Entwicklung des CMS-Spurtriggers für den Hochluminositätsbetrieb des Large Hadron Colliders

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

Dipl.-Phys. Thomas Schuh

Tag der mündlichen Prüfung: 03.11.2017
Referent: Prof. Dr. Marc Weber
Korreferent: Prof. Dr. Ulrich Husemann

Inhaltsverzeichnis

1 Einführung **1**

2 Large Hadron Collider **3**

3 Compact Muon Solenoid Experiment **8**

3.1 Spurdetektor 10

3.2 Kalorimeter 13

3.3 Myon-System 17

3.4 Trigger und Datenakquisition 21

4 Hochluminositätsbetrieb **24**

4.1 High Luminosity Large Hadron Collider 24

4.2 CMS-II Detektor 26

4.3 Äußerer Spurdetektor der Phase-II 27

4.4 Spurtrigger 32

5 Spurrekonstruktion **35**

5.1 Bewegungsgleichungen geladener Teilchen im homogenen Magnetfeld . . . 36

5.2 Ereignis- und Detektorsimulation 42

5.3 Spursuche mit der Hough-Transformation 49

5.4 Spurfit mit der Linearen Regression 51

6 Digitalelektronik **55**

6.1 Configurable Logic Block – CLB 57

6.2 Block RAM – BRAM 59
6.3 Digital Signal Processing – DSP 60
6.4 Transceiver 61

7 Implementierung 62
7.1 Hough-Transformation 62
7.2 Lineare Regression 88

8 Demonstratorarchitektur 102
8.1 Systemarchitektur 103
8.2 Geometric-Processor (GP) 106
8.3 Hough-Transformation (HT) 107
8.4 Kalman-Filter (KF) 108
8.5 Duplicate-Removal (DR) 108
8.6 Demonstrator 110

9 Resultate mit dem Demonstrator 112
9.1 Effizienz der Spurrekonstruktion 113
9.2 Auflösung der Spurparameter 115
9.3 Laufzeiten 119

10 Schlussbemerkung 121

A Appendix 123
A.1 Native Auflösung 123
A.2 Lineare Regression 124
A.3 Zweierkomplement 125

Kapitel 1

Einführung

Der Large Hadron Collider (LHC) [1] am Europäischen Kernforschungszentrum CERN ist der leistungsstärkste Teilchenbeschleuniger unserer Zeit und liefert seit seiner Inbetriebnahme im Jahr 2009 eine Flut an wissenschaftlichen Erkenntnissen. Herausragend ist beispielsweise die Entdeckung des Higgs-Bosons [2, 3] sowie die Vermessung von dessen Spin, Masse und Kopplungskonstanten durch die beiden großen Experimente Compact Muon Solenoid (CMS) [4] und ATLAS [5].

Um auch weiterhin bahnbrechende Erkenntnisse zu gewinnen, beispielsweise durch die Entdeckung von supersymmetrischen Teilchen, bisher unbekannte Raum-Zeit-Dimensionen oder Quark-Substrukturen, sind wesentliche Verbesserungen des Beschleunigers geplant. Der LHC wird voraussichtlich 2026 den Hochluminositätsbetrieb [6] aufnehmen und Protonenstrahlen mit einer Schwerpunktsenergie von 14 TeV bei einer instantanen Luminosität von voraussichtlich $5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, welche fünf- bis siebenmal größer ist als der Designwert des LHC, bereitstellen können.

Für den Hochluminositätsbetrieb des LHC wird unter anderem der äußere Spurdetektor von CMS komplett erneuert. Obwohl der neue Spurdetektor mit rund 500 Millionen Pixel/Streifen eine wesentlich größere Anzahl von elektrischen Kanälen aufweist, können reduzierte

Triggerdaten jedes Ereignisses, also bei der Kollisionsfrequenz des LHC von 40 MHz, mit einer handhabbaren Datenrate von 50 TBit/s ausgelesen werden [7].

Dadurch wird die frühe Spurrekonstruktion von Spuren mit einem Transversalimpuls über 3 GeV ermöglicht. Für CMS sind diese Spurinformationen eine zwingende Voraussetzung, um bei erhöhter Luminosität interessante Ereignisse zu erkennen und auszulesen. Allerdings muss die Spurrekonstruktion in weniger als 4 µs abgeschlossen sein und die resultierenden Spuren der Triggerentscheidung zur Verfügung stehen. Die große Datenrate, zusammen mit der kurzen Laufzeit, stellen eine gewaltige Herausforderung dar. Selbst mit der modernsten Technik im Jahre 2026 ist es unklar, ob dieses Projekt realisiert werden kann. CMS ist das erste Hochenergiephysikexperiment, welches sich dieser Herausforderung stellt.

Obwohl der CMS-Spurtrigger erst in 10 Jahren eingesetzt werden soll, werden heute schon die grundlegenden Architekturen festgelegt. Diese haben massiven Einfluss auf die Umsetzung der Spurrekonstruktion und somit auf die Qualität des Spurtriggers.

In dieser Arbeit untersuche ich, wie man möglichst effektiv unter gegebenen Bedingungen Spuren rekonstruiert. Mein Ziel ist es, die Grundlage der verwendeten Spurrekonstruktion zu bilden und maßgeblichen Einfluss auf die heutigen architektonischen Entscheidungen zu nehmen.

Kapitel 2

Large Hadron Collider

Der Large Hadron Collider ist ein hadronischer Speicherring und befindet sich grob 100 m unter der Erdoberfläche in einem 27 km langen Ringtunnel, welcher bis zum Jahr 2000 von dem Large Electron-Positron Collider (LEP) [8] genutzt wurde. Die Hadronenstrahlen des LHC können sowohl aus Protonen als auch aus schweren Ionen bestehen. Bisher wurden Protonen mit Protonen (auch Protonenmodus genannt), Protonen mit Bleiionen und Bleiionen mit Bleiionen kollidiert. Der Speicherring führt zwei gegenläufige Hadronenstrahlen, welche an vier Stellen gekreuzt werden können, um Kollisionen zu erzeugen.

Besonders interessant sind bei einer Kollision vor allem die inelastische Streuungen der Hadronen, beziehungsweise ihrer Konstituenten, welche sich durch kurze Reichweiten der Wechselwirkung und große Impulsüberträge auszeichnen und auch harte Streuung genannt werden. Die Energie- und Winkelverteilungen der Reaktionspartner sowie der Verlauf der Reaktionsrate für unterschiedliche Kollisionsenergien geben Aufschluss über die Kopplungsstärke sowie die Form des Wechselwirkungspotentials und ermöglichen somit den Nachweis neuer Teilchen. Daher befinden sich an den Kreuzungspunkten die vier Hauptexperimente ALICE, LHCb, ATLAS und CMS, um die Kollisionsprodukte zu detektieren.

Die vorgesehene Schwerpunktsenergie einer Kollision im Protonenmodus beträgt 14 TeV. Um diese hohe Schwerpunktsenergie zu erreichen, ist eine mehrstufige Beschleunigung der Protonenstrahlen durch den Beschleunigerkomplex am CERN erforderlich. Dieser Komplex besteht aus einer Kette verschiedener Beschleunigern und wird in Abb. 2.1 dargestellt. Jeder Beschleuniger erhöht die Energie des Teilchenstrahls und speist diesen in den nächsten Beschleuniger ein.

Die Protonen für den Strahl erhält man durch das Ionisieren von gewöhnlichem Wasserstoffgas. Diese werden von dem Linearbeschleuniger LINAC2 auf 50 MeV beschleunigt, um sie in den PS Booster einzuspeisen. Dort werden sie zunächst auf 1,4 GeV und dann durch das Proton Synchrotron (PS) auf 25 GeV beschleunigt. Das Super Proton Synchrotron (SPS) bildet den letzten Vorbeschleuniger, welcher die Protonen in den LHC mit einer Energie von 450 GeV in Form von Paketen einspeist. Im Normalbetrieb besteht der Protonenstrahl im LHC aus 2808 Paketen mit 10^{11} Protonen pro Paket, bevor die Kollisionen beginnen. Der zeitliche Abstand zwischen zwei Paketen beträgt 25 ns. Der LHC beschleunigt die Protonen derzeit auf eine Energie bis zu 6,5 TeV und erreicht somit eine Schwerpunktsenergie von 13 TeV.

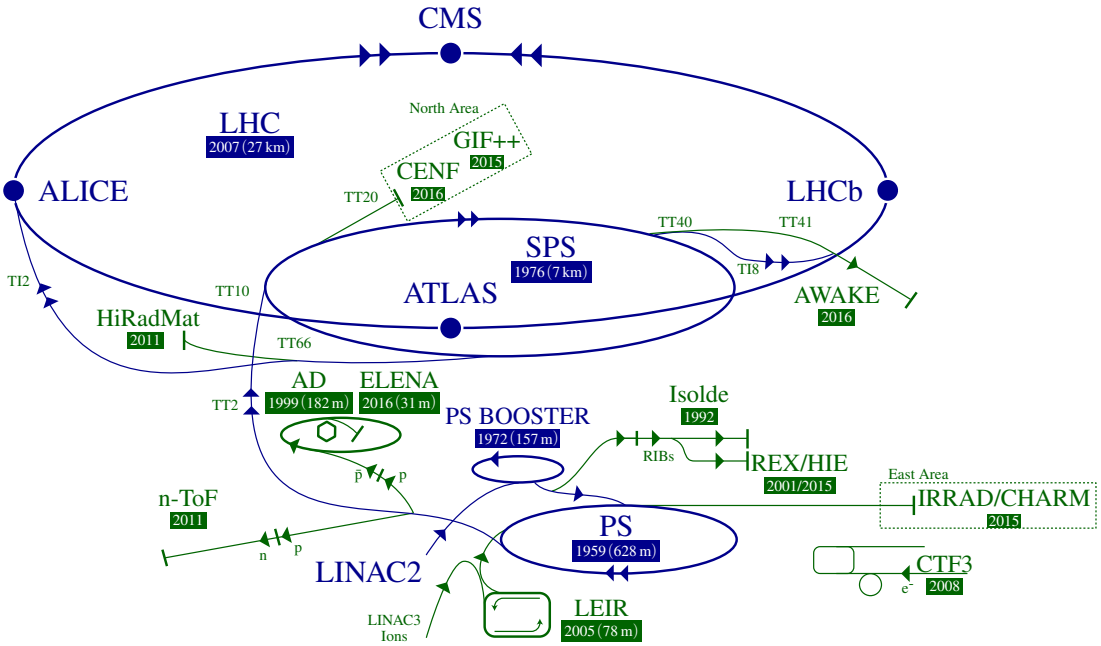

Abbildung 2.1 Der CERN Beschleunigerkomplex. Die für die Protonenstrahlen des LHC relevanten Abschnitte sind in Blau dargestellt. Adaptiert von [9].

Neben der Schwerpunktsenergie ist die Luminosität $\mathcal{L}$ ein wichtiger Parameter eines Colliders. Sie beschreibt die Anzahl der Teilchen pro Zeit und Fläche, welche sich nahe genug kommen, um potentiell zu streuen. Der Designwert für die Luminosität des LHC beträgt 10^{34} cm^{-2} s^{-1}.

Zusammen mit dem Wirkungsquerschnitt σ, der ein Maß für die Streuwahrscheinlichkeit darstellt, ergibt sich die Ereignisrate zu

$$\dot{N} \quad = \quad \mathcal{L} \cdot \sigma \, . \tag{2.1}$$

Der Wirkungsquerschnitt hat die Dimension einer Fläche und wird auch in Barn angegeben, wobei ein Barn einem Wirkungsquerschnitt von 10^{-24} cm^2 entspricht. Wirkungsquerschnitte von ausgewählten harten Prozessen und der totale Proton-Proton-Wirkungsquerschnitt werden in Abb. 2.2 gezeigt.

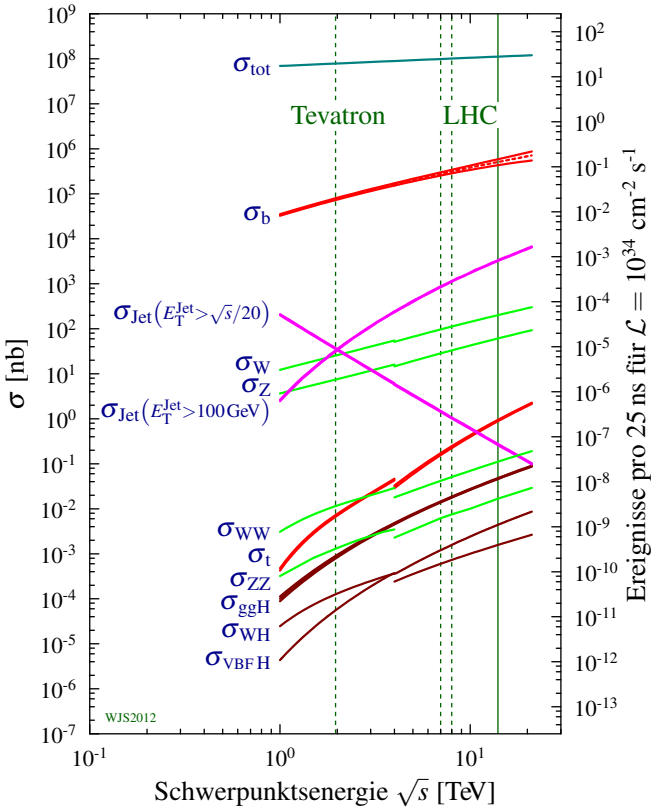

Abbildung 2.2 Ausgewählte Proton-(Anti-)Proton-Wirkungsquerschnitte am LHC (Tevatron) als Funktion der Schwerpunktsenergie. Die rechte Skala zeigt die Ereignisrate am LHC. Adaptiert von [10].

Wie man leicht erkennen kann, ist der Wirkungsquerschnitt von interessanten Prozessen wie beispielsweise Higgs- oder Top-Produktion um viele Größenordnungen kleiner als der totale Wirkungsquerschnitt. Daher rührt das Streben nach immer größeren Luminositäten, um trotz kleiner Wirkungsquerschnitte ausreichend große Ereignisraten zu erzielen.

Abb. 2.2 zeigt darüber hinaus die Anzahl der jeweiligen Ereignisse pro 25 ns, also pro Paketkollision und damit auch pro Aufnahme der Experimente. Aufgrund des großen totalen Wirkungsquerschnitts von Protonen und der hohen Luminosität ereignen sich durchschnittlich 25 unabhängige Streuungen pro Aufnahme. Diese, sich im Detektor überlappende Ereignisse, werden Pile-Up genannt und verkomplizieren den Endzustand. Das ist sozusagen der Preis für eine hohe Luminosität.

Dabei sind sowohl der Anfangs- als auch der Endzustand einer einzelnen inelastischen Proton-Proton-Streuung grundsätzlich komplex. Beim Anfangszustand liegt das daran, dass Protonen aus Quarks und Gluonen, auch Partonen genannt, zusammengesetzt sind und dass der Anfangszustand der harten Streuung aus je einem Parton der beiden kollidierenden Protonen gebildet wird. Auch wenn die Schwerpunktsenergie für die Protonen bekannt ist, gilt dies nicht für die Partonen, da diese nur einen zufälligen Anteil des Impulses der Protonen tragen. Somit sind weder die genaue Teilchenart, noch die Impulse der Teilchen vor dem Stoß bekannt. Allerdings ist die transversale Projektion, also senkrecht zur Strahlachse, der Impulse bekannt. Diese sind aufgrund des kleinen Winkels, mit dem die Protonenstrahlen gekreuzt werden, verschwindend gering. Daher gehören der Transversalimpuls p_{T} und der Azimutalwinkel in der transversalen Ebene ϕ_0 zu den wichtigsten Parametern der Kollisionsprodukte, da die transversale Projektion der Summe aller auslaufenden Impulse aufgrund der Impulserhaltung auch klein sein muss.

Die Schwierigkeit im Endzustand besteht darin, dass in einer harten Streuung häufig Quarks und Gluonen entstehen und diese den Detektor nicht erreichen. Analog zur Bremsstrahlung von Elektronen im Kernfeld strahlen auch Quarks weitere Gluonen ab. Im Gegensatz zur Bremsstrahlung können allerdings auch Gluonen weitere Gluonen oder Quarks abstrahlen. Die emittierten Quarks und Gluonen weisen meist verhältnismäßig kleine Transversalimpulse und Emissionswinkel auf. Dieser Prozess findet innerhalb einiger Femtometer statt und endet damit, dass alle Quarks und Gluonen sich zu Hadronen verbinden. Als Resultat erzeugt jedes

Quark oder Gluon im Endzustand der harten Streuung ein Bündel aus Hadronen, welches auch Jet genannt werden. Daher müssen die Eigenschaften des ursprünglichen Quarks oder Gluons über die Vermessung des Jets bestimmt werden.

Da die harte Streuung und die Bildung von Jets innerhalb kurzer Distanzen erfolgen, erscheint der Ursprung der meisten Kollisionsprodukte einer harten Streuung im Detektor als punktförmig und wird primärer Vertex genannt. Instabile Teilchen mit einer messbaren Zerfallslänge, wie beispielsweise B-Mesonen oder τ-Leptonen, stellen am Zerfallsort auch eine punktförmige Quelle von Teilchen dar, welche sekundärer Vertex genannt wird.

Kapitel 3

Compact Muon Solenoid Experiment

CMS ist einer der beiden großen Mehrzweck-Teilchendetektoren am LHC und wurde konzipiert, um ein breites Spektrum von Physikphänomenen zu erforschen. Daher ist der Detektor nicht nur sensitiv auf spezielle Ereignissignaturen, sondern vermisst weitestgehend sämtliche Kollisionsprodukte. Die wichtigsten Objekte im Endzustand der Kollision sind Elektronen, Myonen (und deren Antiteilchen), Photonen, Jets und fehlende transversale Energie.

In diesem Kapitel wird nur auf den ursprünglichen Detektor eingegangen. Dieser ist insgesamt 21,6 m lang, hat einen Durchmesser von 14,6 m, ist 12 500 t schwer und besteht aus mehreren Teilsystemen. Diese sind in Abbildung 3.1 dargestellt.

Das Kernstück des Detektors ist ein 12,5 m langer supraleitender Solenoidmagnet mit einem Durchmesser von 6 m, welcher imstande ist, ein 3,8 T starkes und homogenes Magnetfeld im Inneren aufzubauen. Das starke Magnetfeld ist von entscheidender Bedeutung, da es die Trajektorien von geladenen Teilchen krümmt und somit die Messung von deren Transversalimpuls ermöglicht. Das 10 t schwere Eisenjoch schließt die Magnetfeldlinien außerhalb der Spule und schafft so ein relativ homogenes Magnetfeld von ungefähr 1,8 T im zentralen Bereich und ein inhomogenes Feld von 2,5 T an den beiden Enden des

Magneten. Im Inneren des Magneten befindet sich ein Spurdetektor, umschlossen von einem Elektromagnetischen und Hadronischen Kalorimeter. Außerhalb befindet sich das im Eisenjoch eingebettete Myon-System.

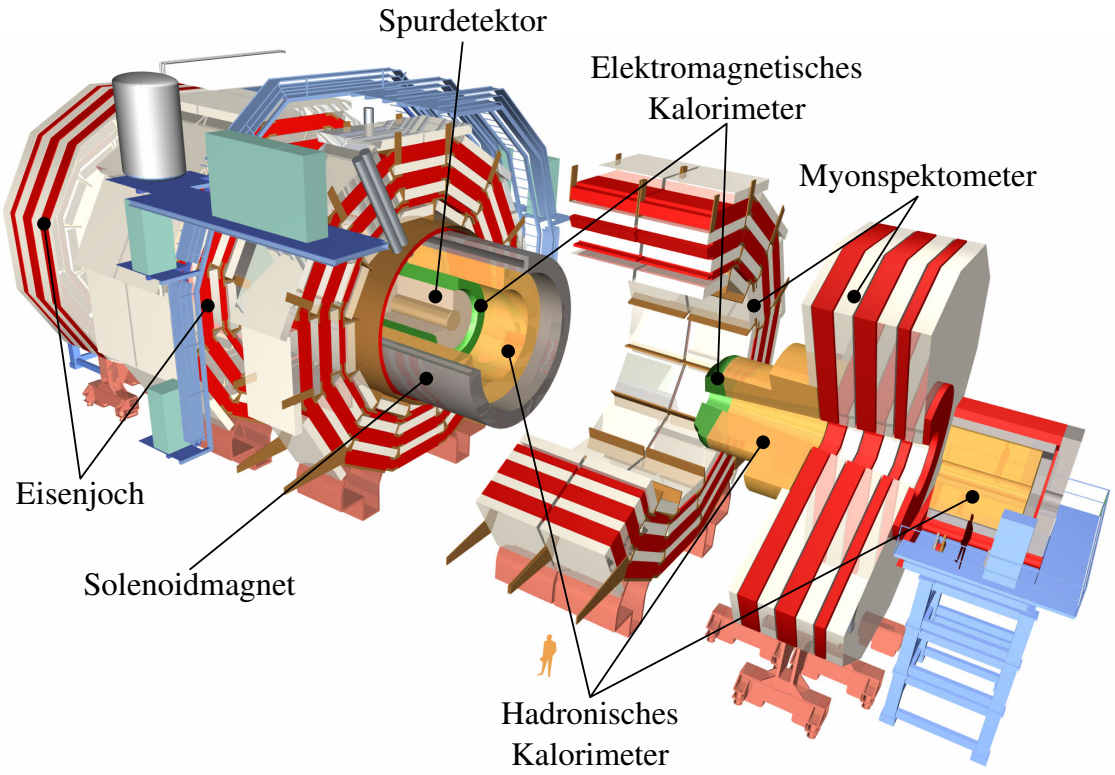

Abbildung 3.1 Explosionszeichnung von CMS mit Beschriftung der Hauptkomponenten. Adaptiert von [11].

Darüber hinaus ist das Triggersystem ein weiterer integraler Bestandteil des Detektors. Da der totale Proton-Proton-Wirkungsquerschnitt bei einer Schwerpunktsenergie von 14 TeV grob 100 mb beträgt, wird CMS bei der Designluminosität des LHC ungefähr 10^9 inelastische Ereignisse pro Sekunde detektieren. Allerdings kann eine so große Rate rein technisch weder ausgelesen, noch gespeichert werden. Zudem ist der Großteil dieser Ereignisse schlicht uninteressant für das Physikprogramm von CMS. Daher erkennt und selektiert das Triggersystem wertvolle Ereignisse, sodass letztendlich handhabbare 100 Ereignisse pro Sekunde dauerhaft gespeichert werden.

Das Koordinatensystem von CMS hat den Ursprung im vorgesehenen Kollisionspunkt. Die x-Achse weist zum Mittelpunkt des Beschleunigers, die y-Achse zeigt aufwärts zur Erdoberfläche und die z-Achse liegt in Strahlrichtung, sodass ein rechtshändiges System entsteht. In der x-y-Ebene wird der Azimutalwinkel ϕ von der x-Achse aus gemessen und der Radius wird mit r bezeichnet. In der r-z-Ebene wird der Polarwinkel θ von der z-Achse aus gemessen, welcher wie folgt mit der Pseudorapidität η im Zusammenhang steht [4]:

$$\eta \quad = \quad -\ln \tan \frac{\theta}{2} \,. \tag{3.1}$$

3.1 Spurdetektor

Der Spurdetektor dient der präzisen und zuverlässigen Vermessung der Trajektorien von geladenen Kollisionsprodukten mit einem Transversalimpuls über 1 GeV innerhalb des Bereiches $|\eta| < 2{,}5$. Da das Ladungsvorzeichen und der Impuls der Teilchen über die Krümmung ihrer Spuren bestimmt werden und die Krümmung sich schlechter bestimmen lässt, je gerader die Spur ist, nimmt die Impulsauflösung typischerweise mit steigendem Impuls ab

$$\sigma\left(p\right) \quad \sim \quad p \,. \tag{3.2}$$

Bei dem Designwert der Luminosität des LHC von 10^{34} cm^{-2} s^{-1} werden durchschnittlich 1000 Teilchen aus 20 primären Vertices den Spurdetektor durchqueren. Dies erfordert eine hohe Granularität und eine gewisse Strahlenhärte. Mit dem kurzen zeitlichen Abstand zwischen zwei Paketkollisionen von 25 ns muss der Detektor darüber hinaus auch schnellansprechend sein.

Aufgrund dieser Anforderungen hat sich CMS für einen Siliziumspurdetektor entschieden. Allerdings führt diese Wahl, für ein Spurdetektor, zu einem eher schweren System. Zum einen aufgrund des Sensormaterials selbst, aber auch wegen der hohen Energiedichte der Ausleseelektronik und des damit einhergehenden Kühlsystems. Die große Materialmenge im Spurdetektor beeinflusst die Flugbahn der Teilchen aufgrund von Vielfachstreuung und erschwert somit deren Rekonstruktion.

Der Spurdetektor hat eine zylindrische Form mit einer Länge von 5,8 m und einem Durchmesser von 2,5 m, mit dem Kollisionspunkt im Zentrum [4]. Das gesamte Volumen befindet sich im homogenen Magnetfeld des Solenoidmagneten. Der Spurdetektor besteht aus zwei Teilen, einem Pixeldetektor im Inneren und einem Streifendetektor, welcher den Pixeldetektor umschließt. Pixel- und Streifendetektor werden aus planaren Sensormodulen zusammengesetzt. Die Module bestehen aus der aktiven Siliziumsensorfläche, welche entweder pixel- oder streifenförmig segmentiert ist und der nötigen Elektronik, um diese auszulesen. Die Anordnung der Sensormodule ist in Abb. 3.2 dargestellt.

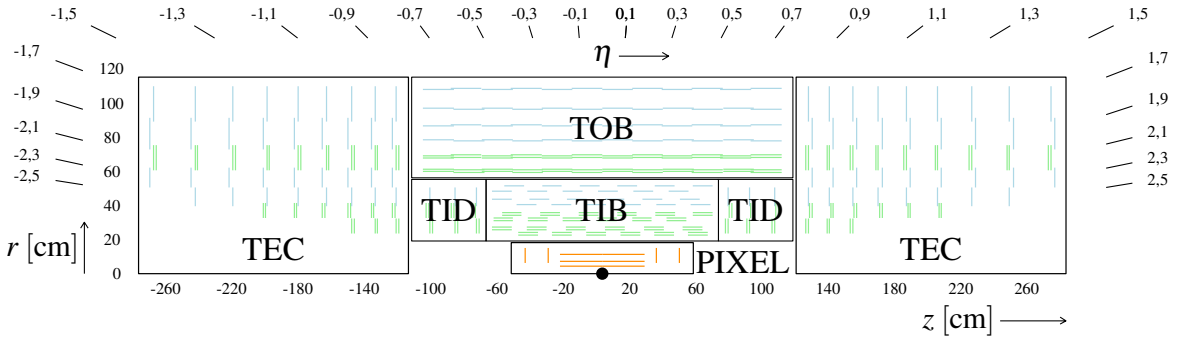

Abbildung 3.2 Detektorlayout des Spurdetektors in der *r*-*z*-Ebene mit Kennzeichnung der einzelnen Teilsysteme. Pixelmodule sind in Orange, Streifenmodule in Blau und doppellagige Streifenmodule in Grün dargestellt. Adaptiert von [4].

Der Pixeldetektor ist dem Kollisionspunkt am nächsten und weist mit einer Pixelgröße von $100 \times 150\,\mu m^2$ die größte Granularität auf. Dadurch wird die Rekonstruktion von primären und sekundären Vertices ermöglicht. Ersteres ist wichtig, um die Pile-Up-Ereignisse voneinander zu trennen und das Zweite ist essentiell, um Jets, die sich aus einem ursprünglichen Bottom- oder Top-Quark entwickelten, zu identifizieren.

