faster than the double brem. This variable is relevant since the CDF detector is triggered by ΣE_t, a closely related quantity. As discussed in section 6.1, the 1987 run CDF trigger results in an effective minimum ΣP_t of 95 GeV. At that value, we expect only 10% of the 4-jet events produced to be due to double parton interactions.

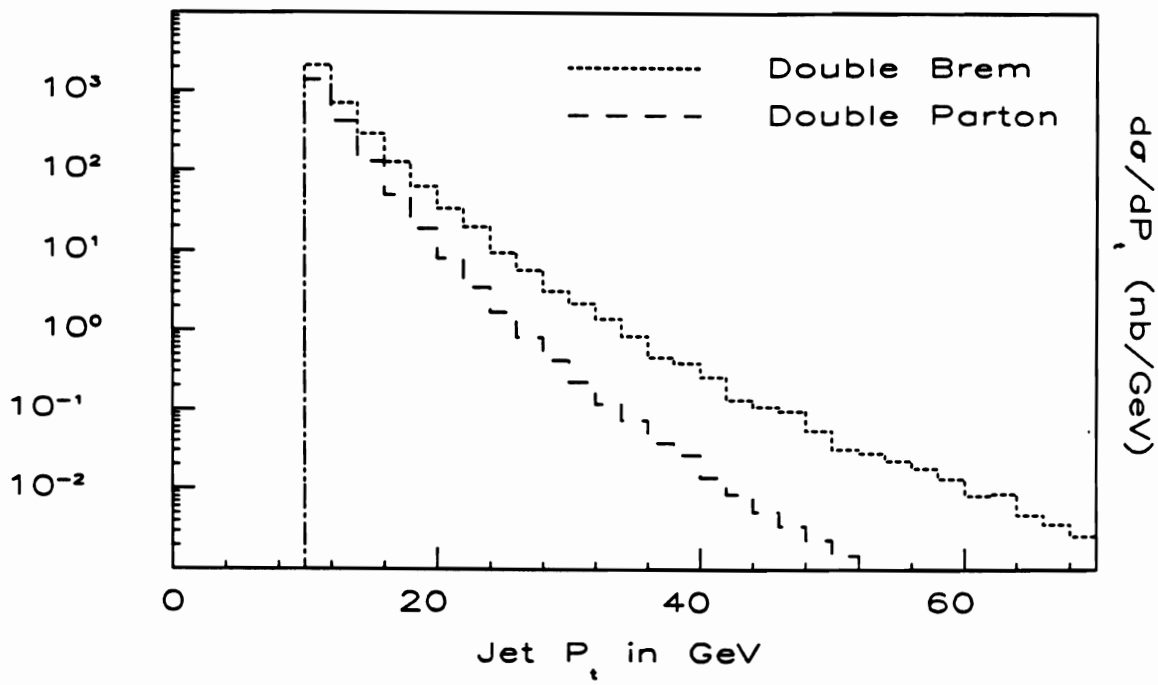


Figure 2.2: *The differential cross section of double brem and double parton events in terms of the P_t of the lowest P_t parton in the event.*

2.5 Double Parton Topologies

From the previous section we know that the expected rate of double parton 4-jet events is much less than that expected for double brem. In this section explore the use of *topological* differences between the two types of 4-jet events to separate them. We use the term topology to refer to the relative size and orientation of parton 4-vectors, independent of their absolute size and orientation.

In section 2.2 we learned that the double parton cross section approximately factorizes into the product of two single interaction cross sections. Therefore double parton 4-jet events should look like two independent dijet events. Because of momentum conservation, we know that the transverse momentum ($\vec{P}_t$) of the jets in a dijet must sum to zero. The net longitudinal momentum of the jets is not constrained, as that depends on the Feynman x values of the colliding partons. This implies that jet $\vec{P}_t$'s will cancel in pairs in double parton 4-jet events. Pairwise $\vec{P}_t$ conservation is not required of double brem 4-jet events, where

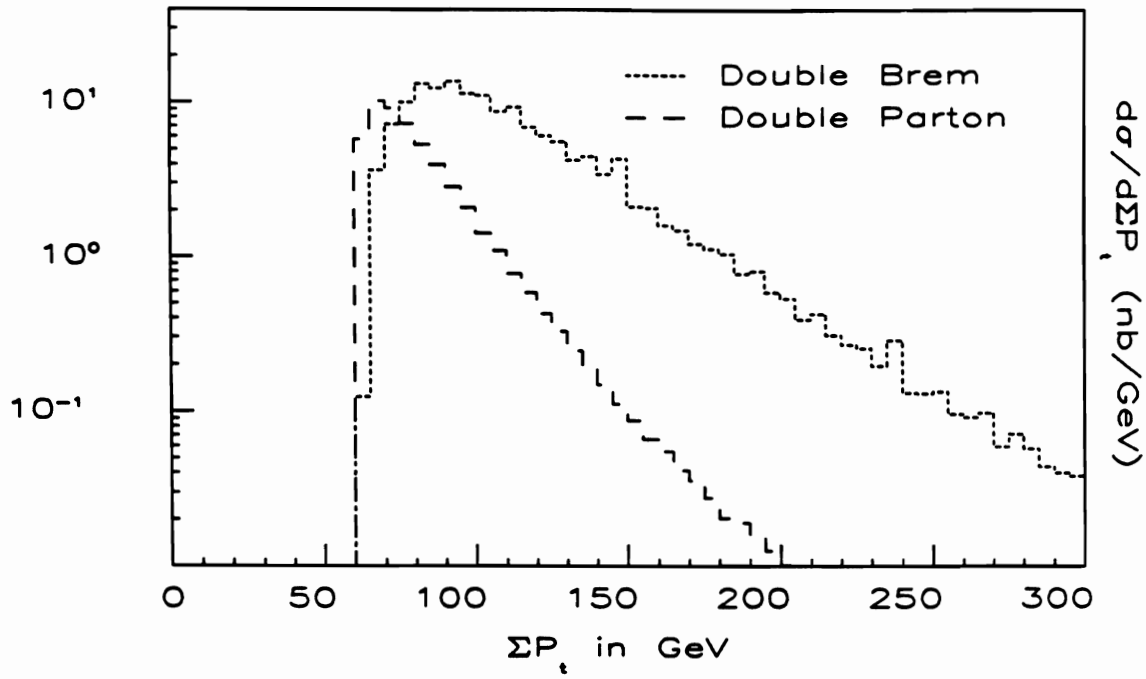


Figure 2.3: *The differential cross section of double brem and double parton events in terms of the sum P_t of all 4 partons.*

only the total $\vec{P}_t$ of all 4 jets is necessarily conserved. This suggests one can separate double parton events from double brem by requiring $\vec{P}_t$ *balancing* of pairs of jets (see figure 2.4). Experimental and non-perturbative effects blur this distinction. Translating the concept of pairwise $\vec{P}_t$ balancing into an experimentally robust variable is done in section 6.3.

More subtle differences between double parton and double brem events are hidden in the matrix elements of the cross sections. Most of these are too delicate to remain useful after experimental effects are included. One simple and useful concept exploits the preferentially small emission angle characteristic of bremsstrahlung. Thus jets produced from double brem tend to be correlated in angle (see figure 2.4). Alternatively, we know the jets from double parton production are un-correlated. Thus the distribution of the angle between jets will be flat for double parton events, and highly peaked for double brem. This distinction is less powerful than pairwise $\vec{P}_t$ balancing, as double parton and double brem distributions will overlap. However, it is relatively independent of the size of the momenta, and so provides a good cross check to pairwise balancing. Translating the concept of anti-correlated jet angles

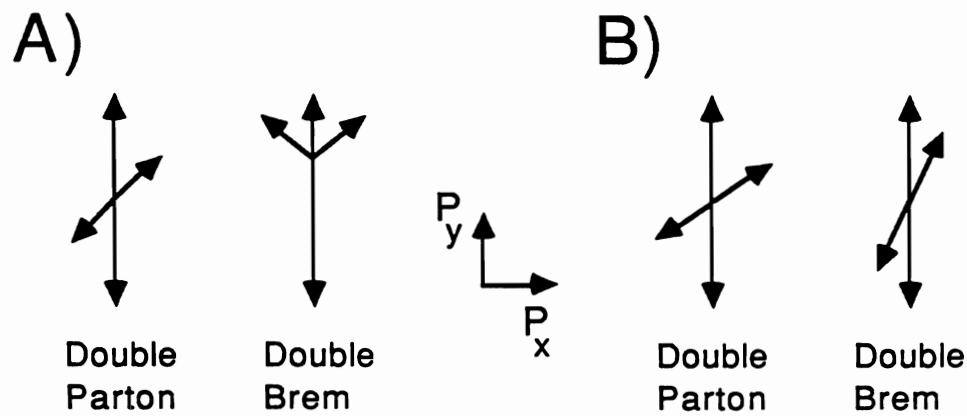


Figure 2.4: *Shown are transverse projections of momentum vectors for hypothetical 4-jet events. A) shows the effect of the extra topological freedom in double brem versus double parton events. B) shows the preference of the double brem matrix element for colinear jets.*

into an experimentally robust variable is done in section 6.3.

2.6 Jet Fragmentation

In the previous sections we dealt with the partonic level of hard $P\bar{P}$ collisions, where perturbative QCD is applicable. This implies a *stick figure* model of jet events as a collection of parton 4-vectors. In reality, partons from a hard scatter fragment into showers of hadrons. Because fragmentation involves 4 momentum transfers of order the QCD scale (~1 GeV), it cannot be calculated perturbatively. Basic arguments have been used to suggest phenomenological parametrizations for describing fragmentation. For all of these forms the parameters must be determined experimentally.

Fragmentation affects experimental jet analysis in two basic ways. First, it makes the association of jet 4-momenta with parton 4-momenta ambiguous. This ambiguity occurs at two levels, one due to the absence of a complete theoretical understanding of fragmentation, and the other due to the limits of particle detection technology. The second basic fragmentation effect concerns the measurement of jet multiplicity. It is conventional to regard fragmentation as occurring long after the parton-level interactions. In reality, the distinction between parton-level gluon radiation and fragmentation is not clear. This im-

plies that the experimental definition of jet multiplicity will depend on fragmentation. This issue is discussed in detail in section 5.5.

Fragmentation is generally thought of as occurring long after the hard collision has determined the parton momenta. One might therefore expect the momenta of the partons to be preserved by fragmentation, as the momentum of a π^+ is preserved in its decay into $\mu^+ + \nu_\mu$. In reality, fragmentation arises from strong interactions between partons, and necessarily affects the 4-momentum of the final particles. This is seen simply in that individual jets have mass, while theoretical partons are described as massless. There is no fundamental theory of how this mass is generated. Fragmentation models implicitly describe mass generation, but each in a different way. The discrepancy between models outlines the theoretical systematic uncertainty in comparing jet measurements with parton calculations.

Fragmentation also affects jet measurements through variations and imperfections in single particle detector response. As described in section 3.1.4, the CDF detector energy response to low energy charged pions is nonlinear. Because fragmentation determines the number of particles a parton evolves into, it couples to this non-linearity and affects the detector jet response. For instance a 10 GeV parton which divides into two 5 GeV hadrons will reconstruct 20% higher in energy than one which divides into five 2 GeV hadrons. To study detector effects to jet response it is therefore important to have an accurate fragmentation model.

When a detailed modeling of jet fragmentation is required in this analysis we use a modified version of the Isajet Monte Carlo [20]. Isajet normally models fragmentation as a combination of gluon radiation and hadronization. Isajet starts each event as a dijet, modeling multi-jet production implicitly by fluctuations in the gluon radiation. Our modification to Isajet suppresses the gluon radiation so as to avoid ambiguities in jet multiplicity.

This modified Isajet model independently transforms individual partons into jets of hadrons. Each parton is sub-divided longitudinally in momentum according to a *fragmen-*

tation function whose form was suggested by theoretical arguments [18]. Hadron identities are decided randomly based on branching fractions measured in e^+e^- jet events. Each hadron is given momentum transverse to the parton direction according to a phenomenological model. After all partons are fragmented, the hadron momenta are *rescaled.* In this procedure the 3-momentum of all particles produced in all jets are multiplied by a single constant near one, calculated so that the total energy of the resulting final particle state equals the total energy of the initial parton state. Because it links the kinematics of the jets, this rescaling implicitly models the effects of parton interaction during fragmentation.

Isajet fragmentation was tuned to reproduce charged particle momenta distributions as measured in the CDF tracking chamber (see section 3.2). Both the longitudinal and transverse momentum spectra of the produced particles relative to the jet direction were adjusted. Tracking data has the advantage of allowing clear identification of individual particles. It has the disadvantage of being blind to neutral particles. We chose to use tracking data to tune fragmentation as we later use the fragmentation model to study the response of the CDF energy detectors (*calorimetry*). The details of the fragmentation tuning study are presented in appendix B.

Chapter 3

The CDF Detector

The CDF detector is a multiple purpose device designed to measure many properties of $P\bar{P}$ collisions [21]. This thesis is concerned mostly with the energy detectors or *calorimeters* which surround the collision point. These detectors register the energy of most particles produced in jets. Analog signals from the calorimeters were used to trigger on the events used in this analysis. Digital signals from the calorimeters were written to tape and processed offline. Reconstructed into jets, these signals were the starting point of the double parton analysis presented in future sections. Tracking chambers, which measure the trajectory and in some cases the momenta of charged particles, were used to check the Monte Carlo but were not extensively used in the jet analysis. Holding all detectors together were online data aquisition and trigger systems. The calorimeters, tracking chambers, and online systems are described below in appropriate detail.

3.1 Calorimetry

Nearly the entire solid angle about the CDF interaction point is covered with both electromagnetic and hadronic shower detectors. These *calorimeters* are finely segmented in solid angle around the collision point and coarsely segmented radially outwards from the collision point (*ie* depth segmentation). Polar angle is segmented in *pseudo-rapidity*, defined as $\eta \equiv -\ln(\tan\theta/2)$, where θ is the angle relative to the beam axis (polar angle). The choice of this segmentation is explained below. Angular coverage about the beam axis (ϕ)

is complete. Coverage in polar angle extends down to about 2° from the beam axis. Small gaps in solid angle coverage occur at structural boundaries and feed throughs.

All calorimeters use *shower sampling* to measure particle energy. The calorimeters are constructed of many layers of a dense *absorber* material (such as steel or lead) interleaved with layers of an *active* material (such scintillator) sensitive to ionization. Particles striking the absorber material undergo energy loss through repeated interactions with both nuclei and atomic electrons. The daughter particles produced in these interactions also interact with the absorber, giving rise to a *shower* of particles. Showers penetrate many layers of absorber before exhausting their energy. As showers pass through the sampling layers they generate a detectable signal, roughly proportional to the number of particles in the shower. Each active layer is therefore said to *sample* the shower. The sum of the signals from all sampling layers is proportional to the energy of the incident particle.

3.1.1 Calorimeter Components

CDF calorimeters are hierarchically segmented. The polar angle coverage is divided into 3 *regions*. Each region contains electromagnetic and hadronic *components*. Components are constructed of many identical *modules*. Each module is subdivided transversely into *towers*. Towers are segmented in *depth*, where each depth of each tower is read out by separate electronics channels.

The beam and physics conditions of the different regions dictated the basic calorimeter design parameters. The *central* region ($0 > |\eta| > 1.1$), being at large polar angle to the beam, is the most important for studying high transverse momentum objects. It also covers the largest piece of solid angle, and experiences the lowest energy flow. Accordingly the central calorimeters were built with scintillator as the active medium. This provides good energy resolution and an adequate luminosity lifetime.

The *plug* ($1.1 > |\eta| > 2.2$) and *forward* ($2.2 > |\eta| > 4.2$) regions are closer to the beam line and hence subject to higher rates and energies. In these regions scintillator would eventually darken and degrade the energy measurement. The plug and forward regions cover

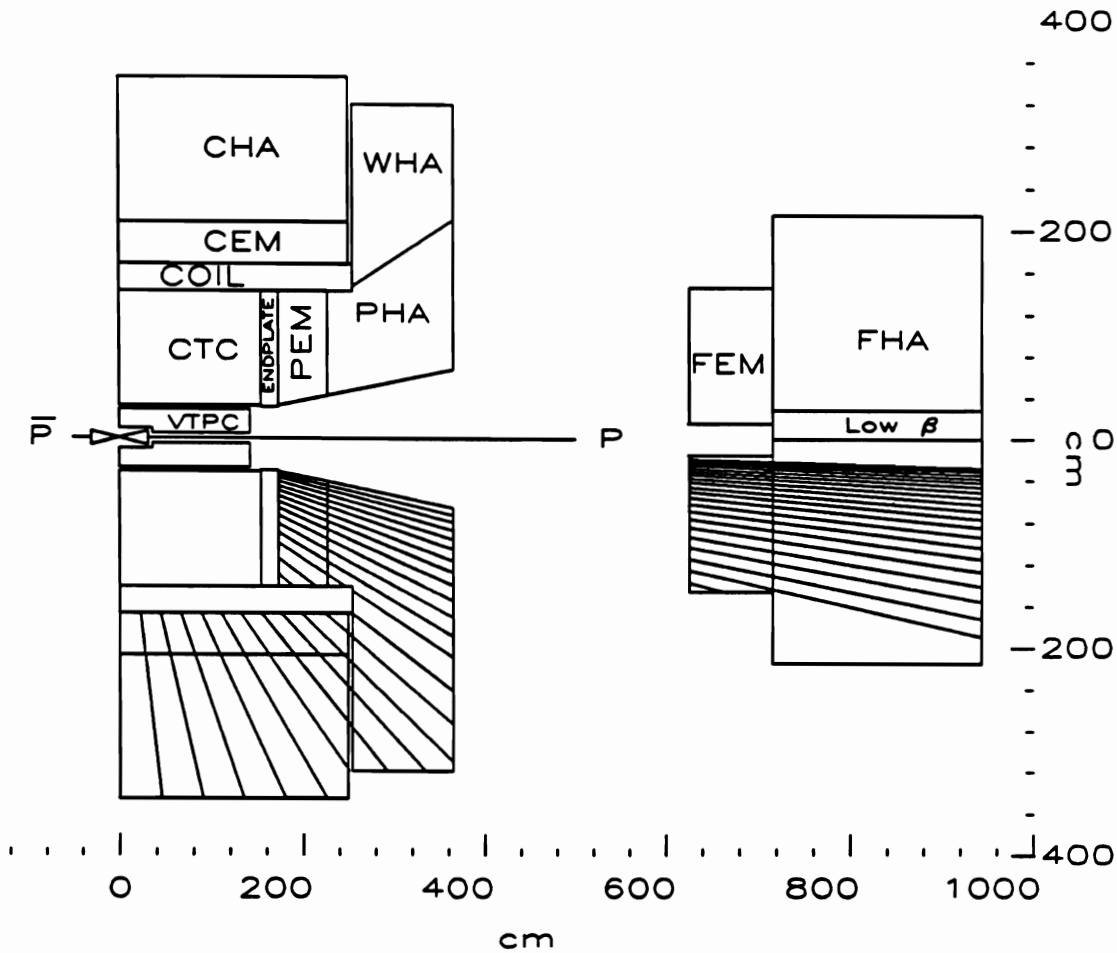


Figure 3.1: *$Y - Z$ view of the CDF detector. The interaction point is at center left. Component and tracking chambers are shown in the positive Y slice, the projective tower η segmentation in the negative.*

large amounts of pseudo-rapidity. Implementing the required transverse segmentation in these regions with scintillator would have been very hard. Accordingly, the plug and forward calorimeters use gas proportional chambers as an active medium. These have poorer energy resolution than scintillator, but have a longer lifetime. The *transverse* energy resolution is approximately uniform throughout the calorimeters. Cathode pad readout of the gas calorimeters easily facilitates the required transverse segmentation.

Each region has separate electromagnetic (*EM*) and hadronic (*HA*) components. The EM detectors use thin lead sheets as absorber, the HA relatively thick steel. The thin lead sheets allow detailed sampling of electromagnetic showers, while the thick steel assures

Comp	η Range	$\Delta\eta \times \Delta\phi$	N_{layers}	Sampler	Absorber	$\sigma_E/\sqrt{E}$
CEM	$\lvert\eta\rvert < 1.1$	$.11 \times 15°$	30	.5 scn[1,2]	.32 Pb[1]	$13\%(e^+)$
CHA	$\lvert\eta\rvert < .8$	$.11 \times 15°$	32	1. scn	2.5 Fe	$70\%(\pi^+)$
WHA	$.8 < \lvert\eta\rvert < 1.3$	$.11 \times 15°$	15	1. scn	5.0 Fe	$80\%(\pi^+)$
PEM	$1.1 < \lvert\eta\rvert < 2.4$	$.09 \times 5°$	34	.7 gas[3]	.27 Pb	$28\%(e^+)$
PHA	$1.3 < \lvert\eta\rvert < 2.4$	$.09 \times 5°$	20	.7 gas	5.1 Fe	$90\%(\pi^+)$
FEM	$2.2 < \lvert\eta\rvert < 4.2$	$.1 \times 5°$	30	.7 gas	.27 Pb	$30\%(e^+)$
FHA	$2.3 < \lvert\eta\rvert < 4.0$	$.1 \times 5°$	12	1. gas	10.2 Fe	$120\%(\pi^+)$

[1] All thicknesses are listed in cm

[2] Scintillator is polystyrene for EM detectors, acrylic for HA

[3] Gas is 50% Argon 50% Ethane with 1-4% alcohol quencher

Table 3.1: Calorimetry Component Properties

containment of hadron showers. The 7 CDF components (3 EM and 4 HA) are shown in figure 3.1. Table 3.1.1 lists design and performance parameters for each.

3.1.2 Readout Segmentation

In order to resolve spatial characteristics of energy deposition, the solid angle coverage of the CDF calorimeters is finely divided. Figure 3.2 shows CDF transverse segmentation in $\eta\phi$ space. Transverse segmentation is adjusted radially to form *towers* which project to the nominal interaction point. There are 5184 towers, read out by 19440 independent electronics channels, of which less than 1% were dead during the runs relevant to this analysis. Dead channels had a negligible impact on the analysis result.

Gas calorimeter transverse segmentation is accomplished with etched cathode pads. Signals from individual plane cathode pads are cabled together in depth before amplification. To aid in electron identification, the readout of the plug and forward EM detectors are divided in depth. Thus the cathode pads from the first and last 15 planes of the FEM are ganged to a single amplifier, as are those from the first 4, middle 26, and last 4 PEM planes. The ϕ segmentation of the gas calorimeters is comparable to their η segmentation , whereas the central ϕ segmentation is more coarse. Central components read each tower with 2 photomultiplier tubes on opposite ϕ sides. The signal balance between tubes allows

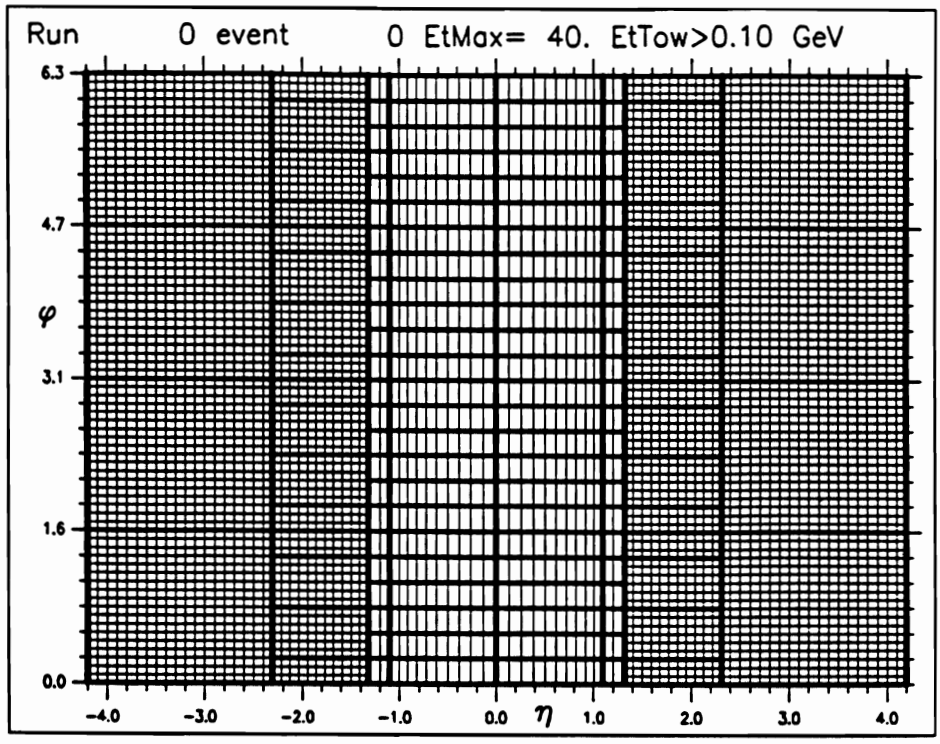


Figure 3.2: *The $\eta\phi$ segmentation of the CDF calorimeters is shown. Light lines are tower boundaries, medium lines module boundaries, and dark lines are component boundaries.*

a determination of the ϕ of a single particle to better than 1°.