Die Pixelsensormodule sind im Zentrum in drei Lagen in der Form von zylindrischen Mänteln konzentrisch um die Strahlachse angeordnet, welche mit je zwei Lagen in Form von Endkappen auf jeder Seite abgeschlossen werden. So durchqueren Teilchen mit einem Transversalimpuls über 1 GeV in der Region $|\eta| < 2{,}5$ stets drei Detektorlagen. Die Oberfläche des Pixeldetektors beträgt $1\,m^2$ und besteht aus 1 440 Pixelmodulen mit insgesamt 66 Millionen Pixeln.

Aufgrund der niedrigeren Teilchendichte und des größeren Hebelarmes im äußeren Bereich wird dort nicht die hohe Granularität des Pixeldetektors benötigt. Daher kommen hier Streifenmodule zum Einsatz, welche entweder zylindrische Mäntel oder Endkappen bilden. Im ersten Fall weisen die Streifen in $\hat{z}$-Richtung und im zweiten in $\hat{r}$-Richtung, um Sensitivität auf den Transversalimpuls der Teilchen zu erhalten. Der Streifendetektor teilt sich in vier Bereiche auf: dem Tracker Inner Barrel (TIB), dem Tracker Outer Barrel (TOB) und den Tracker Inner Disks (TID) im Zentrum und an den Seiten den Tracker End Caps (TEC).

Das innere TIB und TID System erstreckt sich über $20\,\text{cm} < r < 55\,\text{cm}$ und $|z| < 118\,\text{cm}$. Es besteht aus vier zylindrischen Lagen mit je drei Endkappen auf beiden Seiten. Der Streifenabstand der TIB-Module in den beiden inneren Lagen beträgt 80 µm und 120 µm in den beiden äußeren, was zu einer nativen Auflösung, gegeben durch den Streifenabstand geteilt durch $\sqrt{12}$ (siehe Gl. A.3), von 23 µm und 35 µm führt. Der mittlere Streifenabstand der TID-Module variiert zwischen 100 µm und 141 µm.

Der TOB hat dieselbe Abmessung in $\hat{z}$-Richtung, erstreckt sich aber bis zu einem Abstand von 116 cm von der Strahlachse und wird aus sechs Lagen gebildet. Um die Kanalanzahl in dieser äußeren Region aufgrund der großen Oberfläche zu begrenzen, wurde hier eine Streifenlänge von 25 cm gewählt. Die Breite der Streifen beträgt 183 µm in den vier inneren Lagen und 122 µm in den beiden äußeren, sodass sich eine Auflösung von 53 µm und 35 µm ergibt.

Das TEC-System setzt sich aus je neun Endkappen zusammen und erstreckt sich über die Region $22{,}5\,\text{cm} < r < 113{,}5\,\text{cm}$ und $124\,\text{cm} < |z| < 282\,\text{cm}$. Die Endkappen setzen sich aus bis zu sieben Ringen aus Sensormodulen zusammen. Der Streifenabstand variiert zwischen 97 µm und 184 µm.

Um die, aufgrund der langen Streifen, schlechte Ortsauflösung in der r-z-Ebene zu verbessern, bestehen die ersten zwei Lagen und Ringe des TIB, TID und TOB, sowie der erste, zweite und fünfte Ring der TEC aus zwei gestapelten Sensormodulen mit einem Stereowinkel von 100 mrad. Dadurch kann die Auflösung in $\hat{z}$-Richtung auf 230 µm im TIB und 530 µm im TOB verbessert werden.

Der Streifendetektor setzt sich aus insgesamt 15 148 Sensormodulen zusammen und deckt eine Oberfläche von 198 m^2 ab. Die Anzahl der Streifen beträgt 9,3 Millionen.

3.2 Kalorimeter

Im Gegensatz zum Spurdetektor, welcher die Teilchenbahn so wenig wie möglich beeinflussen soll, ist die Aufgabe eines Kalorimeters die Teilchen vollständig zu absorbieren, um ihre Energie vollständig zu messen. Außerdem sind die Messungen des Kalorimeters komplementär zu denen des Spurdetektors, da die Energieauflösung typischerweise mit steigender Energie besser wird

$$\sigma\left(E\right) \sim \frac{1}{\sqrt{E}} \tag{3.3}$$

und auch elektrisch neutrale Teilchen erfasst werden können. Allein die Energien von Neutrinos und Myonen können nicht gemessen werden. Aufgrund der unterschiedlichen Energiepositionen kann sogar zwischen Elektronen/Photonen, Myonen und Hadronen unterschieden werden. Das Kalorimeter von CMS besteht aus zwei Komponenten, dem Elektromagnetischen und dem Hadronischen Kalorimeter.

3.2.1 Das Elektromagnetische Kalorimeter – ECAL

Im Elektromagnetischen Kalorimeter erzeugen hochenergetische Elektronen, Positronen und Photonen einen elektromagnetischen Schauer. Dies geschieht durch zwei grundlegende Prozesse, die Bremsstrahlung und die Paarbildung. Im Fall der Bremsstrahlung strahlen Elektronen und Positronen aufgrund des Abbremsens im Kernfeld Photonen ab und im Fall der Paarbildung konvertiert das Photon im Kernfeld zu einem Elektron-Positron-Paar. Beide Prozesse finden im Kernfeld statt und weisen dieselbe Materialabhängigkeit auf. Mit der Strahlungslänge X_0 und dem Molièreradius R_M lassen sich elektromagnetische Schauer charakterisieren.

Die Strahlenlänge beschreibt die Wegstrecke, nach deren Durchquerung sich die Energie eines hochenergetischen Elektrons aufgrund von Bremsstrahlung um den Faktor e reduziert hat. Gleichzeitig ist die mittlere freie Weglänge von Photonen für Paarbildung gegeben durch $9/7\,X_0$. Bremsstrahlung und Paarbildung können sich solange abwechseln und so den Schauer bilden, bis die Energie der verbleibenden Schauerteilchen unter einen kritischen

Wert fällt, sodass Elektronen und Positronen direkt und Photonen durch Compton-Streuung und Photoeffekt ionisieren. Letztendlich wird die gesamte Teilchenenergie durch Ionisation deponiert, falls das Detektorvolumen ausreichend groß ist. Je nach Energie der Primärteilchen muss dazu das Kalorimeter ca. 15-25 Strahlungslängen lang sein [12].

Der Molièreradius R_{M} beschreibt die laterale Ausdehnung des elektromagnetischen Schauers. Da Bremsstrahlung und Paarbildung nur kleine Ablenkungswinkel erzeugen, wird die laterale Ausdehnung durch Coulomb-Vielfachstreuung dominiert, welche nur bei kleinen Energien wichtig wird. Daher wird der Großteil der Energie in longitudinaler Richtung deponiert und es entsteht ein enger Schauerkern, welcher sich durch einen Zylinder beschreiben lässt. Der Molièreradius ist der Radius des Zylinders, in dem 90 % der Energie enthalten ist.

Für das Elektromagnetische Kalorimeter hat sich CMS für einen feingranularen und homogenen Szintillator aus Blei-Wolframat-Kristallen, $PbWO_4$, entschieden. Homogene Szintillatoren bestehen ausschließlich aus aktivem Detektormaterial und besitzen die bestmögliche Energieauflösung, da die gesamte Energie im aktiven Medium deponiert wird. Eine sehr gute Energieauflösung ist entscheidend für die Entdeckung und Vermessung des Higgs-Bosons im Zerfallskanal nach zwei Photonen. Die Blei-Wolframat-Kristalle zeichnen sich durch ihre hohe Dichte von 8,28 g/cm^3 aus und vereinen mehrere Vorteile:

- Ein kompaktes Design aufgrund einer kurzen Strahlenlänge von 0,89 cm
- Gute Ortsauflösung durch kleinen Molièreradius von 2,2 cm
- Strahlenhärte
- Eine kurze Abklingzeit, 80 % der Anregungsenergie wird innerhalb von 25 ns in Form von blau-grünem Licht emittiert

Allerdings fällt dafür die Lichtausbeute gering aus, was aber aufgrund der vergleichsweise hohen Teilchenenergien am LHC akzeptabel ist.

Die Kristalle haben die Form von gekappten Pyramiden mit quadratischer Grundfläche. Diese sind zu einem zentralen zylindrischen Mantel (EB) angeordnet, welcher sich bis zu $|\eta| = 1{,}5$ erstreckt und zwei Endkappen (EE) mit einer Ausdehnung von $1{,}5 < |\eta| < 3{,}0$. Der Abstand der Innenseite des Mantels zur Strahlachse beträgt 1,29 m und der Abstand der Innenseiten der Endkappen zum Kollisionspunkt in Strahlrichtung beträgt 3,15 m.

Der Mantel besteht aus 61 200 Kristallen. Die Oberfläche der Kristalle an der Innenseite beträgt $22\times22\,\mathrm{mm}^2$ und $26\times26\,\mathrm{mm}^2$ an der Rückseite. Die Länge beträgt 230 mm, beziehungsweise 25,8 Strahlenlängen, was, wie schon erwähnt, ein typischer Wert ist, um den elektromagnetischen Schauer nahezu vollständig zu absorbieren. Als Photodetektor dienen zwei Avalanche-Photodioden (APDs) pro Kristall.

Die Endkappen bestehen aus je 7 324 Kristallen. Deren Abmessung beträgt $29\times29\,\mathrm{mm}^2$ an der Vorderseite und $30\times30\,\mathrm{mm}^2$ an der Rückseite, mit einer Länge von 220 mm ($24{,}7\times X_0$). Hier wird jeder Kristall mit einer Vakuum-Phototriode (VPTs) ausgelesen.

3.2.2 Das Hadronische Kalorimeter – HCAL

Das Hadronische Kalorimeter ist besonders wichtig für die Messung von Energie und Richtung von Jets, sowie fehlender transversalen Energie $E_{\mathrm{T}}^{\mathrm{miss}}$. Eine gute Messung von $E_{\mathrm{T}}^{\mathrm{miss}}$ ist von entscheidender Bedeutung für die Rekonstruktion von W-Bosonen und damit einhergehend für die Erkennung von Jets von einem ursprünglichen Top-Quark, sowie für die Suche nach neuer Physik. Daher ist die Energieauflösung nicht so entscheidend wie beim Elektromagnetischen Kalorimeter. Essentiell ist eine lückenlose und weitreichende geometrische Abdeckung bis zu $|\eta| = 5$. Das Hadronische Kalorimeter besteht aus vier Teilen: Barrel (HB), Endcap (HE), Outer Calorimeter (HO) und Forward Calorimeter (HF). Die Anordnung dieser Teilsysteme ist in Abb. 3.3 dargestellt.

Der HB und HE umschließen das Elektromagnetischen Kalorimeter und befinden sich innerhalb des Solenoidmagneten. Der zylindrische Mantel erstreckt sich über $|\eta| = 1{,}3$ und die Endkappen über $1{,}3 < |\eta| < 3{,}0$. Es handelt sich um Sampling-Kalorimeter, bestehend aus abwechselnden Schichten von Absorber-Material zur Schauerbildung und aktivem Medium zum Schauernachweis. Als Absorber-Material wurde Messing gewählt, da es nicht magnetisch ist und eine kurze Absorptionslänge besitzt. Zwischen den Messingschichten befinden sich Plastikszintillatoren, welche über Wellenlängenschieber und anschließenden Hybridphotodioden (HPD) ausgelesen werden.

Der HF befindet sich 11,2 m vom Kollisionspunkt entfernt und erweitert die Abdeckung bis zu $|\eta| = 5{,}2$, was zu den größten Strahlungsbelastungen führt. Er besteht aus Stahlabsorbern und

szintillierende Quartzfasern. Gemessen wird die Tscherenkow-Strahlung in den Quarzfasern von geladenen Schauerteilchen mit Energien über der Tscherenkow-Schwelle, welche beispielsweise für Elektronen 190 keV beträgt. Ausgelesen werden die Quarzfasern mit Photoelektronenvervielfachern.

Aufgrund der kompakten Bauart ist das Bremsvermögen des HB und EB nicht ausreichend, um hadronische Schauer ausreichend zu absorbieren, wodurch sich die Energieauflösung reduziert. Daher befindet sich der HO, bestehend aus ein bis zwei Lagen Szintillatorplatten, zwischen der Spule und dem Eisenjoch in dem zentralen Bereich $|\eta| < 1{,}3$, um die Energieauflösung wieder ein wenig zu verbessern.

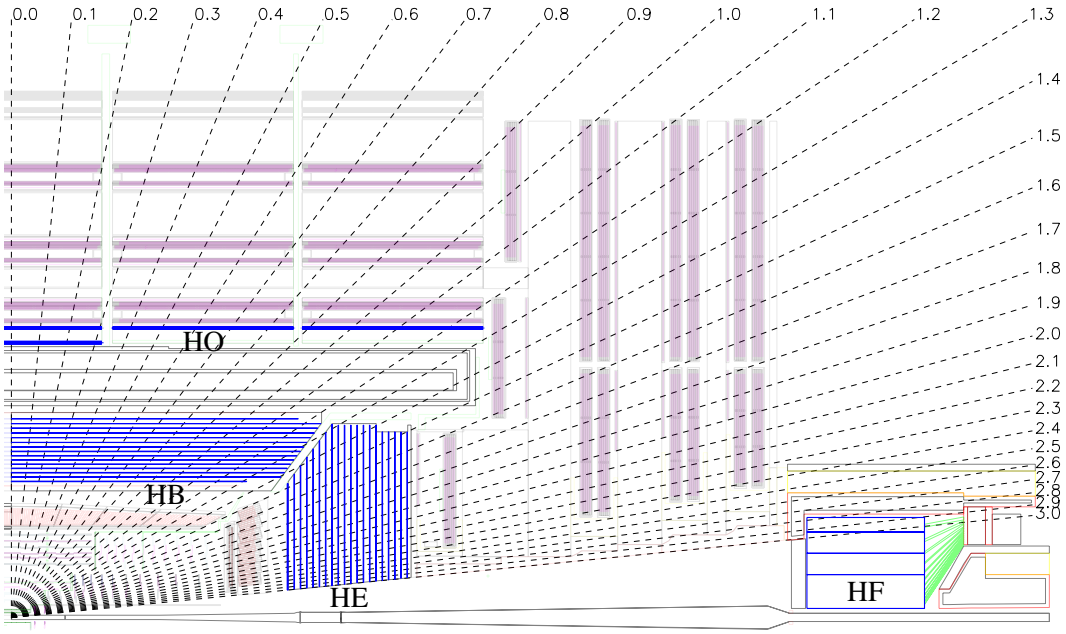

Abbildung 3.3 Layout einer Hälfte von CMS in der *r*-*z*-Ebene mit dem Kollisionspunkt in der unteren linken Ecke. Die Teilsysteme des Hadronischen Kalorimeters sind hervorgehoben und beschriftet. Adaptiert von [4].

3.3 Myon-System

Das Myon-System hat unter anderem, wie auch der Spurdetektor, die Aufgabe, die Trajektorien von geladenen Teilchen zu vermessen. Da sich das Myon-System hinter den Kalorimetern befindet, welche sämtliche Teilchen, außer Myonen und Neutrinos, fast vollständig absorbieren, fällt der Teilchenfluss im Myon-System entsprechend klein aus. Dadurch wird eine wesentlich kleinere Granularität benötigt. Darüber hinaus zeichnen sich Myonen als die geladenen Teilchen aus, welche sich am wenigsten beeinflusst durch den Detektor bewegen, da sie bei LHC-Energien keinen elektromagnetischen Schauer bilden.

Der große Hebelarm macht die Messung des Transversalimpulses vor allem im TeV-Bereich interessant, wo sich die Auflösung des Spurdetektors aufgrund des begrenzten Radius zunehmend verschlechtert, wie in Abb. 3.4 zu sehen ist.

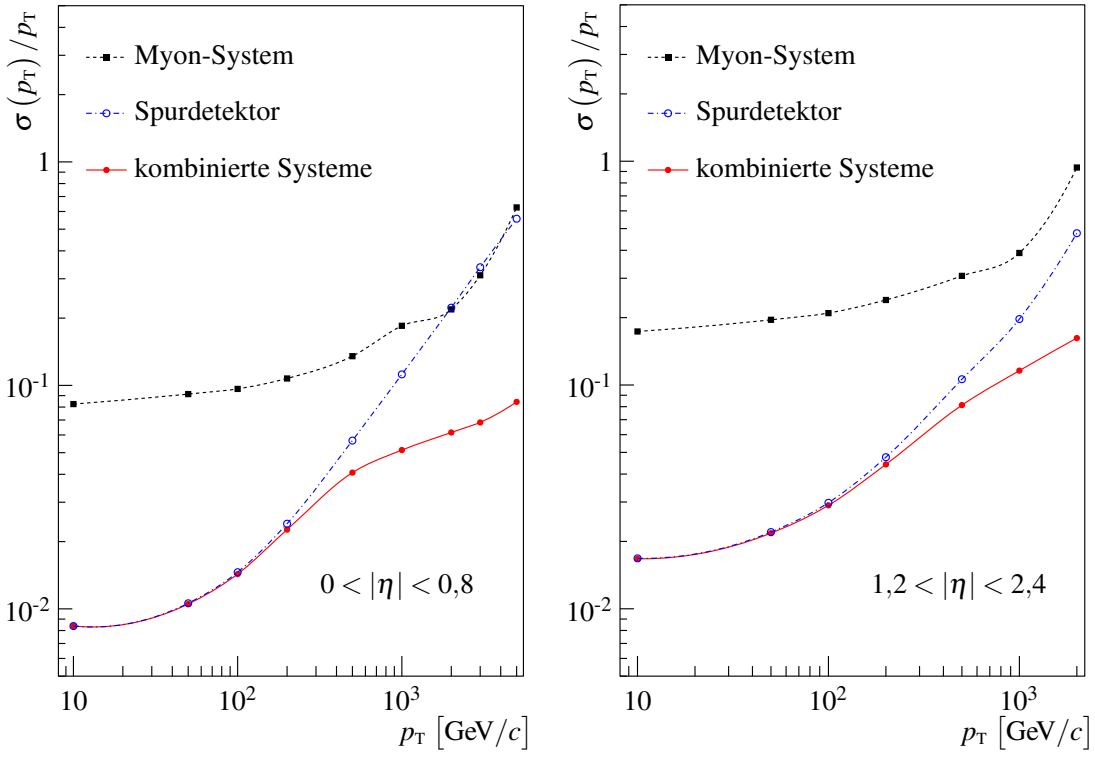

Abbildung 3.4 p_T-Auflösung von Myonen als Funktion des Transversalimpulses von dem Myon-System, dem Spurdetektor und deren Kombination für verschiedene Bereiche in η. Adaptiert von [4].

Im GeV-Bereich kann das Myon-System zwar nicht zur Impulsauflösung des Gesamtsystems beitragen, ist dafür aber im Gegensatz zum Spurdetektor in der Lage, dem Triggersystem sehr früh rekonstruierte Myonen bereitzustellen.

CMS entschied sich für verschiedene gasgefüllte Ionisationsdetektoren, aufgrund des großen Messvolumens und der geringen erforderlichen Granularität. Im zentralen Bereich werden vier zylindrische Stationen aus Driftkammern (DT) gebildet, welche mit vier Endkappen-Stationen aus Kathodenstreifenkammern (CSC) auf beiden Seiten abgeschlossen werden. Jede Station ist in der Lage eigenständig Spursegmente zu rekonstruieren. Zusätzlich sind einige der DT- und CSC-Stationen mit Widerstandsplattenkammern (RPC) versehen, welche hauptsächlich von dem Triggersystem genutzt werden. Die Anordnung der einzelnen Systeme ist in Abb. 3.5 zu sehen.

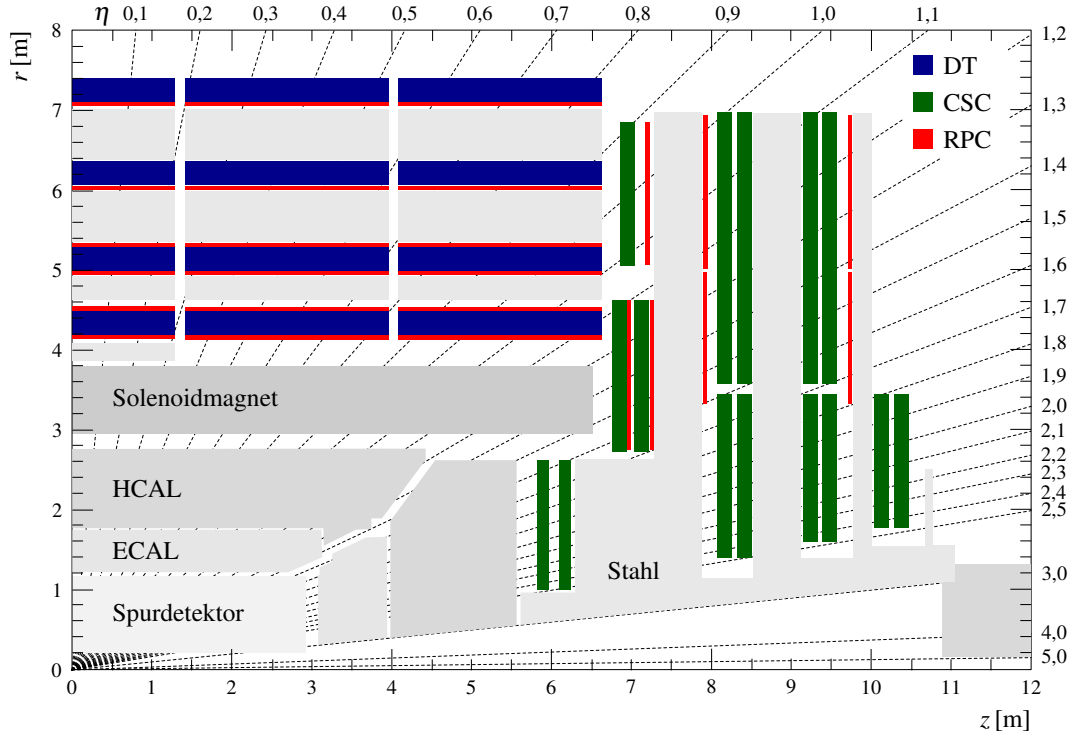

Abbildung 3.5 Skizze einer Hälfte von CMS in der r-z-Ebene mit dem Kollisionspunkt in der unteren linken Ecke. Die Kathodenstreifenkammern sind in Grün, die Driftkammern sind in Blau und die Widerstandsplattenkammern in Rot dargestellt. Adaptiert von [13].

3.3.1 DT-System

Die DT-Stationen setzen sich aus mehreren Lagen von Driftkammern zusammen. Eine Driftkammer besteht aus einem 2,4 m langen gasgefüllten Quader aus Aluminium mit einer Grundfläche von $42 \times 13\,\mathrm{mm}^2$, in dessen Zentrum ein Anodendraht der Länge nach gespannt ist. Durchquerende Myonen ionisieren das Gas und die freigesetzten Elektronen driften zum Anodendraht. Durch das geschickte Anbringen von zusätzlichen Elektroden wird das elektrische Feld so geformt, dass die Driftgeschwindigkeit im gesamten Driftvolumen weitestgehend konstant ist. Sie beträgt ungefähr 55 nm/µs , was zu einer maximalen Driftzeit von 380 ns führt. Die native Ortsauflösung (sie Abschnitt A.1) einer Driftkammer von 12 mm lässt sich durch die Messung der Differenz der Ankunftszeit des Myons und des Anodensignals wesentlich verbessern.

Um die Ankunftszeit des Myons zu bestimmen, werden vier Lagen von Driftkammern mit je einem Versatz von einem halben Anodenabstand gestapelt. Diese Anordnung wird auch Superlage (SL) genannt und liefert durch den Versatz, überwiegend unabhängig von den möglichen Einfallswinkeln der Myonen, eine Zeitauflösung von wenigen Nanosekunden. Dadurch ergibt sich eine Ortsauflösung von ungefähr 250 µm für eine einzelne Driftkammer. Die Superlage rekonstruiert nun einen sogenannten Stub, bestehend aus einem Raumpunkt, gegeben durch den Schwerpunkt des Spursegments und dessen Einfallswinkel.

Die ersten drei DT-Stationen besitzen zwei parallele Superlagen und dazu eine Superlage orthogonal. Die parallelen Superlagen sind in Strahlrichtung ausgerichtet und somit sensitiv auf die r-ϕ-Koordinate der Myonen und die andere Superlage misst die z-Koordinate. Die äußerste DT-Station misst nur in der r-ϕ-Ebene. Die angestrebte Ortsauflösung der Stationen beträgt 100 µm in $\hat{\phi}$-Richtung, 150 µm in $\hat{z}$-Richtung und 1 mrad für den Einfallswinkel. Das DT-System besteht aus insgesamt 172 000 auszulesenden Anodendrähten.

3.3.2 CSC-System

Eine Kathodenstreifenkammer wird aus einem Stapel von sieben Kathodenplatten gebildet. Die Platten sind trapezförmigen und in lange Streifen mit radialer Ausrichtung segmentiert.

Der Abstand der einzelnen Platten beträgt ungefähr 1 cm und zeigt in Strahlrichtung. Die sechs Zwischenräume sind gasgefüllt und mit je einer Reihe Anodendrähte ausgestattet, welche orthogonal zu den Kathodenstreifen verlaufen.

Um eine kurze Driftzeit und damit eine gute Zeitauflösung von 6 ns zu erhalten, wurde ein Anodenabstand von ungefähr 3,2 mm gewählt. Damit die Anzahl der Auslesekanäle handhabbar bleibt, werden die Anodendrähte allerdings nicht einzeln, sondern gruppenweise ausgelesen, was auch die r-Auflösung der Kammer auf 1,9 bis 6 mm reduziert. Darüber hinaus werden auch die Kathodenstreifen ausgelesen. In diesen wird ein positives Signal durch Influenz erzeugt, das sich über mehrere Streifen erstreckt und mit zunehmenden Abstand abnimmt. Durch die Schwerpunktsbildung der Signalstärken benachbarter Streifen kann eine Ortsauflösung in der r-ϕ-Ebene einer einzelnen Kammer von 75 µm bis 150 µm erreicht werden. Das ist weit unterhalb des Streifenabstandes, welcher zwischen 8,4 mm und 16 mm variiert.

Insgesamt besteht das CSC-System aus 220 000 Auslesekanälen für die Kathodenstreifen, die mit 12 Bit digitalisiert werden und 180 000 Auslesekanälen für die 2 Millionen Drähte.

3.3.3 RPC-System

Die Widerstandsplattenkammern bestehen aus vier Lagen von parallelen hochohmigen 2 mm dicken Kunststoffplatten, welche außen zwei gasgefüllte Kammern mit einem Innenabstand von 2 mm bilden. Die äußeren Oberflächen der Kammern sind mit einer leitenden Lage aus Graphit beschichtet. An diesen wird eine Hochspannungsdifferenz angelegt, wodurch ein homogenes Driftfeld im Gas entsteht. Zwischen den beiden Kammern befindet sich eine Lage aus Auslesestreifen.

Wie im CSC-System wird durch Influenz ein Signal in den Auslesestreifen erzeugt, hier wird allerdings die Signalstärke nicht erfasst und kein Schwerpunkt gebildet, da die Ortsauflösung eine weniger wichtige Rolle spielt. Dafür erreichen die Widerstandsplattenkammern eine Zeitauflösung von 3 ns. Die Anzahl der Auslesekanäle beläuft sich auf 109 000.

3.4 Trigger und Datenakquisition

Der LHC lässt im Normalbetrieb und Protonenmodus alle 25 ns zwei Protonpakete kollidieren, also mit einer Rate von 40 MHz. Wirkungsquerschnitt und Luminosität sind so hoch, dass mehrere Protonenpaare gleichzeitig inelastisch streuen. Somit ist jede Paketkollision potentiell interessant und ihre innere Struktur grundsätzlich komplex. Die Rate ist allerdings viel zu hoch, um die Aufnahme der Kollision zu speichern oder vollständig zu analysieren. Daher werden die Aufnahmen erst teilweise analysiert und dann entweder verworfen oder zu weiteren Analysen freigegeben, bis letztlich nur eine Auswahl von Ereignissen für die vollständige physikalische Analyse gespeichert wird.

CMS entschied sich diese Ereignisselektion mit einem zweistufigen Triggersystem zu realisieren. Die erste Triggerstufe (L1) besteht aus maßgeschneiderter Elektronik, welche nur einen Teil der Messdaten aller Ereignisse analysiert. Fällt die erste Triggerstufe eine positive Entscheidung, wird der Detektor vollständig ausgelesen und die Messdaten der zweiten Triggerstufe, dem High-Level-Trigger (HLT), übergeben. Im Gegensatz zur ersten Triggerstufe kann der HLT die Ereignisse mit den vollständigen Messdaten wesentlich detailreicher analysieren. Die positive Triggerentscheidung des HLT führt zum Speichern der vollständigen Messdaten einer Kollision auf Band. Von der anfänglichen Kollisionsrate von 40 MHz werden 100 Hz gespeichert.

Die erste Triggerstufe verwendet nur einen Teil der Messdaten, da der Detektor aufgrund der großen Kanalanzahl nicht mit 40 MHz vollständig ausgelesen werden kann. Daher stellen die meisten Teilsysteme zwei Datensätze bereit, die vollwertige Rohdaten, welche nur bei einer positiven Entscheidung ausgelesen werden und reduzierte Triggerdaten, welche mit 40 MHz ausgelesen werden. Die Rohdaten werden für 3,2 µs, was 128 Kollisionen entspricht, auf dem Detektor gespeichert. Falls Rohdaten ausgelesen werden sollen, muss die erste Triggerentscheidung innerhalb dieser Zeit auf dem Detektor vorliegen.

Die Rekonstruktion der Teilchenspuren im Spurdetektor ist sehr aufwendig, vor allem aufgrund der hohen Teilchendichte im Detektor. Deshalb wird auf die Analyse der Messdaten des Spurdetektors in der ersten Triggerstufe verzichtet. Die analogen Signale der Siliziumsensoren der Pixel- und Streifmodule werden auf dem Detektor gespeichert und nur im Falle der

positiven Triggerentscheidung der ersten Stufe ausgelesen. Die Signale werden dann erst außerhalb des Detektors digitalisiert und weiterverarbeitet. Dadurch muss die Elektronik auf dem Detektor weniger Strahlenhärte aufweisen und benötigt eine kleinere Leistungsaufnahme.

Das Elektromagnetische Kalorimeter digitalisiert die Signale der AVPs und VPTs der einzelnen Kristalle auf dem Detektor und speichert diese in strahlenharten Ringpuffern auf dem Detektor. Als Triggerdaten dienen die Summe von 5×5 benachbarter Kristalle, was der Granularität des Hadronischen Kalorimeters entspricht. Beim Hadronischen Kalorimeter werden die Signale der HPDs auch auf dem Detektor digitalisiert, allerdings können diese mit 40 MHz ausgelesen werden. Daher müssen hier keine Messdaten auf dem Detektor gespeichert werden.

Die Kathodenstreifen- und Driftkammern des Myon-Systems suchen auf dem Detektor innerhalb ihrer Stationen nach Spursegmenten. Diese rekonstruierten Objekte stellen die Triggerdaten dar. Die Widerstandsplattenkammern stellen ausschließlich Triggerdaten bereit. Hier werden die Ankunftszeiten von den Signalen der Auslesestreifen auf dem Detektor digitalisiert und mit 40 MHz ausgelesen.

Außerhalb des Detektors befindet sich die Triggerelektronik. Sie unterteilt sich in den Kalorimetertrigger, den Myontrigger und den globalen Trigger. Kalorimeter- und Myontrigger rekonstruieren aus den jeweiligen Triggerdaten lokale physikalische Objekte wie Myonen, Elektronen und Photonen, Jets, Taus und globale Objekte wie $E_{\mathrm{T}}^{\mathrm{miss}}$ und die transversale Energiesumme. Der globale Trigger trifft eine auf diesen Objekten gegründete Entscheidung und reduziert so die Ereignisrate von 40 MHz auf 100 kHz, sodass die nächste Triggerstufe die Rate um noch drei Größenordnungen reduzieren muss.

Der High-Level-Trigger wird auf einer Computerfarm realisiert. Die Software entspricht prinzipiell der Offline-Rekonstruktion des Ereignisse, allerdings einer vereinfachten und schnelleren Variante. So werden beispielsweise nur bei Bedarf und nur in Teilregionen Spuren mit den Messdaten des Spurdetektors rekonstruiert.

Im Gegensatz zur ersten Triggerstufe, welche lokal und nebenläufig die Triggerdaten analysiert und nur Ergebnisse global zusammenführt, beruht die Analyse des HLT auf den vollen Datensätzen der Ereignisse. Allerdings fließen die Rohdaten des Detektors mit einer gesamten

Datenrate bis zu 100 GB/s aus vielen verteilten Quellen. Daher besteht der erste Schritt in dem Zusammenführen der Messdaten der einzelnen Teilsysteme zu vollständigen Ereignissen und der Verteilung dieser Ereignisse auf Verarbeitungsknoten. Die anschließende Analyse versucht schnellstmöglich, ein uninteressantes Ereignis zu erkennen, um die Verarbeitung dieses Ereignisses abzubrechen und mit dem nächsten zu beginnen. Die Größenordnung der durchschnittlichen Prozessdauer beträgt 100 ms.

Kapitel 4

Hochluminositätsbetrieb

4.1 High Luminosity Large Hadron Collider

Der Betrieb des LHC ist in zwei alternierende Phasen aufgeteilt. In den Run-Phasen werden Hadronenstrahlen erzeugt und deren Kollisionen mit den Detektoren aufgenommen. In den Long-Shutdown-Phasen (LS) werden Beschleuniger und Detektoren ausgebaut. Abb. 4.1 zeigt den Zeitplan des LHC für die nächsten zwei Jahrzehnte.

Die erste Run-Phase erstreckte sich von 2010 bis 2012. In den Jahren 2010 und 2011 betrug die Schwerpunktsenergie im Protonenmodus 7 TeV und es wurde eine integrierte Luminosität von 45 pb^{-1} und 6,1 fb^{-1} für die Messungen von CMS bereitgestellt. Im Jahr 2012 wurde die Schwerpunktsenergie auf 8 TeV erhöht und die integrierte Luminosität betrug in diesem Jahr 23,3 fb^{-1}. Nach dem ersten Run wurde der LHC für zwei Jahre heruntergefahren. In dieser Zeit wurden die Experimente sowie der Beschleuniger ausgebaut, sodass der zweite Run mit einer Schwerpunktsenergie von 13 TeV im Jahr 2015 beginnen konnte. Die integrierte Luminosität für CMS betrug 4,2 fb^{-1} für 2015 und 41,1 fb^{-1} für 2016.

Die instantane Luminosität stieg stetig von $2{,}1 \cdot 10^{32}$ cm^{-2} s^{-1} im Jahr 2010, auf $7{,}7 \cdot 10^{33}$ cm^{-2} s^{-1} im Jahr 2012 auf $1{,}5 \cdot 10^{34}$ cm^{-2} s^{-1} im Jahr 2016. Somit wurde der Designwert der Luminosität

von $1{,}0{\cdot}10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ schon übertroffen, während das Erreichen des Designwertes der Schwerpunktsenergie von 14 TeV noch aussteht.

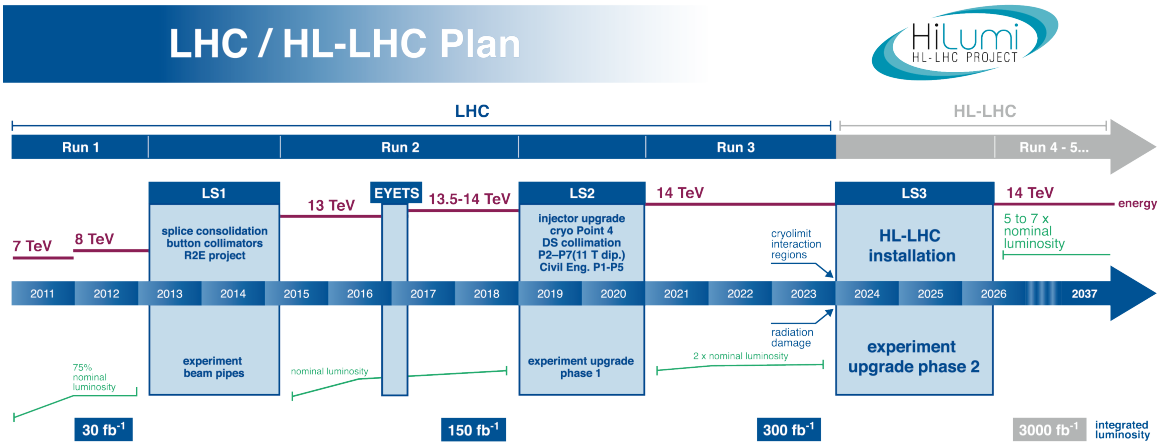

Abbildung 4.1 Zeitplan des LHC. In Rot ist die Schwerpunktsenergie der Proton-Kollisionen dargestellt, in Grün die instantane Luminosität. Die integrierte Luminosität steht in den blauen Boxen am unteren Rand. Stand Februar 2016. [14]

Ohne eine signifikante Erhöhung der instantanen Luminosität nach dem Jahr 2020 wird der statistische Gewinn für die physikalischen Analysen durch das Betreiben des LHC grenzwertig. So würde eine Laufzeit von mehr als zehn Jahren benötigt, um die statistischen Fehler der physikalischen Messgrößen zu halbieren. Daher soll die Luminosität des Beschleunigers maßgeblich erhöht werden, um das Physikpotential des LHC vollständig auszuschöpfen.

Der Hochluminositätsausbau des LHC (HL-LHC) soll während des dritten „Long-Shut-Down“ von 2024 bis 2026 stattfinden. Die Luminosität wird zwischen $5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ und $7{,}5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ liegen, während die Schwerpunktsenergie bei 14 TeV und die Kollisionsfrequenz bei 40 MHz verbleiben sollen. Die integrierte Luminosität soll $300\,\mathrm{fb}^{-1}$ pro Jahr betragen. Dies entspricht der totalen integrierten Luminosität bis dahin.

Dadurch werden Präzisionsvermessungen des Higgs-Bosons und die Suche nach seltenen Prozessen außerhalb des Standardmodells ermöglicht, sowie die Entdeckungsreichweite für neue Teilchen mit kleinen Kopplungen oder mit Massen im Bereich von mehreren TeV

erweitert. Allerdings führt die hohe Luminosität zu 140 bis 200 Überlappereignissen pro Kollision.

Aufgrund der daraus resultierenden Strahlenbelastung und Komplexität der Ereignisse im CMS-Detektor, welche die Genauigkeit der Offline-Rekonstruktion reduziert und vor allem das Trigger-System massiv überfordert, stellt der Hochluminositätbetrieb eine gewaltige Herausforderung für CMS dar. Daher ist ein Ausbau zur sogenannten Phase-II des Experimentes, zu CMS II [7], geplant.

4.2 CMS-II Detektor

Der Ausbau zur Phase-II findet genau wie der Hochluminositätsausbau des LHC im LS-3 statt. Hauptmerkmale des Ausbaus sind die Erhöhung der Strahlenhärte und der Granularität des Detektors, um der hohen Teilchendichte im Hochluminositätbetrieb gerecht zu werden. Das beinhaltet auch verbesserte Ausleseelektronik, vor allem durch erhöhte Bandbreite für größere Datenraten und eine verbesserte erste Triggerstufe, damit die Qualität der Ereignisselektion, trotz komplexerer Ereignisstrukturen aufrecht erhalten werden kann.

Die angestrebte Rate, mit der die erste Triggerstufe Ereignisse akzeptiert, liegt daher bei 750 kHz, was einer Erhöhung proportionalen zur instantanen Luminosität gleichkommt. Allerdings wird die Laufzeit, die Zeitdifferenz, in der die Triggerentscheidung nach der Kollision auf dem Detektor vorliegen muss, auf 12,5 µs erhöht, um die komplexeren Ereignisse detailreicher zu analysieren. Das hat zu Folge, dass die Rohdaten länger auf dem Detektor gehalten werden müssen und die dafür verantwortliche Elektronik ausgetauscht werden muss.

Die Ausleseelektronik der Driftkammern und Kathodenstreifenkammern des Myon-Systems wird komplett mit leistungsfähigerer Elektronik ersetzt. Darüber hinaus werden in den Endkappen-Stationen neue auf Gaselektronenvervielfacher (GEM) basierenden Kammern und verbesserte Widerstandsplattenkammern eingesetzt. Durch die neuen Kammern wird sowohl die Trigger- als auch die Rekonstruktionsqualität verbessert und eine größere Abdeckung in η erzielt.

Die Blei-Wolframat-Kristalle des Elektromagnetischen Kalorimeters im Bereich des Mantels werden auf tiefere Temperaturen gekühlt und die Ausleseelektronik der APDs ersetzt. Die Verbesserung der Ausleseelektronik führt dazu, dass die Messdaten einzelner Kristalle mit 40 MHz ausgelesen werden können, anstatt nur die Summe von 5×5 benachbarten Kristallen. Somit steht der ersten Triggerstufe die volle Granularität zur Verfügung. Die inneren Plastikszintillatoren im zylindrischen Mantel des Hadronischen Kalorimeters werden mit einer strahlenhärteren Variante ausgetauscht. Alle HPDs werden schon vor dem Ausbau zur Phase-II durch Silizium-Photovervielfacher (SiPM) ersetzt.

Die Endkappen des Elektromagnetischen und Hadronischen Kalorimeters werden durch ein kombiniertes feingranulares Sampling-Kalorimeter ersetzt. Es besteht aus zwei Teilen, einem elektromagnetischen nahe des Kollisionspunktes und einem äußerem hadronischen. Als Absorbermaterial wird für ersteres Wolfram und für letzteres Messing verwendet. Das aktive Medium besteht aus Silizium in Strahlnähe und Plastikszintillatoren außen. Es sind 28 Sensorlagen im elektromagnetischen und 12 im hadronischen Teil vorgesehen. Zur Auslese werden SiPMs benutzt. Die feine transversale Granularität von 0,5 cm^2 bis 1 cm^2 und die große Anzahl an Sensorlagen ermöglichen eine exzellente Richtungsauflösung der Schauer, wodurch die Energiedepositionen einfacher den verschiedenen Überlappereignissen zugeordnet werden können.

Der gesamte Spurdetektor wird komplett ausgetauscht. Dies geschieht hauptsächlich aufgrund der Strahlenschäden der Siliziumsensoren nach 15 Jahren Betrieb. Der neue Spurdetektor unterteilt sich wie gehabt in einen inneren und äußeren Spurdetektor. Der innere Teil wird auch mit Pixel-Modulen aufgebaut, allerdings mit einer größeren Abdeckung in η. Sensationell ist, dass der äußere Spurdetektor Triggerdaten für die erste Triggerstufe bereitstellen wird. Dementsprechend wird auch des Trigger-System mit einem Spurtrigger erweitert.

4.3 Äußerer Spurdetektor der Phase-II

Erstmals wird in der Hochenergiephysik ein Spurdetektor konzipiert und umgesetzt, welcher in der Lage ist, der ersten Triggerstufe begrenzte Spurinformationen zu liefern. Durch diese Informationen soll der Trigger in der Lage sein, mit einer Rate unterhalb von

750 kHz Ereignisse zu akzeptieren, ohne dabei seine Sensitivität auf potentiell interessante physikalische Prozesse einzubüßen.

Aufgrund der limitierten Bandbreite ist es nicht möglich, jeden detektierten Teilchendurchgang im Detektor zur 100 m entfernten Triggerelektronik zu senden. Um reduzierte Triggerdaten zu erzeugen, werden deshalb im äußeren Spurdetektor neuartige Sensormodule, sogenannte p_T-Module [15], verwendet.

Diese Module bestehen aus zwei parallelen und nahe beisammen liegenden Siliziumsensoren. Im Gegensatz zu den gestapelten Modulen des ursprünglichen Spurdetektors, weisen die Siliziumsensoren der p_T-Module keinen Stereowinkel auf. Durch die Korrelation der gemessenen Positionen der Teilchendurchgänge können, wie in den Superlagen der Driftkammern des Myon-Systems, Stubs gebildet werden, bestehend aus einem Raumpunkt und dem Einfallswinkel. Aus diesen Informationen kann auf die Krümmung der Teilchenspur und somit auf dessen Transversalimpuls geschlossen werden. Allerdings wird hierbei stets vorausgesetzt, dass der Teilchenursprung auf der Strahlachse liegt.

Wie in Abb. 4.2 veranschaulicht wird, kann durch eine konfigurierbaren Schwelle, welche typischerweise zwischen 2 GeV und 3 GeV liegt, eine Datenreduktion erreicht werden. Diese Triggerdaten können mit 40 MHz ausgelesen werden, da ihr Datenvolumen im Vergleich zu den Rohdaten ungefähr eine Größenordnung kleiner ist [16].

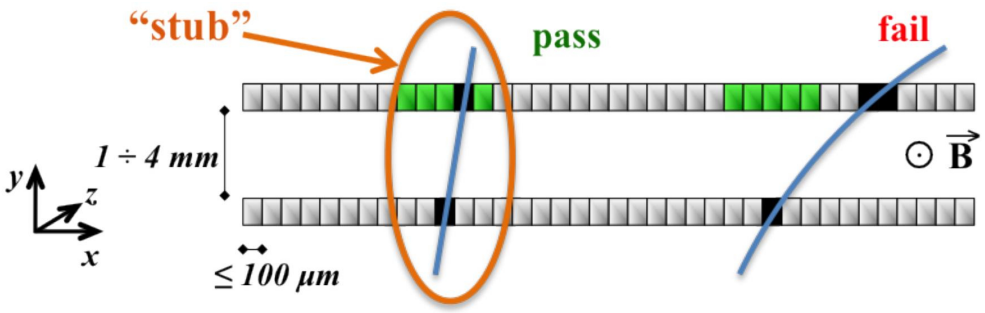

Abbildung 4.2 Stub-Generierung durch p_T-Module im Magnetfeld mit Feldrichtung aus der Blattebene. Die beiden Sensorlagen sind in Grau dargestellt und die gemessenen Teilchendurchgänge sind geschwärzt. Die grünen Blöcke entsprechen dem programmierbaren Fenster, in dem ein Teilchendurchgang liegen muss, damit ein Stub gebildet wird. [7]

Es werden zwei Arten von p_T-Modulen entwickelt, feingranulare Pixel-Streifen-Module (PS-Module) für den inneren Bereich und Streifen-Streifen-Module (2S-Module) außerhalb. In Abb. 4.3 sind beide Module abgebildet und Abb. 4.4 zeigt die geplante Anordnung der Module im Detektor. Diese entspricht zum Großteil dem gewohnten Format eines zylindrischen Mantels im zentralen Bereich, abgeschlossen auf beiden Seiten mit je einer Endkappe.

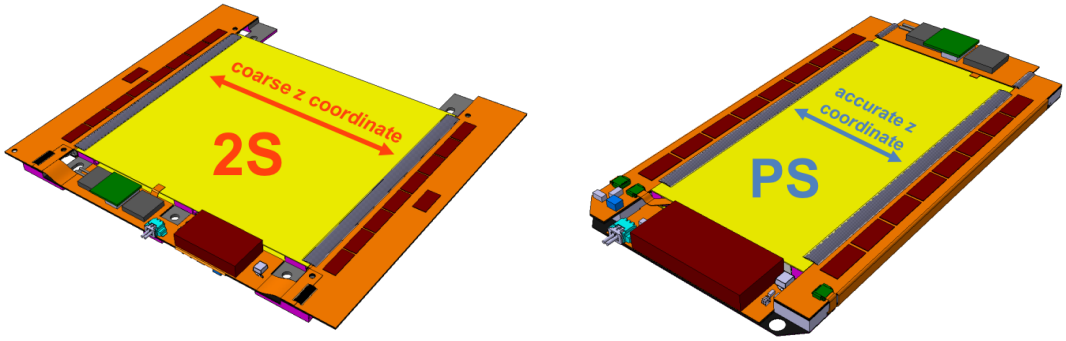

Abbildung 4.3 Sensormodule des äußeren Spurdetektors der Phase-II. Das 2S-Modul ist links und das PS-Modul ist rechts dargestellt. Die Sensorflächen sind gelb. [7]

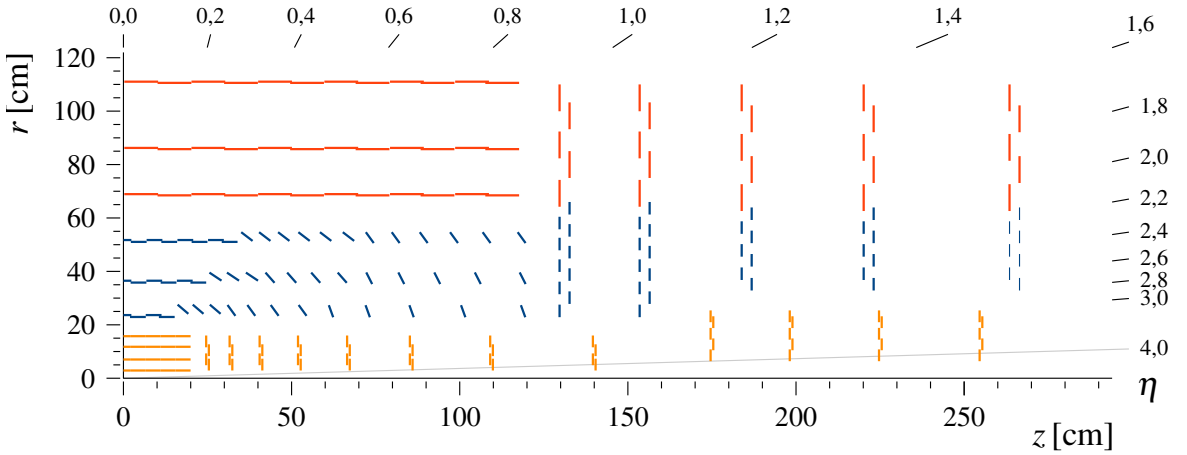

Abbildung 4.4 Layout einer Hälfte des geplanten Spurdetektors der Phase-II in der *r*-*z*-Ebene mit dem Kollisionspunkt in der unteren linken Ecke. PS-Module sind in Blau, 2S-Module in Rot und Pixelmodule in Orange dargestellt. Adaptiert von [17].

Die 2S-Module sind $10 \times 10\,\mathrm{cm}^2$ groß und werden im Bereich $60\,\mathrm{cm} < r < 120\,\mathrm{cm}$ eingesetzt, wo die Teilchendichte im Vergleich zum inneren Bereich geringer ist. Die beiden Sensoren sind identisch und in zwei Reihen aus je 1 016 Streifen segmentiert. Die Streifen haben einen Abstand von 90 µm und die Streifenlänge beträgt 5 cm. Angeordnet sind die 2S-Module in drei Lagen in Form von zylindrischen Mänteln im zentralen Bereich mit $|z| < 120\,\mathrm{cm}$, welche mit je fünf Lagen in Form von Endkappen abgeschlossen werden.

Die PS-Module sind mit $5 \times 10\,\mathrm{cm}^2$ halb so groß wie die 2S-Module und bestehen aus zwei unterschiedlichen Sensoren, einem Streifen- und einem Pixelsensor. Der Streifensensor entspricht im Prinzip den Sensoren der 2S-Module, allerdings sind die Streifen 2,35 cm lang, weisen einen Abstand von 100 µm auf und eine Reihe besteht aus 960 Streifen. Die Pixel im Pixelsensor zeichnen sich durch ihre kurze Länge von 1,47 mm aus. Sie haben wie die Streifen einen Abstand von 100 µm und der Sensor ist in 16 Reihen mit je 690 Pixel segmentiert.

Die hohe Granularität der PS-Module ermöglicht einerseits den Einsatz der Module im Bereich $20\,\mathrm{cm} < r < 60\,\mathrm{cm}$, wo die Strahlendichte höher ist, als im äußeren Bereich. Andererseits wird die Messung von dreidimensionalen Raumpunkten ermöglicht, was für den Spurtrigger entscheidend ist, um Teilchen aus unterschiedlichen primären Vertices zu trennen, da der innere Pixeldetektor keine Triggerdaten liefert.

Um Stubs bilden zu können, muss die Ausleseelektronik eines Modules natürlich mit beiden Sensoren verbunden sein. Aufgrund der einfacheren und zuverlässigeren Umsetzbarkeit werden zwei Sensorhälften getrennt voneinander ausgelesen. Daher kann die Elektronik keine Stubs mit einem räumlichen Schwerpunkt in der Nähe der getrennten Sensormitte und großem Einfallswinkel bilden.

Eine übliche Anordnung der PS-Module zu zylindrischen Mänteln im zentralen Bereich, welches fortan das flache Layout genannt wird, würde an den Randregionen nicht zuverlässig Stubs bilden können, wie in Abb. 4.5 gezeigt wird. Da das Bilden von Triggerdaten eines der Hauptziele des Entwurfs für den äußeren Spurdetektors ist, ist eine solche Anordnung nicht akzeptabel.

Deshalb wurde für den zentralen Bereich ein neuartige Anordnung der Module entwickelt. Diese bestehet aus drei PS-Modullagen in Form von zylindrischen Mänteln, deren Module

in der r-z-Ebene gekippt sind, sodass Teilchen aus dem Ursprung fast rechtwinklig einfallen. Die 2S-Module müssen nicht derartig angeordnet werden, da der Effekt hier nicht so groß ist. Das liegt an den kleineren Einfallswinkeln aufgrund des größeren Abstandes zur Strahlachse, dem kleineren Sensorabstand und den doppelt so großen Sensorhälften.

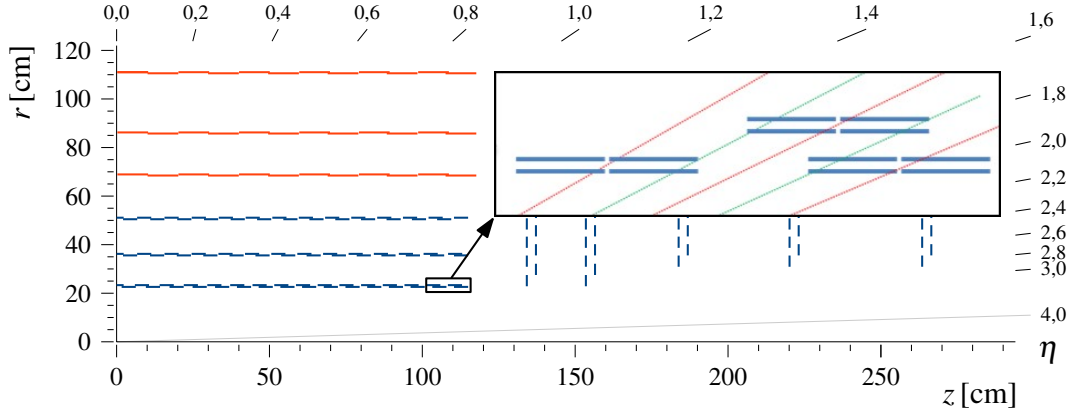

Abbildung 4.5 Das sogenannte flache und inzwischen überholte Layout des geplanten Spurdetektors der Phase-II (siehe Abb. 4.4). Der Kasten in der oberen rechten Ecke zeigt eine Vergrößerung des Rands der innersten zylindrischen Detektorlage. Hier sind Sensorhälften in Blau und Flugbahnen geladener Teilchen, welche Stubs erzeugen können oder nicht, in Grün beziehungsweise Rot dargestellt. Adaptiert von [17].