$\eta\phi$ calorimeter segmentation was chosen for two reasons, one global and the other local. Globally, because partons carry only a fraction of proton beam momentum, $P\bar{P}$ events are not produced at rest in the lab frame. The rest frame of the hard collision is generally boosted relative to the lab frame along the beam direction (Z direction). To avoid biasing the physics analyses, the solid angle segmentation must therefore be invariant under boosts along Z direction.

The *rapidity* of a particle is Lorentz invariant in this sense. Rapidity is defined as $Y \equiv .5\ln(\frac{E+P_z}{E-P_z})$. Under the rules of special relativity, rapidity transforms linearly under a boost, $Y \xrightarrow{boost} Y' = Y + \text{constant}$. Taking the derivative of this proves that segmentation in rapidity would be Lorentz invariant, $\Delta Y \xrightarrow{boost} \Delta Y$. From the definition we see that rapidity is a function of particle mass and polar angle. When considering particles of different masses

there is no single relation between rapidity and polar angle. Consequently, rapidity cannot be used to define angular segmentation.

Pseudo rapidity ($\eta \equiv -\ln(\tan\theta/2)$) is a similar variable to rapidity which depends only on polar angle. For particles with mass much less than their transverse momentum, pseudo-rapidity is a good approximation to rapidity. To see this we expand Y in terms of η and $\alpha \equiv \text{mass}/P_t$ to get;

$$Y = .5\log\left(\frac{\sqrt{\cosh^2\eta+\alpha^2}+\sinh\eta}{\sqrt{\cosh^2\eta+\alpha^2}-\sinh\eta}\right)$$
$$\approx \eta - .5\times\alpha^2\tanh\eta + \mathcal{O}(\alpha^3)$$

For low α particles η nearly equals Y, and a boost is nearly a simple shift in η. Thus, for the high P_t particles of interest to CDF, segmentation in η will be approximately invariant under Z boosts. Being transverse to the Z direction, ϕ is also invariant under Z boosts, and is a logical choice of orthagonal solid angle variable.

Locally, $\eta\phi$ space produces approximately round jets. To see this, consider jet fragmentation as multi-body parton decay. We assume the decay is independent of the parton direction relative to the beamline or any other parton, in which case the decay products are produced symmetrically about the parton direction. In particular, the distribution of momentum transverse to the parton direction will be azimuthally symmetric. It is possible to show that particles produced with a fixed transverse momentum relative to an arbitrary axis describe a circle about that axis in $\eta\phi$ space.

To demonstrate, consider an arbitrary point in $\eta\phi$ space specified by a momentum vector $\vec{P}$. We consider the set of points having a small transverse momentum relative to this vector. We model these points by adding a small transverse kick to the original vector, $d\vec{P} = \Delta\theta\hat{\Theta} + \sin\theta\Delta\phi\hat{\Phi}$. The magnitude of the kick is held constant while its direction is allowed to vary,

$$|d\vec{P}|^2 = \Delta\theta^2 + \sin^2\theta\Delta\phi^2 = \text{constant}.$$

For small $d\vec{P}$, $\sin\theta$ is a constant, being the angle of $\vec{P}$. From the definition of η we can

derive the differential relation;

$$\Delta\eta \approx -\Delta\theta/\sin\theta.$$

This gives the $\eta\phi$ locus of the points described by the kick $d\vec{P}$;

$$\Delta\eta^2 + \Delta\phi^2 \approx \Delta\theta^2/\sin^2\theta + \Delta\phi^2 = |d\vec{P}|^2/\sin^2\theta = \text{constant}.$$

Therefore, jets will on average have a circular P_t distribution in $\eta\phi$ space. This symmetry is exploited by the clustering algorithm described in section 4.3.

In addition to cathode pad readout, the gas calorimeters have anode plane readout of the longitudinal shower development. These are grossly segmented in $\eta\phi$ space, roughly the size of a single module. Anode signals are used in the noise removal algorithms described in section 4.2, and in electron identification.

The CHA and WHA components digitized the time of charge deposition. The resolution of these hadron TDC signals was 1 nanosecond. These signals were used to remove cosmic ray and other backgrounds, as described in section 4.2.

3.1.3 Test Beam Calibration

Modules of each component were put in test beams of electrons and charged pions. The response to beams of momenta from 10 to 200 GeV formed the energy calibration. Shower development and energy resolution were also measured as input to the CDF detector simulation. Calorimeter response was found to be linear over the range of test beam momenta to within a few percent. Scans across modules were used to make response maps. Some boundaries between regions and/or modules were also scanned to study 'crack' response. Figure 3.3 shows the response of central calorimeters (EM + HA) to π^+s and e^+s scanned across a module ϕ boundary. These boundaries occur every 15°. Energy loss occurs when the particles hit the steel skin separating modules. Such response curves were incorporated into the CDF detector simulation.

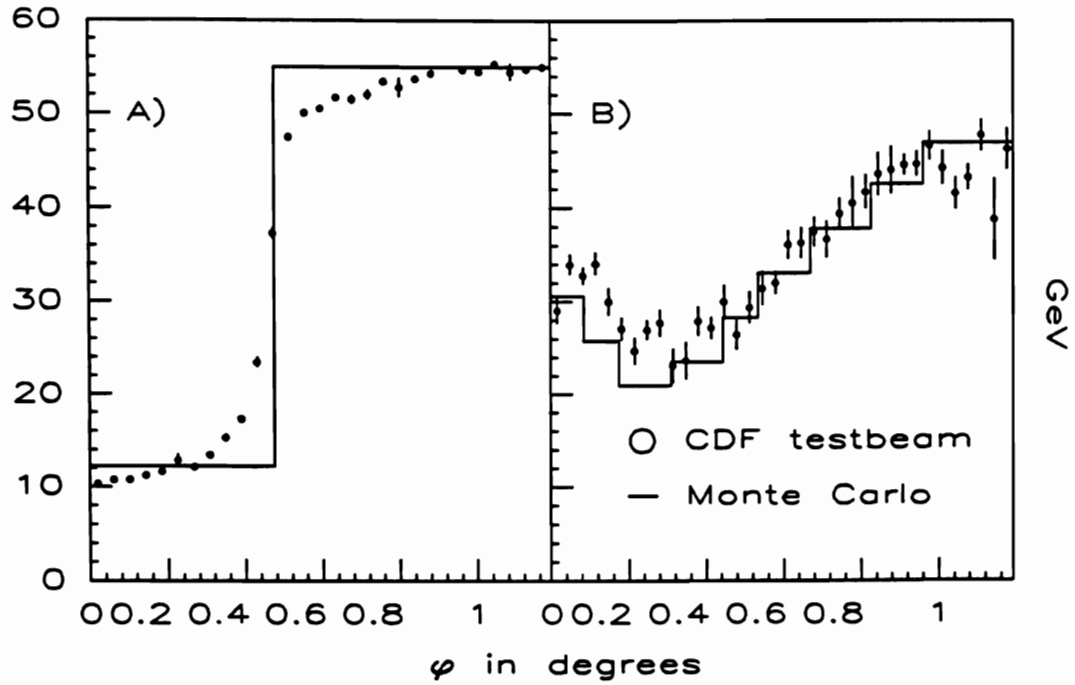


Figure 3.3: *The response at a ϕ module boundary in the central calorimeter to (A) 50 GeV pions and (B) 56 GeV electrons. The overlaid curves are from the CDF detector simulation, which was tuned on this data.*

3.1.4 Calorimeter Non-linearity

In the central calorimeters the response to charged pions between momenta of .5 and 10 GeV was measured using $P\bar{P}$ data. This study used minimum bias triggers, where particle densities are low, and isolated pions could be found. The response was found to be low by around 20 % at low energies [5]. This non-linearity is reproduced in the CDF detector simulation.

Limited tracking information in the plug and forward regions prevented a low energy response measurement in the gas calorimeter. The simulation assumed the gas calorimeter nonlinearity to be the same as in the central. This assumption has very little effect on the subsequent analysis, as very few particles from jets in the gas calorimeter fall in the non-linear region (.5 to 10 GeV). This is because gas calorimeters jets have higher energy than central jets of the same P_t.

3.1.5 Detector Cracks

Due to mechanical constraints, certain regions of the $\sim 4\pi$ calorimeter solid angle coverage of CDF are under-instrumented. Given the $\eta\phi$ segmentation of the calorimeters, these regions are bands in either η or ϕ. Figure 3.2 shows these regions on an $\eta\phi$ grid of the CDF geometry. The narrow bands at constant ϕ in the central calorimeters are due to the steel skins which separate towers. The larger bands at constant η are due to junctions between the components. These regions are capable of absorbing single particles without registering their full energy. However, because a jet consists of many particles, only a fraction of the jet energy will be lost.

Some crack regions of CDF were scanned with test beam particles, and their response mapped out (see section 3.1.3). Due to their complexity, other crack regions were never scanned. To model energy loss in these regions, we rely on parametrizations of the longitudinal and transverse shower development measured at the test beam. The profiles were measured in the main regions of the calorimeters, far from the crack regions [22, 23]. As a function of particle impact point, the CDF detector simulation extrapolates these profiles into the dead regions to model the energy loss.

3.2 Tracking

Near the vertex particles are tracked by a time projection chamber (VTPC), which gives polar angle information but very little ϕ or momentum information. VTPC tracks are used to determine the $P\bar{P}$ collision position. VTPC vertex resolution is $\sim$ .3 cm along the beam direction (Z). The long bunch length of the Tevatron caused the Z vertex position to be distributed about the nominal by a gaussian with a sigma of 45 cm. As previously mentioned, calorimeter towers point back to the nominal Z vertex position. When the actual Z vertex differs from the nominal, the projective tower geometry is distorted (see figure 3.1). To eliminate cases of extreme distortion, events with a Z vertex position more than 60 from the nominal were rejected. The Z vertex position was used to correct for

residual distortions. These distortions altered the η and E_t of the tower relative to their nominal value, being at most .3 in η and 30% in E_t for the 60 cm maximum Z vertex position.

The central tracking chamber (CTC) is a large azimuthally symmetric drift chamber that surrounds the VTPC. This provides good ϕ and momentum measurement of charged particles. Particle θ is measured with stereo layers. The novel feature of this chamber is the 45° tilting of the drift cells, which gives a purely radial drift in the nominal electric and magnetic fields. The jet fragmentation study described in section 2.6 uses CTC tracks. Surrounding the CTC is a superconducting solenoidal coil operated at 1.5 Tesla. The coil provides the track curvature for CTC momentum measurement, which has a resolution of $dp_t/p_t^2 \leq .02GeV^{-1}$. Figure 3.1 shows the tracking chambers and coil relative to the calorimeters.

3.3 Trigger

The trigger for the 1987 run used low-angle scintillator counters ($3.2 > |\eta| > 5.9$) and analog calorimeter signals. The counters were required to have at least 1 hit on each side of the detector in time with a beam crossing to within 15 nanoseconds. The time resolution of the counters is 200 pico-seconds. These counters also integrated the luminosity of the run by recording every valid crossing in a scaler.

Fast calorimeter signals were required to sum to a minimum E_t. Sums were taken over *trigger towers* with $\Delta\eta \times \Delta\phi$ segmentation of $\sim .2 \times 15°$ over the whole detector. EM and HA trigger tower energies were separately required to be above 1 GeV before being included into the sum. The minimum E_t threshold was varied by run between 20 and 45 GeV according to luminosity, keeping the event-to-tape rate constant. Noise problems in the gas hadron calorimeters prevented their inclusion into the calorimeter trigger E_t sum. The noise was removed offline as described in section 4.2. The trigger processing time was less than the time between beam crossings of 7 μ-seconds. Dead time was incurred by digitization and

readout time after a valid trigger, and by inhibiting the trigger during Tevatron booster ring (main ring) injection and extraction cycles. The main ring was used during collider runs to produce $\bar{P}$s for the next *store* of protons and anti-protons. The data used in this analysis were all taken during the 1987 collider run at FNAL which wrote 30 nb^{-1} ($\pm$ 15% systematic) of data to tape. A total of 150,000 calorimeter triggers were written to tape.

3.4 Online Readout and Calibration

The online data acquisition system for the 1987 run included three levels of processing, one analog hardware, one digital hardware, and one digital software. The tracking and calorimeter data were read out through separate systems. The basic structure of each system was very similar. Only the calorimeter readout system is described below.

The *front end* electronics connected directly to the active calorimeter components (either photomultiplier tubes or cathode pads). These used charge-integrating amplifiers to convert calorimeter charge signals into voltage levels. A *sample and hold* system was used to store these voltage levels on capacitors for about 1 micro-second while they waited for digitization. Amplifier output voltages were stored both immediately before and after each beam crossing. The *difference* of these signals was digitized by the ADC, implicitly removing any low frequency ($\nu < 1$ KHz) noise.

Single channel voltages were converted to digital signals with a 16 bit ADC. Signals from many channels (a maximum of 400) were sequentially bussed to the ADCs during readout. A total of 96 ADCs were used to read out all the calorimeters. Channel pedestal values were subtracted from the analog signals by circuitry on the the ADC board before digitization. Analog comparator circuits suppressed digitization of signals less than 10 counts from pedestal (equivalent to between 20 and 80 MeV of energy, depending on the detector). This threshold test sped up the digitization process and reduced the quantity of digital data by a factor of 10. Complete digitization took 40 μ-seconds $\pm$50%, depending on event complexity.

Once digitized, calorimeter signals were read in by a farm of 36 dedicated microprocessors. In addition to readout, the microprocessors multiplied the data by channel dependent electronic and gas gain correction factors. The methods for determining these factors are described below. The microprocessors also controlled the digitization process. Digital data from the events was stored by the microprocessors in 4 buffers to await computer readout.

Microprocessor buffers were read out sequentially by specialized FASTBUS modules. Other FASTBUS modules controlled the communication of trigger signals from the trigger to the microprocessors. Digital data was passed through FASTBUS to a VAX 785 computer. The computer controlled the FASTBUS network. Data was reformatted by the computer, and then written to tape.

Energy calibration was monitored online by radioactive sources. Cs^{137} sources were run past scintillator surfaces of all towers of the central calorimeters to monitor scintillator darkening. Source calibration was performed at the begining and the end of the 1987 run, which showed an energy scale change of less than 2% [24]. Phototube response was monitored weekly with LED flashers in the CEM and with lasers in the WHA and CHA.

Gas calorimeter gains could change by up to 50% in 24 hours, due to atmospheric pressure changes. A system of Fe^{55} source monitors was used to monitor gain changes. Sources were mounted on dedicated proportional tubes run with the same voltage and gas as the main detectors. Source tubes were distributed in redundant pairs around the detector, and placed in thermal contact with the absorber medium. Signals from these tubes were read out through an independent CAMAC online system once every 15 minutes. Gain changes of more than 5% stopped a data run. A new run was started after adjusting the calibration constants to compensate for the gain change. These gain factors were downloaded to the microprocessors prior to starting the new run. These gain factors were also applied to the calorimeter trigger signals.

Electronics channel pedestals and gains were monitored daily. Gains were measured using calibration circuits built on the front end boards. These injected known steps of

charge into the amplifier inputs. The amplifier response was measured using normal digital readout. Electronics gains were found to be stable to better than 1% over the run. The channel-to-channel variations were of order 10%, caused mostly by variations in the source capacitance of the signal lines and detectors. These gain variations were removed during readout as described above.

ADC gains were also monitored daily. Precision DACs built in to the ADC board were used to create known voltage steps. The ADC response to the voltage steps was adjusted by changing the gain of low gain amplifier circuits which buffered the signals from the ADC input. A similar calibration technique was used to adjust the analog pedestal subtraction and comparator circuits.

Chapter 4

Event Reconstruction

The calorimeter data written to tape during a run were stored in a highly compressed format. Raw data was written as integers, scaled to maximize the dynamic range of the 16 bit ADC. To minimize the time spent in writing tape, data were arranged sequentially by the order in which the microprocessors were read out. Offline, this raw integer data was transformed into equivalent GeV of energy. The data were also re-arranged into a physically relevant format. Once re-formatted, noise and unphysical signals were eliminated from the data. Pattern recognition algorithms were then applied to find jets. This chapter describes the chain of algorithms used to reconstruct jets out of raw calorimeter data.

4.1 Energy Reconstruction

The first reconstruction job was removing digital shifts from the raw calorimeter ADC data. As mentioned in section 3.4, channel pedestals were removed online by an analog subtraction circuit on the ADC board. This subtraction was imperfect, and a residual *pedestal shift* of a few tens of counts remained in $\sim 10\%$ of the channels. These shifts appeared in the data as a constant, nonzero signals (either positive or negative) in the affected channels. Positively shifted channels registered a signal even when no energy was present. Both positively and negatively shifted channels changed the value of detected energy. While individual shifts were small, collectively they generated a significant false signal.

Pedestal shifts were calculated offline using $P\bar{P}$ minimum bias triggered data. Minimum

bias events have very low energy flow, with typically 7 particles per unit rapidity [19]. Given the transverse segmentation described in section 3.1.2, we therefore expect a non-zero signal in a given calorimeter tower for only 5% of the minimum bias events. By contrast, a channel with a shifted pedestal registers a non-zero signal nearly every event. Accordingly, pedestal shifts were identified as channels with a high rate (30%) of non-zero signals in minimum bias events. A value for the shift was calculated for these *high occupancy* channels as the mean of that non-zero signal. In keeping with the accuracy of the threshold circuit described in section 3.4, shifts of less than 10 ADC counts were ignored. Channels with shifts of more than 10 ADC counts were corrected.

The pedestal shift correction algorithm simply removed the shift measured in minimum bias events from the ΣE_t triggered data. Data from channels flagged as having a shift were altered each event by subtracting the value of the minimum bias shift. Shifts could be either positive or negative. The typical shift magnitude was a few tens of counts, or equivalently about 50 MeV. The number of channels with shifted pedestals ranged from from 1% in the CEM to 10% in the FHA. The number of shifted channels was approximately constant in time, whereas the identity of shifted channels changed each time the analog pedestal values were remeasured. Because of this variation, pedestal shifts were calculated separately for each run.

The same minimum bias analysis was used to find hot channels. Hot channels were defined as channels with very large pedestal shifts ($\geq$ 200 counts) or widths ($\geq$ 100 counts). Roughly, 200 counts corresponds to about 1 GeV of energy. These channels were suppressed offline. Channels which failed to respond to electronic calibration were also suppressed. Less than 1% of all calorimeter channels were suppressed in 1987 data.

Once correct at the ADC level the data were converted to energies by multiplying the ADC data with detector dependent calibration constants. Section 3.1.3 describes how these constants were derived from testbeam data and maintained by sources and other calibrations throughout the run.

Because of high voltage breakdown problems, some layers of the gas calorimeters were turned off. This reduced the signal seen in towers including a dead chamber. To compensate for this loss, energies from towers including a dead chamber were multiplied by a correction factor. This factor was determined from CDF jet data by measuring the *longitudinal profile* of the energy deposited in the gas calorimeter. This profile was measured using anode signal readout of the detectors, and only in regions without any dead chambers. The profile from each jet was normalized to the total jet energy, giving the fraction of energy appearing in each layer. Averaged over many jets, this profile described the average fraction of jet energy appearing in a given layer. The correction factor for a tower with dead chambers was calculated as the inverse of the sum of the average normalized longitudinal profile over all active chambers;

$$\text{Correction} = 1/\Sigma_{active\ chambers} P_i.$$

For the worst case, this factor was 20% from unity.

4.2 Noise Removal

In the scintillator calorimeters it was necessary to remove *spike* signals associated with HV discharges at phototube bases. Single phototube spikes were also caused when particles showered in the light guides or wave shifters. Although relatively rare, these spikes accounted for 5% of all the calorimeter triggers taken in the 1987 run. Single phototube spikes were eliminated offline by exploiting the redundant phototube readout of each tower. In a normal particle shower, both tubes register roughly the same energy. In a spike, only one tube sees any signal. Spikes were removed by requiring both tubes to have measured the same energy. Towers with an energy match 4σ outside the observed testbeam response were removed from the data. Events which had contained spikes were then analyzed normally.

In the gas calorimeters there were other kinds of spikes. Most common were large isolated deposits of energy in the gas hadron calorimeters. This energy appeared without associated tracks or electromagnetic energy, and so was known to be noise. The depositions

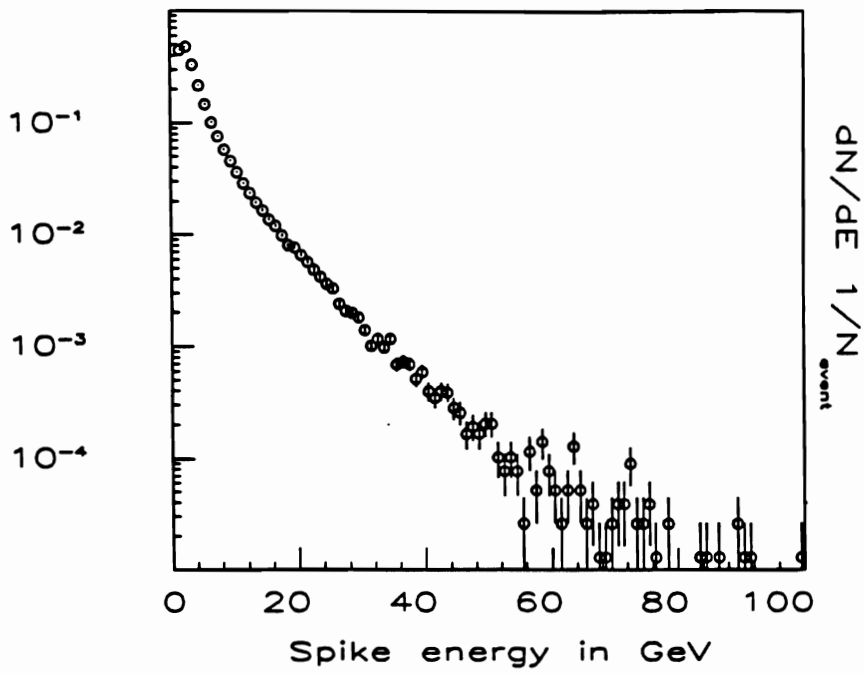


Figure 4.1: *The effective texas tower energy spectrum from the PHA calorimeter is shown*

were also much more spatially contained than normal particle showers. These so called *texas tower* spikes were caused by low energy neutrons. Neutrons are produced as a normal part of hadronic showers. At low energy, neutrons have a small cross section with lead and steel, and so penetrate deeply into the calorimeters. Low energy neutrons have a much larger cross section with hydrogen. This enables them to transfer their kinetic energy to protons in the ethane of the chamber gas. The protons lose that energy through ionization loss, ranging out in just a few centimeters of the gas. This energy generates a large calorimeter signal in a single layer and tower of the sampler.