Wie im originalen äußeren Spurtrigger sind die Module so ausgerichtet, dass die Streifen in den Endkappen weitestgehend in $\hat{r}$-Richtung und im zentralen Bereich weitestgehend in $\hat{\phi}$-Richtung zeigen, um die größte Sensitivität auf den Transversalimpuls zu erhalten. Die Abdeckung in $|\eta|$ wird im Vergleich zum Vorgängerdetektor von 2,5 auf 3 erweitert. Jedoch werden die Module außerhalb von 2,4 keine Stubs zur Triggerelektronik senden, da Teilchen aus der primären Wechselwirkungszone in diesem Bereich nicht genügend Detektorlagen durchqueren, um sie rekonstruieren zu können.

Alle Teilchen aus der primären Wechselwirkungszone durchqueren innerhalb der Region $|\eta| < 2{,}4$ mindestens sechs Detektorlagen, mit Ausnahme einer kleinen Lücke in der Übergangsregion von den Mantellagen zu den Endkappenlagen bei $|\eta| \approx 1$, wo nur fünf Detektorlagen durchquert werden können. Dies ist die erforderliche Mindestanzahl an

Detektorlagen, zur effizienten und robusten Rekonstruktion von Spuren aus Stubs [18]. Mit robust ist gemeint, dass eine Spur rekonstruiert werden kann, trotz des Versagens einer Detektorlage. Da mindestens zwei dreidimensionale Messpunkte vonnöten sind, um die Flugbahn der Teilchen präzise zu bestimmen, gibt es drei PS-Modullagen.

4.4 Spurtrigger

Der Spurtrigger ist der neuer Bestandteil der ersten Triggerstufe von CMS II und verarbeitet die Triggerdaten des Spurdetektors. Die Aufgabe besteht in der Rekonstruktion und Parameterbestimmung der Flugbahnen von elektrisch geladenen Teilchen aus primären Vertices mit einem Transversalimpuls über 2 GeV bis 3 GeV.

Die Bereitstellung dieser Spurinformationen für den globalen Trigger eröffnet viele neue Möglichkeiten, die Ereignisselektion zu optimieren. Beispielsweise könnte die p_T-Auflösung für eine Vielzahl von Objekten verbessert werden oder zumindest ansatzweise eine Zuordnung von Objekten zu den verschiedenen primären Vertices stattfinden. Dies sind zwingende Voraussetzungen, um die Ereignisrate trotz 200 Überlappereignissen von 40 MHz auf 750 kHz zu reduzieren, ohne dabei einen signifikanten Verlust von interessanten Ereignissen hinnehmen zu müssen. Der Datenfluss von den Sensormodulen bis zur ersten Triggerstufe, sowie die geplanten Laufzeiten der Prozesse sind in Abb. 4.6 abgebildet.

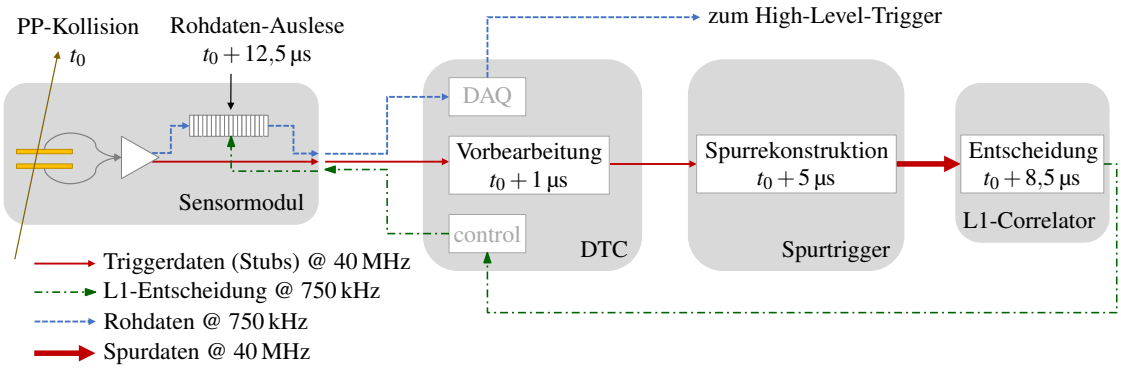

Abbildung 4.6 Übersicht der Grundarchitektur des Spurtriggers. Eingetragen sind die Pfade der Trigger-, Roh- und Spurdaten, der Pfad der Triggerentscheidung sowie die geplanten Laufzeiten. Adaptiert von [19].

Die Latenz der ersten Triggerstufe ist auf 12,5 µs begrenzt. Die Triggerelektronik, welche die rekonstruierte Spuren mit den rekonstruierten Objekten von den Kalorimetern und Muonkammern korreliert und eine Triggerentscheidung fällt, wird auch L1 Correlator genannt und benötigt schätzungsweise 3,5 µs. Der Transfer dieser Entscheidung zum Detektor benötigt voraussichtlich 1 µs, wobei weitere 3 µs zur Sicherheit reserviert sind. Auch der Datenweg von den Sensormodulen zum ersten Elektroniksystem außerhalb des Detektors, das Data, Trigger and Control (DTC) System, zusammen mit der Logik zur Bildung von Stubs auf dem Sensormodul, benötigt voraussichtlich 1 µs. Daraus ergibt sich ein Zeitrahmen von 4 µs, um aus den Stubs Spuren zu rekonstruieren.

Da die Spurrekonstruktion in der ersten Triggerstufe für die gesamte Hochenergiephysik Neuland ist, ist es noch offen, was und wie genau im Spurtrigger und im L1-Correlator geschieht. Allerdings sind die Abläufe im Sensormodul und TDC gut spezifiziert.

Die Sensormodule werden mittels Lichtwellenleiter bidirektional mit dem DTC-System verbunden. Die inneren Module werden mit einer totalen Bandbreite von 10,24 Gb/s und die äußeren Module mit 5,12 Gb/s angebunden. Aufgrund von Fehlerkorrektur und Protokoll der Datenübertragung ergibt sich eine effektive Bandbreite von 8,96 Gb/s beziehungsweise 3,84 Gb/s für die Roh- und Triggerdaten [20]. Der Datenstrom der Triggerdaten, also der Stubs, wird voraussichtlich dreimal so groß sein, wie der Datenstrom der Rohdaten.

Das Datenformat der Stubs hängt von der Modulart ab. Für beide Arten wird eine 11-Bit-Adresse verwendet, welche der Position des Teilchendurchgangs in einem der beiden Sensoren, dem sogenannten Seeding-Sensor, entspricht. Der Seeding-Sensor ist in den PS-Modulen der Pixelsensor und in den 2S-Modulen einer der Streifensensoren. Die Position wird aufgrund der Schwerpunktsbildung von benachbarten aktivierten Streifen/Pixeln in Einheiten von halben Streifen/Pixel angegeben. Ausschließlich für PS-Module wird eine zusätzliche 4-Bit-Adresse angegeben, welche der Pixelnummer in Längsrichtung entspricht.

Die Positionsdifferenz des Teilchendurchgangs in den beiden Sensoren, relativ zu einem programmierbaren Versatz, wird „Bend“ genannt und bildet einen weiteren Teil des Datenformats. Der Versatz korrigiert den Fakt, dass die Sensormodule eben sind und somit nicht alle Streifen oder Pixel denselben Abstand zur Strahlachse haben. Dadurch wird der Bend proportional zur lokalen Krümmung der Teilchenspur. Der Bend wird in Einheiten von

ganzen Streifen/Pixel gemessen und mit 3 Bit für PS-Module oder mit 4 Bit für 2S-Module angegeben. Falls zwei Teilchendurchgänge in dem Nicht-Seeding-Sensor gemessen werden, wird der Stub mit dem kleineren Bend gebildet.

Das Auslesen der Stubs soll auf statischen und zu den Kollisionen des LHC synchronen Paketen beruhen. Aufgrund der statistischen Fluktuationen in der Anzahl von Stubs pro Modul und Kollisionen werden die Stubs auf dem Sensormodul erst über acht Kollisionen, also 200 ns, gesammelt und dann ausgelesen, um die Verluste aufgrund der statischen Paketgröße zu minimieren. Das führt dazu, dass 2S-Module 16 Stubs, äußere PS-Module 17 Stubs und innere PS-Module 35 Stubs pro Sensorhälfte und acht Kollisionen liefern können. Falls mehr Stubs generiert werden, werden die Stubs mit dem kleinerem Bend ausgelesen, um die zuverlässige Rekonstruktion von elektrisch geladenen Primärteilchen mit den größten Transversalimpulsen zu gewährleisten.

Das DTC-System wird voraussichtlich aus 256 ATCA-Karten (Advanced Telecom Computer Architecture) bestehen, welche von CMS entwickelt werden. Eine DTC-Karte wird die Trigger- und Rohdaten von bis zu 72 Sensormodulen extrahieren. Die Rohdaten werden an das Datenakquisitionssystem (DAQ) weitergeleitet und die Triggerdaten werden für den Spurtrigger vorbereitet und an diesen gesendet. Darüber hinaus hat das DTC-System auch die Aufgabe, die Sensormodulen zu konfigurieren und zu kalibrieren.

Die optische Anbindung an den Spurtrigger soll insgesamt eine Datenbandbreite von 600 Gb/s pro DTC-Karte zur Verfügung stellen. Dabei ist die genaue Anzahl der Lichtwellenleiter beziehungsweise die Datenrate pro Leiter nicht spezifiziert. Die gesamte Datenbandbreite sollte mehr als ausreichend sein, um die Stubs an den Spurtrigger zu übertragen. Bevor wir uns näher mit dem Spurtrigger beschäftigen, wenden wir uns nun erst der Spurrekonstruktion und der Digitalelektronik zu.

Kapitel 5

Spurrekonstruktion

In diesem Kapitel erarbeiten wir uns die Algorithmen, mit denen der Spurtrigger Spuren rekonstruieren soll. Dazu leiten wir uns die Bewegungsgleichungen der Teilchen her, die wir rekonstruieren möchten. Da die Anforderungen an den Spurtrigger aufgrund der kurzen Laufzeit von bis zu 4 µs und der großen zu verarbeitenden Datenrate von $\mathcal{O}(100\,\mathrm{Tb/s})$ sehr herausfordernd sind, achten wir bei der Herleitung der Bewegungsgleichung auf größtmögliche Einfachheit. Um sicherzustellen, dass die Spuren trotz Vereinfachungen gut genug beschreiben werden, betrachten wir die Spuren, die wir rekonstruieren möchten und analysieren diese mithilfe von Simulationen.

Die Spurrekonstruktion teilen wir in zwei Schritte auf, die Spursuche, welche grobe Spurkandidaten mit den dazugehörenden Stubs bildet, und den Spurfit, der die Stubs der Spurkandidaten bereinigt und die genauen Spurparameter bestimmt. Auch bei diesen Algorithmen achten wir auf größtmögliche Einfachheit.

5.1 Bewegungsgleichungen geladener Teilchen im homogenen Magnetfeld

Betrachten wir ein homogenes Magnetfeld mit einer Flussdichte $\vec{B}$. Aufgrund der Symmetrie bieten sich zylindrische Koordinaten (r, ϕ, z) an. Die Basisvektoren sind gegeben durch

$$\hat{r}(\phi) = \begin{pmatrix} \cos\phi \\ \sin\phi \\ 0 \end{pmatrix}, \quad \hat{\phi}(\phi) = \begin{pmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix}, \quad \hat{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{5.1}$$

Wir legen unser Koordinatensystem so, dass $\vec{B}$ in $\hat{z}$-Richtung zeigt

$$\vec{B} = B\hat{z}. \tag{5.2}$$

Während sich ein geladenes Teilchen in die $\hat{z}$-Richtung gleichförmig bewegt, wird die Teilchenbahn durch das Magnetfeld in der r-ϕ-Ebene gekrümmt. Daher beschreiben wir zunächst die Bewegung in der r-ϕ Ebene und erweitern dann unser Bild auf den dreidimensionalen Raum. Der Krümmungsradius R ist proportional zum Transversalimpuls p_{T} des Teilchens und antiproportional zu dessen elektrische Ladung q, sowie der magnetischen Flussdichte, siehe z.B. [12]

$$R = -\frac{10}{3} \cdot \frac{p_{\mathrm{T}}}{qB} \left[\frac{\mathrm{meT}}{\mathrm{GeV}/c} \right]. \tag{5.3}$$

Das Vorzeichen des Radius gibt die Umlaufrichtung der Flugbahn an. Um die simpelsten Bewegungsgleichungen zu erhalten, führen wir insgesamt drei Vereinfachungen ein.

Erste Annahme: Die Wechselwirkungen der Teilchen mit dem Detektormaterial und dem Strahlrohr, also Vielfachstreuung und Bremsstrahlung, vor allem die damit einhergehenden Ablenkungen der Teilchenbahn, seien vernachlässigbar. Bei Elektronen, den leichtesten geladenen Teilchen, sowie Teilchen mit kleinen Transversalimpulsen, ist diese Annahme kritisch. Es ist von vornherein davon auszugehen, dass die Wahrscheinlichkeit Elektronen

erfolgreich zu rekonstruieren kleiner ist, als die von anderen Teilchenarten. Mit der Vereinfachung ergeben sich allerdings einfache Kreise als Trajektorie $\vec{r}$ und es gilt die Kreisgleichung

$$R^2 = \left(\vec{r} - \vec{M}\right)^2 , \tag{5.4}$$

wobei $\vec{M}$ zum Kreismittelpunkt zeigt.

Zweite Annahme: Der Impaktparameter d_0 sei klein. Der Impaktparameter ist gegeben durch den minimalen Abstand in der r-ϕ-Ebene vom Ursprung, also der Strahlachse, zur vervollständigten Kreisbahn des Teilchens und ist nicht mit dem Abstand zum Vertex zu verwechseln. Zu den Teilchen mit kleinem d_0 gehören also auch die, deren Ursprung in der primären Wechselwirkungszone liegen, welche für die erste Triggerstufe besonders interessant sind. Dadurch werden wir aber nicht alle Teilchenzerfälle von relativ langlebigen Teilchen, wie beispielsweise B-Mesonen, finden. Diese Beschränkung dient nicht nur der Vereinfachung des Problems, es ist schlicht nicht möglich ohne Pixeldetektor, welcher der Wechselwirkungszone am nächsten ist und die beste Ortsauflösung besitzt, den genauen Zerfallsort solcher Teilchen effizient zu rekonstruieren. Durch diese Beschränkung werden die Trajektorien weiter zu Kreisen durch den Ursprung angenähert, welche sich durch zwei Parameter beschreiben lassen, zum Beispiel dem Azimutwinkel der Trajektorie am Ursprung ϕ_0 und dem Radius des Kreises R. Außerdem wird $\vec{M}$ eingeschränkt

$$\vec{M} = R\,\hat{\phi}\left(\phi_0\right) \tag{5.5}$$

und die Kreisgleichung 5.4 entwickelt sich zu

$$R^2 = r^2 - 2\,r R\,\hat{r}\left(\phi\right)\cdot\hat{\phi}\left(\phi_0\right) + R^2 . \tag{5.6}$$

Somit erhalten wir folgende Bewegungsgleichung für die r-ϕ-Ebene

$$r\left(\phi\right) = 2R\sin\left(\phi - \phi_0\right) . \tag{5.7}$$

Betrachten wir nun die r-z-Ebene. Wie schon erwähnt bewegt sich das Teilchen in $\hat{z}$-Richtung gleichförmig.

$$z(t) = z_0 + s(t) \cdot \cot\theta \,, \tag{5.8}$$

wobei z_0 der Schnittpunkt der extrapolierten Trajektorie mit der z-Achse ist, s ist die zurückgelegte Wegstrecke in der r-ϕ-Ebene und θ ist der Polarwinkel, mit dem das Teilchen produziert wurde. Als Parametrisierung wurde der Öffnungswinkel t der Kreisbahn gewählt. Abb. 5.1 veranschaulicht diesen Parameter und hilft die beiden folgenden Zusammenhänge leicht zu erkennen

$$s(t) = Rt \overset{5.8}{\Rightarrow} t = \frac{z - z_0}{R\cot\theta} \,, \tag{5.9}$$

$$\sin\frac{t}{2} = \frac{r}{2R} \,. \tag{5.10}$$

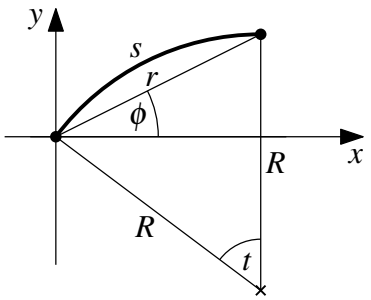

Abbildung 5.1 Skizze der perfekten Flugbahn eines Teilchens mit $d_0 = 0$ in der r-ϕ-Ebene. Eingezeichnet sind der Krümmungsradius R, die zurückgelegte Strecke s, die Ortskoordinaten r und ϕ, sowie der Öffnungswinkel t der Kreisbahn.

Daraus ergibt sich die folgende Bewegungsgleichung für die r-z-Ebene

$$r(z) = 2R\sin\left(\frac{z - z_0}{2R\cot\theta}\right) \,. \tag{5.11}$$

Dritte Annahme: Der Transversalimpuls sei groß. Daher entwickeln wir die beiden Bewegungsgleichungen für große Transversalimpulse ($R \to \infty$ und $\phi \to \phi_0$) bis zur ersten Ordnung:

$$r\left(\phi\right) = 2R \cdot \left(\phi - \phi_0\right) , \tag{5.12}$$

$$r\left(z\right) = \tan\theta \cdot \left(z - z_0\right) . \tag{5.13}$$

Als Resultat erhalten wir zwei Geradengleichungen, um die Trajektorie im dreidimensionalen Raum zu beschreiben. Da die Parameter der Geraden, die Steigung und der Y-Achsenabschnitt, nicht beschränkt sind, bieten sich deren Umkehrfunktionen an:

$$\phi\left(r\right) = \phi_0 + r \cdot \frac{1}{2R} , \tag{5.14}$$

$$z\left(r\right) = z_0 + r \cdot \cot\theta . \tag{5.15}$$

Dies sind unsere vorläufigen Bewegungsgleichungen mit den vier Spurparametern S

$$S = \left(\frac{1}{2R},\, \phi_0,\, \cot\theta,\, z_0\right) . \tag{5.16}$$

Für CMS ergibt sich aus Gl. 5.3 und der magnetischen Flussdichte des Solenoidmagneten von 3,8 T folgender Zusammenhang

$$\frac{1}{2R} = \frac{-1}{1{,}14} \cdot \frac{q}{p_{\mathrm{T}}} \left[\frac{\mathrm{GeV}/c}{\mathrm{m\,e}}\right] . \tag{5.17}$$

Darüber hinaus lässt sich Gl. 3.1 umformen zu

$$\cot\theta = \sinh\eta . \tag{5.18}$$

Natürlich stellt sich nun die Frage, ob diese Annahmen und Vereinfachungen anwendbar sind, um die Bewegung der Teilchen zu beschreiben, welche man mit dem Spurtrigger rekonstruieren möchte. Damit eine Spur überhaupt rekonstruierbar ist, muss es Stubs geben, welche auf der Flugbahn liegen. Der Spurdetektor selbst oder besser gesagt die Logik, welche die Stubs bildet, muss allerdings die ersten beiden Annahmen voraussetzen, da nur die lokale Krümmungsinformation mit zwei Raumpunkten vorliegt. Daher ist davon auszugehen, dass

Teilchen, auf die die ersten beiden Annahmen nicht zutreffen, aufgrund von unzureichenden Triggerdaten, nicht rekonstruierbar sind.

Allerdings muss der Spurdetektor die Flugbahnen nicht für große Transversalimpulse annähern. Der minimale Transversalimpuls wird aber durch die erforderliche Datenreduktion auf 2 GeV bis 3 GeV begrenzt. Im Folgenden gehen wir von einer Begrenzung auf 3 GeV aus. Um uns zu vergewissern, dass die Entwicklung der Bewegungsgleichungen für großer Transversalimpulse bis zur ersten Ordnung ausreichend ist, betrachten wir zunächst die korrekten und die genäherten Trajektorien von zwei Teilchen im äußeren Spurdetektor, auf die die ersten beiden Annahmen zutreffen. Die Spurparameter der beiden Teilchen sind in Tab. 5.1 festgehalten, die Trajektorien in der r-z-Ebene sind in Abb. 5.2 und in der r-ϕ-Ebene in Abb. 5.3 dargestellt.

Tabelle 5.1 Spurparameter der beiden Testtrajektorien A und B.

Teilchen	p_T	q	ϕ_0	η	z_0
A	$1\,\mathrm{GeV}/c$	$+e$	0	0,6	0
B	$3\,\mathrm{GeV}/c$	$-e$	$\frac{\pi}{4}$	1,7	7 cm

Wie man einerseits erkenn kann, unterscheidet sich die Trajektorie eines Teilchens mit einem Transversalimpuls von 1 GeV stark von der Trajektorie eines Teilchens mit 3 GeV. Andererseits ist die Abweichung zwischen der korrekten und der genäherten Flugbahn für das 3 GeV Teilchen, vor allem im Inneren des Detektors, recht klein. Um zu entscheiden, ob diese Abweichung vernachlässigbar ist, betrachten wir als nächstes die zu erwartende Detektorauflösung der Stubs. Dazu bedienen wir uns physikalischer Simulationen der Ereignisse und der resultierenden Detektorantwort.

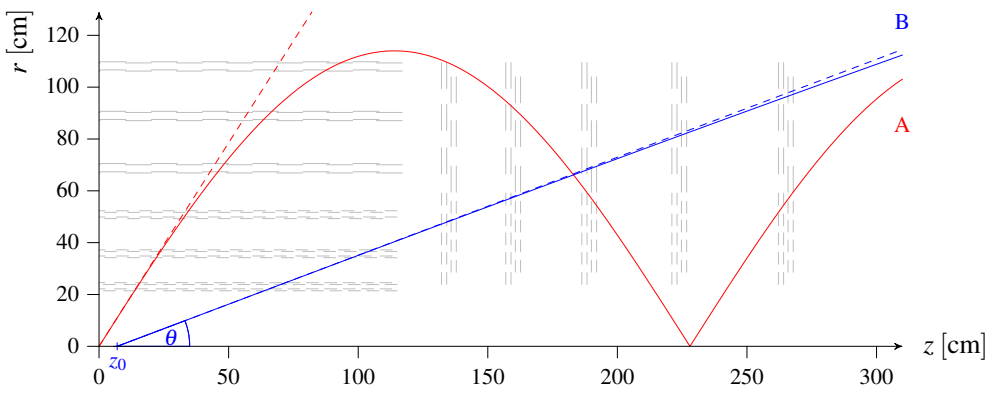

Abbildung 5.2 Skizze einer Hälfte des äußeren Spurdetektors in der r-z-Ebene mit dem Kollisionspunkt in der unteren linken Ecke. Eingezeichnet sind die korrekten und angenäherten (gestrichelt) Flugbahnen der beiden Testtrajektorien A (rot) und B (blau) aus Tab. 5.1, sowie die Spurparameter z_0 und θ von B.

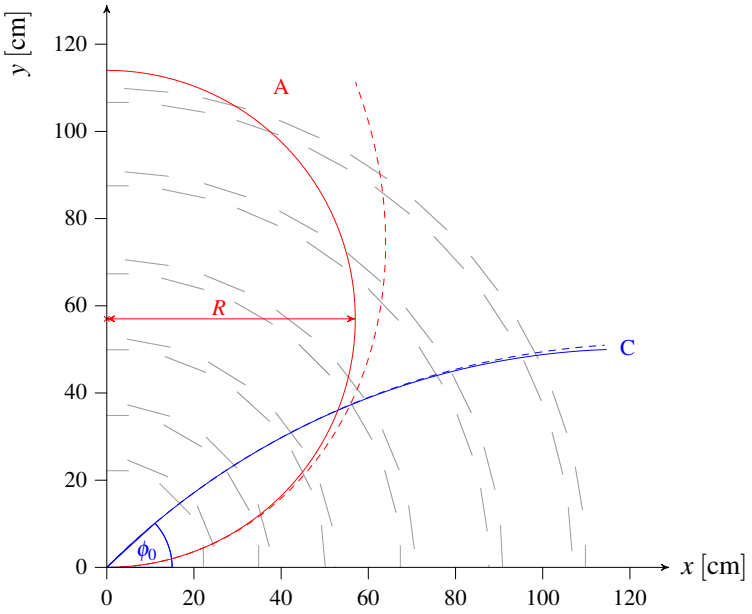

Abbildung 5.3 Skizze eines Viertels des äußeren Spurdetektors in der r-ϕ-Ebene mit dem Kollisionspunkt in der unteren linken Ecke. Eingezeichnet sind die korrekten und angenäherten (gestrichelt) Flugbahnen der beiden Testtrajektorien A (rot) und B (blau) aus Tab. 5.1, sowie die Spurparameter R von A und ϕ_0 von B.

5.2 Ereignis- und Detektorsimulation

Für [17] wurden von CMS Monte-Carlo-Simulationen diverser Ereignisse des HL-LHC und der Simulation der Detektorantwort des CMS II Detektors erstellt, allerdings für das inzwischen überholte flache Layout des äußeren Spurdetektors. Für die Simulation wurde die offizielle CMS-Software CMSSW verwendet, welche auch für die Simulation, sowie die Ereignisrekonstruktion und die Analyse der Messdaten des gegenwärtigen Detektors benutzt wird. Die Software verwendet Geant4 [21], um die Effekte der Energiedeposition, Vielfachstreuung und Schauerbildung in den aktiven und inaktiven Detektormaterialien zu simulieren. Dabei wird auch die elektrische Detektorantwort, mitsamt elektronischem Rauschen, erzeugt.

Die Simulation bildet weitestgehend Stubs nach dem gegenwärtigen Verständnis der zukünftigen Ausleseelektronik. Die Transversalimpulsschwelle für die Stubgenerierung in den Sensormodulen beträgt 3 GeV. Aufgrund des flachen Layouts des äußeren Spurdetektors werden in der Simulation aber auch Stubs gebildet, deren gemessenen Teilchendurchgänge nicht auf derselben Sensorhälfte liegen. Darüber hinaus fehlt in der Simulation das Sammeln der Stubs über acht Kollisionen, um die Auslesedatenrate zu glätten und das Verlieren von Stubs aufgrund der limitierte Bandbreite der Ausleseelektronik.

Als Ereignissignatur für unsere Studien haben wir die Top-Quark-Antiquark-Produktion ($t\bar{t}$) mit zusätzlichen 200 Überlappereignissen gewählt. Wie wir in Abb. 5.2 gesehen haben, können elektrisch geladene Teilchen mit einem Transversalimpuls unterhalb von 1 GeV aufgrund ihrer schraubenförmigen Bewegung lange Strecken im Detektor zurücklegen. Zusammen mit dem kurzen zeitlichen Abstand zwischen zwei Kollisionen von 25 ns können sich Teilchen aus den vorangegangenen Kollisionen immer noch im Detektor befinden und Signale verursachen. Diese Form von überlappenden Ereignissen wird „out-of-time Pile-Up“ genannt und wurde auch simuliert.

Als Resultat erhalten wir für jede Kollision eine Liste von Teilchen mit Produktionsparametern, also den Spurparametern am Vertex und eine Liste von Stubs, sowie die Information, welcher Stub von welchen Teilchen stammt. Tab. 5.2 zeigt unter anderem die Teilchenselektion für die Studien der Qualität des Spurtriggers, auf die sich CMS geeinigt hat.

Tabelle 5.2 Definition diverser Stub- und Teilchenkategorien. Der mittlere Bereich der linken Spalte zeigt die charakterisierenden Parameter, wobei v_{xy} den Abstand des Vertices zur Strahlachse und v_z den Abstand zum Ursprung in Strahlrichtung beschreibt. Die Anzahl der Detektorlagen, in denen das Teilchen mindestens einen Stub produziert hat, ist durch n gegeben. Die letzten beiden Zeilen zeigen pro Ereignis die resultierende mittlere Anzahl an Teilchen und die mittlere Anzahl der Stubs, welche von diesen Teilchen stammen, wobei Stubs nicht doppelt gezählt werden. Es wurden 2 000 $t\bar{t}$-Ereignisse mit jeweils 200 zusätzlichen Überlappereignissen betrachtet.