The signal generated by neutron-induced protons is much larger than that generated by a shower of the same energy. For instance, a 1 MeV knock-on Proton in the PHA calorimeter deposits charge equivalent to a 28 GeV pion shower. To understand this, we must review the calorimeter calibration procedure. Testbeam calibration constants were defined by equating the signal observed in the active medium of the calorimeter with the energy of the incident particle. Most shower energy is lost inside the absorber material. The

Table 4.1: Relative calorimetry absorber/sampler energy absorption

Detector	CEM	CHA	WHA	PEM	PHA	FEM	FHA
MIP absorber/sampler	4.1	14.5	29.0	1800	31,000	3200	63,000

This table lists the ratio of energy deposited by a minimum ionizing particle in the absorber and sampler of the different CDF calorimeter components. This ratio approximates the factor by which energy from neutron-induced protons (texas towers) is over-estimated.

calibration therefore extrapolates the observed signal, scaling it up to compensate for signal lost in the absorber. This calibration procedure is necessary in all sampling calorimeters. Unlike showers, neutron-induced protons range out in the gas, depositing all their energy in the sampling medium. Interpreted using a shower-generated calibration, the computed 'energy' of this signal therefore over-estimates the true proton energy. The amount of over-estimation is given by the value of the calibration scaling factor.

The neutron sensitivity of a detector is proportional to the factor by which observed signal is scaled to account for absorbed signal. This factor can be estimated by modeling a particle shower as a collection of minimum ionizing particles (MIPs). In that model, the calibration scaling factor is simply the ratio of energy deposited by a MIP in the absorber to that in the sampler. Table 4.1 lists this ratio for all the CDF components. The larger the value, the more susceptible the calorimeter is to low energy neutrons. The gas hadron calorimeters were clearly the most susceptible. Figure 4.1 plots the observed texas tower 'energy' spectrum for the PHA gas calorimeter. Given the estimated scaling factor for these detectors, the data indicates a fast falling neutron energy spectrum with a range of a few MeV. The high rate and 'energy' of texas towers in the gas hadron calorimeters prevented their inclusion into the ΣE_t trigger. While texas towers were observed in the gas EM calorimeters, they did not present a serious problem in the trigger. No texas tower effects were seen in the scintillator calorimeters.

In the PEM detector spikes were also caused by high voltage breakdown at the wire feedthroughs of the chambers. This type of spike accounted for 30% of all calorimeter triggers taken in the 1987 run. A similar problem in the FEM occurred at a much lower

rate. Like texas towers, these spikes deposit their signal at a single point in the calorimeter. Because both types of gas calorimeter spikes gave a similar signature in the data, both were removed by the same algorithm. The gas spike removal algorithm is described in detail in appendix C. The principle was to remove deposits of energy that were too localized to be particle showers. Both the pad and the (redundant) anode gas calorimeter data were used in this algorithm. After texas tower removal, events which had contained texas towers were analyzed normally.

The efficiency of the gas spike removal algorithm was tested using a sample of verified texas towers in ΣE_t triggered events. The verification algorithm required the texas tower to be *isolated* in $\eta\phi$ space from all other energy deposits. Isolation allowed tight cuts on the anode energy (*longitudinal profile*), resulting in a very pure texas tower sample. Unfortunately, the isolation requirement precluded testing the efficiency of the algorithm when the texas tower overlapped with a jet. Figure 4.2 plots the isolated spike removal efficiency as a function of spike energy in the PHA component. The plot shows that over 95% of all isolated texas towers above 20 GeV in energy are eliminated.

In addition to texas towers, the PHA calorimeter was afflicted with electronics noise caused by ground loops in its signal cables. The noise was coherent across all the channels in the cable. Each cable carried signals for 12 channels. Cable noise appeared in the data as a uniform signal across a 3×4 *plateau* of towers. This topology was used to eliminate the noise offline. The redundant anode readout of the PHA was also used to eliminate cable noise. In cable noise, only the pads register energy. In particle showers, the pads and anodes measure approximately the same energy. PHA cables with ≥ 8 of 12 towers above an energy of 100 Mev, and showing a corresponding anode signal of less than 1 GeV, were eliminated. Events which had contained cable noise were then analyzed normally.

Events were filtered based on TDC information from the CHA and WHA components. This eliminated triggers on cosmic rays and particle spray from Tevatron booster rings. Events where ≥ 8 towers each reading ≥ 2 GeV more than 5 nanoseconds outside the beam

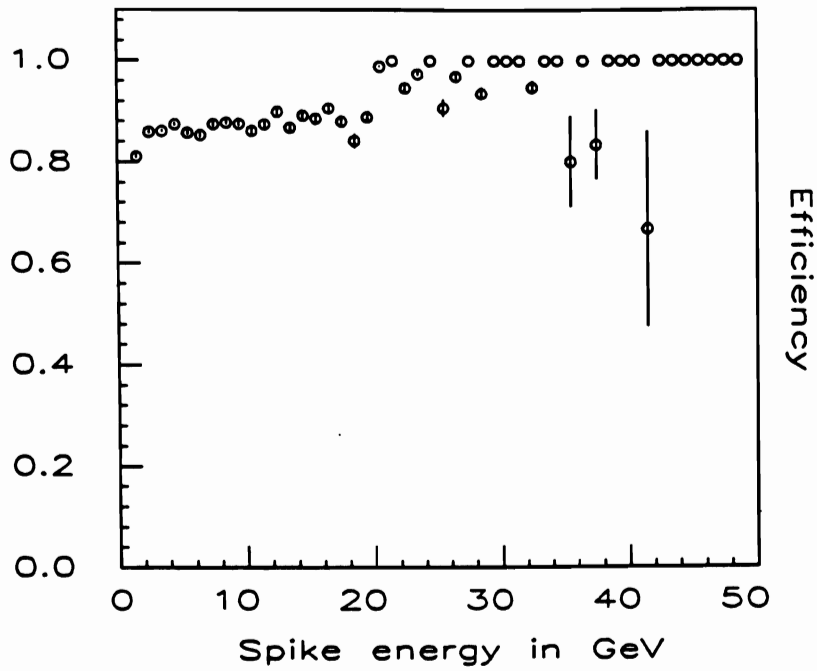


Figure 4.2: *Efficiency for removal of isolated texas towers as a function of energy in the PHA detector.*

crossing time were removed.

4.3 Jet Clustering

Once the calorimeter data were fully cleaned, they were *clustered* to reconstruct jets. In clustering, the information from the thousands of separate tower energy measurements is combined and simplified. Nearby towers are associated together by an *algorithm* based on their proximity and energy, as shown in figure 1.2. Conceptually, clustering attempts to undo parton fragmentation, recombining the energies deposited by the many particles produced. Clustering determines the jet interpretation of events, describing the number of jets in an event, as well as their direction and energy. The 4-momentum of each cluster is computed assuming the energy in each tower was deposited by a massless particle striking tower center. The cluster 4-momentum is simply the sum of the 4-momenta of all the towers in that cluster. The relationship between cluster 4-vectors and parton 4-vectors of theoretical calculations is discussed in the next chapter.

Because fragmentation is subject to large fluctuations, no single jet looks exactly like any other. It is therefore unrealistic to expect any single clustering algorithm to perfectly undo the effects of fragmentation for every jet. Instead, a good clustering algorithm must be flexible enough to provide a reasonable jet interpretation of all events. The response of the algorithm should therefore be uniform over a reasonable range of fragmentation. Additionally, a good algorithm should be simple, and incorporate our physical understanding of both fragmentation detector response.

The CDF algorithm uses a *fixed circle* method, collecting together towers lying within a fixed radius in $\eta\phi$ space around a selected seed point. This circle is displayed in figure 1.2. Initial seed points were selected as centers of regions of high E_t density in the calorimeters. As described in section 3.1.2, jets are on average circular in $\eta\phi$ space. Therefore the CDF algorithm collects towers in a uniform, unbiased way. The details of the jet clustering algorithm are presented in appendix D.

This algorithm was selected over others (most notably algorithms which use the *adjacency* of towers) based on tests using $P\bar{P}$ data. These tests attempted to quantify the criteria described above. The clusters produced by a given algorithm were examined for having missed energy clearly associated with a jet, having included spurious energy, or having improperly divided energy into separate jets. As these qualities are difficult to define objectively, judgements focused on the *consistency* of algorithm response over a common sample of random jet events. Another test examined the sharpness of jet angular separation resolution. In all these tests, the fixed circle algorithm was found to be superior.

The single parameter of the jet clustering algorithm is the radius, which was chosen to be .7 $\eta\phi$ units for this analysis. Clustering with a different radius can give different jet energies, or even a different numbers of jets in an event. This ambiguity cannot be resolved theoretically; *ie* there is no single 'right' value of radius to use. The .7 value was chosen based on studies using CDF data, where it was found to give both reasonable jet P_t resolution, and reasonable jet angular separation. A large cone gives better P_t resolution by

including more of the fragmented particles. A small cone allows good jet angular separation at smaller values of centroid separation (see section 5.7). To be consistent, the .7 value of circle radius was used throughout all the CDF data and most of the Monte Carlo studies presented for this analysis.

Chapter 5

Cluster Correction and Selection

In chapter 2 we discussed the calculation of jet cross sections using perturbative QCD and the parton model. In chapter 4 we discussed the procedure used to reconstruct jets from CDF data. Our ultimate goal is to quantitatively compare the QCD calculations with CDF data. Three problems prevent simply equating the 4-vectors of clusters with partons. First, as mentioned in section 2.6, there is a fundamental theoretical ambiguity in equating 4-vectors since jets have mass and partons do not. Secondly, calorimeter cracks and non-linear response can reduce the measured energy and momentum of a jet from its true value. Reconstruction and energy flow effects also systematically change the measured jet 4-vector from its true value. Lastly, not all clusters deserve to be called jets. Very low transverse momentum clusters may be reconstructed around fluctuations in the energy flow of an event. These *fake* clusters must be eliminated from the events in order to correctly measure the jet multiplicity.

This chapter describes the use of a detailed Monte Carlo model of jet creation and detection to quantitatively study the last two effects listed above. The jet model used is described in section 2.6, the detector model throughout chapter 3. Using this model, a correction function which adjusts parton 4-vectors to compensate for calorimeter energy mis-measurement is made. Cluster selection criteria for minimizing the ambiguity involved in counting jets are also developed. These corrections avoid the theoretical question of how jets get mass. Instead, they correct only for detector and reconstruction response effects.

Jet mass generation is modeled explicitly in the Monte Carlo, as described in appendix E. Thus the final comparison between data and Monte Carlo is made between corrected clusters and massive partons.

5.1 Underlying Event Contribution

When partons undergo a hard collision, the remaining parts of the proton and anti-proton interact, fragmenting into hadrons. Most of theses daughter products go down the beam pipe. A few are emitted at large angles, and deposit energy in the calorimeter. This *underlying* energy is distributed nearly uniformly throughout the detector, much like the energy deposited in a minimum bias event. Some fraction of this energy overlaps with the jet energy from the hard scattering. To obtain a true measure of jet energy and momentum, this underlying event contribution must be subtracted from the jet. It is impossible to determine whether an individual particle arose from the hard scattering or the underlying event. Consequently, it is impossible to extract the underlying event from each jet particle by particle. Instead, we subtract from each jet an estimate of the *average* underlying event contribution. This will degrade the jet energy and momentum resolution, but will on average give the correct value for the jet momentum and energy.

We calculate the average value of the underlying event contribution using $P\bar{P}$ dijet data. To measure the underlying event energy we must differentiate it from jet energy. To do this we exploit the uniform distribution of underlying event relative to the highly collimated energy distribution of jets. Figure 5.1 shows the transverse energy flow in a sample of $P\bar{P}$ dijet events. The events were selected to have *exactly* 2 clusters of $P_t \geq 15$ GeV (*ie* no third cluster above 15 GeV P_t). The plot shows E_t density by ϕ relative to the ϕ of the leading jet (averaged over η). The plot averages the E_t flow of many events. The peak near $\phi = 0$ is from the leading jet, the peak near $\phi = \pi$ from the balancing jet. The balancing jet peak is smeared by initial state radiation effects, which make the jets not be opposite in ϕ. These effects are described in appendix E.

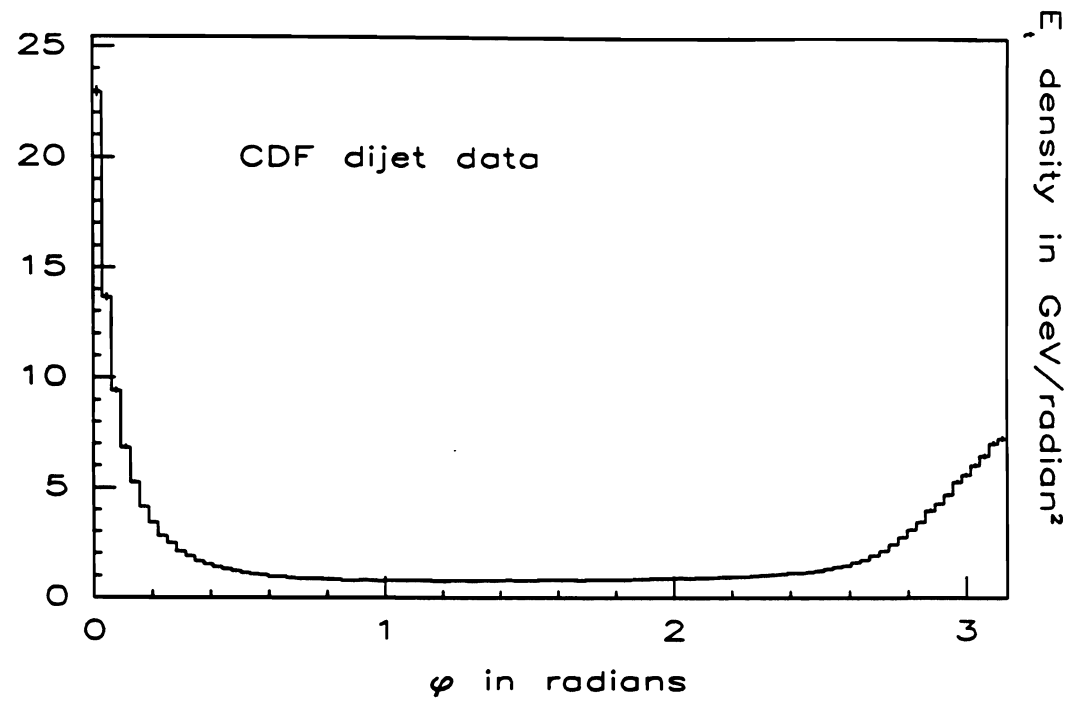


Figure 5.1: *The average E_t density in CDF dijet events relative to the ϕ position of the leading jet is shown.*

The wide, flat valley of E_t between the 2 jets is due to underlying event. The flatness of this valley suggests that it extrapolates smoothly under the jets. The flatness also shows that jet energy does not extend very far from the jet centers. This implies that we can estimate the underlying event contribution in jets by measuring the E_t density well away from the jets in ϕ.

The E_t contribution of the underlying event was measured by summing the E_t in a circle in $\eta\phi$ space. This circle imitates the one used in the jet clustering, having a radius of .7 units. The circle center was positioned at a random η, 90° in ϕ away from the leading jet ϕ direction. This E_t was averaged over many events to define the average underlying event E_t contribution. The average value was about 1.5 GeV of E_t. No dependence of this average on the kinematics of the jet system was found. A slight dependence of underlying E_t on the η position of the circle was found. This η dependence is shown in figure 5.2, and is included in the parametrizations used in the cluster correction algorithm.

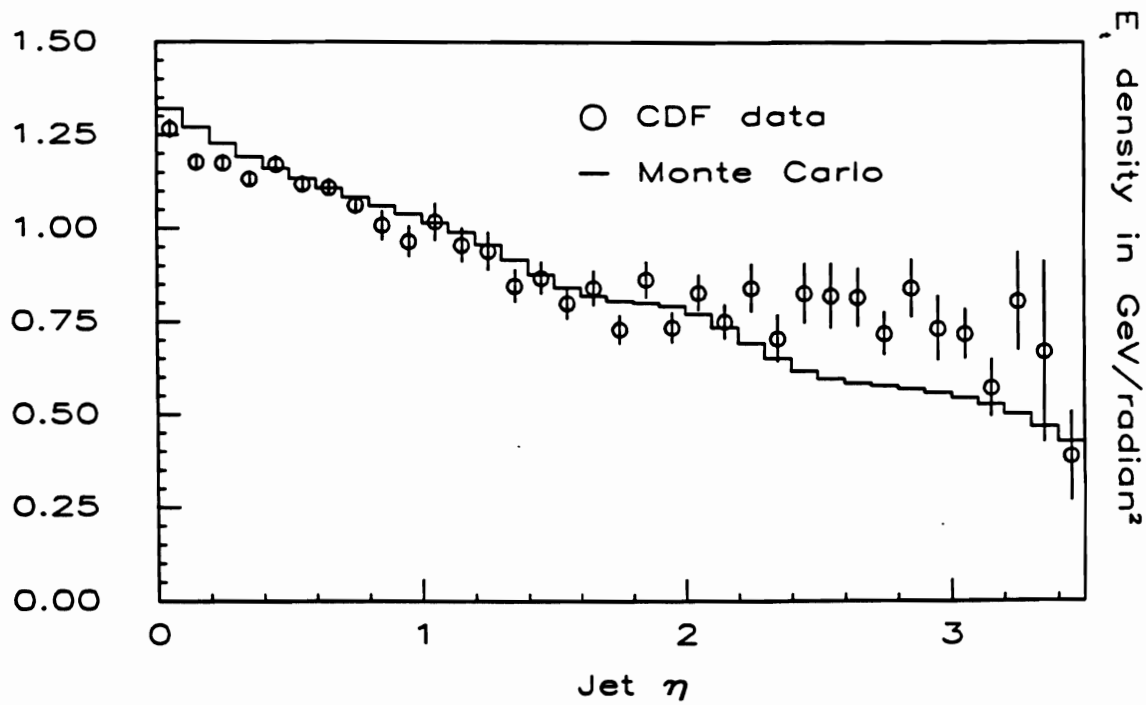


Figure 5.2: *The average E_t density as a function of jet η. Overlaid is the response of the tuned Isajet Monte Carlo, measured using the same analysis. The slight discrepancy in the forward calorimeters ($\eta > 2.2$) is probably due to residual detector noise.*

5.2 Energy Loss Mechanisms

The largest loss of jet energy comes from the non-linear response of the central calorimetry to low energy pions (see section 3.1.4). The second largest loss comes from particles hitting calorimeter 'cracks', as described in section 3.1.5. In these regions, the calorimeters do not register the full particle energy. Energy loss also occurs from curl-up of low momentum ($\leq$ 400 MeV) charged particles in the central magnetic field. Other losses come from the weak decays of strange hadrons (mostly kaons) into neutrinos and/or muons, which leave little or no calorimetry signal. These single particle effects are modeled in CDF detector simulation. The details of their measurement are presented in other chapters (see section 3.1.4 and section 3.1.5).

Cluster reconstruction reduces the amount of jet energy collected. Because each tower included in the jet is required to have a minimum E_t, energy from particles which deposit less than this threshold is lost. Energy is also lost from particles emitted at large angles

from the jet axis, whose energy is not collected within clustering circle. These losses can be modeled simply by applying the same reconstruction algorithm to Monte Carlo data as was applied to CDF data.

The impact of single particle energy response effects on the net jet response depends on the model of fragmentation used. For instance, the average number of pions in the non-linear energy range (.5 to 10 GeV) is determined by the fragmentation model. A jet that fragments mostly to pions in this range will give a smaller net energy than one composed of higher energy pions. Similarly, a jet that fragments into many particles can better afford the loss of the energy of single particle to a crack. This underscores the importance of using a correct fragmentation model when studying jet response.

5.3 Correction Algorithm

The jet fragmentation model and the detector response simulation were combined into a Monte Carlo program. We used this Monte Carlo program to study jet response. We started by generating artificial gluon scattering, producing 2 back-to-back gluons in the final state. The gluon pairs were produced with flat spectra in both P_t and η. Flat spectra were necessary to generate a statistically adequate number of events at all P_t and η. The gluons were fragmented into hadrons using the tuned Isajet model described in section 2.6. No underlying event was generated with the gluons. Particle energy response was modeled by the CDF detector simulation. The simulated energy data was reconstructed and clustered with the standard algorithms described in chapter 4.

The resulting clusters were associated back to their gluon ancestors by requiring their separation in $\eta\phi$ space to be less then .6 units. Gluons with no matching cluster were ignored. If a gluon matched more than 1 cluster, only the higher P_t cluster was associated. Un-associated clusters and partons are the subject of section 5.5.

Jet response was measured in P_t, being the kinematic variable most relevant to the double parton analysis. Jet P_t is proportional to jet energy, so all the above statements about

jet energy loss apply equally to jet P_t loss. The P_t of the matched cluster was compared to the sum P_t of all *final particles* which derived from the matched gluon. The final particle P_t is used to compare against since that is what a perfect detector and reconstruction algorithm would measure. We *did not* compare to the original gluon P_t, as that would presume our model of jet mass generation to be correct. The comparison between cluster and final particle P_t was made as a function of cluster P_t and η. P_t dependence was necessary to measure the changing contribution of the non-linear pion response to the jet response. η dependence was needed to account for the effects of the η cracks.

For each matched cluster-gluon pair, the ratio of final particle P_t to cluster P_t was computed. This ratio was computed for many events, and histogramed as a function of cluster P_t and η. Each histogram of the final particle P_t to cluster P_t ratio was fit to a gaussian. The mean of the gaussian fit was defined as the *jet response factor* for that cluster P_t and η. A gaussian fit was used instead of a simple mean to suppress the effect of jets with catastrophic energy loss. These were rare, high energy jets which lost over half their energy to cracks. The presence of these in the jet response measurement sample was an artifact of the flat P_t generation spectrum. In real $P\bar{P}$ data, the low rate of high P_t jets naturally suppresses such events.

The ratio of final particle to cluster P_t was fit for each bin of cluster P_t and η. The mean of these gaussian fits was then smoothed and interpolated between the values of cluster P_t and η to yield a continuous function of cluster P_t and η. The value of this function at a given value of cluster P_t and η is thus the factor by which to multiply an observed cluster P_t to get the most likely final particle P_t. Figure 5.3 plots the response function as a function of η for different values of cluster P_t. The effects of η cracks and the non-linear pion response can be clearly seen.

The jet response function was combined with results of the underlying event analysis described in section 5.1 to give a total cluster correction function. To apply the cluster correction function to $P\bar{P}$ data, the cluster P_t was first adjusted by subtracting the average

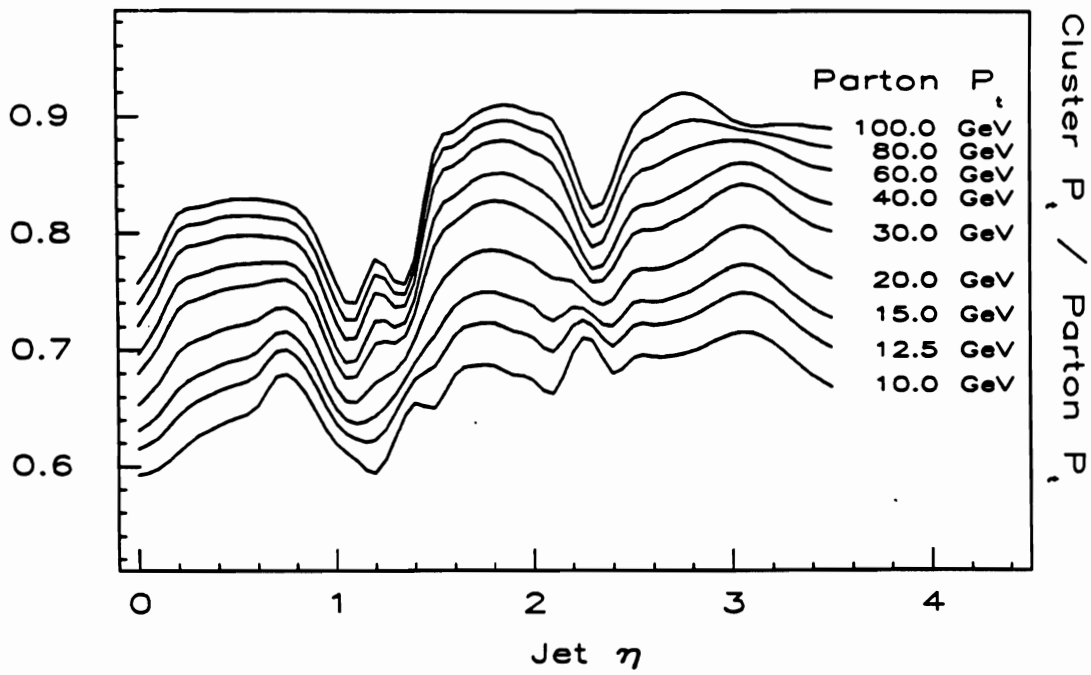


Figure 5.3: *The result of the Monte Carlo study of jet response is shown. The average value of the ratio of observed cluster to final particle P_t is plotted against the cluster η for clusters of different P_t.*

underlying event contribution (at the appropriate value of cluster η). The result was then multiplied by the value of the response function (at the appropriate cluster P_t and η). This procedure corrected the cluster P_ts so that their *average* values would agree with what a perfect jet detector would measure. The corrected P_t of a cluster is then the *most likely* post-fragmentation P_t for the particles which generated the calorimeter signals.

Corrected clusters are still not directly comparable to theoretical partons, simply because corrected clusters still have mass. In the end, a comparison between clusters and partons must involve a direct modeling of fragmentation. The role of the cluster correction function is to remove as many detector and reconstruction effects from clusters as possible. This isolates the effects of fragmentation modeling to simply generating mass for the jets. The mass generation model used in this analysis is based on $P\bar{P}$ data, and is described in appendix E. Comparisons between clusters and partons are made and described in chapter 6.

5.4 Cluster Correction Testing

The correction function described above is based on a long chain of measurements and assumptions. To insure the correctness of the final result we tested the correction function using $P\bar{P}$ data. Our test applied the principle of momentum conservation to dijet events. Because the pre-collision parton system has nearly zero $\vec{P}_t$, we expect the final jet system to also have zero $\vec{P}_t$. Jet resolution and higher order QCD effects will vary the P_t of an individual event, but the P_t averaged over many events should be zero. We apply this principle by testing whether the average $\vec{P}_t$ in dijet events is zero. This allows us to test the *relative* jet P_t response. It does not test the *absolute* jet response energy scale. The uncertainty in the absolute jet energy scale is included as a systematic error in the double parton result.

The P_t balancing study used an inclusive dijet event sample (at least 2 clusters with $P_t >$ 15 GeV). One jet was required to be well contained in the central calorimetry ($-.8 < \eta < .8$) while the other jet (*probe* jet) was allowed anywhere in the detector ($-3.5 < \eta < 3.5$). These events formed a laboratory for comparing the energy scale of all parts of the detector with the central detector. The central detector was chosen as standard first as the best understood of the CDF calorimeters, and second as the only calorimeter fully participating in the ΣE_t trigger.

Central ΣE_t triggering created a bias in the balancing study. To understand this, imagine a dijet event with one central jet and a probe jet outside the central region. Further suppose that the E_t of the central jet is just below the the ΣE_t trigger threshold. A resolution fluctuation that boosts the central jet energy would then trigger the event. An analogous boost to the probe would not, since only the central participated in the trigger. Thus dijet data from the central ΣE_t trigger is biased to include central jets with higher P_t than probe jets.

The central trigger bias was removed by making an η-uniform hardening cut on the sample. Analogous to the central-only hardware ΣE_t trigger, an offline cut on the ΣP_t of

the 2 jets was made without regard to the η values of the jets. The value of the ΣP_t cut was determined from the hardware trigger threshold as;

$$\Sigma P_t \geq 2 \times \text{Trigger } E_t + 15\, GeV.$$

The factor of 2 guaranteed that the central jet alone would satisfy the hardware threshold. The offset of 15 GeV accounts for resolution smearing.

The vector sum of the event $\vec{P_t}$ was formed for events passing this cut. The $\vec{P_t}$ of all towers above the jet threshold (100 MeV) were added vectorially, assuming the energy in each tower was deposited by a massless particle striking the tower center. The sum was taken over the entire event as opposed to jets to minimize the sensitivity to clustering effects. The resultant vector was projected along the ϕ direction of the probe jet. The magnitude of this projection was interpreted as the P_t imbalance caused by the probe jet. This imbalance was divided by the average P_t of the 2 jets, giving a *fractional* imbalance. Averaged over many events, this fractional imbalance gives the average jet energy scale difference between the probe region and the central calorimetry.

Figure 5.4 plots the average fractional imbalance as a function of probe η. The data are shown without jet response map corrections. For events with both jets in the central calorimetry one is randomly called the probe. The sign of the data are chosen to show energy loss in the probe region relative to the central as positive. Regions of low probe response relative to the central are seen near η's of 0, 1, and 2. These correspond to the crack regions of figures 3.2 and 5.3. The plug and forward regions are seen to have a higher response than the central (negative imbalance fraction) as expected due to the larger non-linearity of the central relative to the plug and forward.

Figure 5.4 also plots the average fractional imbalance as a function of probe η after application of the jet response map. We see that the variations in the P_t response, previously of order 10%, have been reduced to less than 5%.

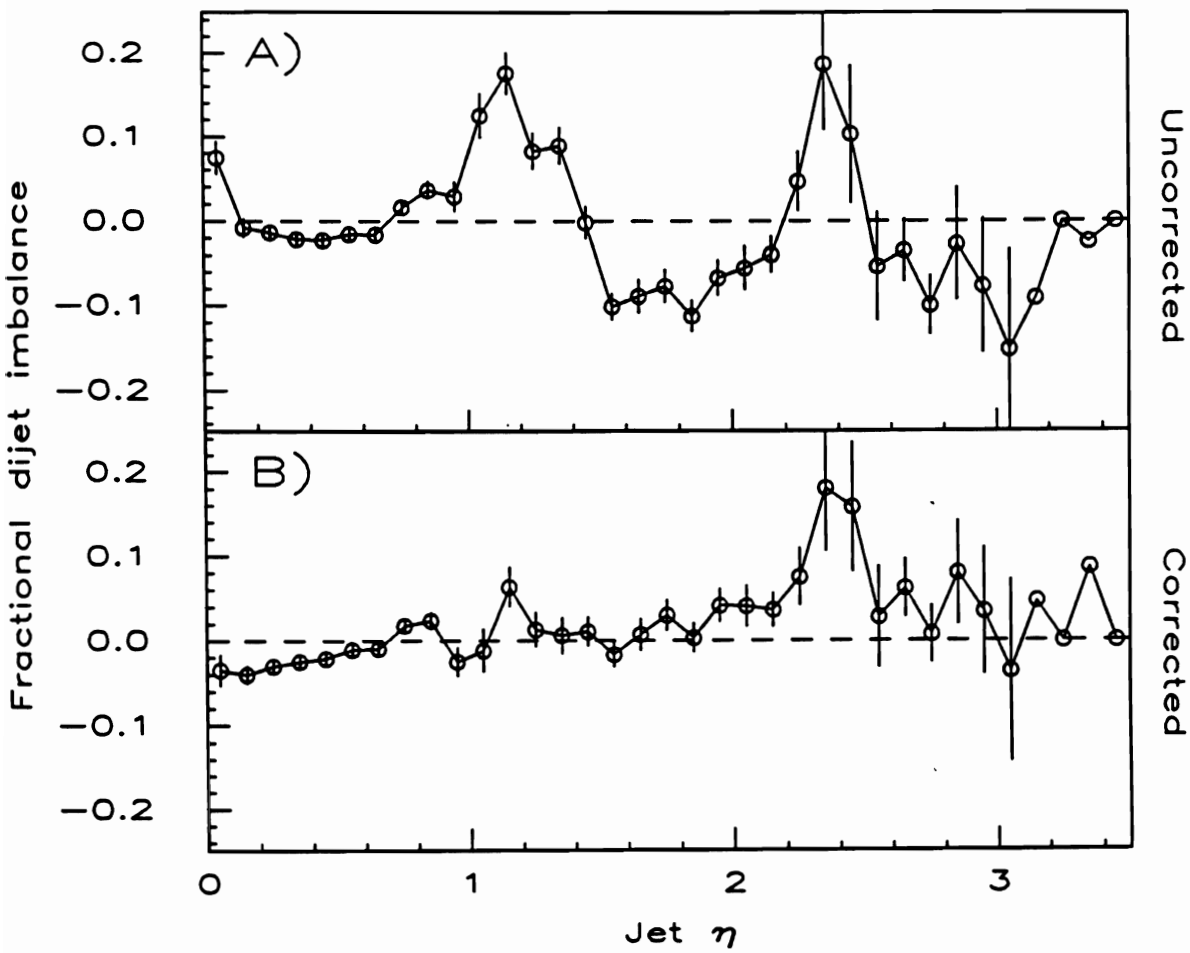


Figure 5.4: *A) The average fractional P_t imbalance in central-probe dijet events is plotted by probe η. This dimensionless quantity measures the effective energy scale difference between jets in central and probe regions. B) The same, after applying correction functions to jet P_ts*

5.5 Fake Clusters

Not every reconstructed cluster can be matched to a gluon. These *fake* clusters arise from fluctuations in underlying event or fragmentation. We expect the rate of such fakes to fall rapidly with increasing P_t, as the fluctuation scale becomes small compared to the jet energy. Fake clusters form a background to multi-jet events, artificially increasing the jet multiplicity above the parton multiplicity. Avoiding such fakes is the main function of cluster selection cuts.

We study fake clusters with the same simulation used to study jet response. Unlike the

response study, we include a model of underlying event with the parton fragmentation. The model used comes from Isajet, which views the residual $P\bar{P}$ collision as *pomeron* exchange [20]. These models are known to be imperfect [25]. Figure 5.2 plots the measured underlying event E_t density for both the data and the Monte Carlo. Both the real and simulated data were analyzed as described in section 5.1. The agreement shown is adequate for the double parton analysis.

As in the response study, events were generated as gluon dijets, fragmented by Isajet, simulated, and reconstructed. A cluster was called fake if it had no matching gluon within .6 $\eta\phi$ space units. The .6 value for the matching was chosen from a small Monte Carlo study of jet angular resolution. .6 $\eta\phi$ units corresponds to $\geq 3\sigma$ of the angular resolution for 15 GeV P_t partons.

Figure 5.5 plots the fraction of fake clusters against their *corrected* P_t. Even though fakes have no parent parton, they are experimentally indistinguishable from real clusters. Their P_ts are therefore corrected with the correction function to compare their rate with the parton rate. The plot indicates that 10% of the clusters at a P_t of 15 GeV are fakes. The plot sums over all η. Variations in fake rate with η were found to be negligible.

This 10% *fake rate* results in a background to 4-jet events. For instance, dijet events with two fake clusters, or three jet events with a single fake cluster, make fake 4-jet events. This background can be estimated from the absolute jet rates. Roughly, the two:three:four jet event rates go as 10:3:1. This implies that roughly 30% of the 4-jet events could be background. This fraction of fake 4-jet events isn't substantially reduced by increasing single jet P_t cut, as, at high P_t, the real 4-jet rate falls as fast as the fake rate. 30% is an over-estimation of the fake 4-jet event background, as the Monte Carlo fragmentation model includes the effects of parton level gluon radiation. To some extent, these radiation effects are modeled by the matrix element used in the Monte Carlo, and thus account for some of the 'fake' 4-jet events in the CDF data.

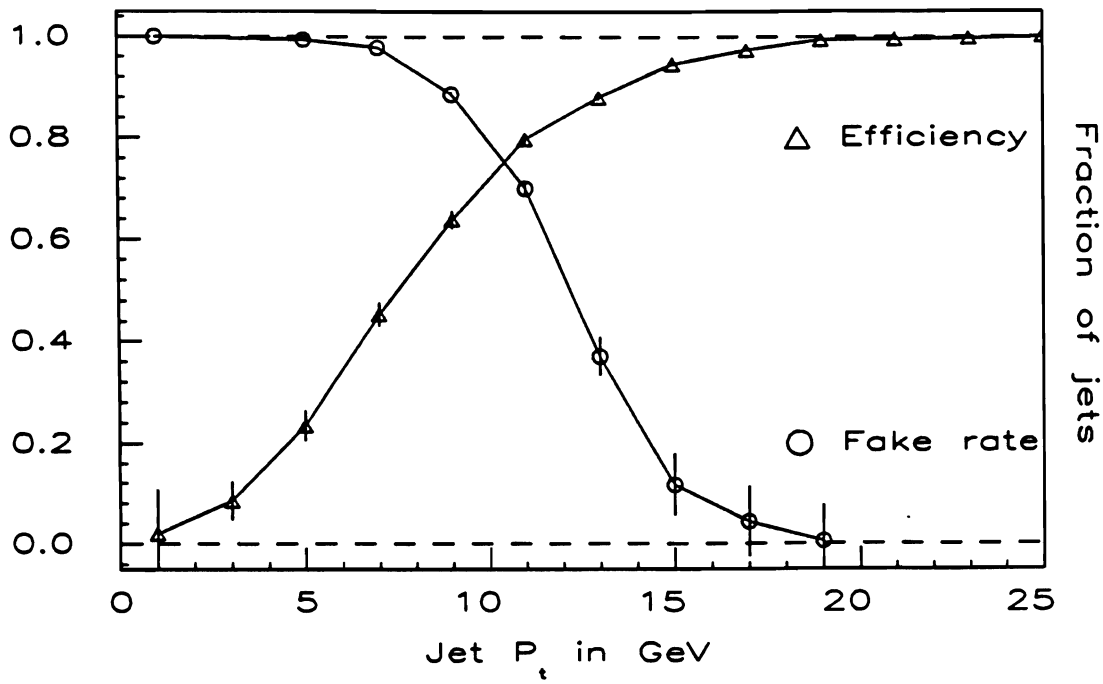


Figure 5.5: *Shown are the relative frequency for clusters with non-partonic origin (fake clusters), and the partonic cluster reconstruction efficiency.*

5.6 Reconstruction Efficiency

Not every generated gluon was reconstructed as a cluster. These lost partons represent inefficiency in the detector and/or the reconstruction algorithm. As both the detector and algorithm performance improve with increasing P_t, we expect this inefficiency to vanish at large P_t. The data for this study were produced identically to the fake cluster study. A parton was defined as *reconstructed* if a cluster was found within .6 $\eta\phi$ units.

Figure 5.5 also plots the fraction of partons reconstructed as clusters as a function of the P_t of the parton. The plot averages over all η. The plot indicates a reconstruction efficiency of > 95% for 15 GeV P_t partons. η variation of reconstruction efficiency was found to be negligibly small.

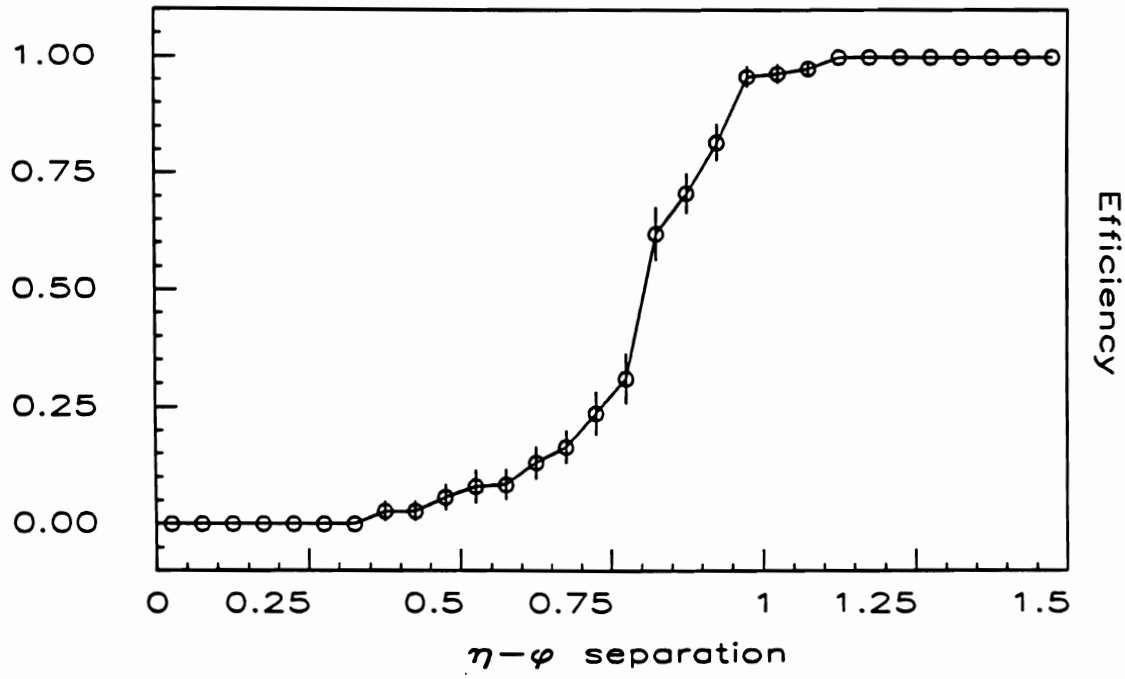


Figure 5.6: *Shown is the jet reconstruction $\eta\phi$ separation efficiency.*

5.7 Cluster Separation Efficiency

Fragmentation and reconstruction also limit the ability to resolve nearby partons. This spatial resolution has an important effect on multi-jet reconstruction. We study spatial separation efficiency using $P\bar{P}$ data. First, 2 events with a basic dijet structure are reconstructed. The raw data from one cluster in one of these events is combined with the raw data of the other entire event. This *merged* event is clustered normally, and the clusters found are associated back to those in the original events. Merged events where the added jet is separately identified are considered to be correctly separated. Separation efficiency is then measured as the fraction of such successful events as a function of the minimum $\eta\phi$ separation between the original clusters. As shown in figure 5.6, this is a sharp function with a transition at about .8 units. We obtain > 95% efficiency for reconstructing clusters separated by ≥ 1 unit.

5.8 Summary

In this chapter we described studies of CDF detector response to jets. Detailed models of jet fragmentation and the CDF detector indicate that the measured jet P_t is lower than the actual, final particle P_t. We used the Monte Carlo model to derive a correction algorithm for this P_t loss. The algorithm was tested with $P\bar{P}$ data and found to adequately describe P_t response as a function of η. We also used the Monte Carlo model to study reconstruction efficiency. We found that requiring a minimum P_t of 15 GeV per jet gives good reconstruction efficiency and background rejection.

Chapter 6

4-Jet Analysis

In this chapter the CDF 4-jet data sample is defined. The generation and simulation of a double brem and double parton Monte Carlo data sample is also described. Variables sensitive to the presence of double parton scattering in the 4-jet data are introduced. The double parton cross section is determined by comparing the CDF data with the Monte Carlo data in these variables. The systematic errors on the cross section are estimated. The statistics of the 1987 CDF data sample are inadequate to decisively prove the existence of double parton scattering, though all the results are consistent with a cross section near the expected value. Based on the estimated errors, an upper limit on the double parton cross section is set.

6.1 The CDF 4-Jet Data Sample

In order to keep the data selection process understandable and reproducible, a minimum of cuts are used to define the 4-jet data sample. These cuts serve three basic purposes. First, the cuts select events with 4 jets. Second, cuts must be made to remove any bias coming from the online trigger used to select the events. Lastly, some cuts are made to insure event quality.

4-jet events are first selected as having *exactly* 4 clusters whose corrected P_t is above 15 GeV. Events with 5 or more clusters above 15 GeV are rejected. This makes the data set comparable to leading order $2 \rightarrow 4$ calculations. Applying the cluster correction insures

that the P_t cut is applied uniformly over the detector. Below 15 GeV P_t The association between clusters and partons is ambiguous (see section 5.5). The threshold is set as low as possible so as to include as many events as possible.

Recall from section 3.3 that the online calorimetry trigger for CDF required a minimum E_t in the central calorimetry. This threshold was set to values between 20 and 45 GeV. In order to simplify the simulation of equivalent data sets, this cut offline was hardened so that all accepted events would have triggered at the highest threshold. The selection cut was made on jet P_t to avoid having to simulate the trigger. At least two jets within the central calorimeter (centroid within $-.9 < \eta < .9$) were required. The sum of the corrected P_t of these central jets was required to be at least 65 GeV. Figure 6.1 plots the central jet P_t sum against the online trigger E_t sum. The plot uses CDF 4-jet events triggered at 30 GeV E_t. The 65 GeV cut is 97% efficient for selecting events with trigger E_t above the highest trigger threshold of 45 GeV. The spread between trigger E_t and jet P_t comes from a variety of sources, including jet reconstruction and offline correction of tower E_t by the Z vertex position (see section 3.2).

All jets are required to be separated in $\eta\phi$ space from their nearest neighbors by at least 1 unit. This cut removes events of uncertain angular resolution efficiency. From section 5.7 we know that jets separated by 1 unit are correctly reconstructed with over 95% efficiency. All jets are required to be within a range $-3.5 < \eta < 3.5$. Clusters with centroid within $|\eta| < 3.5$ are guaranteed to have the entire .7 radius clustering cone within the $|\eta| < 4.2$ limit of the calorimeters.

An event selection cut is made on the Z position of the collision point as measured by the VTPC. Events with a vertex more than 60 cm away from the nominal point are rejected. As discussed in section 3.2, shifted Z vertices distort the projective tower geometry of the calorimeters. The 60 cm cut insures that this distortion is kept within correctable limits.

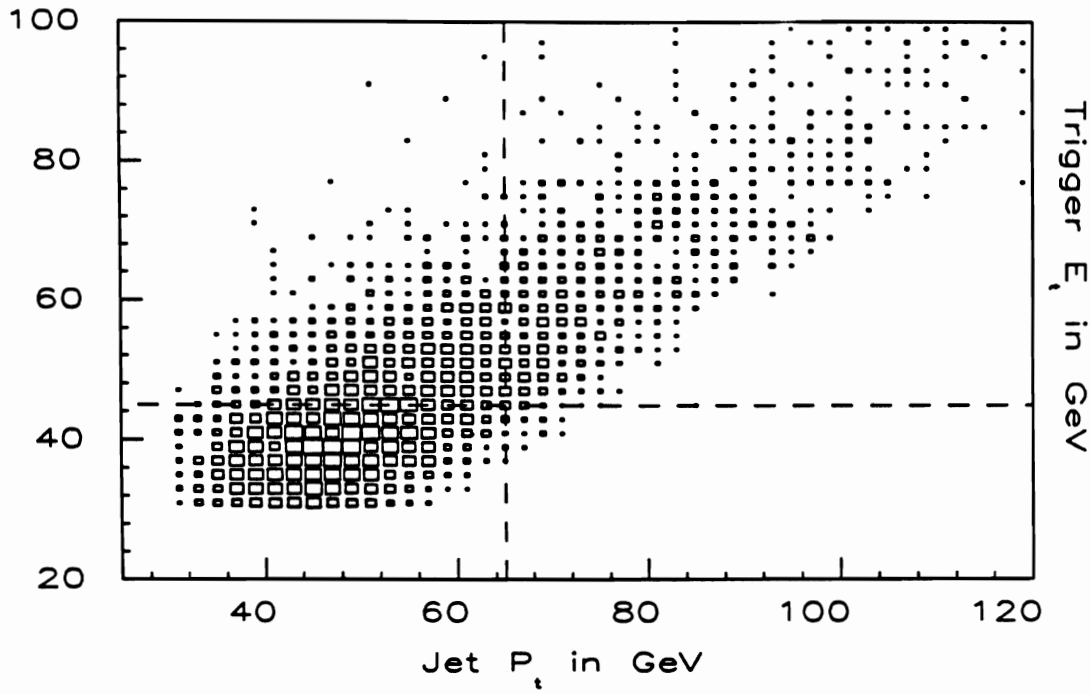


Figure 6.1: *Trigger E_t is plotted against the central jet P_t sum for CDF 4-jet events. The absence of events below 30 GeV E_t comes from the online cut. The offline cut at 65 GeV P_t and the online cut at 45 GeV E_t are shown.*

6.2 The 4-Jet Monte Carlo Data Sample

The 4-jet Monte Carlo data sample was generated using the Papageno Monte Carlo program. As described in appendix A, this program produces parton level events according to first order QCD matrix elements. Double brem events are produced directly using an approximate matrix element. Double parton events are produced by combining two independent dijet events. The procedures used to generate 4-jet Monte Carlo events are discussed in section 2.4.