	Gesamtzahl Stubs	Relevante Stubs	Relevante Teilchen	Davon Rekonstruierbar	Selektion für Qualitätsstudien
$p_T >$	$\sim$3 GeV	3 GeV	3 GeV	3 GeV	3 GeV
$\lvert\eta\rvert <$	–	2,4	2,4	2,4	2,4
$v_{xy} <$	–	–	1 cm	1 cm	1 cm
$\lvert v_z\rvert <$	–	–	30 cm	30 cm	30 cm
$n \geq$	–	–	0	**4**	4
nur $t\bar{t}$	–	–	nein	nein	**ja**
# Teilchen	–	–	424	58	18
# Stubs	19 337	15 041	452	440	135

Auffällig ist, dass von den 19 337 Stubs, die der Detektor liefert, nur 15 041 für den Spurtrigger relevant sind. Das liegt einmal daran, dass, wie schon erwähnt, Teilchen aus der primären Wechselwirkungszone im Bereich $|\eta| > 2{,}4$ nicht genügen Detektorlagen durchqueren, um sie rekonstruieren zu können. Daher sind die Stubs aus diesem Bereich für die Spurrekonstruktion nicht nützlich. Allerdings müsste der Wert von 2,4 ein wenig korrigiert werden, um der Unsicherheit des Kollisionsortes in $\hat{z}$-Richtung gerecht zu werden.

Eine zweite Ursache für die nicht relevanten Stubs ist detektorbedingt. Die Abschätzung des Transversalimpulses eines Teilchens in einem Sensormodul beruht auf der gemessenen Positionsdifferenz des Teilchendurchgangs in den beiden Siliziumsensoren relativ zu einem programmierbaren Versatz. Dieser Versatz kann auf dem 2S-Modul nur für Blöcke von 128

benachbarten Streifen eingestellt werden. In der Triggerelektronik kann dieser Versatz für alle Positionen korrekt gesetzt und somit der Transversalimpuls genauer abgeschätzt werden. In unserer Simulation liegen drei Viertel der nicht relevanten Stubs im Bereich $|\eta| > 2{,}4$ und das verbleibende Viertel weist einen genauer abgeschätzten Transversalimpuls unter 3 GeV auf.

Die Selektionskriterien für die rekonstruierbaren Teilchen sind etwas ungeschickt gewählt. Die wichtigste Eigenschaft, um die Rekonstruktion zu ermöglichen, ist die Anzahl der Detektorlagen, in denen das Teilchen mindestens einen Stub produziert hat. Die Beschränkung der Pseudorapidität führt aber dazu, dass die Teilchen, die trotz großem η, aufgrund von $z_0 \neq 0$, genügend Detektorlagen durchqueren und somit rekonstruiert werden könnten, als nicht rekonstruierbar klassifiziert werden. Die Einschränkung der Vertexposition hat denselben Effekt. Vor allem die Beschränkung von v_{xy} ist irritierend, da die Logik der Stubbildung sensitiv auf d_0 ist und unabhängig von v_{xy} ist, solange sich der Vertex vor dem Sensormodul befindet und nicht dahinter.

Wie schon erwähnt werden mindestens zwei PS-Modullagen benötigt, um die Spurparameter der r-z-Ebene akkurat zu bestimmen. Daher wäre es sinnvoll, mindestens zwei assoziierte PS-Modullagen zu fordern. Nichtsdestotrotz verwenden wir dieselbe Teilchenselektion, um die Konsistenz zu [17] zu wahren.

Weiter auffällig ist, dass von den 15 041 Stubs, die für den Spurtrigger relevant sind, nur 452 Stubs von relevanten Teilchen stammt. Durch elektrisches Rauschen in der Ausleseelektronik der Siliziumsensoren, können fälschlicherweise Teilchendurchgänge gemessenen werden, welche fortan auch Störtreffer genannt werden. Diese Stubs könnten durch zwei Störtreffer in beiden Siliziumsensoren des p_{T}-Moduls, oder aufgrund einer falschen Abschätzung des Transversalimpulses gebildet worden sein.

Eine Quelle für die falsche Abschätzung ist gegeben durch das Bilden eines Stubs aus einem gemessenen Teilchendurchgang in einem Sensor und einem Störtreffer oder gemessenen Teilchendurchgang eines anderen Teilchens im zweiten Sensor. Ist der Sensor mit beiden Treffern der Seeding-Sensor, können zwei Stubs gebildet werden. Davon ist einer korrekt und der andere hat einen kleineren oder größeren abgeschätzten Transversalimpuls. Ist der Sensor mit dem einzelnen Treffer der Seeding-Sensor, so wird ein Stub mit einem zu großen abgeschätzten Transversalimpuls gebildet, da die Stubbildelogik sich so verhält.

Eine zweite Quelle für die falsche Abschätzung ist durch einen aufgrund von Vielfachstreuung vergrößerten Impaktparameter gegeben, da die Abschätzung des Transversalimpulses auf der Annahme eines kleinen Impaktparameters beruht.

Weiter auffällig ist, dass von den 424 relevanten Teilchen nur 58 ausreichend Stubs erzeugen, um eine Grundlage für die Spurrekonstruktion zu bilden. Die charakterisierenden Parameter, welche für die Selektion verwendet werden, sind die Produktionsparameter der Teilchen. Daher werden die Teilchen, die zwischen Produktionsort und Messort so viel Transversalimpuls verlieren, dass ihr Transversalimpuls unterhalb der Schwelle liegt oder so viel Impaktparameter gewinnen, sodass ihr abgeschätzter Transversalimpuls unterhalb der Schwelle liegt, möglicherweise nicht in genügend Detektorlagen Stubs erzeugen, um rekonstruiert werden zu können. Diese detektorbedingte Ineffizienz ist in den Qualitätsstudien des Spurtriggers nicht zu sehen und wird von CMS hoffentlich nicht übersehen.

Kommen wir nun zur Validierung der hergeleiteten Bewegungsgleichungen und der Abschätzung der Detektorauflösung. Abb. 5.4 zeigt für jede Detektorlage den mittleren Abstand der assoziierten Stubs eines Teilchens aus der Selektion für Qualitätsstudien zur perfekten Teilchenspur, beruhend auf den Produktionsparametern.

Der Abstand wurde in der r-ϕ-Ebene und in der r-z-Ebene getrennt bestimmt. In der r-ϕ-Ebene wurde der Abstand in $\hat{\phi}$-Richtung gemessen und in der r-z-Ebene wurde der Abstand in $\hat{z}$-Richtung für Stubs in den Mantellagen, beziehungsweise in $\hat{r}$-Richtung für Stubs in den Endkappen, gemessen. Die Teilchenposition beruht auf den nicht genäherten Gleichungen 5.7 und 5.11. Die Abstände sind gegeben durch

$$\Delta_{r\text{-}\phi} = \left| \phi_0 - \arcsin\frac{r}{2R} - \phi \right| \cdot r \,, \tag{5.19}$$

$$\Delta_{r\text{-}z}^{\text{Mantel}} = \left| z_0 + 2R\cot\theta \cdot \arcsin\frac{r}{2R} - z \right| \,, \tag{5.20}$$

$$\Delta_{r\text{-}z}^{\text{Endkappe}} = \Delta_{r\text{-}z}^{\text{Mantel}} / \cot\theta \,, \tag{5.21}$$

wobei die Stubposition durch r, ϕ und z gegeben ist. In Abb. 5.4 sind sehr große Abstände zu sehen, besonders für die äußerste PS-Endkappe, in der manchmal Stubs zu zentralen und mehreren Metern entfernten Teilchen zugeordnet werden. Dies ist höchst wahrscheinlich ein Programmierfehler in der Detektorsimulation.

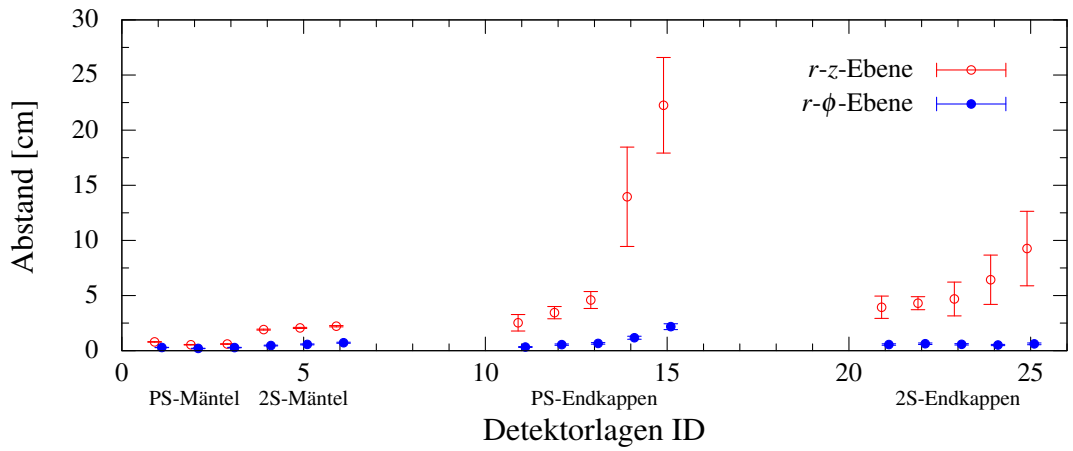

Abbildung 5.4 Mittlere absolute Abweichung von den assoziierten Stubs eines Teilchens zur perfekten Teilchenspur für jede Detektorlage in der r-ϕ-Ebene und in der r-z-Ebene. Die Detektorlagen werden von innen nach außen beziffert. Der Bereich 1-6 wird für die Mäntel, 11-15 wird für die PS-Module der Endkappen und 21-25 wird für die 2S-Module der Endkappen verwendet. Es wird nicht zwischen den linken und rechten Endkappen unterschieden.

Um dennoch ein Gefühl für die Detektorauflösung zu bekommen und zu entscheiden, ob eine Näherung der Bewegungsgleichungen bis zur ersten Ordnung gerechtfertigt ist, bereinigen wir die Zuordnung, welche Stubs von welchen Teilchen stammt. Dazu entfernen wir die Stubs, welche in einer der beiden Ebenen mehr als 10 cm von der perfekten Flugbahn entfernt sind.

Die Teilchen, die dadurch keine Stubs aus vier oder mehr Detektorlagen aufweisen, werden nicht mehr berücksichtigt, was auf 6 ‰ der für Qualitätsstudien selektierten Teilchen zutrifft.

Abb. 5.5 und 5.6 zeigen die neuen Resultate nach der Bereinigung für die korrekten und für die genäherten Bewegungsgleichungen. In der r-z-Ebene können wir keinen Unterschied zwischen der korrekten und der genäherten Rechnung feststellen. Die Auflösung in den 2S-Modulen und die Auflösung in den PS-Modulen sind relativ nahe an der nativen Auflösung, gegeben durch den Streifenabstand geteilt durch $\sqrt{12}$ (siehe Gl. A.3).

In den 2S-Modulen ist die Auflösung etwas besser aufgrund der hermetischen und überlappenden Anordnung der Module. In den PS-Modulen ist sie deutlich schlechter.

Das ist nicht verwunderlich, da die native Auflösung eine Abschätzung ist und keinerlei Detektoreffekte berücksichtigt. Die Detektoreffekte können von den PS-Modulen aufgelöst werden, während die 2S-Module zu grob sind.

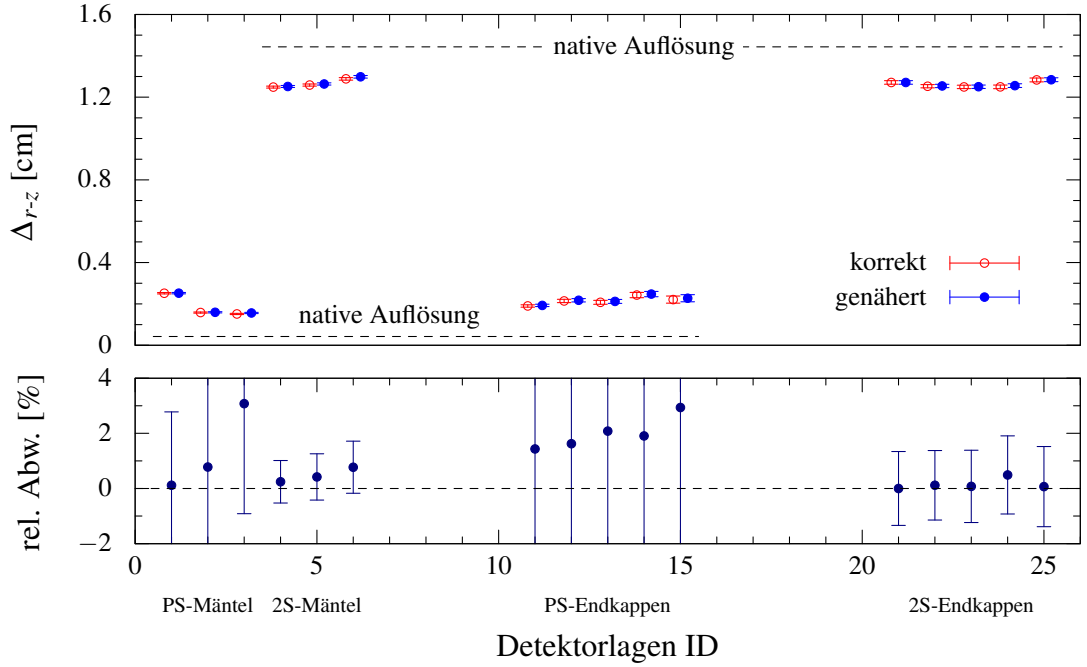

Abbildung 5.5 Mittlere absolute Abweichung von den bereinigten assoziierten Stubs eines Teilchens zur perfekten Teilchenspur für jede Detektorlage in der r-z-Ebene. Die Detektorlagen sind wie in Abb. 5.4 nummeriert. Die Abweichung, beruhend auf der korrekten Bewegungsgleichung, ist in Rot und beruhend auf der genäherten Bewegungsgleichung in Blau eingezeichnet. Zusätzlich sind die nativen Auflösungen eingetragen. Der untere Abschnitt zeigt die relative Abweichung zwischen den genäherten zu den korrekten Werten.

Auch in der r-ϕ-Ebene können wir keinen Unterschied zwischen den korrekten und den genäherten Bewegungsgleichungen feststellen. Allerdings ist in dieser Ebene nicht die Granularität des Detektors zu grob, vielmehr ist die Beeinflussung der Flugbahn durch Vielfachstreuung und nicht verschwindendem Impaktparameter der begrenzende Faktor.

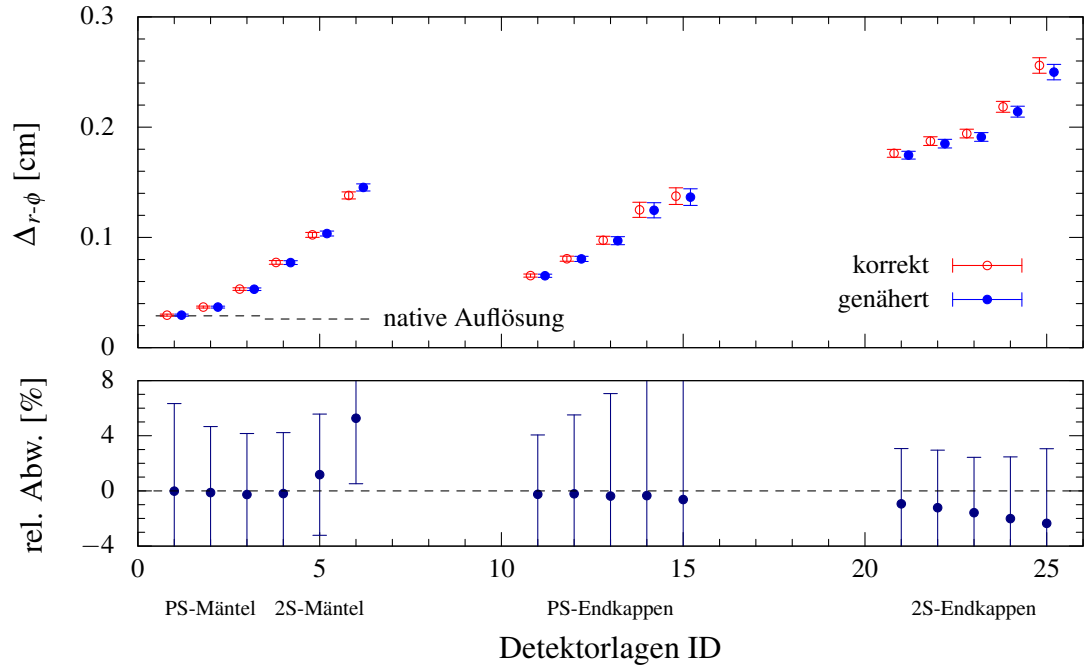

Abbildung 5.6 Wie Abb. 5.5, aber in der r-ϕ-Ebene.

Zum Vergleich wurden in Abb. 5.7 nur Myonen mit einem Transversalimpuls über 20 GeV selektiert. Hier ist die gemessene Auflösung mit der nativen Auflösung verträglich.

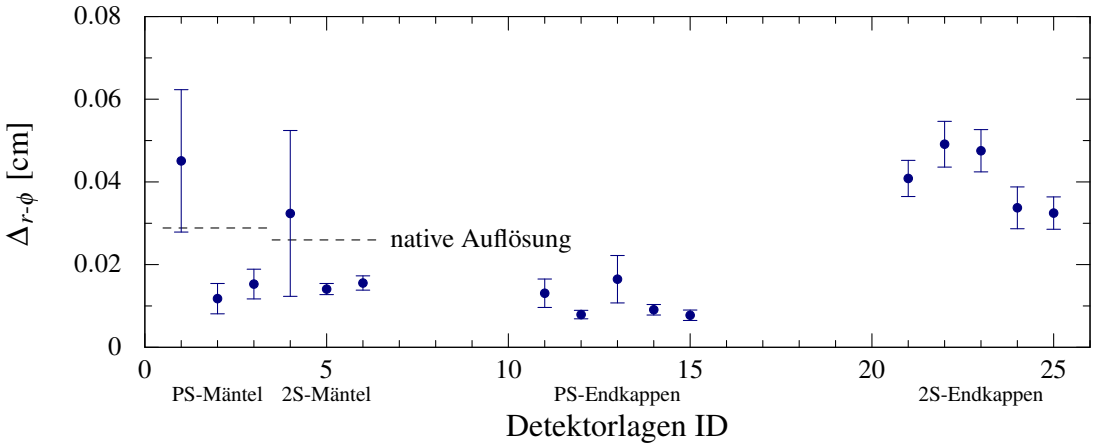

Abbildung 5.7 Wie Abb. 5.5, aber nur für Myonen mit $p_{\mathrm{T}} > 20\,\mathrm{GeV}$.

Zusammenfassend stellen wir fest, dass die Flugbahnen der Teilchen, welche Stubs in mehreren Detektorlagen erzeugen, gut durch unsere Bewegungsgleichungen beschrieben werden.

5.3 Spursuche mit der Hough-Transformation

Die Hough-Transformation ist eine weit verbreitete Methode, um geometrische Merkmale in digitalen Bildern zu entdecken [22]. Wir verwenden die Methode, um die Stubs eines Ereignisses auszuwählen, deren Positionen mit der Flugbahn eines geladenen Teilchens aus der primären Wechselwirkungszone mit einem Transversalimpuls über 3 GeV kompatibel sind. Um das Problem zu vereinfachen, beschränken wir uns auf die r-ϕ-Ebene, in der der Detektor die bessere Auflösung besitzt.

Betrachten wir zunächst einen einzelnen gemessenen Stub mit den Koordinaten r und ϕ und stellen Gl. 5.14 um. Anstatt mit gegebenen Spurparametern $1/2R$ und ϕ_0 die Raumkoordinaten r und ϕ miteinander zu korrelieren, gehen wir umgekehrt vor. Somit erhalten wir

$$\phi_0\left(\frac{1}{2R}\right) \quad = \quad \phi - r \cdot \frac{1}{2R}\,. \tag{5.22}$$

Dieses Resultat zeigt, dass die Raumkoordinaten r, ϕ der Stubs zu Geraden im Parameterraum $\phi_0, 1/2R$ transformiert werden können.

Erweitern wir das Bild eines einzelnen Stubs auf mehrere Stubs von einem Teilchen aus dem Ursprung. Da der Krümmungsradius eines geladenen Teilchens mit einem Transversalimpuls über 3 GeV größer ist als der äußere Radius des Spurdetektors, wird das Teilchen den gesamten Spurdetektor durchqueren und in allen Detektorlagen einen Stub erzeugen können.

Zeichnen wir für jeden Stub die transformierte Gerade in einen Graphen ein, so sehen wir, dass diese sich in einem Punkt schneiden. Die Koordinaten des Schnittpunktes beschreiben einen Kreis, welcher mit allen Stubs und dem Ursprung vereinbar ist und stellen somit die gesuchten Spurparameter in der r-ϕ-Ebene dar.

Die Spursuche funktioniert also wie folgt: Man berechne für sämtliche Stubs eines Ereignisses die entsprechenden Geraden im Parameterraum, zeichne diese in einen Graphen ein und suche dann nach Häufungspunkten. Diese Häufungspunkte stellen Spurkandidaten dar, bestehend aus den Stubs, welche zur Häufung führen, und den Koordinaten des Häufungspunktes.

Da nur positive Radien existieren, würden auch sämtliche Geraden im Graphen eine positive Steigung aufweisen. Dies erschwert die Lokalisierung des Häufungspunktes. Daher nehmen wir eine Translation von r um T vor

$$r \;\rightarrow\; r_T \;=\; r - T\,. \tag{5.23}$$

Dadurch transformiert sich Gl. 5.22 zu

$$\phi_T\left(\frac{1}{2R}\right) \;=\; \phi - r_T \cdot \frac{1}{2R}\,, \tag{5.24}$$

wobei ϕ_T dem Azimutalwinkel des Raumpunktes der Trajektorie bei dem Radius T entspricht. Durch das Legen von T zwischen die innerste und äußerste Detektorlage können nun die Steigungen auch negative Werte annehmen, wie in Abb. 5.8 veranschaulicht wird.

Die transformierten Geraden von Stubs einer nicht perfekten Teilchenbahn werden sich nicht genau schneiden. Deshalb muss bei der Lokalisierung der Häufungspunkte auf eine gewisse Grobheit geachtet werden, um auch durch Vielfachstreuung oder nicht verschwindendem Impaktparameter leicht beeinflusste Flugbahnen zu finden. Wir bekommen also nicht die bestmöglichen Spurparameter in der r-ϕ-Ebene, und über die Spurparameter der r-z-Ebene haben wir bis jetzt noch keinerlei Information erhalten. Daher werden die endgültigen Spurparameter von gefundenen Spurkandidaten mit anderen Methoden bestimmt.

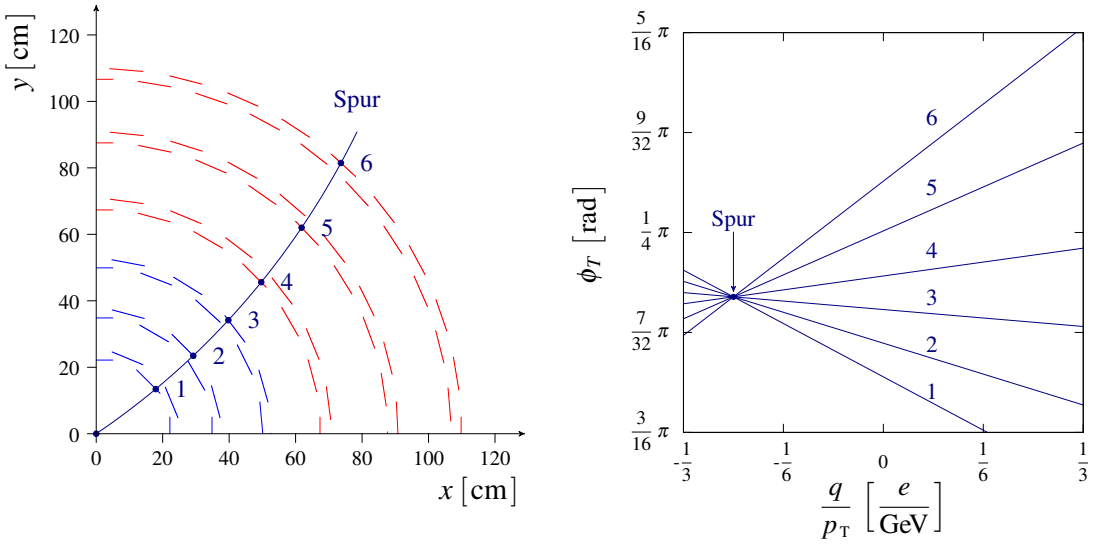

Abbildung 5.8 Illustration der Hough-Transformation. Das linke Bild zeigt die x-y-Ebene des äußeren Spurdetektors im Bereich der Mantellagen. Eingezeichnet sind die Spur eines Teilchens und deren Stubs mit Nummern. Rechts ist die Spurparameterebene zu sehen. Dort sind die transformierten Stubs eingezeichnet.

5.4 Spurfit mit der Linearen Regression

Die Aufgabe des Spurfits besteht in erster Linie in der Berechnung der besten Spurparameter aus den Stubs eines Spurkandidaten. Beginnen wir mit einem Spurkandidaten, welcher ausschließlich Stubs von einem rekonstruierbaren Teilchen aufweist. Jeder Stub stamme von einer anderen Detektorlage i und die Anzahl der Stubs sei $n \geq 4$. Die Bewegungsgleichungen des Teilchens sind linear und unabhängig voneinander in der r-ϕ-Ebene und der r-z-Ebene. Sie haben in beiden Ebenen die Form

$$y(x) \quad = \quad m \cdot x + c \; . \tag{5.25}$$

Um die Parameter m und c zu bestimmen, minimieren wir die Summe der Fehlerquadrate, welche gegeben ist durch

$$\sum_{i=0}^{n-1} r_i^2 \;=\; \sum_{i=0}^{n-1} \left(m \cdot x_i + c - y_i\right)^2 \,. \tag{5.26}$$

Um eine bessere Übersicht zu erhalten, führen wir folgende Kurzschreibweise ein

$$\sum_{i=0}^{n-1} x_i \;=\; \bar{x}\,, \qquad \sum_{i=0}^{n-1} x_i x_i \;=\; \overline{xx}\,, \qquad \sum_{i=0}^{n-1} x_i \cdot \sum_{j=0}^{n-1} x_j \;=\; \bar{x}\cdot\bar{x}\,, \tag{5.27}$$

welche nicht mit dem arithmetischen Mittel zu verwechseln ist. Als Resultat der Minimierung, welche in Abschnitt A.2 zu finden ist, erhalten wir die vier Spurparameter

$$\frac{1}{2R} \;=\; \frac{n\overline{r\phi} - \bar{r}\cdot\bar{\phi}}{n\overline{rr} - \bar{r}\cdot\bar{r}}\,, \tag{5.28}$$

$$\phi_0 \;=\; \frac{\bar{\phi}\cdot\overline{rr} - \bar{r}\cdot\overline{r\phi}}{n\overline{rr} - \bar{r}\cdot\bar{r}}\,, \tag{5.29}$$

$$\cot\theta \;=\; \frac{n\overline{rz} - \bar{r}\cdot\bar{z}}{n\overline{rr} - \bar{r}\cdot\bar{r}}\,, \tag{5.30}$$

$$z_0 \;=\; \frac{\bar{z}\cdot\overline{rr} - \bar{r}\cdot\overline{rz}}{n\overline{rr} - \bar{r}\cdot\bar{r}}\,. \tag{5.31}$$

Dies sind die besten Spurparameter, die wir finden können, unter der Annahme, dass die Flugbahnen perfekt sind und dass die Detektorauflösungseffekte vernachlässigbar sind. Darüber hinaus sind die Formeln, abgesehen von der Division, äußerst simpel. Um den großen Unterschied der Detektorauflösung zwischen den PS- und 2S-Modulen in der r-z-Ebene zu berücksichtigen, verwenden wir in Gl. 5.30 und 5.31 nur die Stubs von PS-Modulen.