As discussed in appendix A, the statistical errors from Papageno do not go simply as $\sqrt{N}$. Instead, the errors for 4-jet production are approximately 1000 times larger. Given the approximately 1000 events in the CDF 4-jet data sample, this implies that about 1 million Monte Carlo 4-jet events are needed to match the statistical accuracy of the CDF data. Simulating this many events with the detailed fragmentation and detector simulation used in section 5.3 would require about one thousand hours of VAX 8650 CPU time. Instead,

a fast *parton level* simulation was used. This required about one hundred CPU hours to simulate the same number of events.

The parton level jet simulation is philosophically different from the detailed, particle level simulation described in section 2.6 and 3.1. Instead of modeling the intermediate processes of fragmentation, detector response, and reconstruction, the parton level simulation models the entire *transformation* of partons into clusters in a single step. Each generated parton is separately transformed into a single cluster. This process gives mass to the jets, and smears and reduces their energy according to measured detector responses. The transformation is tuned to reproduce the jet response measured in CDF dijet events. The details of the mechanisms and tuning of the parton level jet simulation are presented in appendix E.

It is legitimate to ask if a parton level jet simulation is realistic enough to compare with real detector data. The answer depends on the measurement being made. The key issue is whether the data can be interpreted in terms of parton dynamics, without specifically accounting for fragmentation or other soft QCD effects. For instance, if one is studying jet fragmentation, or the emergence of true jets from the low E_t cluster background, a particle level simulation is essential. Such a simulation was used for the fragmentation and fake cluster analyses presented in sections 2.6 and 5.5. If, however, one is studying high P_t jets, where the effects of non-perturbative QCD are indirect, a parton level simulation suffices. Soft QCD processes are important for determining jet P_t *resolution*, but they do not affect the identification of jets, or their average P_t. Sections 5.5 and 6.1 prove that the CDF 4-jet sample is chosen so that the jets are directly related to parton level processes. Consequently, the parton level simulation is adequate for the 4-jet analysis.

Once generated and simulated, the P_ts of the Monte Carlo jets were corrected with the same response function used to correct the P_ts of CDF jets. The only difference between the function used on the Monte Carlo data and the function used on CDF data was the absence of underlying event subtraction for the Monte Carlo algorithm. This is because no

underlying event was generated for the the simulated data, so none needed to be removed.

The Monte Carlo 4-jet events were required to satisfy the same selection cuts applied to the CDF data sample. In summary, these consisted of requiring a minimum corrected P_t of 15 GeV for each jet, requiring an $\eta\phi$ separation between jets of at least 1 unit, and requiring at least 2 jets in the central calorimeter with a minimum ΣP_t of 65 GeV. The vertex selection cut was applied to a Z vertex position generated according to the distribution observed in the CDF 4-jet sample. This cut reduced the number of events by 10%, but had no other selection effect.

6.3 Double Parton Variables

As mentioned in section 2.5, two basic topological features can be used to discriminate double parton events from double brem events. The most powerful of these exploits the double dijet nature of double parton events, requiring that pairs of jets independently balance in P_t. The second exploits the preferentially small angle between jets emitted through bremsstrahlung. For first-order parton level models, these effects distinguish double parton from double brem absolutely. Experimentally, finite resolution and higher order QCD effects limit their resolving power. In this section variables to extract information about pairwise P_t balancing and angular correlations from 4-jet events are developed. These explicitly include detector effects, but are conceptually equivalent to the parton level principles.

6.3.1 Pairwise Significance

To develop a useful variable for detecting P_t balanced pairs of jets within a 4-jet event, we start with dijet events themselves. We define the *significance* ($\mathcal{S}$) of a dijet event as the magnitude of the sum of the $\vec{P_t}$s of the two jets divided by the square root of their scalar sum or,

$$\mathcal{S}_{dijet} \equiv \frac{|\vec{P_{t_1}} + \vec{P_{t_2}}|}{\sqrt{P_{t_1} + P_{t_2}}}. \tag{6.1}$$

Significance has units of $\sqrt{GeV}$. This equation is inspired by the fact that calorimeter resolution generally goes as $\frac{\sigma_E}{\sqrt{E}} \sim$ constant. Since jet P_t is determined from calorimeter energy, we expect jet P_t resolution to behave similarly. Based on dijet studies (see appendix E), we believe the $\vec{P_t}$ imbalance in jets is dominated by detector resolution effects. Therefore, $|\vec{P_{t_1}} + \vec{P_{t_2}}|$ is determined by the single jet P_t resolution (σ_{P_t}). Because the P_t resolutions of the two jets are uncorrelated, this actually measures $\sqrt{2}$ times the single jet σ_{P_t}. Similarly, $\sqrt{P_{t_1} + P_{t_2}}$ is $\sqrt{2}$ times the average single jet P_t. Thus, $\mathcal{S}_{dijet}$ measures the single jet P_t resolution constant, analogous to the calorimeter energy resolution constant. The distribution of $\mathcal{S}_{dijet}$ should therefore be independent of the P_t of the jets.

Figure 6.2 plots the significance for CDF central dijet events (2 jets above 15 GeV P_t, both in the central calorimeter). The peak near 1 translates into a single jet P_t resolution constant of $100\%/\sqrt{P_t}$. The tail at large values of $\mathcal{S}_{dijet}$ comes from poorly measured dijet events and residual three jet events. This curve defines the experimental accuracy obtainable for measuring dijet balancing. The overlaid curve comes from a parton level simulation of Papageno dijet events. The agreement between these demonstrates that the Monte Carlo is adequate for simulating dijet balancing.

The idea of dijet significance can be extended to 4-jet events in two ways. Trivially, we can add the appropriate terms to equation 6.1 and define a *4-jet significance*,

$$\mathcal{S}_{4jet} \equiv \frac{|\vec{P_{t_1}} + \vec{P_{t_2}} + \vec{P_{t_3}} + \vec{P_{t_4}}|}{\sqrt{P_{t_1} + P_{t_2} + P_{t_3} + P_{t_4}}}. \tag{6.2}$$

As with the dijet significance, the 4-jet significance measures the single jet P_t resolution constant. Figure 6.3 plots the 4-jet significance for the CDF 4-jet sample, as defined in section 6.1. The distribution is nearly identical to the dijet significance of figure 6.2, showing that jet P_t resolution of 4-jet events is the same as in dijet events. The overlaid curve comes from a parton level simulation of Papageno 4-jet events. The agreement between these demonstrates that the Monte Carlo is adequate for simulating $\vec{P_t}$ resolution effects in 4-jet events.

To look for double dijets in 4-jet events, we combine the ideas behind equations 6.1 and

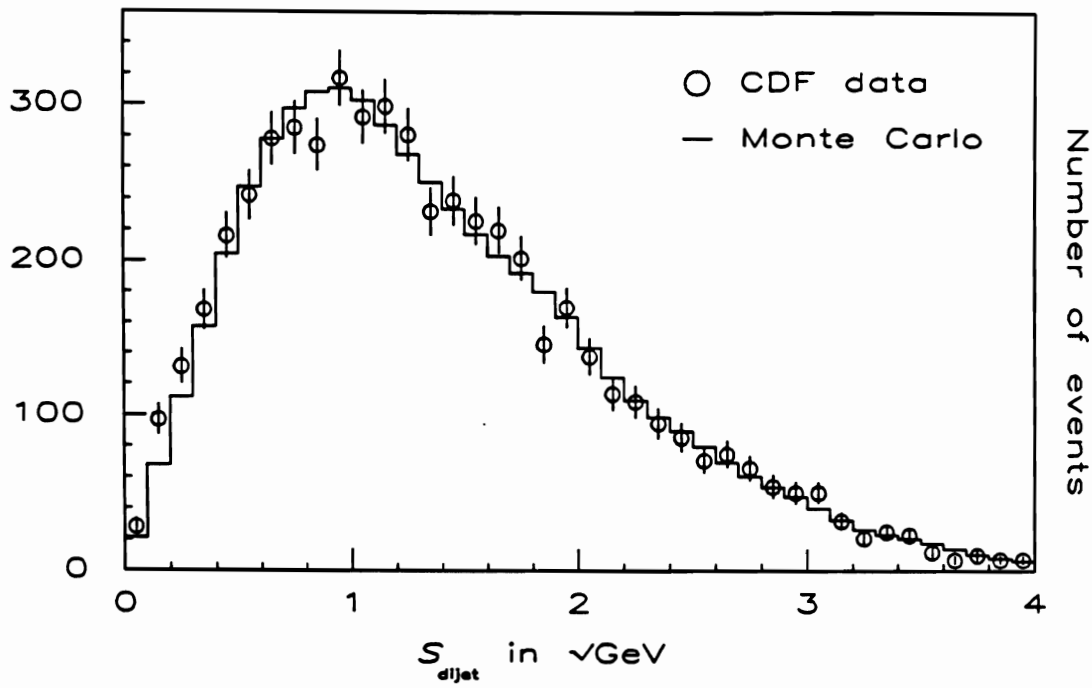


Figure 6.2: *The jet P_t significance is plotted for CDF dijet events. Events were selected from the ΣE_t ¿ 30 GeV trigger sample. The events were required to have exactly 2 clusters above 15 GeV P_t, both in the range $-.9 < \eta < .9$. The Monte Carlo event sample was generated by Papageno using the lowest order QCD $2 \rightarrow 2$ matrix element, and simulated with the parton level simulation.*

6.2. First, we divide the 4 jets into two groups of two jets (*ie* jets 1+2 and jets 3+4). For this *pairing* we define the *pairwise significance* as;

$$\mathcal{S}_{pair}(1+2,3+4) \equiv \sqrt{\left[\left(\frac{|\vec{P_{t_1}}+\vec{P_{t_2}}|}{\sqrt{P_{t_1}+P_{t_2}}}\right)^2 + \left(\frac{|\vec{P_{t_3}}+\vec{P_{t_4}}|}{\sqrt{P_{t_3}+P_{t_4}}}\right)^2\right]/2} \tag{6.3}$$

This is then computed for each of the three possible pairings of 4 jets into two sets of 2 jets (*ie* (1+2,3+4), (1+3,2+4) or (1+4,2+3)). We then define a single pairwise significance for each 4-jet event as the *minimum value* of the three specific significances, or;

$$\mathcal{S}_{pair} \equiv \min[\mathcal{S}_{pair}(1+2,3+4), \mathcal{S}_{pair}(1+3,2+4), \mathcal{S}_{pair}(1+4,2+3)]$$

This implicitly defines the *minimum pairing*, being that pairing which gives the minimum pairwise significance. As with the 4-jet significance, pairwise significance has units of $\sqrt{GeV}$.

Comparing equation 6.1 to equation 6.3 we see that $\mathcal{S}_{pair}$ is simply the quadratic average of two dijet significances. For double parton events, where the P_t resolution of the pairs is

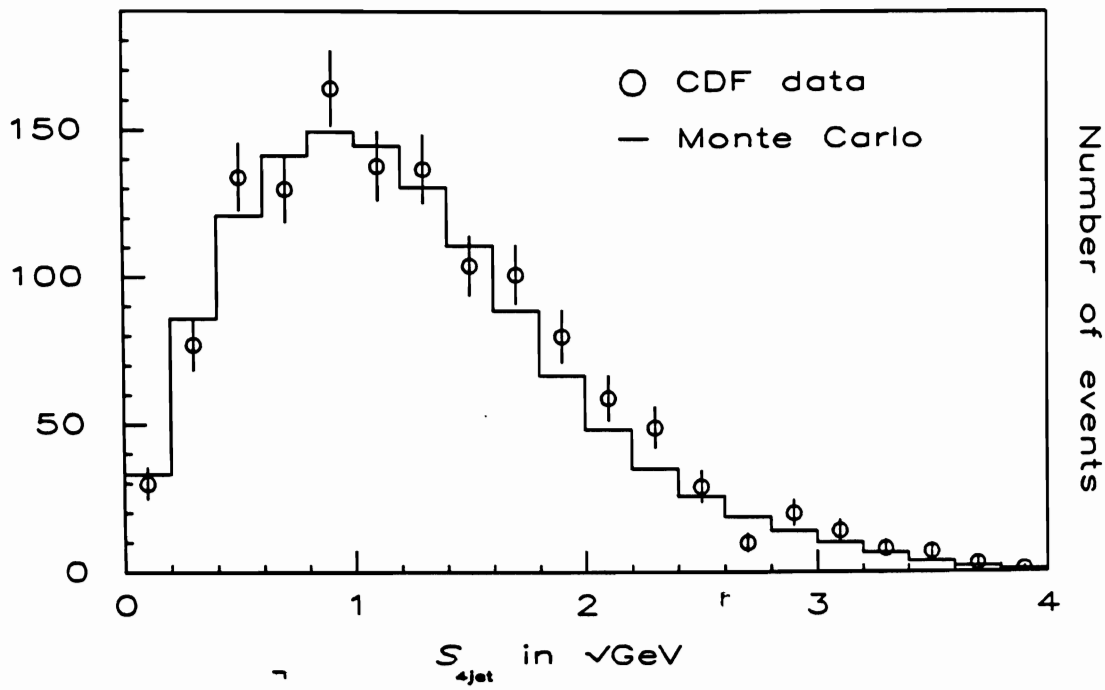


Figure 6.3: *The jet P_t significance is plotted for CDF 4-jet events. The Monte Carlo event sample was generated by Papageno as a combination of double parton ($\sigma_{eff} = 17mb$) and double brem events, simulated with the parton level simulation.*

uncorrelated, S_{pair} therefore has a distribution identical to that of single dijets, peaking at small values. For double brem events, where the jets do not naturally balance in pairs, it will typically have a large value. Pairwise significance is therefore sensitive to double dijet events within the limitations of jet P_t resolution.

To demonstrate S_{pair}, figures 6.4 and 6.5 present two high P_t 4-jet events from the CDF data sample. Both events are well measured, as shown by the small value of their S_{4jet}. However, the topologies of the events are very different. In event A, the jets fall naturally into balanced pairs. Jets 1 and 2 are nearly opposite in ϕ, and have similar values of P_t, as are jets 3 and 4. Mentally removing jets 1 and 2 (or 3 and 4) from the event leaves a perfectly reasonable looking dijet event. Because the jet pairs are opposite in ϕ and balance in P_t, their $\vec{P}_t$ sum is small. Consequently, event A has a small value (.8) of S_{pair}, completely compatible with the distribution of dijet significance shown in figure 6.2. Thus event A is a likely candidate for double parton scattering.

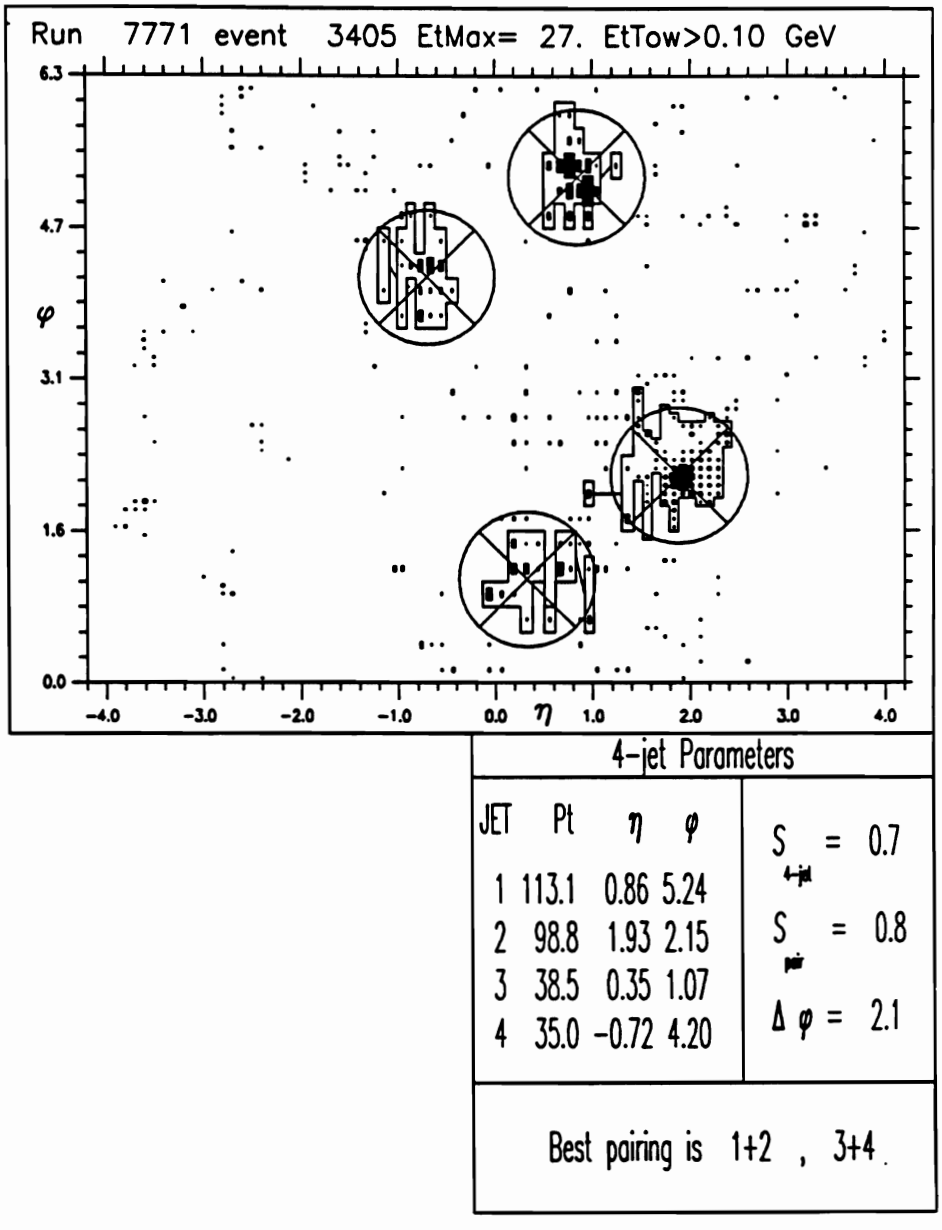
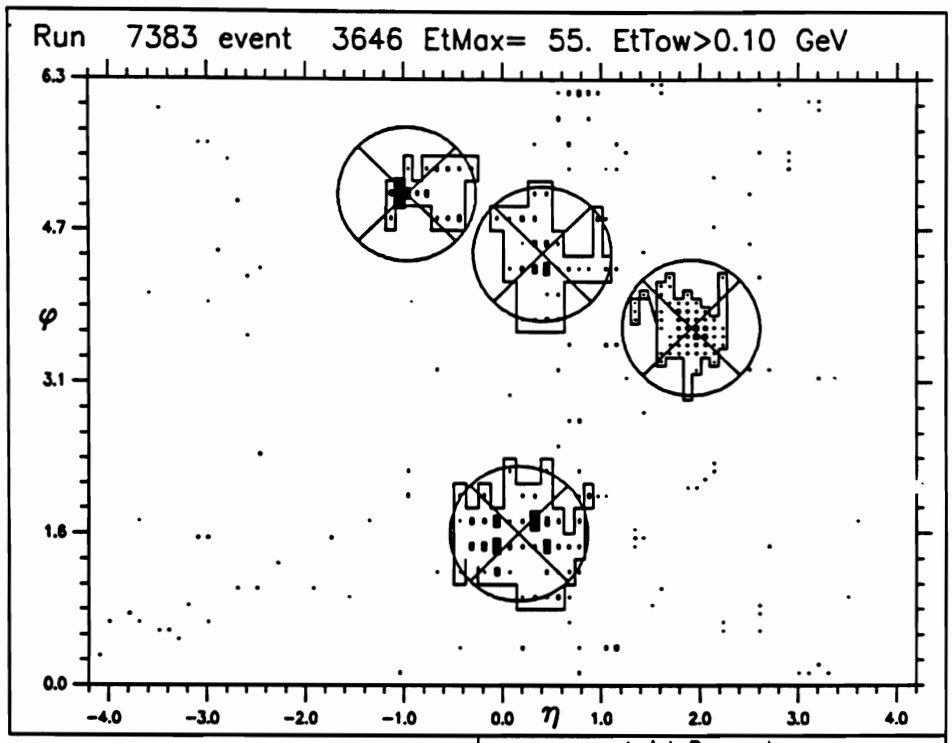


Figure 6.4: *A 4-jet event from the CDF data (event A).*



4-jet Parameters

JET	Pt	η	φ
1	130.3	0.18	1.58
2	98.7	-0.98	5.10
3	41.1	0.43	4.48
4	33.5	1.94	3.71

$S_{4\text{-jet}} = 1.1$

$S_{pair} = 6.2$

$\Delta\varphi = 2.9$

Best pairing is 1+2 , 3+4

Figure 6.5: *A 4-jet event from the CDF data (event B).*

In contrast, event B does not naturally fall into pairs. The jets balance in $\vec{P_t}$ more in the pattern of three versus one instead of two pairs of 2. Applying equation 6.3 to this event gives a large value of $\mathcal{S}_{pair}$ (6.2). This value is inconsistent with the dijet distribution shown in figure 6.2. Event B is therefore probably a double brem event. Thus $\mathcal{S}_{pair}$ gives a numerical measure of how well a given 4-jet event balances into pairs.

6.3.2 Pairwise ϕ Difference

As discussed in section 2.5, the inter-jet angles of double brem events are expected to be small, simply because bremsstrahlung favors small angles. In contrast, no inter-jet angle correlation is expected of double parton events. Thus the processes can be statistically distinguished using their inter-jet angular distributions.

In this analysis, inter-jet angles are measured only in the transverse plane (*ie* jet ϕ angles). This is because the polar angle provides no additional leverage for separating double parton events from double brem. In double parton events, the pairs of jets are produced independently, and consequently the ϕ angle between the pairs of jets has a flat distribution. This simple spectrum is only slightly affected by event selection cuts, and can be easily distinguished from the double brem spectrum. By contrast, the polar angle distribution of both double parton and double brem jets has a non-trivial distribution [6]. Further, this spectrum is highly distorted by the 4-jet event selection cuts, particularly the central jet ΣP_t requirement. These distortions make it impractical to use polar angle distributions to distinguish double parton from double brem events.

The strength the inter-jet angle 4-jet separation method is its independence from the pairwise P_t balancing method. Care must therefore be taken to define an inter-jet angular measure that is independent of the definition of $\mathcal{S}_{pair}$. Some angular information is used to compute the pairwise significance of an event. Specifically, the $|\vec{P_{t_1}} + \vec{P_{t_2}}|$ terms depend on the angle between jets within a given pair. To be independent of $\mathcal{S}_{pair}$, the inter-jet angle measure can therefore depend only on the angles between *pairs* of jets. We use the minimum pairing defined in section 6.3.1 to define jet pairs, as this is the most likely pairing

for double parton events. The minimum pairing is well defined for double brem events even though it has no physical interpretation.

The above arguments give basic restrictions to the definition of an inter-jet angular correlation measure. To be useful in the double parton search, this measure must depend only on the ϕ angle between jet pairs. Angular correlations between jets are then measured as the magnitude of the difference between the ϕs of the two jet pairs. We therefore define the *pairwise ϕ difference* as;

$$\Delta\phi = |\phi_{pair_1} - \phi_{pair_2}| \tag{6.4}$$

For double brem, the $\Delta\phi$ distribution will peak near 0 or π, depending on the definition of ϕ_{pair} (see figure 2.4). For double brem, the distribution of $\Delta\phi$ will be flat.

Some latitude exists in equation 6.4 for defining the ϕ angle for a pair of jets. In the course of this investigation, a number of definitions were tried, including averaging the ϕ angles of the jets ($\phi_{pair} = (\phi_1 + \phi_2 - \pi)/2$), and using the ϕ angle of their vector $\vec{P_t}$ difference ($\phi_{pair} = \phi(\vec{P_1} - \vec{P_2})$). The final definition was chosen based on its ability to separate Monte Carlo double parton events from double brem, and for its simplicity. We define the ϕ angle of a jet pair as the ϕ angle of the highest P_t jet of the two;

$$\phi_{pair} = \phi_{max\ P_t} \tag{6.5}$$

For double parton events, the choice of the highest P_t jet is random, and thus the $\Delta\phi$ spectrum is still flat between 0 and π. For double brem events, the choice of the highest P_t jet is biased, as momentum conservation will tend to put the highest P_t jets of each pair opposite each other in ϕ. Thus the $\Delta\phi$ spectrum of double brem events will be highly asymmetric, peaking near π, corresponding to small inter-jet angles. The scarcity of double brem events at low $\Delta\phi$ angles provides a clean place to look for double parton events.

6.4 4-jet Data Comparison

Having defined all the relevant variables, we now compare the CDF 4-jet data with Monte Carlo data. Previous sections have demonstrated the agreement of the detector simulation

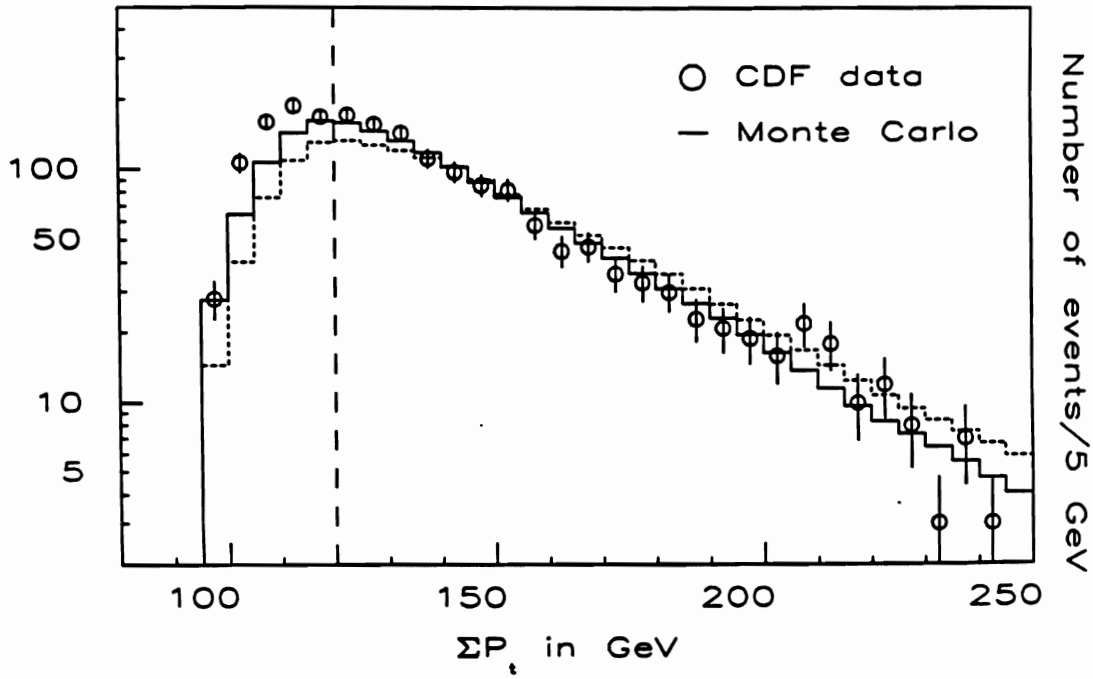


Figure 6.6: *The sum of the P_t of all 4 jets in the CDF 4-jet sample is plotted. The overlaid Monte Carlo curves include both double parton and double brem 4-jet events. The spread between the two histograms shows the systematic errors on the shape of the distribution. Monte Carlo rate was normalized to the CDF data in the region* $120 < \Sigma P_t < 250$ *GeV.*

with CDF data, so we now concentrate on comparing 4-jet production dynamics with QCD. The qualitative comparisons made in this section are followed by quantitative comparisons in the next.

Figure 6.6 shows the ΣP_t spectrum for 4-jet events selected as described in section 6.1. ΣP_t is defined as the sum of the P_ts of all 4 jets in the event. This variable is important as it defines the range of topological freedom available for separating double parton events from double brem, as described below. The overlaid Monte Carlo curves result from two representative choices of structure functions (EHLQ1 and DO2, see section 6.5.3). Monte Carlo double parton and double brem events have been added together to produce the curves, using $\sigma_{eff} = 17$ mb for double parton production. Both Monte Carlo curves were independently normalized to the number of events above 120 GeV in the CDF data, so as to compare the shapes of the distributions.

The cutoff in the ΣP_t spectrum at 95 GeV comes from the combination of the trigger

hardening cut (which requires the central jet $\Sigma P_t \geq 65$ GeV) with the ≥ 15 GeV single jet P_t cut. The falling spectrum at high ΣP_t results from the high x falloff of the structure functions. Both Monte Carlo curves agree within errors for the shape of the high ΣP_t spectrum. However, at low ΣP_t they disagree significantly, and neither Monte Carlo curve agrees well with the data. This discrepancy may be due to inadequate modeling of very low x structure functions. This is possible, since no experimental data exists to compare with the functions used. The discrepancy may also be due to higher order QCD effects. Qualitatively similar effects have been seen in higher order calculations of other QCD processes, where the new terms result in a larger cross section near ΣP_t thresholds [26]. Since a higher order calculation of double brem is unlikely in the lifetime of this experimenter, it was decided to eliminate the low ΣPt events from the 4-jet sample. Therefore a final cut was made on both the data and Monte Carlo samples removing all events with $\Sigma P_t < 120$ GeV. This cut removes about half the events from the final CDF 4-jet sample. While the cut worsens the statistical errors on the double parton result, by eliminating regions of large theoretical uncertainty it improves the systematic error.

The ΣP_t cut does not substantially affect the double parton cross section measurement. To demonstrate this, figure 6.7 plots $\mathcal{S}_{pair}$ for 4-jet events divided below and above 120 GeV ΣP_t. Overlaying the CDF data are Monte Carlo double brem and double parton data, divided into the same ΣP_t regions. The Monte Carlo curves are normalized to the CDF data sample. As shown, the CDF data and the predicted $\mathcal{S}_{pair}$ distributions look nearly identical above and below the cutoff. This indicates a similar double parton presence in both samples. Thus the discrepancy between the CDF data and the Monte Carlo ΣP_t spectrum is not due to an enormous double parton signal. Figure 6.8 demonstrates the same conclusion using the $\Delta\phi$ distributions above and below 120 GeV ΣP_t. The deficit of double parton events at $\Delta\phi = 0$ and π is caused by the jet $\eta\phi$ separation event selection cut, described in section 6.1.

Figure 6.9 plots $\mathcal{S}_{pair}$ against $\Delta\phi$ for CDF data, and for double brem Monte Carlo data.

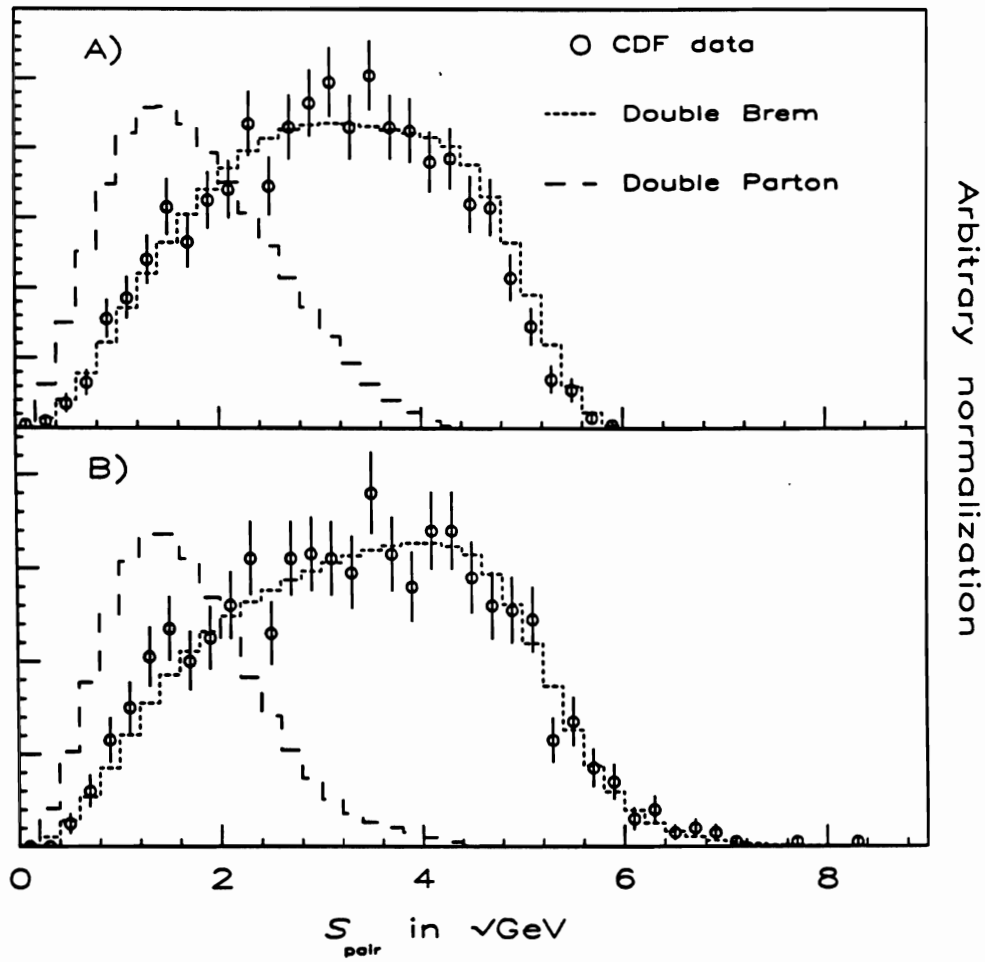


Figure 6.7: *The pairwise significance of selected CDF 4-jet events A) below 120 GeV ΣP_t and B) above 120 GeV ΣP_t are shown. The overlaid Monte Carlo curves are normalized to the CDF data.*

The Monte Carlo distribution is primarily smooth, but shows large *spikes* in regions of otherwise low density. Monte Carlo histogram spikes arise from statistical fluctuations in the event weights (see appendix A). Spikes are absent from the one dimensional Monte Carlo plots as the number of bins is much smaller. A smooth two dimensional plot would require many more Monte Carlo events than the 5 million used to create figure 6.9. Because the CDF data shows no strong correlation between S_{pair} and $\Delta\phi$, and because of the practical limitations of the Monte Carlo, those correlations are ignored, and this analysis uses only one dimensional distributions.

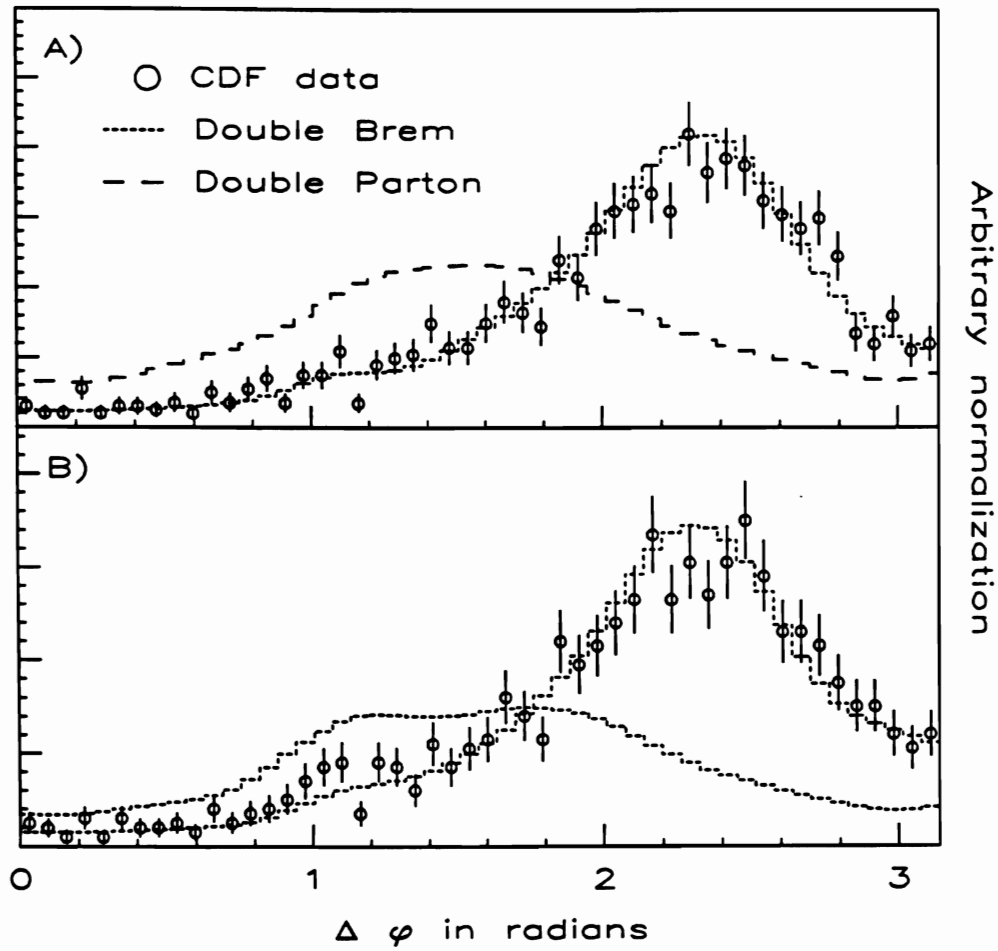


Figure 6.8: *$\Delta\phi$ of selected CDF 4-jet events A) below 120 GeV ΣP_t and B) above 120 GeV ΣP_t are shown. The overlaid Monte Carlo curves are normalized to the CDF data.*

6.5 Double Parton Cross Section Limits

In this section the CDF and Monte Carlo 4-jet data samples are used to determine the double parton cross section. This results in a measurement of the effective cross section (σ_{eff} , see equation 2.3). σ_{eff} is determined in two stages. First, the measured 4-jet $\mathcal{S}_{pair}$ and $\Delta\phi$ distributions are compared to the Monte Carlo double parton and double brem predictions. Using the shape of these, a fit is made to the *relative admixture* of double parton to double brem events in the CDF data. This relative admixture value is then interpreted in terms of σ_{eff}.

There are two reasons why the determination of σ_{eff} is divided into two steps. First,

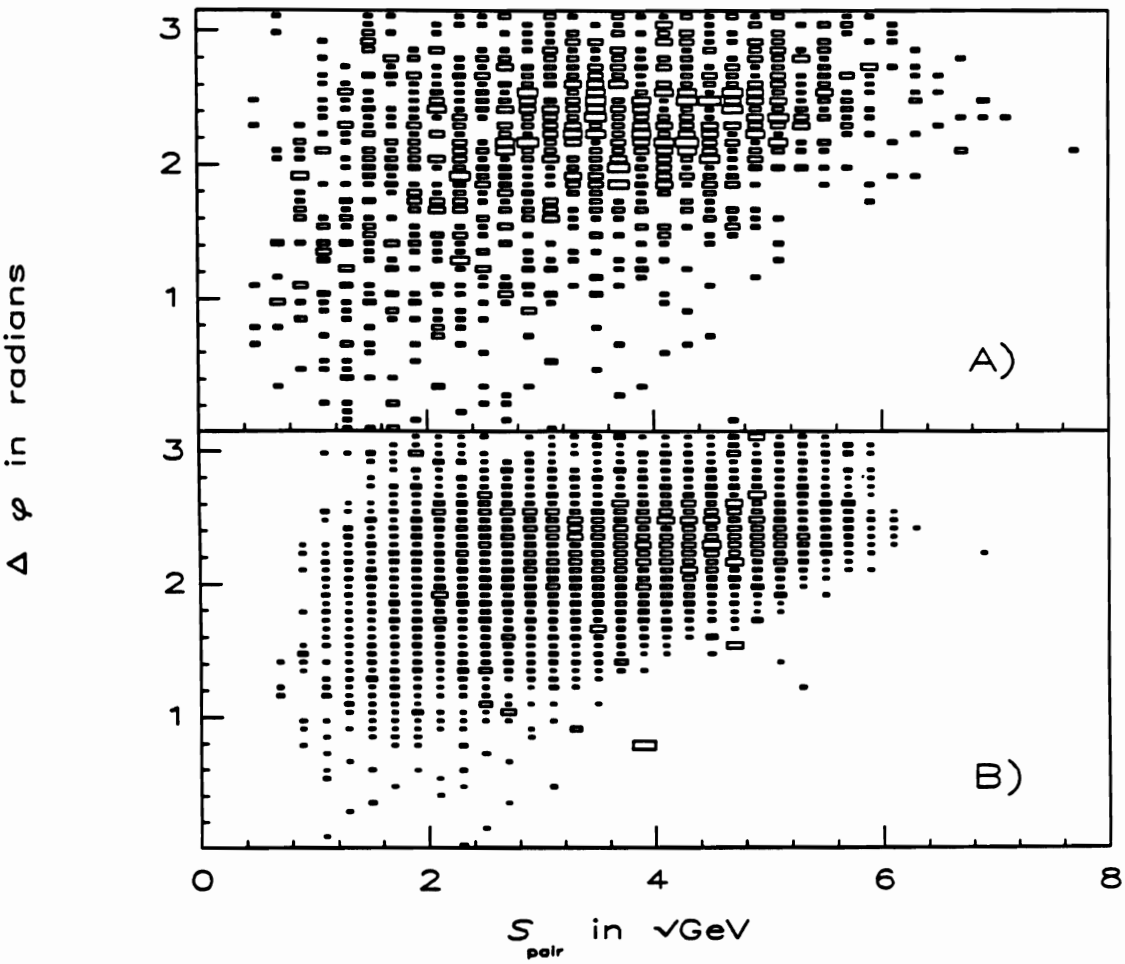


Figure 6.9: *$\Delta\phi$ versus S_{pair} for A) CDF 4-jet events and B) double brem Monte Carlo events. Only bins containing 1 or more events are shown in the Monte Carlo plot.*

the division provides a distinction between those parts of the analysis which are based on the *shapes* of distributions, and those based on *absolute rates*. This distinction is convenient because the *systematic errors* on shapes and rates are mostly independent, and arise from different sources. Thus the systematic errors can be calculated independently and later added.

Second, the separation is expedient for the calculation of statistical errors. At its root, measuring σ_{eff} involves counting the number of events believed to come from double parton scattering. The statistical errors on such a number are easily determined, and are basically symmetric. However, σ_{eff} is *inversely* related to the double parton cross section, and

therefore the number of double parton events. Thus, fitting directly for σ_{eff} would be somewhat unnatural, and would give highly asymmetric statistical errors. On the other hand, the relative double parton to double brem admixture is proportional to the double parton cross section, and hence the number of events. It is therefore a more natural variable to use in a fit. The details of how fits are made and errors are calculated are presented in the subsequent sections.

6.5.1 σ_{DP}/σ_{DB} Ratio Fitting

To measure the relative admixture of double parton to double brem events, we define a quantity $\mathcal{R}$ as;

$$\mathcal{R} \equiv \frac{\sigma_{DP} \times A_{DP}}{\sigma_{DB} \times A_{DB}} \tag{6.6}$$

A_{DP} and A_{DB} represent the acceptance effects of the 4-jet selection cuts on double parton and double brem events. In principle the acceptances can be different, but in practice the selection cuts are so general that they are nearly identical.

$\mathcal{R}$ can be fit for directly from the measured $\mathcal{S}_{pair}$ and $\Delta\phi$ distributions. To do this, separate histograms of these are made using CDF data, double brem, and double parton Monte Carlo events. A fit is then made to determine the admixture of double parton and double brem distributions which best agrees with the shape of the CDF distributions. Both the $\mathcal{S}_{pair}$ and the $\Delta\phi$ distributions are fit simultaneously to a single admixture value. The fit has two free parameters, one being $\mathcal{R}$ itself, the other being the absolute normalization of the double parton and double brem admixture. The fit value of the absolute normalization was not used. The fit minimized the value of a χ^2 computed using poisson statistics. The statistical errors on the value of $\mathcal{R}$ were determined as the change in $\mathcal{R}$ necessary to increase the χ^2 from its minimum value by 1.

Figure 6.10 plots the $\mathcal{S}_{pair}$ and $\Delta\phi$ distributions for the final CDF and Monte Carlo data samples. The overlaid Monte Carlo curves are scaled as dictated by the fit to the relative admixture of double parton and double brem events. Fits were performed for many

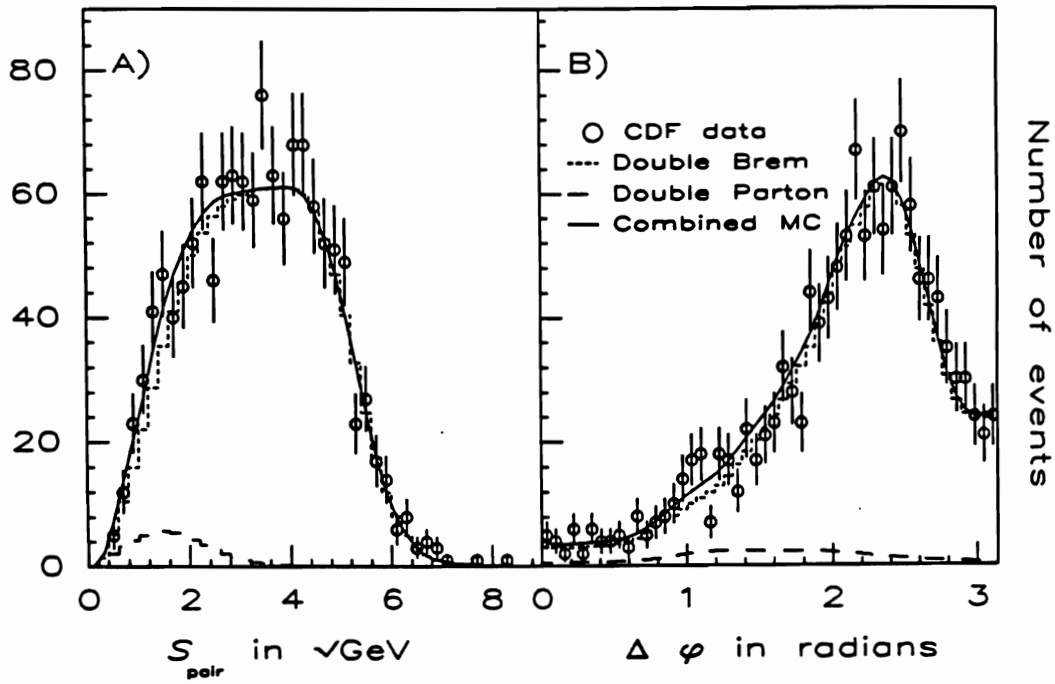


Figure 6.10: *A) the S_{pair} and B) the $\Delta\phi$ spectra for selected 4-jet events. The relative double parton and double brem admixture, and the absolute Monte Carlo rate result from a fit to the CDF data.*

different Monte Carlo data samples, as described in section 6.5.3. The average value of $\mathcal{R}$ from these is;

$$\mathcal{R} = .056 \pm .017 (\text{statistical errors only})$$

The average χ^2/DOF for the fits is .85, indicating a good match between the CDF and Monte Carlo data. The non-zero value of $\mathcal{R}$ indicates a small double parton contribution is necessary to best describe the CDF data. The statistical error on this contribution is about 30% of its value, and is thus barely significant at the 3σ level. The inclusion of systematic errors in subsequent sections will unfortunately reduce this result beyond the limit of statistical significance. Systematic errors are discussed in section 6.5.3

6.5.2 σ_{eff} Determination

By combining equations 2.3 and 6.6 we can solve for the effective proton cross section as;

$$\sigma_{eff} = \frac{\sigma^2_{dijet}}{\sigma_{DP}} = \frac{1}{\mathcal{R}} \times \frac{\sigma^2_{dijet} \cdot A_{DP}}{\sigma_{DB} \cdot A_{DB}}. \tag{6.7}$$

This determines σ_{eff} from the measured the ratio of double parton to double parton rate ($\mathcal{R}$), and the dijet and double brem cross sections and acceptances. $\mathcal{R}$ can be determined from CDF data with relatively small systematic errors. Since the CDF 4-jet data is clearly dominated by double brem events, $\sigma_{DB} \cdot A_{DB}$ can be determined from the CDF data. σ_{dijet} and A_{DP} are both determined from the Monte Carlo, while $\mathcal{R}$ is measured directly from the data. The propagation of statistical and systematic errors of these components into the determination of σ_{eff} is described in the following section.