Eine weitere Aufgabe des Spurfits besteht im Erkennen und Bereinigen von Stubs, welche dem Spurkandidaten fälschlicherweise zugeordnet wurden und nicht auf der Spur liegen. Diese Stubs werden fortan Störstubs genannt.

Erweitern wir unser Bild, indem wir einen Störstub zum Spurkandidaten hinzufügen, welcher der einzige Stub seiner Detektorlage ist. Je größer die Distanz vom Störstub zur Flugbahn ist,

desto mehr werden die Spurparameter verfälscht. Da die Lineare Regression die Summe der Fehlerquadrate minimiert und die Anzahl der Stubs, welche auf einer Linie liegen überwiegt, wird das Residuum des Störstubs von allen Stubs am größten sein. Daher bietet es sich an, nach der Berechnung der Spurparameter die Residuen zu berechnen, um Ausreißer zu erkennen und zu entfernen. Dabei sollte die gemessene Detektorauflösung aus Abb. 5.5 und 5.6 als Maß dienen.

Wenn wir die Zahl der Störstubs, welche die einzigen Stubs auf ihren Detektorlagen sind, erhöhen, wird es schwieriger, die Störstubs zu erkennen. Kritisch wird es, wenn die Zahl der Detektorlagen mit Störstubs überhand gewinnt. In der r-ϕ-Ebene ist dieser Fall glücklicherweise ausgeschlossen, da rekonstruierbare Teilchen Stubs von mindestens vier und nicht mehr als sieben Detektorlagen aufweisen. Das heißt, der größte Ausreißer wird wahrscheinlich immer noch ein Störstub sein und kann entfernt werden. Allerdings muss nun die Berechnung der Spurparameter und Residuen wiederholt werden, um den nächsten Störstub zu finden.

Betrachten wir nun den Fall, dass sich zwei Stubs auf einer Detektorlage befindet. Es eröffnen sich zwei Lösungsansätze, welche wir fortan die lokale und die globale lineare Regression nennen. Bei der **lokalen** Variante wird der Spurkandidat in zwei Spurkandidaten aufgeteilt, wobei der eine Kandidat alle Stubs des ursprünglichen Kandidaten bis auf einen der zwei Stubs von derselben Detektorlage aufweist und der andere Kandidat alle Stubs bis auf den anderen Stub beinhaltet. Diese beiden Spurkandidaten durchlaufen den Prozess des Entfernens der Ausreißer. Am Ende wird wieder einer der beiden Spurkandidaten entfernt und zwar der, der die größte Unsicherheit aufweist. Als Unsicherheit könnte die Summe der quadrierten gewichteten Residuen geteilt durch die Anzahl der Freiheitsgrade dienen.

Befinden sich im ursprünglichen Kandidaten viele Stubs auf einer Detektorlage oder weisen mehrere Detektorlagen mehr als einen Stub auf, kann die Zahl der temporär erzeugten Spurkandidaten und der damit einhergehende Rechenaufwand schnell explodieren. Allerdings weist die lokale lineare Regression die ultimative Robustheit gegenüber Störstubs auf, falls der Rechenaufwand bewältigt werden kann.

Bei der **globalen** Variante wird für die Berechnung der Spurparameter für jede Detektorlage ein virtueller Stub erzeugt. Die Koordinaten des virtuellen Stubs einer Detektorlage berechnet

sich aus einem robusten Mittelwert der Koordinaten der Stubs, die sich auf dieser Detektorlage befinden, gegeben durch das arithmetische Mittel der jeweiligen Maximal- und Minimalwerten. Für die Berechnung der Residuen verwenden wir allerdings die ursprünglichen Stubs.

Dieser Algorithmus weist gegenüber der lokalen linearen Regression nicht den potentiell explodierenden Rechenaufwand auf, könnte aber bei starker Störung eher versagen. Welcher der beiden Algorithmen oder ob sogar eine Kombination der beiden sinnvoll ist, hängt stark von den Fähigkeiten des verwendeten Rechenwerkes, den Anforderungen des Systems an die Spurrekonstruktion und der Qualität der gefundenen Spurkandidaten ab. Daher betrachten wir als nächstes die Digitalelektronik, auf der wir die Algorithmen umsetzen wollen.

Kapitel 6

Digitalelektronik

Das wichtigste Merkmal eines Triggersystems ist die robuste und zuverlässige Selektion von interessanten Ereignissen. Dies bedingt im Falle des Spurtriggers eine robuste und zuverlässige Spurrekonstruktion, deren Laufzeit auf 4 µs beschränkt ist. Daher ist ein Echtzeitsystem als Rechenwerk ein klarer Favorit, um die Fertigstellung der Spurrekonstruktion innerhalb des vorgegeben Zeitraums zu garantieren. Die zu bewältigende Datenrate von bis zu 150 Tb/s beschränkt die Auswahl an Echtzeitsystemen auf ASICs (application-specific integrated circuits) und FPGAs (field progammable gate arrays).

Ein ASIC ist ein elektronischer Schaltkreis, welcher eine festgelegte Funktion erfüllt. Falls dieser Schaltkreis eine Konfiguration vorsieht, kann die Schaltung entsprechend konfiguriert werden, allerdings kann der Schaltkreis selbst nicht mehr verändert werden. Typischerweise ist daher die Konfigurierbarkeit und Anpassungsfähigkeit von ASICs äußerst beschränkt. Da die genauen Anforderungen an die Spurrekonstruktion nicht klar sind und sich diese auch während des Betriebes durchaus weiterentwickeln könnten, ist dies der Grund für uns auf ASICs nach Möglichkeit zu verzichten. Somit sind FPGAs hochinteressant.

Ein FPGA ist im Gegensatz zum ASIC ein programmierbarer integrierter Schaltkreis. Die grundlegende Struktur ist gegeben durch eine zweidimensionale Matrix von Logikblöcken.

Bei modernen FPGAs können die Logikblöcke jede beliebige 6-stellige binäre Funktion abbilden. Die Verbindungsmöglichkeiten der Logikblöcke sind sehr flexibel. Sowohl die abzubildenden binären Funktionen der Logikblöcke, als auch die Verbindungen der Blöcke sind konfigurierbar. Dadurch können hochkomplexe digitale Schaltungen realisiert und jederzeit verändert werden.

Das Programmieren einer hochkomplexen digitalen Schaltung unterscheidet sich vom Programmieren eines Computerprogramms, da nicht nur zeitliche Abläufe, sondern auch die Schaltungsarchitektur beschrieben werden muss. Wir beschreiben digitale Schaltungen textbasiert mit der Hardwarebeschreibungssprache VHDL (Very High Speed Integrated Circuit Hardware Description Language) [23], welche sowohl maschinen- als auch menschenlesbar ist. Ein Synthesewerkzeug übersetzt die Hardwarebeschreibung in eine Konfigurationsdatei für den FPGA, falls die beschriebene Schaltung realisierbar ist. Um eine realisierbare und effiziente Hardwarebeschreibung für einen FPGA zu erzeugen, ist eine gute Kenntnis des Aufbaus des FPGAs, auf dem die Beschreibung synthetisiert werden soll, unerlässlich.

Daher betrachten wir nun einen FPGA im Detail und zwar den XC7VX690T. Das ist ein Virtex-7 FPGA der Firma Xilinx, auf dem wir auch die Algorithmen realisiert haben. Würde man heute den Spurtrigger bauen, wäre dieser FPGA ein möglicher Kandidat. Der letztendlich verwendete FPGA ist wahrscheinlich eine Weiterentwicklung des Virtex-7, dessen Komponenten sich in ihrer Natur typischerweise kaum ändern, sondern deren Anzahl und Geschwindigkeit sich erhöht.

Der Virtex-7 enthält mehrere unterschiedliche Ressourcen. Davon betrachten wir die Ressourcenarten, die für die Realisierung der Algorithmen und den Aufbau der Architektur des Spurtriggers relevant sind. Dabei handelt es sich um:

- Configurable Logic Block (CLB)
- Block RAM (BRAM)
- Digital Signal Processing (DSP)
- GTH Transceiver

Im Virtex-7 bilden die Ressourcen eine lange Reihe, auch Spalte genannt, sodass jedes Element zwei Nachbarn besitzt, mit dem es typischerweise auch verbunden ist. Der FPGA

setzt sich aus einer Komposition von Spalten unterschiedlichen Typs zusammen, um eine gewisse Homogenität zu erhalten.

6.1 Configurable Logic Block – CLB

Die CLBs [24] sind die fundamentale Ressource des FPGAs und enthalten die schon erwähnten programmierbaren Logikblöcke. Strukturiert ist ein CLB in zwei sogenannte Slices. Die wichtigsten Elemente eines Slices sind vier Lookup-Tabellen (LUTs) und acht Register (FFs). Eine Lookup-Tabelle des Virtex-7 besitzt sechs binäre Ein- und zwei binäre Ausgänge. Sie kann jede beliebige 6-stellige binäre Funktion, zwei unabhängige 3-stellige Funktionen oder zwei 5-stellige Funktionen abbilden, wobei im letzten Fall beide Funktionen dasselbe Argument verwenden müssen.

Realisiert ist die Lookup-Tabelle durch einen Speicherbaustein, einem SRAM (static random-access memory). In der Konfigurationsphase des FPGAs wird die Lookup-Tabelle mit den gewünschten Werten beschrieben. Im Betrieb, nach der Konfiguration, kann die Lookup-Tabelle typischerweise nur gelesen werden. Dazu verwendet der Speicherbaustein das Eingangssignal als Adresse und liest das dort gespeicherte Wort aus, welches nach einer gewissen Zeit ($\sim$50 ps) am Ausgang erscheint. Der Leseprozess ist asynchron, also unabhängig von einem Takt.

Falls ein synchrones Lesen gewünscht ist, können die vorhandenen Register verwendet werden. Die Register können bei einer steigenden Taktflanke je ein Bit speichern, wobei das vorherige Bit stets überschrieben wird. Die Register können durch externe Signale gesteuert werden, sodass das Schreiben ausgesetzt werden kann oder das Register auf eine Konstante gesetzt wird. Alle Register in einem Slice teilen sich die Steuersignale, sodass es sich empfiehlt, möglichst gleich gesteuerte Register zu beschreiben, damit das Syntheseresultat, also das Produkt der Hardwarebeschreibung, möglichst wenige Slices benötigt. Die internen Verbindungen eines Slices sind konfigurierbar und werden in der Konfigurationsphase festgelegt.

In einem Virtex-7 gibt es zwei Arten von Slices, das SLICEL und das SLICEM. Beide besitzen die eben beschriebene Funktionalität. Die Besonderheit des SLICEM besteht darin,

dass der Speicherbaustein der Lookup-Tabelle auch im Betrieb beschrieben werden kann. Diese Slices können auch als vollwertige Speicher konfiguriert werden, wobei auch mehrere Lookup-Tabellen eines Slices kombiniert werden können. Diese Speicher werden „Distributed Memory“ genannt.

Grundsätzlich charakterisieren wir Speicher durch ihre Breite, Tiefe, Anzahl und Art der Zugänge, auch Ports genannt. Bei der Art eines Ports unterscheiden wir zwischen einem Port der nur lesen, nur schreiben oder lesen und schreiben kann. Des Weiteren unterscheiden wir zwischen asynchronem und synchronem Lesen. Beim Schreiben müssen wir diese Unterscheidung nicht treffen, da alle Speicher auf dem Virtex-7 nur synchron schreiben können.

Betrachten wir einen einzelnen Speicher mit einem schreibenden Port. Dieser Speicher kann pro Takt ein Datenwort an eine Position im Speicher schreiben, wobei die Anzahl der Bits des Wortes der Breite des Speichers entspricht. Die Anzahl der Wörter, die der Speicher ablegen kann, ist durch seine Tiefe gegeben, welche stets durch eine Zweierpotenz gegeben ist. Die Position eines gespeicherten Wortes wird Adresse, oder auch Adresswort, genannt. Die Anzahl der Bits des Adresswortes ist durch den Logarithmus zur Basis zwei der Tiefe gegeben. Typische Signale eines schreibenden Portes sind also ein Datenwort, ein Adresswort, ein Takt und ein weiteres Kontrollsignal, oft „write-enable“ genannt, welches angibt, ob man im aktuellem Takt schreiben möchte.

Wie bei einer Lookup-Tabelle ist das Lesen asynchron, wobei das gelesene Wort bei Bedarf in den Registern synchron gespeichert werden kann. Das asynchrone Lesen führt dazu, dass das gleichzeitige Lesen und Schreiben eines Portes reibungslos funktioniert, das alte gespeicherte Wort wird gelesen und mit der nächsten steigenden Taktflanke überschrieben.

Die Ports des Distributed Memory können wie folgt konfiguriert werden:

- Single Port: ein Port für synchrones Schreiben und gleichzeitiges asynchrones Lesen
- Dual Port: ein Single Port und ein Port für asynchrones Lesen
- Simple Dual Port: ein Port für synchrones Schreiben und ein Port für asynchrones Lesen
- Quad Port: ein Single Port und drei Ports für asynchrones Lesen

Tab. 6.1 zeigt, welche Konfigurationen des Distributed Memory eines Slices möglich sind. In unserem 7VX690T FPGA befinden sich 64 750 SLICEL und 43 550 SLICEM, also insgesamt 433 200 Lookup-Tabellen und 866 400 Register. Die maximale Speicherkapazität des Distributed Memory beträgt 10 888 Kb, wobei ein Kb 1024 Bits entspricht.

Tabelle 6.1 Distributed Memory Konfigurationsmöglichkeiten eines SLICEM.

Port	Tiefe	Breite	# LUTs
Single Port	32	1	1
Dual Port	32	1	2
Quad Port	32	2	4
Simple Dual Port	32	6	4
Single Port	64	1	1
Dual Port	64	1	2
Quad Port	64	1	4
Simple Dual Port	64	3	4
Single Port	128	1	2
Dual Port	128	1	4
Single Port	256	1	4

6.2 Block RAM – BRAM

Der Block RAM [25] ist ein statischer Speicher, wie das Distributed Memory. Die Speicherkapazität eines Block RAMs beträgt 36 Kb, wobei jeder BRAM auch als zwei unabhängige 18 Kb Speicher verwendet werden kann.

Die Speicher besitzen zwei Ports, welche als True Dual Port oder als den uns schon bekannten Simple Dual Port konfiguriert werden können. Der True Dual Port besteht aus zwei Single Ports, kann also an zwei Adressen lesen und gleichzeitig schreiben. Daraus ergibt sich das Problem, dass beide Ports an dieselbe Stelle schreiben wollen könnten, was zu nichtdeterministischem Verhalten führt und vermieden werden muss.

Im Gegensatz zum Distributed Memory kann ein BRAM nur synchron, also zur nächsten steigenden Taktflanke, gelesen werden. Zusätzlich befinden sich weitere Register im Block RAM, sodass bei Bedarf das gelesene Wort nochmals synchron gespeichert werden kann. Damit werden für das Lesen zwar zwei Takte benötigt, allerdings kann so eine höhere Taktrate erreicht werden.

Die Tiefen und Breiten der Speicherbausteine sind relativ flexibel und reichen von $32\,\mathrm{K} \times 1\,\mathrm{b}$ bis $1\,\mathrm{K} \times 36\,\mathrm{b}$ beziehungsweise $512 \times 72\,\mathrm{b}$. Im 7VX690T FPGA befinden sich 15 Spalten mit je 100 RAM Blöcken mit kleinen Unregelmäßigkeiten, sodass es insgesamt nur 1 470 Blöcke sind. Die akkumulierte Speicherkapazität beträgt 52 920 Kb.

6.3 Digital Signal Processing – DSP

Das DSP-Slice [26] ist ein dediziertes Rechenwerk, welches Ganzzahlen addieren und multiplizieren kann. Dabei wird das Zweierkomplement (siehe Abschnitt A.3) als binäre Darstellung der Ganzzahlen vorausgesetzt. Abb. 6.1 zeigt ein vereinfachtes Blockschaltbild des DSP-Slices.

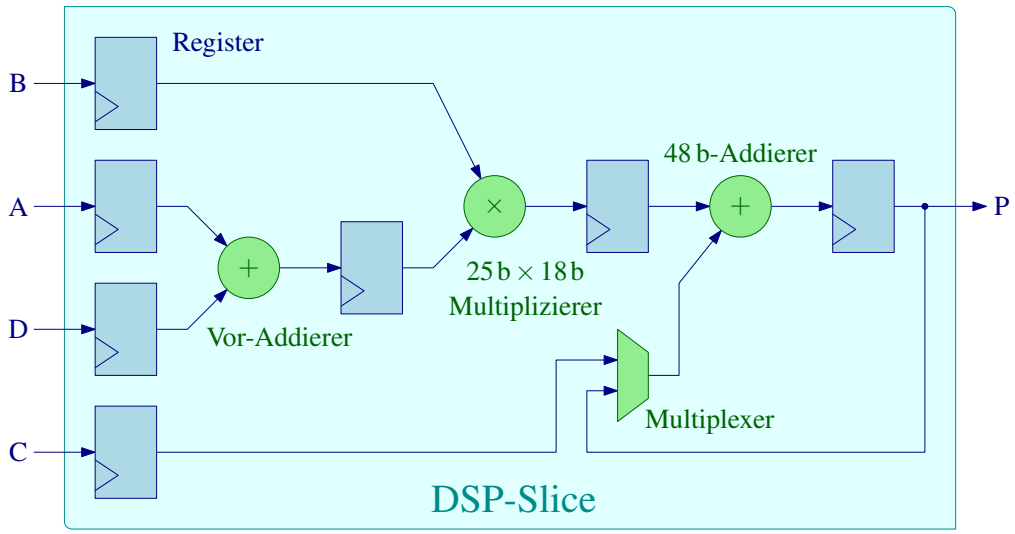

Abbildung 6.1 Blockschaltbild der Basisfunktionalität eines DSP-Slices. Die synchronen Elemente sind in Blau und die asynchronen Elemente sind in Grün dargestellt. Sämtliche Register sind optional. Adaptiert von [26].

Die wichtigsten Komponenten des DSP-Slices sind ein 25 b×18 b-Multiplizierer und ein 48 b-Addierwerk, sowie diverse optionale Register und weitreichende und konfigurierbare Verbindungsmöglichkeiten, sowohl innerhalb eines Slices als auch mit den beiden Nachbarn. Im Prinzip benötigen wir für die Hough Transformation und die Lineare Regression nur die beiden Operationen

$$p = a \cdot b + c \tag{6.1}$$

und

$$p = a \cdot b + p\,, \tag{6.2}$$

welche sich mit einem DSP-Slice effizient abbilden lassen. Es befinden sich 18 Spalten mit 200 Slices, also insgesamt 3 600 DSP-Slices auf dem 7VX690T FPGA.

6.4 Transceiver

Die Hochgeschwindigkeitstransceiver [27] sind serielle Sender und Empfänger, welche vor allem für die FPGA-zu-FPGA-Kommunikation gedacht sind. Diese Komponente verändert sich am stärksten über die FPGA-Generationen. Während die CLBs, BRAMs und DSP-Slices durch kleinere Strukturgrößen kleiner werden, erhöht sich typischerweise die Anzahl dieser Ressourcen in neueren FPGAs. Die Transceiver hingegen werden vor allem schneller.

In unserem Virtex-7 befinden sich zwei Spalten, mit je 40 Hochgeschwindigkeitstransceiver, an den Rändern, welche mit einer maximalen Bandbreite von 13,1 Gb/s senden und empfangen können. Damit liegt die theoretische gesamte Bandbreite des FPGAs bei $\sim$1 Tb/s.

Kapitel 7

Implementierung

Für den Phase-I Ausbau des Kalorimetertriggers wurde von CMS die MP7-Karte [28] entwickelt. Diese Karte ist mit einem 7VX690T FPGA bestückt und verfügt darüber hinaus über eine ausreichende optische Anbindung, um 72 der 80 Hochgeschwindigkeitstransceiver des FPGAs zur Datenübertragung mittels Lichtwellenleiter zu nutzen. Besonders angenehm ist eine existierende Hardwarebeschreibung von grundlegenden Funktionen im FPGA, zukünftig auch Infrastruktur genannt, die zum Beispiel eine einfache Benutzung der Datenein- und ausgänge ermöglicht. Daher nutzen wir diese Karte als Plattform für unsere Implementierung der Spurrekonstruktion.

7.1 Hough-Transformation

Die Spursuche mit der Hough-Transformation beruht auf der Transformation der Koordinaten der Stubs (r, ϕ) zu Geraden in der Spurparameterebene (q/p_{T}, ϕ_T) und der Suche nach Schnittpunkten in dieser Ebene.

In Anbetracht der vorhandenen FPGA-Ressourcen bietet es sich an, die Spurparameterebene als zweidimensionales Histogramm mit einer gewissen Granularität abzubilden. Dabei

verwenden wir die Spalten für q/p_{T} und die Zeilen für ϕ_T. Für [17] haben wir 32 Spalten und 64 Zeilen gewählt.

Die Stubs eines Ereignisses werden in das Histogramm gefüllt und die Suche nach Schnittpunkten wird realisiert durch eine Suche nach Zellen, die Stubs von einer gewissen Anzahl verschiedener Detektorlagen beinhalten. Diese Zellen nennen wir aktivierte Zellen. Für [17] haben wir weitestgehend eine Schwelle von fünf gewählt, auf Ausnahmen wird an entsprechender Stelle hingewiesen.

Die Stubs in einer aktivierten Zelle stellen einen gefundenen Spurkandidaten dar. Die Zeilen- und Spaltennummer der aktivierten Zelle entsprechen den gefundenen Spurparametern. Abb. 7.1 zeigt ein befülltes Hough-Histogramm für die Testtrajektorie aus Abb. 5.8.

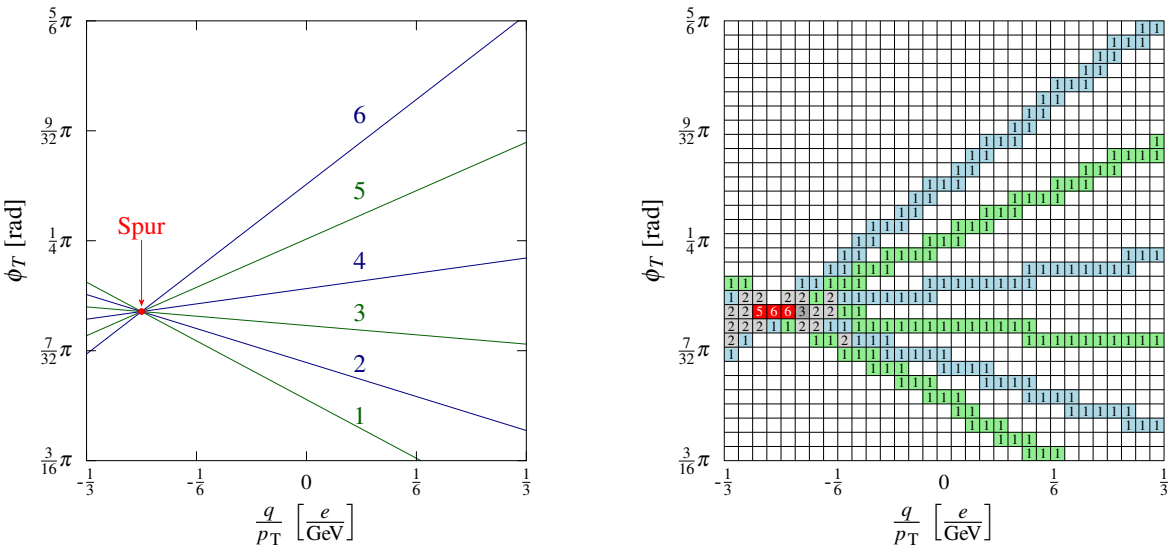

Abbildung 7.1 Illustration des Hough-Histogramms. Das linke Bild ist identisch mit dem rechten Bild von Abb. 5.8. Es zeigt die Spurparameterebene, mitsamt den transformierten Stubs eines Testteilchens. Rechts ist das entsprechende befüllte Hough-Histogramm zu sehen. In den Zellen ist die Anzahl der Stubs von unterschiedlichen Detektorlagen n eingetragen. Die Zellen mit $n = 1$ sind in grün oder blau eingefärbt, die Zellen mit $n > 1$ sind grau und die Zellen mit $n \geq 5$ sind rot. Die roten Zellen stellen gefundene Spurkandidaten dar.

Das Histogramm teilen wir in die einzelnen Spalten auf, welche weitestgehend unabhängig voneinander und parallel agieren. Die Zeilenstruktur realisieren wir durch Speicherelemente innerhalb der Spalten, da diese Bausteine äußerst kompakt sind und somit möglichst viele Histogramme auf einem FPGA untergebracht werden können. Der größte Nachteil von Speichern ist ihre limitierte Bandbreite, aufgrund der begrenzten Anzahl an Speicherzugriffe pro Takt.

Die benötigte Anzahl an Speicherzugriffe pro Stub und Spalte lässt sich begrenzen, indem der Absolutwert des Gradienten der transformierten Geraden in Zeilen- und Spalteneinheiten den Wert Eins nicht überschreitet. Dadurch muss ein Stub pro Spalte in höchstens zwei Zeilen eingefügt werden.

Damit ein Schnittpunkt von Geraden möglichst wenige Zellen aktiviert, sollten die möglichen Gradienten eine größtmögliche Variation aufweisen. Zusammen mit einer gegebenen Detektorgeometrie und der daraus folgenden Variation der möglichen Radien der Stubs und einem gegebenen Parameter T aus Gl. 5.23 ergibt sich ein optimales Verhältnis von Zeilen zu Spalten.

Für die tatsächliche Granularität lässt sich nur schwer ein Optimum bestimmen. Man wird hier stets zwischen der Effizienz und der Qualität der Spursuche abwägen müssen. Hierbei definieren wir die Effizienz der Spursuche als das Verhältnis der Zahl von gefundenen Teilchen zu der Anzahl von zu findenden Teilchen, wobei die betrachteten Teilchen aus der Selektion für Qualitätsstudien aus Tab. 5.2 stammen. Ein Teilchen gilt nach [17] als gefunden, wenn es mindestens einen Kandidaten gibt, welcher gemeinsame Stubs aus mindestens vier Detektorlagen ausweist.

Die Qualität der Spursuche bezieht sich auf die Auflösung der Spurparameter, die Anzahl von den bereits erwähnten Störstubs, in den Kandidaten und die Anzahl von Kandidaten, welche nur eine zufällige Kombination aus unabhängigen Stubs darstellen.

Wie wir später sehen werden, wird das Histogramm nicht mehr als einen Stub pro Takt verarbeiten können. Als Takt verwenden wir denselben Takt, welcher von der Infrastruktur verwendet wird. Diese Taktfrequenz beträgt 240 MHz und entspricht der sechsfachen Frequenz des LHC.

Die durchschnittliche Anzahl von zu verarbeitenden Stubs in $t\bar{t}$-Ereignissen mit 200 Überlappereignissen beträgt nach Tab. 5.2 durchschnittlich 15 041 Stubs, was 2 507 Stubs pro 240 MHz Takt entspricht. Daher werden tausende Histogramme benötigt, welche unabhängig voneinander und parallel operieren.