In principle, σ_{dijet} could have been measured from CDF dijet data. In practice, none of the trigger ΣE_t thresholds used in the 1987 run were low enough to trigger on 15 GeV (corrected) P_t jets. Without data, σ_{dijet} must be determined from Monte Carlo calculations. This dependence on the Monte Carlo causes additional systematic errors on σ_{eff}, as described in the next section.

6.5.3 Systematic Errors

Important systematic uncertainties in the measurement of $\mathcal{R}$ and σ_{eff} come from theoretical assumptions made within the Monte Carlo. To study these, many different Monte Carlo data samples were generated, each with different assumptions. The variation in input parameters was chosen to cover a reasonable range of theoretical uncertainty. The $\mathcal{S}_{pair}$ and $\Delta\phi$ distributions from each set were compared to the CDF data, and used to determine a value of $\mathcal{R}$. The variation in $\mathcal{R}$ values obtained was interpreted as the systematic error on $\mathcal{R}$ due to theoretical uncertainties. The one σ systematic error on the value of $\mathcal{R}$ was estimated as the *standard deviation* of these Monte Carlo results. Systematic errors on $\sigma^2_{dijet} \cdot A_{DP}$ were estimated in the same way. This method was used to propagate uncertainties in the Q^2 scale, the structure function parametrization, and the double brem matrix element of the Monte Carlo into the $\mathcal{R}$ and σ_{eff} measurements. Monte Carlo systematic errors are listed in table 6.1, and described in detail below.

Some systematic errors arose from uncertainties in the energy scales of the CDF calorimeter reconstruction. These mostly affect the absolute rate of 4-jet production. The

Table 6.1: Errors on Double Parton Results

	$\mathcal{R}$	σ_{eff}
Statistical	30%	30%
Structure function	20%	50%
Q^2 scale	0%	50%
$\|M\|^2$ approximation	20%	0%
Relative energy scale	5%	10%
Absolute energy scale	0%	65%
Luminosity	0%	15%
$\mathcal{R}$ determination	-	30%
Total (stat $\oplus$ sys)	30% $\oplus$ 30%	30% $\oplus$ 100%

1 σ systematic error estimates on $\mathcal{R}$ and σ_{eff} are listed for the various contributing sources. The values are rounded to the nearest 5%

uncertainty in the energy scale of the gas calorimeters relative to the central calorimeters can in principle affect double parton events differently than double brem events, and thereby contribute to the errors in the $\mathcal{R}$ measurement. The effect of uncertainties in the calorimeter energy scales was tested by reconstructing the CDF data with modified scale constants. $\mathcal{R}$ and σ_{eff} were then determined from the modified data, and the *standard deviation* in values was interpreted as the one σ systematic error. Note from equation 6.7 that energy scale uncertainty would not have effected the measurement of σ_{eff} if σ_{dijet} could have been determined from CDF data. Energy scale systematic errors are listed in table 6.1, and described in detail below.

To estimate the systematic error due to Q^2 scale, Monte Carlo events were generated using both $Q =< P_t >$ and $Q =< P_t > /2$ parametrizations. Q^2 determines α_s, which enters multiplicatively into the 4-jet cross sections. Since α_s enters to the same order (fourth power) in both double parton and double brem diagrams, their relative rate (and hence $\mathcal{R}$) is unaffected by the Q^2 scale. Additionally, $\mathcal{S}_{pair}$ and $\Delta\phi$ are unaffected, since event topology is insensitive to Q^2.

A more radical Q^2 parametrization was also tried. This calculated an independent value of α_s for each of the Feynman graph vertices using $Q = P_t/2$ of the parton at that vertex. This gave the same rate as $Q =< P_t > /2$ generation, and had no effect on the $\mathcal{S}_{pair}$ or

$\Delta\phi$ distribution shapes, for either double parton or double brem events. Consequently, systematic errors on $\mathcal{R}$ due to Q^2 parametrizations were found to be negligible. The change in $\sigma^2_{dijet} \cdot A_{DP}$ over this range of Q^2 values contributed 50% to the systematic error in σ_{eff}.

To estimate the systematic errors associated with the choice of a structure function, four different parametrizations were used. These were the Eichten, Hinchliffe, Lane, and Quigg (EHLQ) 1 and EHLQ2 sets, and the Duke and Owens 1 (DO1) and DO2 sets [12]. The main difference between the EHLQ and DO sets is in the functional form of the low x extrapolation of DIS data. The difference between the variants of both is in the choice of Λ_{QCD}, which affects the Q^2 evolution of the parametrizations. As mentioned in section 2.4, the structure function choice mainly affects the relative double parton to double brem rate, not event topologies. These choices gave only 20% variation in $\mathcal{R}$ values, but 50% variations in $\sigma^2_{dijet} \cdot A_{DP}$ (and hence σ_{eff}).

A systematic error arises because Papageno uses an approximate double brem matrix element. This error is estimated by comparing with a slightly different approximation. The comparison approximation is simpler than the one described in section 2.3.2. The simple approximation uses the exact $gg \rightarrow gggg$ matrix element, and models sub-processes involving quarks as having the same functional form as this, but *scaled* by an overall factor of 4/9 for every quark replacing a gluon. [27]. The 4/9 factor results from tuning the approximation to agree with the exact calculation.

The choice of approximation technique couples to the choice of structure function, as the different approximations react differently to parton identities. Consequently, sets of data were generated for both approximations with all four structure functions. A small variation in $\mathcal{R}$ was found for the different approximations, giving a 20% systematic error. This error does not directly affect the determination of σ_{eff}, but does enter indirectly through the dependence of σ_{eff} on $\mathcal{R}$.

A more fundamental error enters simply because the matrix elements in Papageno are calculated only to first order. Because a second order QCD $2 \rightarrow 4$ calculation does not

exist (and probably never will), there is no way to reliably estimate this error. Since the CDF data and the Monte Carlo agree on the shape of the $\mathcal{S}_{pair}$ and $\Delta\phi$ distributions, we conclude that higher order terms have a negligible impact on the value of $\mathcal{R}$. The effect of higher order contribution uncertainty on the absolute 4-jet rate (and hence σ_{eff}) is implicit in the variation of Q^2 scale discussed above.

The determination of $\sigma_{DB} \cdot A_{DB}$ involves a very small ($\sim 5\%$) error which arrises from the approximation that the entire CDF 4-jet sample comes from double brem. A 15% error comes from the uncertainty in the total luminosity of the CDF data [28]. Both of these propagate directly into the determination of σ_{eff}

A large error on $\sigma_{DB} \cdot A_{DB}$ comes from the uncertainty in the absolute energy scale of the CDF calorimeters. The absolute energy scale uncertainty has been estimated to be at most 12% for jets [5]. This is mostly caused by uncertainty in the low energy non-linearity of the calorimeters, and in the measurement of jet fragmentation. To estimate the effect of energy scale uncertainty on σ_{eff}, the CDF 4-jet data was re-analyzed using calorimeter reconstruction scale factors $\pm$ 12% from their nominal (see section 4.1). The number of selected 4-jet events was found to change by roughly a factor of 2, giving a net 65% uncertainty on $\sigma_{DB} \cdot A_{DB}$, and hence σ_{eff}. Absolute energy scale variation has no effect on the value of $\mathcal{R}$.

A smaller error comes from the uncertainty in the relative energy scale of the CDF calorimeters. Based on the analysis presented in section 5.4, this is estimated as $\pm 5\%$. The effect of this uncertainty was tested by reconstructing the CDF data with a $\pm 5\%$ change in the scale factor of the gas calorimeters. The number of selected 4-jet events was found to change by 10%, giving a net 10% uncertainty on $\sigma_{DB} \cdot A_{DB}$ and hence σ_{eff}. This also resulted in a 5% change in the fit value of $\mathcal{R}$.

6.6 Final Results and Conclusions

To compute a 'most likely' value of σ_{eff}, we average the individual values determined from each Monte Carlo set, and from each energy scale factor reconstruction. The systematic errors on this value are propagated assuming they are un-correlated (*ie* they are added in quadrature). The breakdown of systematic errors into individual sources is listed in table 6.1. The one σ errors are expressed as a percent of the average value. The average value and errors are then found to be;

$$\sigma_{eff} = 39 \text{ mb } \pm 30\% \text{ (statistical) } \pm 100\% \text{ (systematic)}$$

Recall from section 2.2 that a simple proton model predicted $\sigma_{eff} \sim 17$ mb, well within the uncertainties of this measurement.

95% confidence level limits on σ_{eff} were also determined. To account for systematic errors we adopt the policy of using the *worst case* of the individual values of σ_{eff} calculated for the different Monte Carlo and energy scale factor samples. The 95% limit is then calculated as a 3 σ statistical error deviation from this worst value. From equation 6.7, we see that, to determine an *upper* limit on σ_{eff}, the worst case involves choosing the smallest value of $\mathcal{R}$, the smallest value of $\sigma_{DB} \cdot A_{DB}$, and the largest value of $\sigma^2_{dijet} \cdot A_{DP}$. Conversely, a *lower* limit on σ_{eff} comes from choosing the largest value of $\mathcal{R}$, the largest value of $\sigma_{DB} \cdot A_{DB}$, and the smallest value of $\sigma^2_{dijet} \cdot A_{DP}$. Applying this method gives the limits;

$$\sigma_{eff} > 5.14 \text{ mb (95\% confidence level)}$$

No 95% confidence upper limit on σ_{eff} can be determined from this data (*ie* the limit is infinite). Translated into the language of equation 2.3, the absence of an upper limit on σ_{eff} means that the existence of double parton scattering *has not* been proven. The lower limit on σ_{eff} is effectively an *upper* limit on the double parton scattering cross section.

The CDF value of σ_{eff} and its lower limit can be compared with previous experimental results. These have been performed at both the Intersecting Storage Ring (ISR) PP collider,

and at the Super Proton Synchrotron (SPS) $P\bar{P}$ collider, both at CERN. Even though these machines operate at substantially different energies than the Tevatron (63 GeV for the ISR, 630 GeV for the SPS), the simplicity of the double parton cross section (see equation 2.3) allows a direct comparison of the measured values of σ_{eff}. In other words, σ_{eff} is primarily a property of the proton, and its value is independent of the experiment used to determine it. This universality breaks down when parton correlations start to influence the double parton cross section.

The Axial Field Spectrometer (AFS) collaboration working at the ISR claims to see a double parton signal, and measures $\sigma_{eff} = 5$ mb (without quoted errors) [29]. This value is excluded by the CDF limit. The AFS result is based on an analysis very similar to this one, save that the jets in the AFS sample are only required to have 4 GeV of P_t. While the fake jet background is surely smaller at ISR energies than the Tevatron, the studies done for this thesis indicate that 4 GeV P_t is insufficient to allow a comparison between clusters and parton level QCD predictions (see section 5.5). Additionally, the theoretical uncertainties in the AFS result are stated to be large. The discrepancy between the AFS and CDF results is probably due to one of these effects.

The UA2 collaboration has also performed a double parton search at the SPS [30]. Their analysis is virtually identical to the one presented in this thesis, except in using 10 Gev P_t jets. The UA2 detector has a small η acceptance region ($-1 < \eta < 1$). Because of this reduced acceptance, the UA2 4-jet sample coming from 742 nb^{-1} is only half as large as the CDF sample from 30 nb^{-1}. Using this sample, UA2 found no evidence for a double parton signal. Their limit of $\sigma_{eff} \geq 10$ mb (90% confidence level) is completely compatible with the CDF limit, and also excludes the AFS result.

6.7 Future Analysis

The prospects for making this measurement with data from the 1988 run of CDF are good. A special trigger was designed for double parton events, requiring a ΣE_t of 80 GeV, taken

over all the calorimeters (including the gas calorimeters). Most of the events passing this requirement are dijets. To remove these dijet events, online topology cuts are applied. In hardware, events are kept only if the ΣE_t is substantially larger than the sum of the E_ts of the two largest clusters;

$$\Sigma E_t - E_{t_1} - E_{t_2} > 40\text{GeV}.$$

Cluster E_t is determined using a hardware clustering algorithm, which operates on the analog trigger signals generated by the calorimeters (see section 3.1). This removes most of the dijet events. A cut on the number of hardware clusters is not made, as the efficiency for reconstructing low E_t clusters in hardware is poor. Events passing the hardware cut are then read out, and reconstructed. Reconstruction is performed by a farm of dedicated microcomputers operating in parallel, running the standard offline reconstruction algorithms (see chapter 4). Before being written to tape, a final cut is made requiring at least 4 offline clusters above 15 GeV P_t (uncorrected) in the event. This removes the three jet events, and the residual dijet events. These cuts reduce the raw $\Sigma E_t > 80$ GeV rate by a factor of 10, and have an estimated 80% efficiency for 4-jet events. Introduced late in the run, this trigger has currently accumulated 100 inverse nanobarns of data, and will hopefully reach 500 inverse nanobarns by the end of the 1988 run. From the best fit value of σ_{eff} presented above, this implies a sample of over 500 double parton events.

The 1988 data sample will include some data taken at very low ΣE_t (30 GeV). This data will allow a direct determination of the low P_t dijet cross section, thus eliminating most of the σ_{eff} systematic errors listed in table 6.1 . This plus the increase in 4-jet statistics should allow the positive identification of double parton scattering, and possibly a measure of parton correlation functions.

Appendix A

The Papageno Monte Carlo

The Papageno Monte Carlo [17] simulates $P\bar{P}$ collisions at the parton-level using first order QCD matrix elements. As input it requires the definition of a specific final state. For example, one definition could be 2 jets, both within $-3.5 < \eta < 3.5$, each jet having at least 10 GeV of P_t. For a given run, Papageno will produce only that final state. The final state definition must be chosen to avoid the singularities in the matrix elements. Unlike many other jet Monte Carlos, Papageno does not produce a mixture of jet multiplicities. Mixing events of different jet multiplicity is inconsistent with the use of first-order matrix elements, since higher multiplicities are higher order in α_s.

Papageno requires other inputs such as a choice of Q^2 scale for evaluating α_s, and a structure function parametrization. As output Papageno produces *events*. Each event consists of the N 4-vectors of the partons, and an overall weight which expresses the probability of having obtained those 4-vectors according to the QCD model. All partons are assumed massless.

In quantum mechanics, the differential interaction probability from an initial state to a final state can be written as [31];

$$dW = \rho_f \times |M|^2. \tag{A.1}$$

$|M|^2$ is the matrix element which describes the interaction dynamics. In field theory this is calculated as the square of the sum of all contributing Feynman diagrams. ρ_f describes the density of states available for the given final state. This *phase space* factor is uniquely

specified by the number and masses of the final state particles.

Papageno uses equation A.1 to generate events and weights. Initially, the events are randomly selected to evenly populate the space of possible N-body topologies [32]. The events are scaled to have a fast falling total mass spectrum described by $\mathcal{P}(M) \sim 1/M^3$. The events are boosted in the beam direction to give a Feynman x distribution described by $\mathcal{P}(x) \sim 1/x$.

Event weights are calculated by multiplying the matrix element for the given topology with the structure functions for the given Feynman x values. The effect of the non-uniform mass and x generation is removed by multiplying the weight by a jacobian factor. In this scheme the random choice of topology expresses the phase density term of equation A.1, while the weight expresses the matrix element term.

In general, the QCD matrix element does not want to uniformly populate the space of possible topologies. For instance, in multi-jet production, QCD has a strong bias for producing co-linear jets, which are produced only rarely by phase space. The event weight compensates for this skewing, giving co-linear events a larger weight than separated ones. However, this method is statistically inefficient, since only a few events are produced with the most likely topology. Similarly, the functions used to describe event mass and x do not match perfectly with the structure functions. To improve the Papageno event generation efficiency, the event topology and x generation are *relaxed* to correspond more closely with the matrix element. This is done by numerically adjusting the functions which transform a set of random numbers into an event. The adjustment is driven by a feedback loop which tries to minimize the discrepancy between event weights [33]. This procedure reduces the spread in weights by a factor of 10.

Because the events are weighted, statistical errors on Papageno calculations do not go simply as the square root of the number of events. From basic arguments, it's possible to show that the statistical errors on weighed events go as;

$$\text{Error} = \sqrt{N(\overline{w}^2 + \sigma_w^2)}.$$

$\overline{w}$ is just the average weight, while σ_w is the standard deviation of weight values. Notice this gives the traditional $\sqrt{N}$ answer for unweighted events ($\overline{w} = 1$ and $\sigma_w = 0$). For complicated processes, the statistical errors on Papageno calculations are dominated by the σ_w term. For example, to obtain $\sqrt{N}$ statistical accuracy for 4-jet production, one needs $1000 \times N$ events.

Appendix B

Jet Fragmentation Tuning

Events used in the fragmentation tuning were selected from the same triggers used in the double parton analysis. Calorimetry information was reconstructed as described in section 3.1. Tracks in the CTC were reconstructed and selected using standard algorithms [34]. Jets were clustered with a large cone size (1 $\eta\phi$ unit). Events were required to have a dijet structure (exactly 2 clusters with $P_t > 15$ GeV), with at least one of the jets in the CTC. To sharpen the trigger, the jet(s) in the central were required to have sufficient E_t to have alone triggered the event. Events were separated according to dijet system mass, making tracking acceptances and efficiencies uniform.

Figure B.1 shows the match between the Isajet and the CDF data for a *longitudinal* fragmentation variable. This variable, called Z', measures the fraction of charged momentum along the *average charged momentum direction* of each charged particle. Plotted is a frequency distribution of this variable, averaged over dijet events with mass between 100 and 120 GeV. The data are shown without acceptance or efficiency corrections, merely to indicate the similarity between data and model.

Figure B.1 shows the match for a *transverse* fragmentation variable, namely the momentum perpendicular to the average momentum direction of each charged particle. Again, the data is shown without acceptance or efficiency corrections, merely to indicate the similarity between data and model.

Agreement between data and model holds equally well at other masses. This agreement

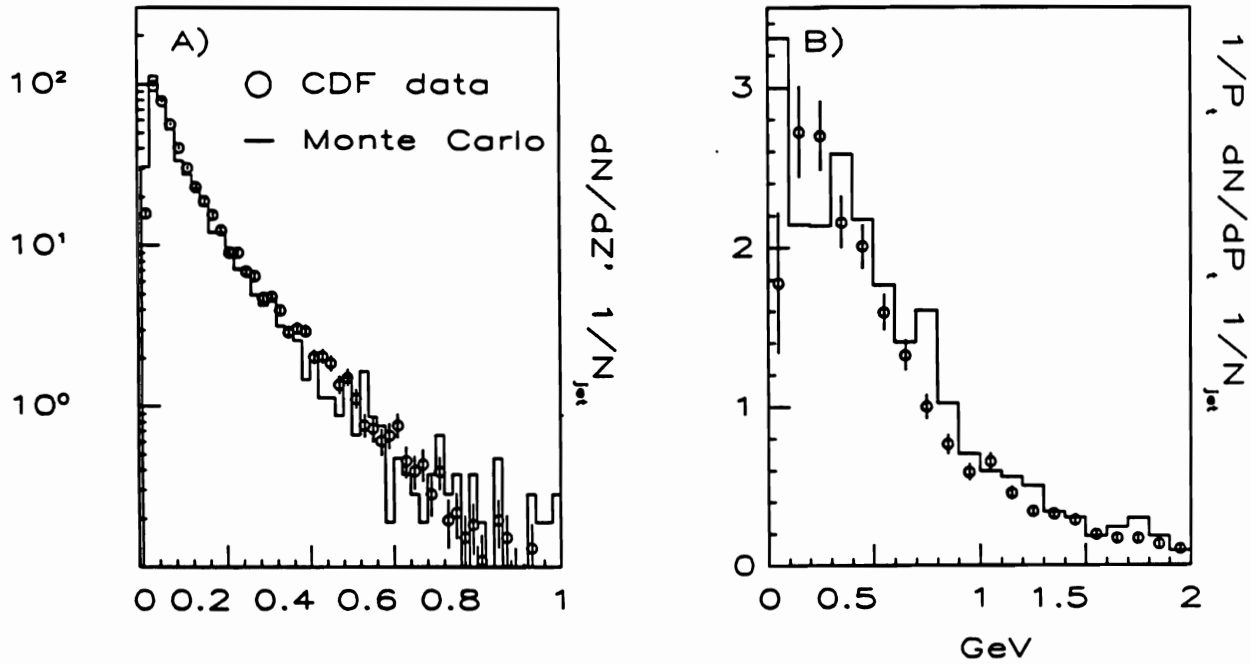


Figure B.1: *A) The average fraction of charged momentum carried by a single track in a jet, or Z'. B) The transverse momentum of charged particles relative to the calorimetry jet axis. Both plots are normalized to the number of jet events analyzed*

is adequate for the double parton analysis. We assume the model describes equally well the fragmentation of multiple jet events.

Appendix C

Gas Spike Removal Algorithm

The philosophy of the gas spike removal algorithm was to remove energy deposits that were too narrow to be due to particle showers. Narrowness was measured using the transverse and longitudinal distribution of energy deposits. The same removal algorithm was applied to all 4 gas calorimeters with a slightly different set of cuts.

The first step was to look for *clumps* of energy in $\eta\phi$ space. These were defined with a simple adjacency clustering algorithm. This clump clustering was done separately in the plug and forward regions, and not at all in the central. The algorithm gathered together towers above a threshold *energy* (EM + HAD) which touched on a side. Clumps were *seeded* by towers above a second, higher threshold. The tower energies were required to decrease monotonically going outward from the seed (*ie* the clumps were not allowed to contain energy valleys). The thresholds for the plug and forward clump clustering are listed in table C.1. These were set to remove all spikes which might serve as a seed for jet clustering (*ie* above 1 GeV E_t).

The main job of the spike removal algorithm was to decide which clumps were spikes and which were showers. The separation was done using the longitudinal and transverse distribution of energy within individual clumps. The transverse energy distribution was measured simply as the number of towers in the clump. Spikes are expected to be very narrow and hence have a small number of towers, while showers tend to be wider and contain more towers.

Table C.1: Clump clustering parameters

Parameter	Forward	Plug
Tower threshold (GeV)	2.5	.75
Seed threshold (GeV)	5.0	1.5

The longitudinal energy distribution was measured using either the depth segmentation of the pad readout or the independent anode readout. Because the anode longitudinal segmentation is much finer than the pad, the anode readout provides more information. However, the transverse segmentation of the anode readout is very poor, and overlap ambiguities limited their usefulness. These ambiguities were only a problem when the clump energy was low. For this reason, clumps were divided according to energy into *high* and *low* samples. The energy thresholds for dividing high and low clumps are listed by detector in table C.2.

Low energy clumps used the depth segmentation of the calorimeters to measure the longitudinal energy distribution of the clump. The plug has 4 depth segments (3 in the PEM plus the PHA), the forward 3 (2 in the FEM plus the FHA). Spikes were defined as clumps with more than a given fraction of their energy in a single depth segment of the calorimeter.

High energy clumps used the anode readout of the calorimeters to measure the longitudinal energy distribution of the clump. Unlike low energy spikes, high energy spikes were not required to have a small number of towers. This was because the neutron-induced proton of a texas tower sometimes ran down a wire of the chamber, making a very long and energetic 'spike'. High energy spikes were identified as clumps having a match between the clump energy and the wire energy in a single layer. The match was made in the ratio of the single layer wire energy to pad energy. This ratio is about 1 for spikes, because all the energy is deposited in a single layer, and small for showers, where the energy is spread between many layers. An additional cut was made to eliminate high energy clumps with no wire energy. These were caused when the spike energy exceeded the range of the wire

Table C.2: Spike identification cuts

	FEM	FHA	PEM	PHA
Low/high energy (GeV)	30	80	30	20
Number of towers (low)	≤ 2	≤ 4	≤ 3	≤ 2
Depth fraction (low)	$\geq .98$	$\geq .90$	$\geq .98$	$\geq .95$
Single wire fraction (high)	$\geq .4$	$\geq .6$	$\geq .3$	$\geq .5$
Total wire fraction (high)	$\leq .7$	$\leq .7$	$\leq .7$	$\leq .7$

readout circuits, causing them to read zero. This cut was made on the ratio of the *total* wire energy (*ie* summed over all layers) to pad energy.

The spike selection cut values for the different calorimeters are listed in table C.2. Clumps which satisfy these spike identification cuts were removed from the calorimeter data.

Appendix D

Jet Clustering Algorithm

Seed sites for the fixed circle jet clustering algorithm were selected based on the density of E_t deposition. To insure uniformity in the seed selection across the detector, the $\eta\phi$ segmentation was unitized in software. The energies of the 3 towers of extra ϕ segmentation of the gas calorimetry were summed together, giving an effective tower size of $\sim .1 \times 15^\circ \eta\phi$ over the entire detector. Effective towers containing over 1 GeV (EM + HA) of E_t were used as seeds.

Before being used in circle clustering, the seeds themselves were pre-clustered using a simple adjacency algorithm. Seed towers were associated if they touched on a side or corner in $\eta\phi$ space. Associated seed towers were required to fall monotonically in energy going out from the highest energy tower to the lowest. An 'energy valley' between groups of adjacent towers forced pre-cluster separation. The E_t weighted $\eta\phi$ centroid of each pre-cluster of seeds was computed, and used as the initial site for fixed circle clustering.

The fixed circle algorithm used the full $\eta\phi$ segmentation of CDF. The algorithm collected towers above 100 MeV E_t (EM + HA) inside a circle of fixed radius in $\eta\phi$ space. This radius was typically chosen to be .7 units. Circles were initially centered about each precluster centroid. The center was then iterated until it coincided with the E_t weighted $\eta\phi$ centroid of the enclosed towers. The towers of the original pre-cluster were retained during iteration regardless of whether they lay inside the circle. This provided a mild hysteresis, preventing the clusters from wandering too far from their original center. The iterations ended when the

towers enclosed remained unchanged. Convergence usually occurred within 2 or 3 iterations.

Each cluster was developed independently, creating the possibility of more than one cluster claiming the same tower. The final stage in the algorithm separated any such overlapping clusters. First, the E_t of the towers held in common by the two overlapping clusters was computed, and compared to the total E_t of each cluster. If the common towers accounted for more than half the E_t of either cluster, the two were merged into a single cluster. If less, they were separated. Separation was an iterative process, like the centroid determination. First, all towers in common were divided by removing them from the cluster further in in $\eta\phi$ space, keeping them in the nearer. The E_t weighted centroid of the now separated clusters were then re-computed, and the original set of common towers again divided. The iteration ended when the tower separation didn't change. Convergence usually occurred within 2 or 3 iterations.

Appendix E

The Parton Level Jet Simulation

The parton level jet simulation is based in equal measures on studies using the particle level simulation and on studies using the CDF data. Where ever possible, CDF data was used to define and/or tune the simulation. The fundamental limits discussed in section 2.6 on the accuracy of jet fragmentation models apply equally to this simulation.

The first job of the parton level simulation is to model fragmentation by giving mass to the individual partons. The mass of a jet clearly depends on its P_t, with higher P_t jets having larger mass. To avoid the largest part of this dependence mass is generated according to the distribution of *obesity* $\Omega \equiv \frac{M^2}{2P_t}$. Simple models of fragmentation predict a logarithmic dependence of obesity on P_t [35]. Figure E.1 plots the obesity distribution for jets in CDF central dijet events (exactly two jets above 15 GeV P_t, both in the central calorimeter). The slight dependence of obesity on jet P_t is ignored. The CDF data histogram is used to functionally generate the Monte Carlo obesity distribution. Once all the partons have mass, their 4-vectors are *rescaled* by a single constant to restore the event $\hat{S}$ to its pre-parton-mass value. This process is identical to the rescaling done by Isajet (see section 2.6).

Next, the entire event is given a Lorentz boost transverse to the beam direction at a random ϕ angle. This so-called K_t *kick* simulates the effect of initial state radiation. The magnitude of the K_t is described as the sum of two gaussians, with widths of 5 GeV and 10 GeV. These are added together in equal admixture. The K_t kick was tuned to reproduce measured $\vec{P}_t$ balancing distributions for CDF central dijet events. This tuning

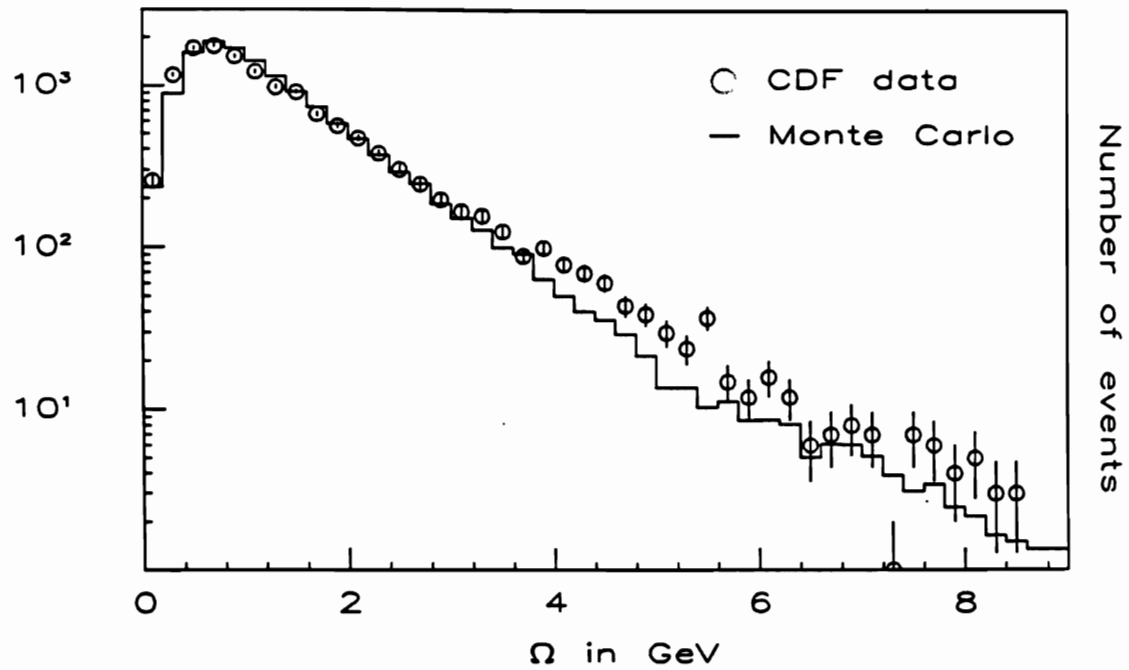


Figure E.1: *Plotted is the obesity distribution for central CDF dijets. This variable is equivalent to the jet mass, and hence measures fragmentation effects. This plot defines the distribution of jet mass used in the parton level simulation.*

is also sensitive to P_t and angular resolution effects described below. The tuning analysis is described at the end of this section. When simulating double parton events, this kick is applied to each dijet event independently.

The next step of the parton level simulation is to invert the jet response function described in section 5.3. This function is based on the particle level simulation, and corrects on average for the effects of detector response and reconstruction on the P_t of measured jets. Inverted, this function transforms parton P_t into the most likely value of cluster P_t. Because of detector non-uniformities and non-linearities, this transformation depends on both the parton P_t and its η position.

Next, the transformed parton P_t values are smeared to simulate the finite energy resolution of the calorimeters. Each parton P_t is independently distributed about its transformed value by a gaussian function. The width of the gaussian was measured using the particle level simulation in the process of calculating the response function (see section 5.3). The

variation of gaussian width with parton P_t and η was parametrized to form a continuous function similarly to how the response function was made. A quadratic function describes the variation of width with P_t, and a smoothed, splined fit describes the η variation. Given that calorimetry resolution for single particles is described by $\sigma \sim C \cdot \sqrt{E}$, one might expect jet $\sigma^2_{P_t}$ to vary linearly with P_t. Interestingly, it was found that σ_{P_t} itself varied linearly with P_t. This implies additional P_t smearing due to the varied composition of the jet. This same result was found in Monte Carlo studies of the UA1 detector [36].

Finally, the $\eta\phi$ directions of the partons are smeared about their nominal by gaussian functions. This smearing simulates the *pointing resolution* of the cluster reconstruction and detector response. This was studied using the particle level simulation, comparing the initial $\eta\phi$ direction of a parton with the $\eta\phi$ direction of the resulting cluster. The η and ϕ smearings were measured separately, as the η smearing was larger. The larger η smearing resulted from residual effects of the Z vertex shift correction described in section 3.2. Both η and ϕ smearing were measured as a function of jet P_t, as lower P_t jets showed larger smearing. The sigma of $\eta\phi$ smearing reached an asymptotic minimum value of .03 (.01) in η (ϕ) for jets above 100 GeV P_t, and a maximum value of .2 (.1) in η (ϕ) for 15 GeV P_t jets.

To test the effects of the three different kinds of smearing (K_t, P_t, and $\eta\phi$), a comparison was made between simulated Papageno dijet events and CDF dijet events. This study used $\vec{P}_t$ balancing of dijet events to test the simulation. $\vec{P}_t$ balancing was also used in section 5.4 to test the cluster response correction. At the parton level, we expect the $\vec{P}_t$ of the two jets in a dijet event to sum to zero. Detector smearing and higher order QCD effects (K_t) give non-zero balancing. Figure E.2 shows the momentum vectors of an idealized dijet event. $\vec{P}_t$ imbalance in the dijet system has occurred from P_t loss through resolution effects along the jet direction, and from a very low P_t jet emitted close to the beamline. The two effects can be separated by projecting the net $\vec{P}_t$ of the dijet system along axes parallel and perpendicular to the jet directions. Imbalance transverse to the jet direction will be

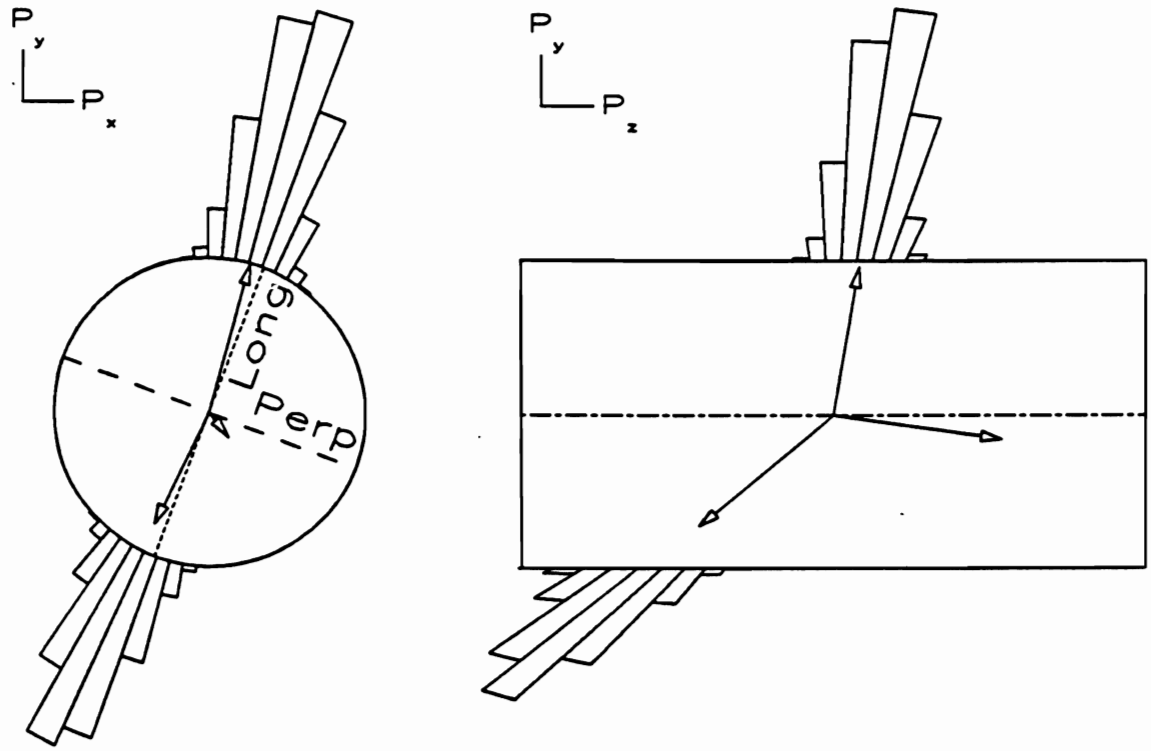


Figure E.2: *Momentum vectors of an idealized dijet event are shown in the X-Y and Y-Z projections. The third jet escapes the detector volume, giving a K_t kick to the detected jet vectors. The detected momenta are also resolution smeared. Axes transverse and longitudinal to the jet directions are shown.*

dominated by K_t effects, while parallel imbalance will be dominated by P_t resolution effects.

Figure E.3 plots the *transverse* and *longitudinal* projections of the dijet $\vec{P}_t$ for both CDF data and Papageno dijets simulated with the parton level simulation. The transverse distribution is narrower because it measures only K_t and ϕ smearing, being transverse to calorimeter resolution loss. The longitudinal distribution is sensitive to K_t and to P_t resolution effects. The agreement for both projections between the CDF data and the Monte Carlo data indicates that all the smearings are adequately simulated.

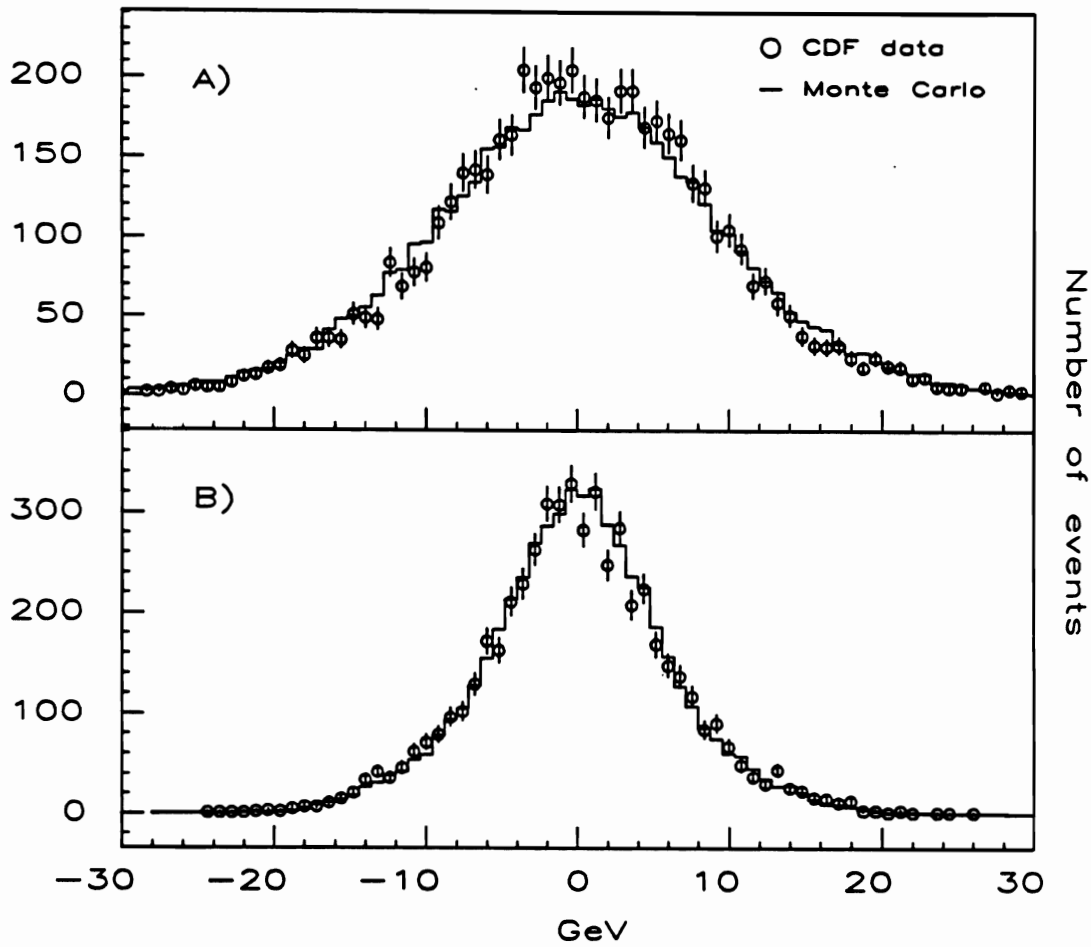


Figure E.3: *Projections of the $\vec{P}_t$ sum of the jets in CDF central dijet events are plotted. Overlaid are histograms of Monte Carlo data.*

Bibliography

[1] G. Hanson *et al.*, *Phys. Lett.* **35** 1609 (1977)

[2] The TASSO Collaboration, *Zeitschrift fur Physics* C **22** 307 (1984)

[3] The UA2 Collaboration, Observation of Very Large Transverse Momentum Jets at the CERN $P\bar{P}$ Collider, *Phys. Lett.* B **118** 203 (1982)

[4] R. St. Denis, Dijet Angular Distributions in Proton-Antiproton Collisions at the Tevatron, *Harvard University PhD Thesis* (1988)

[5] The CDF collaboration, Measurement of the Inclusive Central Jet Cross Section in $P\bar{P}$ Collisions, *Phys. Rev. Lett.* **62** 613 (1989)

S. E. Kuhlmann, Inclusive Central Jet Production at $\sqrt{S} = 1.8$ TeV, *Purdue University, PhD Thesis* (1988)

[6] R. Harris, Two Jet Differential Cross Section and Structure Functions in $P\bar{P}$ Collisions at $\sqrt{S} = 1.8$ TeV, *U.C. Berkeley PhD Thesis* **LBL-27246** (1989)

[7] The UA2 Collaboration, A Study of Three Jet Events at the CERN $P\bar{P}$ Collider, *Zeitschrift fur Physics* C **30** 341 (1986)

[8] D. Perkins, Introduction to High Energy Physics, *Addison-Wesley Inc.* 348 (1982)

[9] J. D, Bjorken, Asymptotic Sum Rules at Infinite Momentum, *Phys. Rev.* D **179 5** 1547 (1969)

[10] G. Altarelli and G. Parisi, *Nucl. Phys.* B **126** 298 (1977)

[11] V. Barger and R. Phillips, Collider Physics, *Addison-Wesley Inc.* 164 (1987)

[12] E. Eichten, I. Hinchliffe, K. Lane, and C. Quigg, Supercollider Physics, *Reviews of Modern Physics* **56** 579 (1984) erratum **58** 1065

D. Duke, J. F. Owens, Q^2-dependent Parameterizations of Parton Distribution Functions, *Phys. Rev.* D **30 1** 49 (1984)

[13] R. D. Carlitz, Parton Correlation Functions, *University of Pittsburgh Note* **PITT-88-09** (1988)

[14] . I. Hinchliffe, Personal Communication

[15] Z. Xu, D. H. Zhang, and L. Chang, *Nucl. Phys.* B **291** 392 (1987)

J. Gunion, Z. Kunszt, *Phys. Lett.* B **159** 167 (1985)

S. Parke, T. Taylor, *Nucl. Phys.* B **269** 410 (1986)

Z. Kunszt, *Nucl. Phys.* B **271** 333 (1986)

J. Gunion, Z. Kunszt, *Phys. Lett.* B **176** 163 and 477 (1986)

J. Gunion, J. Kalinowski, *Phys. Rev.* D **34** 2119 (1986)

S. Parke, T. Taylor, *Phys. Rev.* D **35** 313 (1987)

F. A. Berends, W. Giele, *Nucl. Phys.* B **294** 700 (1987)

M. Mangano, S. Parke, Z. Xu, *Nucl. Phys.* B **298** 653 (1988)

M. Mangano, S. Parke, *Nucl. Phys.* B **299** 673 (1988)

[16] C. J. Maxwell, *Phys. Lett.* B **192** 190 (1987)

[17] I. Hinchliffe, PAPAGENO, a QCD Monte Carlo, (in preparation)

[18] R. D. Field and R. P. Feynman, *Nucl. Phys.* B **136** 1 (1978)

[19] F. Snider, the CDF Collaboration, η Distribution of Charged Particles Produced in $P\bar{P}$ Interactions, *Proc. XXIV Rencontre de Moriond* (1989)

[20] F. E. Paige and S.D. Protopopescu, Isajet: A Monte Carlo Event Generator for PP and $P\bar{P}$ Interactions, *BNL Theoretical Paper* **37066** (1985)

[21] The CDF Collaboration, The CDF Detector: An Overview, *Nuclear Instruments and Methods* A **271** 387 (1988)

[22] G. Brandenburg*et al.*, An Electromagnetic Calorimeter for the Small Angle Regions of the Collider Detector at Fermilab, *Nuclear Instruments and Methods* A **267** 257 (1988)

[23] L. Balka*et al.*, The CDF Central Electromagnetic Calorimeter, *Nuclear Instruments and Methods* A **267** 272 (1988)

[24] S. R. Hahn *et al.* Calibration Systems for the CDF Central Electromagnetic Calorimeter, *Nuclear Instruments and Methods* A **267** 351 (1988)

[25] I. Hinchliffe, M. D. Shapiro, Report of the QCD Working Group, *Proc. of the 1988 Summer Study on HEP in the 1990's*

[26] R. K. Ellis and P. Nason, *Fermilab preprint* **FNAL 88/54-T** (1988)

[27] Z. Kunszt and W. J. Stirling, Multijet Cross Sections in Hadronic Collisions, *Phys. Rev.* D **37 9** 36 (1988)

[28] T. M. Liss, Luminosity Monitoring and Beam-Beam Counter Performance, *CDF Note* **552**

[29] The Axial Field Spectrometer Collaboration, Double Parton Scattering in PP Collisions at $\sqrt{S}$=63 GeV, *Zeitschrift fur Physics* C **34** 163 (1987)

[30] The UA2 Collaboration, A Study of Four-Jet Events at the CERN $P\bar{P}$ Collider, *Proc. XXI Rencontre de Moriond* (1986)

[31] L. I. Schiff, Quantum Mechanics, *McGraw-Hill, New York* 197 (1955)

[32] R. Kleiss *et al.*, A New Monte Carlo Treatment of Multiparticle Phase Space at High Energies, *Computer Physics Comm.* **40** 359 (1986)

[33] G. P. Lepage, VEGAS, and Adaptive Multidimensional Integration Program, *Cornell University preprint* **80 447** 30 (1980)

[34] B. Hubbard, the CDF Collaboration, Jet Fragmentation Properties at CDF, *Proc. XXIV Rencontre de Moriond* (1989)

[35] R. Schwitters, From Clusters to Jets, *CDF Note* (Unpublished)

[36] R. Sobie, A Remake of the Jet Correction Table with New Isajet, *UA1 note* **UA1-TN-86-77** (1986)