Die einfachste Methode zur Parallelisierung ist das sogenannte Zeitmultiplexverfahren. Hier werden mehrere Histogramme verwendet, welche jeweils die Stubs eines Ereignisses prozessieren. Allerdings würde jedes Histogramm ein ganzes Ereignis verarbeiten, was mit gegebener Prozessgeschwindigkeit von nicht mehr als einen Stub pro 240 MHz Takt zu einer minimalen Prozessdauer von $\mathcal{O}(100\,\mu s)$ führe.

Daher müssen auch die Stubs einzelner Ereignisse in unabhängige Mengen unterteilt werden. Dazu teilen wir den Detektor in Sektoren in der r-ϕ- und in der r-z-Ebene. Damit diese Teilmengen unabhängig voneinander prozessiert werden können, werden die Stubs in der Nähe von Sektorgrenzen allen angrenzenden Sektoren im Vorfeld zugeordnet. Dadurch vergrößert sich zwar die Anzahl der zu verarbeitenden Stubs, allerdings wird so eine parallele und vor allem auch einfachere Verarbeitung eines Ereignisses ermöglicht.

Die Größe dieser Grenzregionen hängt in der r-ϕ-Ebene von der maximalen Krümmung eines zu findenden Teilchens, also dem minimalen Transversalimpuls, ab. In der r-z-Ebene wird die Grenzregion von der Unsicherheit der Ausmaße der primären Wechselwirkungszone in $\hat{z}$-Richtung bestimmt.

Die Sektorunterteilung in der r-z-Ebene stellt eine sehr grobe Spursuche in der r-z-Ebene dar. Da die Hough-Transformation nur in der r-ϕ-Ebene nach Spuren sucht, wird durch die Sektorunterteilung eine dreidimensionale Spursuche erzeugt, wodurch sich die Qualität der Spursuche verbessert. In der r-ϕ-Ebene hat die Sektorunterteilung den Nebeneffekt, dass das Histogramm sich nicht mehr über $2\,\pi$ erstrecken muss, sondern nur noch über einen Teil, sodass weniger Zeilen benötigt werden. In dieser Ebene bietet sich die Segmentierung in gleichgroße ϕ Sektoren an und eine Translation der ϕ-Koordinaten der Stubs eines Sektors um S, der ϕ-Koordinate der Sektormitte

$$\phi \quad \rightarrow \quad \phi_S \;=\; \phi - S\,. \tag{7.1}$$

Dadurch transformiert sich Gl. 5.24 zu

$$\phi_T\left(\frac{q}{p_\mathrm{T}}\right) \quad = \quad \phi_S - r_T \cdot \frac{q}{p_\mathrm{T}}\,, \tag{7.2}$$

wobei ϕ_T ab jetzt der Azimutalwinkel des Raumpunktes der Trajektorie bei dem Radius T relativ zur Sektormitte ist.

Für [17] haben wir den Detektor in 16 Sektoren in der r-ϕ-Ebene (Abb. 7.2), 18 Sektoren in der r-z-Ebene (Abb. 7.3) unterteilt und prozessieren 36 Ereignisse parallel. Dafür werden 10 368 Histogramme benötigt. Die Granularität der einzelnen Histogramme haben wir auf 32 Spalten und 64 Zeilen festgelegt. Die Anzahl an Stubs von verschiedener Detektorlagen beträgt fünf in allen Sektoren bis auf diejenigen, in denen ein Teilchen aus der primären Wechselwirkungszone nur fünf Detektorlagen durchqueren kann. Die Sektoren in der r-z-Ebene sind derart zugeschnitten, dass dieser Umstand nur auf zwei der 18 Sektoren zutrifft. In diesen Sektoren beträgt die Schwelle vier Stubs von verschiedenen Detektorlagen.

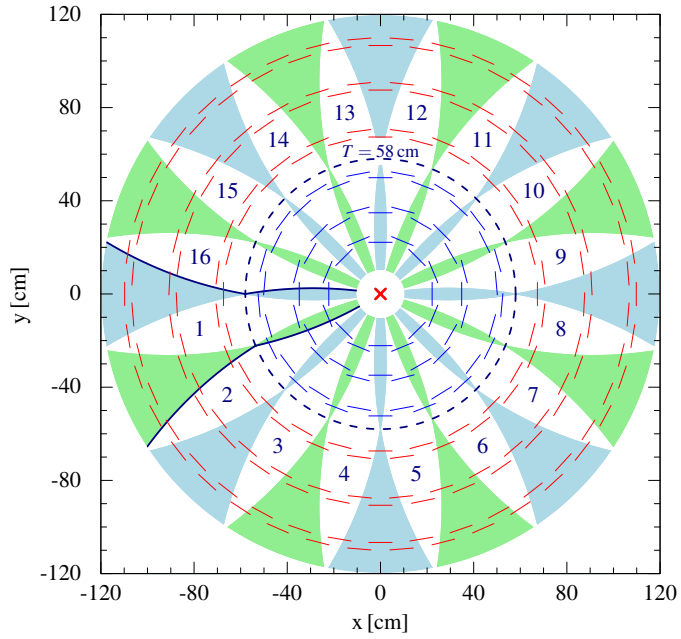

Abbildung 7.2 Sektorunterteilung des äußeren Spurdetektors in der r-ϕ-Ebene. In Rot sind die 2S-Sensormodule, in Blau sind die PS-Sensormodule und der Parameter T ist als gestrichelte Linie eingezeichnet. Die Überlappregionen sind alternierend in grün und in blau dargestellt. Die Sektoren sind durchnummeriert und die Grenzen des ersten Sektors sind hervorgehoben.

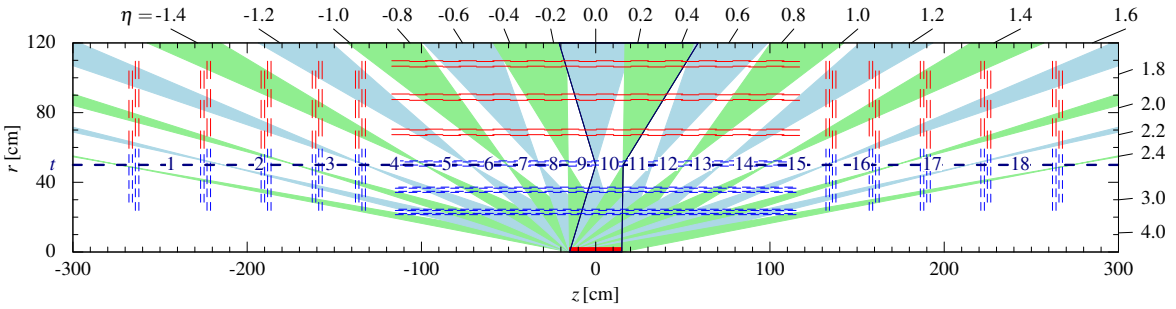

Abbildung 7.3 Sektorunterteilung des äußeren Spurdetektors in der r-z-Ebene. In Rot sind die 2S-Sensormodule, in Blau sind die PS-Sensormodule und der Parameter t ist als gestrichelte Linie eingezeichnet. Die Überlappregionen sind alternierend in grün und in blau dargestellt. Die Sektoren sind durchnummeriert und die Grenzen des zehnten Sektors sind hervorgehoben.

Die benötigte Anzahl an Stubs von verschiedenen Detektorlagen, um einen Spurkandidaten zu finden, beträgt typischerweise fünf, kann aber bei Bedarf für ein Hough-Histogramm auf vier gesetzt werden. Das Heruntersetzen der Schwelle eines Hough-Histogramms ist vor allem interessant, wenn in dem entsprechenden Sektor aufgrund einer technischen Störung eine Detektorlage ausgefallen ist, oder wenn ein Teilchen aus der primären Wechselwirkungszone aufgrund der Detektorgeometrie in diesem Sektor nur fünf Detektorlagen durchqueren kann.

7.1.1 Datenformate

Da sich auf unserem FPGA nur Ganzzahloperationen gut abbilden lassen, stellen wir alle Variablen mit Ganzzahlen dar. Eine Ganzzahl zeichnet sich durch ihre Breite, also der Anzahl an Bits, ihren numerischen Wert und ihre Basis aus. Weiter unterscheiden wir zwischen Ganzzahlen mit und ohne Vorzeichen. Für die Darstellung einer vorzeichenlosen Ganzzahl nutzen wir das Dualsystem und für eine vorzeichenbehaftete Ganzzahl das Zweierkomplement.

Die Basis entspricht der kleinsten darstellbaren reelle Zahl. Mit der folgenden Transformation wechseln wir von der reellen Darstellung in die ganzzahlige

$$x \quad \rightarrow \quad \text{floor}\left(\frac{x}{\text{Basis}\,(x)}\right) , \tag{7.3}$$

wobei mit floor das Abrunden zur nächsten Ganzzahl gemeint ist. Die Rücktransformation ist gegeben durch

$$x \quad \rightarrow \quad \text{Basis}\,(x) \cdot \left(x + 0,5\right) \ . \tag{7.4}$$

Der Wertebereich von ϕ_T erstreckt sich über ein Achtel π. Da das Histogramm 64 Zeilen aufweist, wählen wir für ϕ_T eine Breite von 6 Bit und wählen folgende Basis

$$\text{Basis}\,(\phi_T) \quad = \quad \frac{2\,\pi}{16 \cdot 64} \quad = \quad 6,14\,\text{mrad} \ . \tag{7.5}$$

Der Wertebereich von q/p_T wird durch den minimalen Transversalimpuls von 3 GeV beschränkt. Dieser Bereich wird im Histogramm in 32 Spalten aufgeteilt. Daher verwenden wir 5 Bits für q/p_T und wählen als Basis

$$\text{Basis}\left(\frac{q}{p_\text{T}}\right) \quad = \quad \frac{3}{1000}\frac{\text{GeV}}{\text{cm}\,\text{T}} \cdot \frac{3,8112\,\text{T}}{3\,\text{GeV} \cdot 32} \quad = \quad 1,191 \cdot 10^{-4}\,\text{cm}^{-1} \ . \tag{7.6}$$

Diese beiden Basen sind durch äußere Definitionen wie der Granularität des Histogramms festgelegt und stellen optimale Basen dar, da nur diese eine minimale Breite benötigen, um den gewünschten Wertebereich mit der gewünschten Granularität darzustellen.

Die Basen für r_T und ϕ_S wählen wir nun so geschickt, dass sich daraus minimaler Aufwand für den FPGA ergibt, um die Hough-Transformation auszuführen. Grundsätzlich versuchen wir beliebige Basiswechsel zu vermeiden, da sie eine Multiplikation benötigen. Diese würden Ressourcen und Zeit kosten, sowie unter Umständen weitere Rundungsfehler mit sich bringen. Allerdings sind Basiswechsel um eine beliebige Zweierpotenz kostenfrei, da sie durch eine Anpassung der Verkabelung der Ressourcen realisiert werden können.

In Hinblick auf Gl. 5.24 wählen wir die Basis von ϕ_S so, dass der Unterschied zur Basis von ϕ_T durch eine Zweierpotenz gegeben ist. Da wir die Detektorauflösung nicht beschränken wollen, sollte die Basis kleiner sein als der kleinste messbare Winkelunterschied zwischen zwei Stubs. Mit einem äußersten Radius von 108 cm der Sensormodule und dem Streifenabstand von 90 µm ergibt sich eine maximale Basis von 83 µrad. Gewählt haben wir

$$\text{Basis}\,(\phi_S) \quad = \quad 2^{-7} \cdot \text{Basis}\,(\phi_T) \quad = \quad 48\,\mu\text{rad} \ . \tag{7.7}$$

Der Wertebereich von ϕ_S muss größer sein als der von ϕ_T, aufgrund der Überlappregionen in der Sektorunterteilung. Wir nehmen an, dass diese Sektorunterteilung nie so fein wird, dass sich zwei benachbarte Überlappregionen überlappen, also Stubs in der r-ϕ-Ebene höchstens verdoppelt werden müssen. Daher kann der Wertebereich für ϕ_S höchstens doppelt so groß sein, wie der von ϕ_T und wir benötigen 14 Bits, um ϕ_S darzustellen.

Die Basis für r_T wählen wir so, dass der Unterschied der Basis des Produkts mit q/p_T und der Basis von ϕ_T durch eine Zweierpotenz gegeben ist. Um die Detektorauflösung nicht zu sehr zu beschränken, wählen wir als maximale Basis die Pixellänge von 1,47 mm. Daraus ergibt sich

$$\mathrm{Basis}\,(r_T) \;=\; 2^{-9} \cdot \frac{\mathrm{Basis}\,(\phi_T)}{\mathrm{Basis}\,(q/p_\mathrm{T})} \;=\; 0,1006\,\mathrm{cm}\,. \tag{7.8}$$

Der Wertebereich hängt von der Detektorgeometrie und der Wahl von T ab. Für [17] haben wir $T = 58\,\mathrm{cm}$ gewählt. Damit benötigen wir 10 Bits um r_T darzustellen.

7.1.2 Hough-Histogramm

Wir setzen voraus, dass im Vorfeld die Stubs eines Ereignisses in Sektoren gruppiert, mit unseren Koordinaten dargestellt und paketweise an die Hough-Histogramme gesendet werden. Dabei wird ein Stub pro Takt gesendet, bei einer Taktfrequenz von 240 MHz.

Die Paketgröße ist statisch und wird durch die 36 zeitlich parallel verarbeiteten Ereignisse beschränkt. Weiter benötigt die Infrastruktur des FPGAs eine zeitliche Lücke von sechs Takten zwischen zwei Paketen. Daher beträgt die Paketgröße 210 Takte, dies entspricht 210 Stubs oder 875 µs.

Die Stubs werden an die 32 Spalten verteilt, welche unabhängig voneinander, jeweils unter der Annahme unterschiedlicher q/p_T - Werte, nach Spuren suchen. Die Spursuche in einer Spalte wird durch das Sortieren der Stubs nach ϕ_T realisiert.

Da ein Sortiervorgang erst sicher und korrekt abgeschlossen ist, wenn der letzte Stub einsortiert wurde, besteht der Prozess aus zwei Schritten. Der erste Schritt ist das Befüllen

des Histogramms und der zweite Schritt besteht in der Auslese der Zellen, welche aufgrund ihrer Stubkomposition als Spurkandidaten erkannt wurden.

Da wir prinzipiell einen unendlichen Strom an Paketen erwarten, müssen beide Schritte zeitgleich geschehen. Während die gefunden Spurkandidaten eines Sektors des ersten Ereignisses ausgelesen werden, wird mit den Stubs desselben Sektors des 37. Ereignisses nach Spuren gesucht. Die Paketgröße der auszulesenden Stubs ist identisch mit der Eingangspaketgröße. Die Stubs im Ausgangspaket bilden einen kontinuierlichen Block und sind in ihrer Reihenfolge zu Kandidaten gruppiert.

Falls ein Stub in mehreren Spurkandidaten vorkommt, wird er entsprechend mehrmals ausgelesen. Daher könnte die Anzahl der auszulesenden Stubs größer sein als die Zahl der einzulesenden Stubs. Allerdings begrenzt die Paketgröße die Zahl der auslesbaren Stubs auf 210. Dementsprechend sollte man bei der Unterteilung des Detektors in Sektoren und der Wahl der Paketgröße auf Engpässe sowohl am Eingang, als auch am Ausgang der Hough-Transformation achten.

Die einzelnen Spalten im Histogramm bilden eine lange Kette und können nur in eine Richtung kommunizieren. Die Verteilung der Stubs auf die Spalten geschieht also, indem die Stubs durch alle Spalten propagieren. Dabei reicht jede Spalte einen Stub pro Takt weiter. Abb. 7.4 zeigt den Aufbau eines Histogramms.

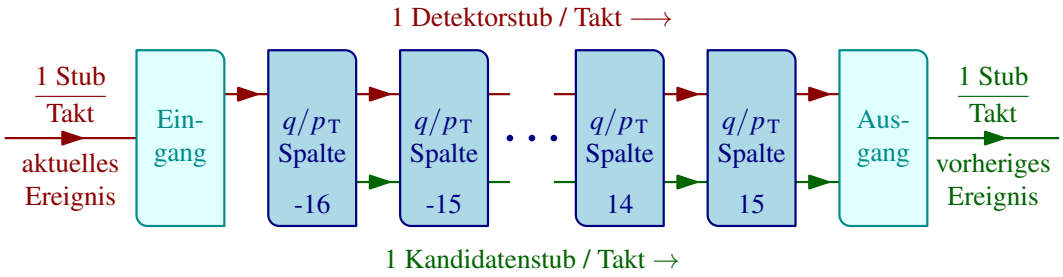

Abbildung 7.4 Blockschaltbild eines Hough-Histogramms. Die ganzzahligen numerischen Werte von q/p_T sind in die einzelnen Spalten eingetragen. Die Spalten mit den Werten von -14 bis 13 wurden aus Gründen der Übersicht ausgelassen.

Die Aufgabe des Eingangblocks besteht im Erkennen eines neuen Eingangpakets und im Zählen der Stubs. Dieser Zählwert dient als Stubkennung und ist ein Bestandteil der

Stubinformation, welche an die Spalten weitergereicht wird. Um die Stubs eines Pakets eindeutig zu kennzeichnen benötigen wir 8 Bit.

Der Ausgangsblock erhält die eingehenden Stubs nach einer Verzögerung von 32 Takten, da diese durch die 32 Spalten mit einer Geschwindigkeit von einer Spalte pro Takt propagieren. Diese werden in einen Speicher abgelegt, den wir Stubspeicher nennen. Dabei wird die Stubkennung nicht gespeichert, sondern als Schreibadresse verwendet.

Darüber hinaus erhält der Ausgangsblock die Stubs der gefundenen Spurkandidaten nach einer Verzögerung von 32 + 216 + 8 = 286 Takten, was 1,19s µs bei einer Traktfrequenz von 240 MHz entspricht. Dabei entsprechen die 32 der Spaltenanzahl und die 216 der Paketgröße. Die 8 Takte entsprechen der Laufzeit von einer Spalte.

Wie wir später sehen werden, wird ein Stub eines gefundenen Spurkandidaten nur aus seiner Stubkennung und den Spurparametern q/p_{T} und ϕ_T bestehen. Die Stubkennung wird als Leseadresse des Stubspeichers verwendet, sodass die vollständigen Stubinformation aus dem Speicher wiedergewonnen wird. Dadurch werden viele Ressourcen in den Spalten eingespart.

Damit die Stubs des zweiten Pakets nicht die des ersten Pakets überschreiben, besteht der Speicher aus zwei Seiten. Die eine Seite wird nur für Stubs von geraden Paketen und die andere nur für Stubs von ungeraden Pakete verwendet.

7.1.3 q/p_{T} - Spalte

Abb. 7.5 zeigt das Blockschaltbild einer Spalte des Hough-Histogramms. Der erste Arbeitsschritt ist, neben dem Weiterreichen des Stubs, die Berechnung von ϕ_T nach Gl. 5.24. Da wir die Stubs in alle Zeilen füllen wollen, welche von der transformierten Geraden gekreuzt werden, berechnen wir ϕ_T am linken und rechten Rand der Spalte. Konstruktionsbedingt können sich die beiden ϕ_T-Werte höchstens um Eins unterscheiden.

Berechnung der Histogrammzeile

Für die Berechnung haben wir zwei Möglichkeiten. Entweder benutzen wir zwei DSP-Slices und führen für beide Randwerte je eine Multiplikation aus oder wir nutzen die propagierende Struktur der Spalten und berechnen ϕ_T iterativ mit

$$\phi_T\left(q/p_{\mathrm{T}} \oplus 1\right) \;=\; \phi_T\left(q/p_{\mathrm{T}}\right) + r_T \qquad \text{mit} \quad \phi_T\left(0\right) \;=\; \phi_S\,, \tag{7.9}$$

wobei wir $\oplus$ verwenden, um zu verdeutlichen, dass der ganzzahlige numerischen Wert der Variablen addiert wird. Eine Spalte würde in diesem Fall das Ergebnis der vorherigen Spalte ausnutzen, wodurch nur noch eine Addition anstelle einer Multiplikation notwendig wäre. Allerdings müssten benachbarte Spalten in diesem Fall auch benachbarte q/p_{T} Werte aufweisen und die Operation dürfte nicht länger als einen Takt benötigen, da sich sonst die Propagationsgeschwindigkeit der Stubs durch die Spalten reduzieren würde.

Im Falle der Multiplikation sind die Spalten unabhängiger voneinander und man könnte mehrere Takte für die Rechnung verwenden, wodurch man leichter höhere Taktraten erreichen könnte. Für [17] verwenden wir die Multiplikation und berechnen ϕ_T mit

$$\phi_T \;=\; \left(2^2 \cdot \phi_S \oplus q/p_{\mathrm{T}} \otimes r_T\right) \cdot 2^{-9}\,, \tag{7.10}$$

wobei die Multiplikationen mit den Zweierpotenzen Schiebeoperationen darstellen, wofür der FPGA weder Ressourcen, noch Zeit aufwenden muss. Das Symbol $\otimes$ verdeutlicht analog zu $\oplus$ die Multiplikation der ganzzahligen numerischen Werte der Variablen.

Zwischenspeicher

Nun ergeben sich drei Möglichkeiten, entweder muss ein Stub in zwei Zeilen, in eine oder in keine einsortiert werden. Die Sortiermaschine, welche gleich erklärt wird, wird nur einen Stub pro Takt in eine Zeile einsortieren können. Daher muss im ersten Fall der Stub verdoppelt werden. Da wir mit jedem weiteren Takt mit einem weiteren Stub rechnen müssen, werden die zu verdoppelten Stubs zwischengespeichert.

Im zweiten Fall wird der Stub einfach mit dem berechneten ϕ_T weitergeleitet und im dritten Fall ist nichts zu tun, stattdessen wird ein Stub vom besagten Zwischenspeicher weitergeleitet, falls dort zwischengespeicherte Stubs vorhanden sind.

Ein Stub wird in keine Zeile einsortiert, falls ϕ_T für diese Spalte außerhalb des Wertebereiches liegt, der stubeigene Transversalimpuls nicht mit dem Transversalimpuls der Spalte vereinbar ist oder wenn der Stub keine Gültigkeit hat. Ein ungültiger Stub stellt eine Lücke im Eingangspaket dar, welche entweder entsteht, wenn die Quelle die Stubs nicht als kontinuierlichen Block senden kann oder das Paket nicht vollständig gefüllt ist.

Wenn ein neues Paket eintrifft, werden die Stubs im Zwischenspeicher verworfen. Daher ist die Notwendigkeit eines Zwischenspeichers sehr unschön, da eine vollständige Prozessierung der eingehenden Stubs nicht gewährleistet ist. Daher werden später alternative Histogrammformen vorgestellt, welche keinen Zwischenspeicher benötigen.

Identifikation von Histogrammzeilen als Spurkandidaten

Der nächste Schritt besteht im Sortieren der Stubs in ϕ_T, sowie der Identifikation von Spurkandidaten. Mit jedem Takt wird potentiell ein Stub mit einem beliebigen ϕ_T und von einer beliebigen Detektorlage i prozessiert. Für die Identifikation merken wir uns für jede Zelle, welche Detektorlagen aktiviert wurden. Dazu verwenden wir einen Distributed Memory, den wir Detektorlagenspeicher nennen.

Durch die Verwendung eines Speichers erhalten wir ein möglichst kompaktes Design. Das Distributed Memory hat den Vorteil, dass sich solch ein kleiner Speicher effizient realisieren lässt und das asynchrone Lesen vereinfacht darüber hinaus den Prozessablauf.

Der Detektorlagenspeicher ist 8 Bit breit, 64 tief und ist als Single Port konfiguriert. Jedes Bit eines gespeicherten Wortes des Detektorlagenspeichers repräsentiert eine Detektorlage und alle Bits des Speichers sind mit ’0’ initialisiert. Eigentlich wären nur 7 Bit vonnöten, da eine Spur aus der primären Wechselwirkungszone nicht mehr als sieben Detektorlagen durchqueren kann.

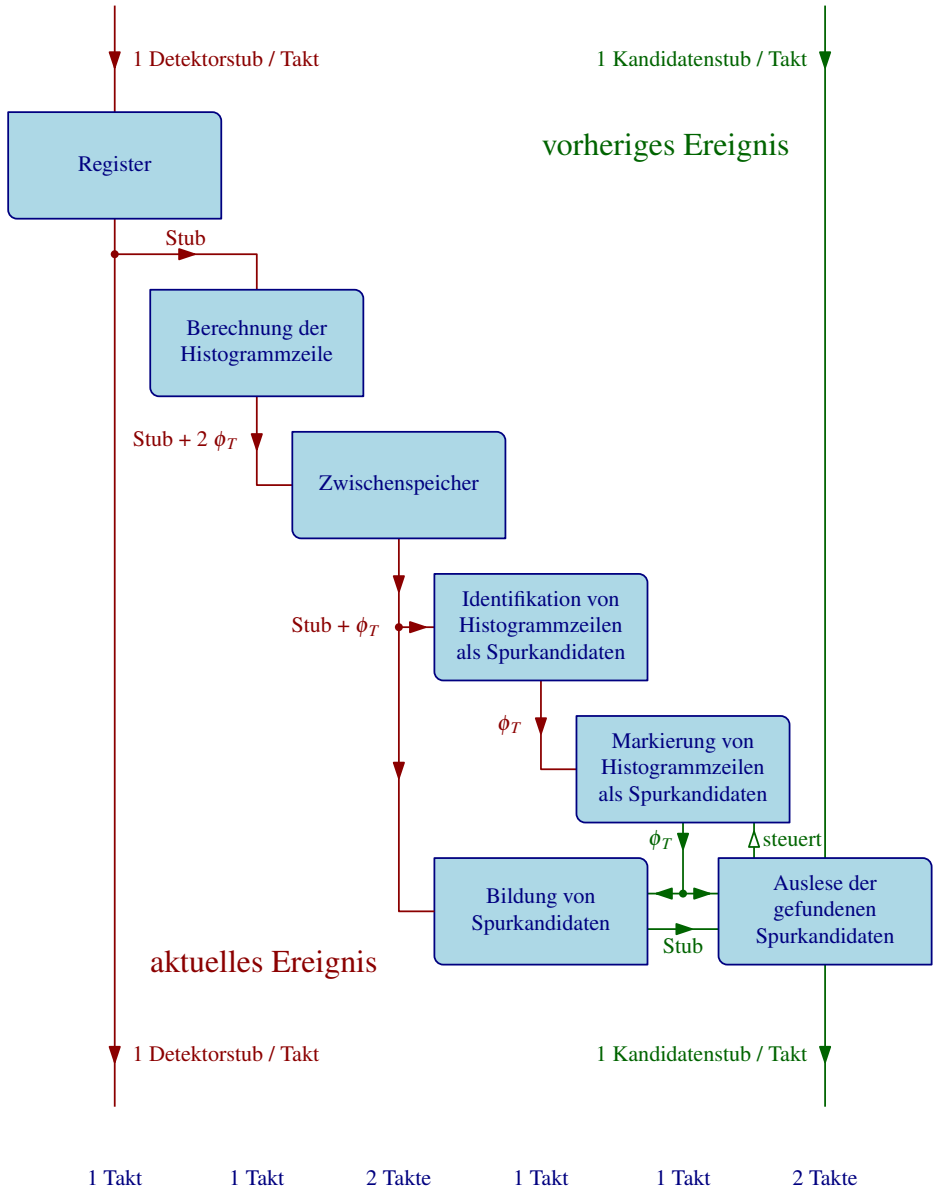

Abbildung 7.5 Blockschaltbild einer q/p_{T} - Spalte. Unten sind die Laufzeiten der einzelnen Prozessschritte eingetragen.

ϕ_T wird als Adresse verwendet. Geschrieben wird das asynchron gelesene Wort, dessen i-tes Bit auf '1' gesetzt wird. Gleichzeitig wird die Zahl der aktivierten Detektorlagen für diese Histogrammzelle, durch das Zählen der Einsen, ermittelt. Falls diese Zahl über einer festgelegten Schwelle liegt, wird dieser ϕ_T-Wert als Spurkandidat identifiziert.

Ein fundamentales Problem mit Speichern ist die limitierte Bandbreite mit der gelesen und geschrieben werden kann, da nur ein oder zwei Ports zur Verfügung stehen. Daher kann ein Speicher nicht auf einmal zurücksetzen werden. Da wir die aktivierten Detektorlagen mit dem nächsten Paket wieder von neuem zählen wollen, müssen wir aber alle Bits des Speicher auf '0' zurücksetzen.

Eine Methode dieses Ziel zu erreichen, wäre den Detektorlagenspeicher zu verdoppeln. Den einen Speicher würden wir wie gewohnt verwenden, während der andere Speicher Zeile für Zeile zurückgesetzt wird. Sobald ein neues Paket beginnt, würden die beiden Detektorlagenspeicher ihre Rolle tauschen, sodass ein zurückgesetzter Speicher für das neue Paket zur Verfügung steht. Diese Methode wird „Ping Pong“ spielen genannt.

Um Ressourcen zu sparen, verdoppeln wir nicht den Detektorlagenspeicher, sondern führen einen weiteren Distributed Memory ein, welchen wir Rücksetzspeicher nennen und Ping Pong spielen lassen. Dieser Speicher ist aber nur ein Bit breit, hat eine Tiefe von 64 und ist als Single Port konfiguriert. Für jede Zeile des Hough-Histogramms enthält dieser Speicher die Information, ob dieses ϕ_T zum ersten Mal in diesem Paket prozessiert wird. Falls dem so ist, schreibt der Detektorlagenspeicher ein Wort, welches bis auf das i-te Bit nur aus Nullen besteht. Somit wird der Detektorlagenspeicher im normalen Betrieb zurückgesetzt. Wichtig ist, dass die Anzahl der Zeilen des Hough-Histogramms nicht die Anzahl der Takte, welche die Paketgröße definieren, überschreitet.

Markierung von Histogrammzeilen als Spurkandidaten

Falls der Detektorlagenspeicher einen Spurkandidaten identifiziert, wird das entsprechende ϕ_T einmalig in einem weiteren Speicher abgelegt, welchen wir den Kandidatenspeicher nennen. Dies ist ein als Simple Dual Port konfigurierter Distributed Memory mit einer Breite von 6 Bit und einer Tiefe von 32.

Befüllt wird der Kandidatenspeicher von unten nach oben, die Schreibadresse wird mit jedem neuen Kandidaten inkrementiert. Die Differenz der aktuellen Schreibadresse und der letzten Schreibadresse des vorangegangenen Pakets entspricht der aktuellen Anzahl an gefundenen Kandidaten in der Spalte des Hough-Histogramms.

Um das ϕ_T eines Spurkandidaten einmalig zu speichern, müssen wir die erste Identifikation eines Spurkandidaten erkennen. Dazu verwenden wir einen als Single Port konfigurierten Distributed Memory, welcher 64 tief und 1 Bit breit ist.

Wieder nutzen wir ϕ_T als Adresse. Geschrieben wird eine '1', falls dieser Wert als Spurkandidat erkannt wurde, ansonsten eine '0'. Daher können wir die erste Identifikation erkennen, wenn wir eine '0' lesen und eine '1' schreiben. Da erst mehrere Stubs mit demselben ϕ_T zur Identifikation eines Spurkandidaten führen, ist dieser Speicher nicht von der Initialisierung abhängig und muss nicht zurückgesetzt werden.

Zum Sortieren der Stubs in die Zeilen des Hough-Histogramms haben wir zwei mögliche Speicherstrukturen implementiert, der „segmentierte Speicher“ und die „verlinkte Liste“.

Bildung von Spurkandidaten mit dem segmentierten Speicher

Der Segmentierte Speicher ist ein als Simple Dual Port konfigurierter halber BRAM, die kleinste nutzbare BRAM Einheit. Die Breite beträgt 9 Bit und der Speicher ist 2 k tief. Der Adressraum ist in $2 \cdot 64$ Seiten unterteilt, welche je 16 Zeilen tief sind, wobei jede Zeile einen Stub speichern können soll. Das heißt, wir können nur bis zu 16 Stubs pro Spurkandidaten finden.

Da ein Stub aus mehr als 9 Bit besteht, speichern wir nicht die volle Stubinformation, sondern eine Adresse, an der die volle Information gespeichert ist. Die Adresse generieren wir mit einem Zähler, welcher jeden Stub eines Paktes eindeutig beziffert. Da die Paketgröße 210 beträgt, werden nur 8 der 9 Bit benötigt.

Jede Seite im segmentierten Speicher steht für eine Zeile des Hough-Histogramms. Damit das gleichzeitige Einlesen des aktuellen Pakets und das Auslesen des letzten Pakets ermöglicht wird, verwenden wir die doppelte Anzahl an Seiten. Die eine Hälfte wird für gerade und die andere für ungerade Pakete verwendet. Während der schreibende Port in der einen Hälfte operiert, um das Histogramm zu befüllen, operiert der lesende Port in der anderen Hälfte, um das Histogramm auszulesen.

Die einzelnen Seiten werden von unten nach oben aufgefüllt. Für jede Seite gibt es eine aktuelle Schreibposition, welche der aktuellen Anzahl von Stubs in den jeweiligen Seiten entspricht. Daher werden 64 Zählstände benötigt, welche jeweils aus einem 4 Bit breiten Wort bestehen.

Diese speichern wir in einem Distributed Memory, welches wir im Folgenden den Adressenspeicher nennen. Der Adressenspeicher ist 128 tief, 4 Bit breit und als Dual Port konfiguriert, wobei hier auch die eine Hälfte des Adressraumes für gerade Pakete und die andere für ungerade Pakete verwendet wird.

Mit jedem Takt prozessieren wir potentiell einen neuen Stub mit einem beliebigen ϕ_T. Der Wert dient als Adresse für den Port des Adressenspeichers nutzen, welcher lesen und schreiben kann. Das asynchron gelesene Wort wird asynchron inkrementiert und synchron an dieselbe Adresse geschrieben. Gleichzeitig dient das gelesene Wort aus dem Adressenspeicher zusammen mit dem ϕ_T-Wert des Stubs als Adresse des schreibenden Ports des segmentierten Speichers.

Da wir die Seiten mit dem nächsten Paket wieder von unten bis oben befüllen wollen, müssen wir die Zählstände zurücksetzen. Dazu wird auch der bereits erwähnte Rücksetzspeicher verwendet. Falls dieser das erste Vorkommen des aktuellen ϕ_T im Paket erkennt, verwendet der segmentierte Speicher den Wert Null als Teiladresse und der Adressenspeicher schreibt den Wert Eins. Somit werden diese beiden Speicher im normalen Betrieb zurückgesetzt.

Bildung von Spurkandidaten mit der verlinkten Liste

Die verlinkte Liste besteht aus einem halben BRAM und ist als Simple Dual Port Speicher konfiguriert. Der Speicher ist 512 tief und 36 Bit breit. Der Adressraum ist auch zweigeteilt, für gerade und ungerade Pakete. Beschrieben wird der Speicher von unten nach oben.

In der verlinkten Liste bestehen die gespeicherten Wörter aus zwei Teilen, dem Zählwert, der die Stubs eindeutig kennzeichnet, und einer Adresse, an der der letzte Stub mit demselben ϕ_T gespeichert wurde. Da eine Adresse nur aus einem Wort mit 8 Bit besteht, wird effektiv nur eine Breite von 16 Bit benötigt.

Um diese Adresse zu erhalten, verwenden wir einen weiteren Speicher. Dies ist ein als Dual Port konfigurierter Distributed Memory mit einer Tiefe von 128 und einer Breite von 8 Bit. Auch dieser Speicher verwendet je einen halben Adressraum für gerade und ungerade Pakete. Wir nennen diesen Speicher Listenanfangsspeicher. In einem halben Adressraum existiert für jedes ϕ_T eine Zeile, welche für dieses ϕ_T die Speicherposition des letzten Subs in der verlinkten Liste enthält.

Wie bei dem segmentierten Speicher wird mit jedem Takt potentiell ein Stub mit einem beliebigen ϕ_T einsortiert. Das ϕ_T wird als Adresse des Ports des Listenanfangsspeichers verwendet, der lesen und schreiben kann. Das gelesene Wort dient zum Teil dem zu schreibenden Wort der verlinkten Liste. Geschrieben wird die Schreibadresse der verlinkten Liste, welche mit jedem neuen Stub inkrementiert wird.

Der erste Stub, welcher in eine Zeile des Hough-Histogramms einsortiert wird, mit einer speziellen Adresse versehen, welche das Ende eines Kandidaten markiert. In diesem Fall wird das gelesene Wort des Listenanfangsspeicher ignoriert und die verlinkte Liste schreibt den Stubzähler und die Zahl 255. Das erstmalige Prozessieren eines ϕ_T-Wertes in einem Paket wird mit dem Rücksetzspeicher erkannt.

Hiermit ist der Prozess des Einsortierens der Stubs in eine Spalte des Hough-Histogramms mit dem segmentierten Speicher und der verlinkten Liste beschrieben und wir wenden uns dem Auslesen der Spurkandidaten zu.

Auslese der gefundenen Spurkandidaten

Der Ausleseprozess der einzelnen Spalten wird derartig konstruiert, dass das gesamte Histogramm einen Stub pro Takt ausliest. Die einzelnen Spalten reichen die Stubs der Spurkandidaten der vorangegangenen Spalten weiter und hängen ihre eigenen gefundenen Kandidaten am Ende des Datenstromes an. So entsteht beim Auslesen ein kontinuierlicher Block aus Stubs.

Betrachten wir nun eine einzelne Spalte und beginnen mit dem Kandidatenspeicher. In diesem wurden die ϕ_T-Werte der gefundenen Spurkandidaten gespeichert. Sobald ein neues Paket

und somit das nächste Ereignis beginnt, wird die aktuelle Leseadresse auf den Wert der letzten Schreibadresse des vorangegangenen Paktes gesetzt und der Wert der letzten Schreibadresse des vorangegangenen Paktes wird auf den Wert der aktuellen Schreibadresse gesetzt. Falls sich die neue Leseadresse und die letzte Schreibadresse des vorangegangenen Paktes nicht unterscheiden, wurden keine Spuren gefunden und diese Spalte wird nur die auszulesenden Stubs der vorangegangenen Spalten zur nächsten Spalte weiterreichen.

Im Falle, dass sich die Adressen unterscheiden, werden die auszulesenden Spurkandidaten bis zum letzten Kandidaten weitergereicht und dann beginnt die Auslese der Spalte. Die Parameter des Kandidaten q/p_{T} und ϕ_T sind durch die Spaltennummer und das gelesene Wort des Kandidatenspeichers gegeben.

Im Falle des segmentierten Speichers nutzt der lesende Port dieses ϕ_T und einen Zähler als Adresse. Mit jedem Takt wird der Zähler inkrementiert und der nächste Stub ausgelesen. Die Anzahl der Stubs des Kandidaten ist im Adressenspeicher an der Stelle ϕ_T abgelegt. Sobald der Zähler diese Anzahl erreicht, wird die Leseadresse des Kandidatenspeichers inkrementiert.

Wird dadurch der Wert der letzten Schreibadresse des vorangegangenen Pakets erreicht, ist die Auslese der Spalte abgeschlossen und die Spalte teilt der nächsten mit, dass diese nun mit der Auslese beginnen kann. Ansonsten wird der nächste Kandidat ausgelesen.

Im Falle der verlinkten Liste, nutzt der lesende Port im ersten Takt den Listenanfang des ϕ_T-Wert des Kandidaten als Adresse. Dieser ist im Listenanfangsspeicher an der Stelle ϕ_T abgelegt. In den folgenden Takten dient der dafür vorgesehene Teil des gelesenen Wortes als Adresse. So wird mit jedem Takt ein Stub eines gefundenen Spurkandidaten ausgelesen, in der umgekehrten Reihenfolge mit der sie eingelesen wurden. Das Ende eines Kandidaten erkennt die verlinkte Liste anhand der speziellen Adresse 255.

Beide Speicherstrukturen unterbrechen den Ausleseprozess, sobald ein neues Paket die Spalte erreicht und beginnen mit der Auslese des nächsten Ereignisses. Die noch nicht ausgelesenen Kandidaten des vorangegangenen Ereignisses sind verloren.

Grundsätzlich erscheint die verlinkte Liste als die bessere Speicherstruktur, da die maximale Anzahl der Stubs pro Spurkandidat nicht begrenzt ist, der Ressourcenbedarf vergleichbar

ist und besser mit der Anzahl an Zeilen des Hough-Histogramms skaliert. Allerdings würde es mit der verlinkten Liste schwieriger werden höhere Taktraten zu erreichen, da man den optionalen Ausgangsregister des BRAMs nicht verwenden kann. Da die Leseadresse vom zuvor gelesenen Wort abhängt, würde man zwei Takte pro auszulesenden Stub benötigen, wenn man diesen verwenden würde. Für [17] haben wir den segmentierten Speicher gewählt.

Virtuelle Untersektoren

Die Identifikation von Histogrammzeilen als Spurkandidaten ist kostengünstiger als das Bilden von Spurkandidaten. Eine Möglichkeit diesen Umstand auszunutzen besteht in einer feineren Sektorunterteilung in der r-z-Ebene, virtuelle Untersektoren genannt, welche nur in der Spuridentifikation berücksichtigt werden.

Im Vorfeld wird ein Stub den Sektoren und Untersektoren zugeordnet. Während die Sektorunterteilung dazu führt, dass der Stub zum entsprechenden Hough-Histogramm transportiert wird, wird die die feinere Unterteilung als Datenwort den Stubinformationen hinzugefügt.

In den q/p_{T} - Spalten wird nun für jeden Untersektor ein Detektorlagenspeicher generiert, welche in Abhängigkeit dieser Stubinformation befüllt werden. Eine Zelle wird als Spurkandidat erkannt, sobald ein Untersektor die nötige Stubkomposition aufweist. Als Resultat steigert sich die Qualität der Spursuche in dem Sinn, dass weniger Kandidaten, welche nur eine zufällige Kombination aus unabhängigen Stubs darstellen. Die Zahl der Störstubs in den gefundenen Kandidaten wird allerdings nicht reduziert, da die Stubbildung die feinere Sektorunterteilung ignoriert, um Ressourcen zu sparen. Für [17] haben wir jeden Sektor in der r-z-Ebene in zwei virtuelle Untersektoren unterteilt.

7.1.4 Resultate

Implementiert haben wir die Hough-Transformation sowohl in Hardware als auch in der Simulationssoftware von CMS. Die Simulationssoftware muss weder Laufzeiten einhalten, mit begrenzten Speicherressourcen auskommen, noch mit Ganzzahlen operieren. Daher stellen

die Softwareresultate stets eine Art Obergrenze dar, welche von der Hardware nie ganz erreicht werden.

Beginnen wir mit den Eingangsdaten der Hough-Transformation. Hierbei handelt es sich um die relevanten Stubs aus Tab. 5.2. Diese werden in 16 Sektoren in der r-ϕ-Ebene und 18 Sektoren in der r-z-Ebene eingeteilt. An den Sektorgrenzen werden die Stubs den angrenzenden Sektoren zugeordnet und somit vervielfacht, sodass jeder Sektor unabhängig von den anderen Sektoren nach Spuren suchen kann. Abb. 7.6 zeigt die Häufigkeitsverteilung der Anzahl der Stubs, die einem Sektor zugeordnet wurden.

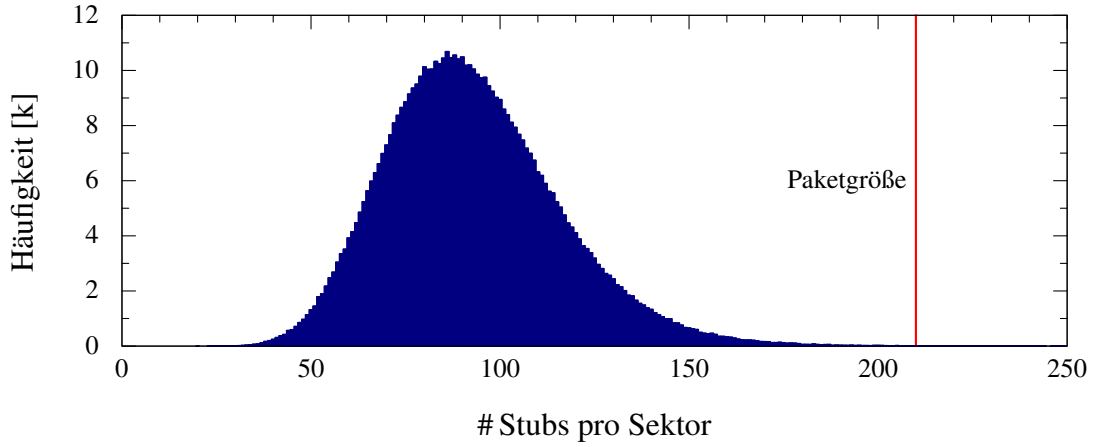

Abbildung 7.6 Häufigkeitsverteilung der Anzahl der Stubs eines Sektors. Die Paketgröße von 210 Stubs ist eingetragen.

Im Mittel werden 93 Stubs einem Sektor zugeordnet. Allerdings schwankt diese Anzahl stark von Ereignis zu Ereignis. Diese mittlere Anzahl passt sehr gut in unsere Pakete mit einer Größe von 210 Stubs. Ein Stub wird im Mittel 1,8 Sektoren zugeordnet.

Kommen wir nun zur Effizienz der Spursuche. Wir messen die Effizienz relativ zu den generierten Teilchen der Selektion für Qualitätsstudien aus Tab. 5.2. Ein Teilchen gilt als erfolgreich rekonstruiert, wenn mindestens ein Spurkandidat gefunden wurde, welcher mindestens vier Stubs aus verschiedenen Detektorlagen aufweist, die mit dem Teilchen assoziiert sind. Spurkandidaten, die sich mit keinen rekonstruierbaren Teilchen aus Tab. 5.2 assoziieren lasen, nennen wir Fakes. Falls mehrere Spurkandidaten diese Bedingungen für

ein rekonstruierbares Teilchen erfüllen, so werden diese Spurkandidaten, bis auf einen, als Duplikat gewertet.

Pro Ereignis werden durchschnittlich 334 Spurkandidaten gefunden, welche im Mittel 7,4 Stubs enthalten. Der durchschnittliche Anteil an Fakes unter den gefundenen Spurkandidaten beträgt 32 %. Der durchschnittliche Anteil an Duplikaten der Spurkandidaten abzüglich der Fakes beträgt 43 %. Daher scheint die Spursuche zunächst höchst ineffektiv zu sein, allerdings ist die Anzahl an Stubs nach der Spursuche um eine Größenordnung kleiner, als vor der Spursuche und das bei einer Effizienz von 97 %. Abb. 7.7 und 7.8 zeigen die Effizienz über diverse kinematische Größen und Abb. 7.9 und 7.10 verdeutlichen den Effekt der Datenreduktion.

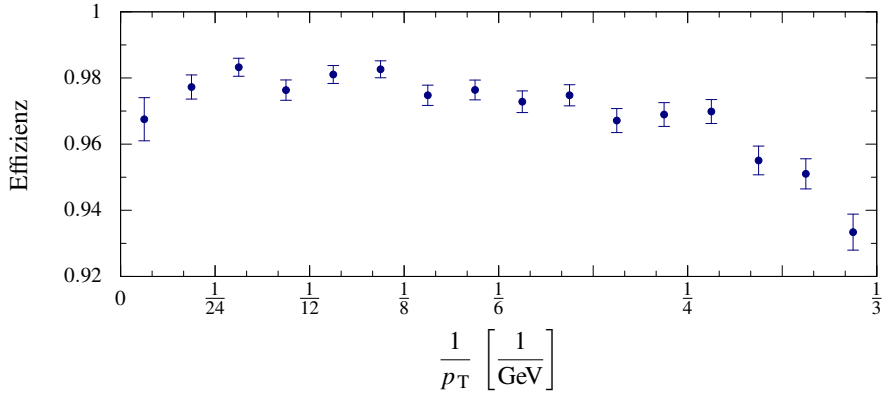

Abbildung 7.7 Effizienz der Spursuche aufgetragen über dem inversen Transversalimpuls.

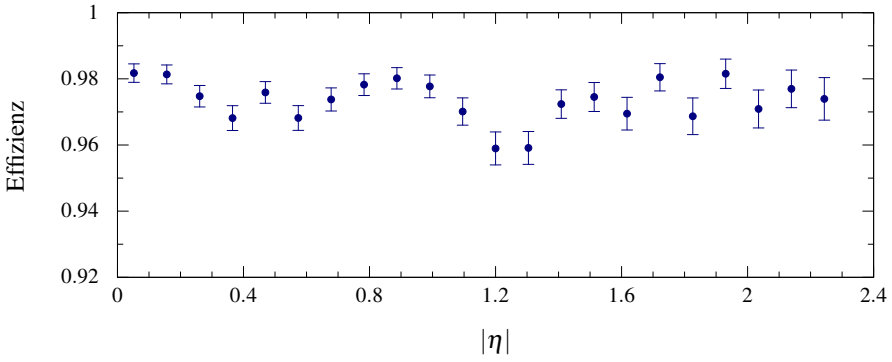

Abbildung 7.8 Effizienz der Spursuche aufgetragen über der Pseudorapidität.

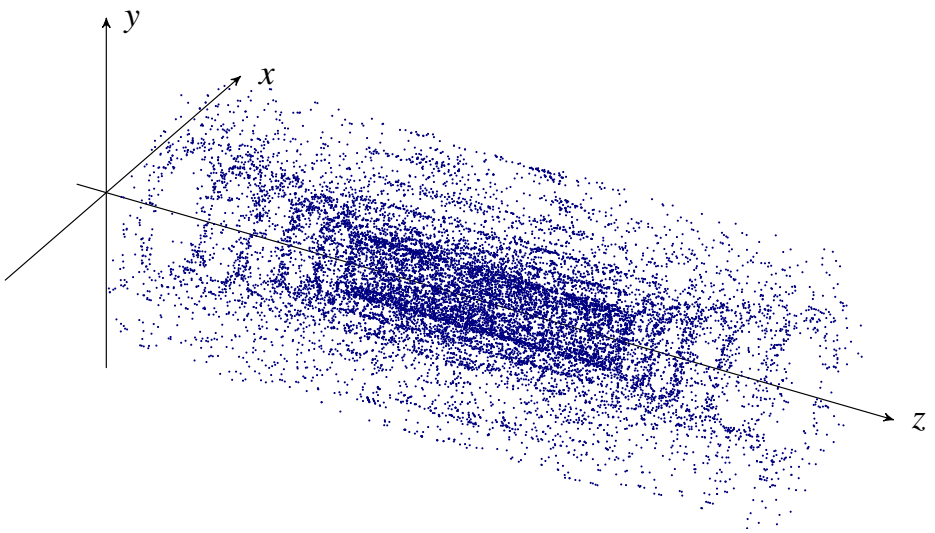

Abbildung 7.9 Darstellung der relevanten Stubs aus Tab. 5.2 ($t\bar{t}$ + 200 PU) eines willkürlichen Ereignisses im dreidimensionalen Raum. Dieses Ereignis enthält 16 149 relevante Stubs und 59 rekonstruierbare Spuren.

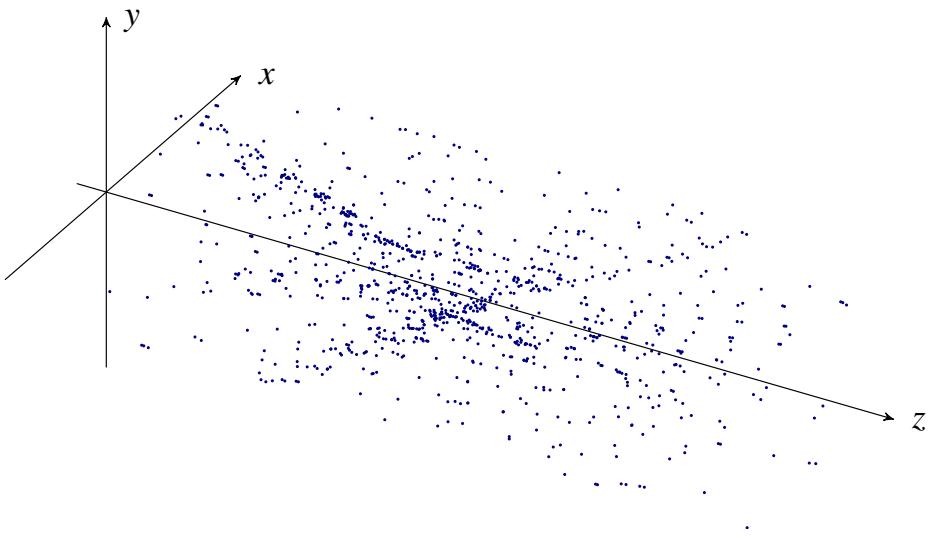

Abbildung 7.10 Darstellung der Stubs aus Fig. 7.9, die von der Hough-Transformation zu gefundenen Spurenkandidaten gruppiert werden. Es handelt sich um 351 Spurkandidaten und 2 690 Stubs, wobei mehrfach vorkommende Stubs auch mehrfach gezählt wurden. 1 074 Stubs sind einzigartig.

Der Ursprung der größten Effizienzverluste mit der Hough-Transformation liegt in der Definition von rekonstruierbaren Teilchen aus Tab. 5.2. CMS definiert ein Teilchen als rekonstruierbar, wenn es mit Stubs von mindestens vier Detektorlagen assoziiert werden kann. Unsere Hough-Transformation bildet aber weitestgehend nur Spurkandidaten, wenn Stubs von fünf Detektorlagen mit einer Spur vereinbar sind. Daher sinkt die Effizienz der Spursuche in den p_{T}- und η-Bereichen, in denen die Wahrscheinlichkeit, dass ein Teilchen nur Stubs von genau vier Detektorlagen generiert, steigt.

Dies sind die Bereiche, in denen der Detektor besonders ineffizient ist. Der Detektor hat Schwierigkeiten Stubs von Teilchen zu bilden, deren Flugbahn stark vom Detektor selbst oder dem Strahlrohr aufgrund von Vielfachstreuung beeinflusst wurden, oder sich innerhalb von Jets befinden, wegen der damit einhergehenden Komplexität überforderten Ausleseelektronik. Die Effizienz der Hough-Transformation sinkt somit mit kleiner werdendem Transversalimpuls und für große Transversalimpulse. Letzteres aufgrund der hohen Wahrscheinlichkeit, dass sich ein Teilchen des $t\bar{t}$-Systems mit hohem Transversalimpuls innerhalb eines Jets befindet.

7.1.5 Mögliche Erweiterungen

Inhomogene Granularität

Die Qualität der Hough-Transformation lässt sich durch eine feinere Granularität steigern. Dies geht allerdings zu Lasten der Effizienz, da weniger imperfekte Spuren, aufgrund der Auswirkungen der Vielfachstreuung auf die Trajektorie, gefunden werden. Allerdings ist die Beeinflussung der Teilchenbahn durch das Strahlrohr und den Detektor abhängig vom Transversalimpuls des Teilchens, sodass eine feinere Granularität in der Region des Histogramms, welche hohen Transversalimpulsen entspricht, äußerst interessant erscheint.

Als Beispiel haben wir die Granularität der inneren Hälfte des ursprünglichen Histogramms verdoppelt, was in Abb. 7.11 angedeutet wird. Das Histogramm besteht aus 8 Spalten mit 64 Zeilen, 16 Spalten mit 128 Zeilen und wieder 8 Spalten aus 64 Zeilen. Somit weist der Transversalimpulsbereich von 3 GeV bis 6 GeV die ursprüngliche Granularität auf, während die Granularität für Transversalimpulse über 6 GeV doppelt so groß ist.

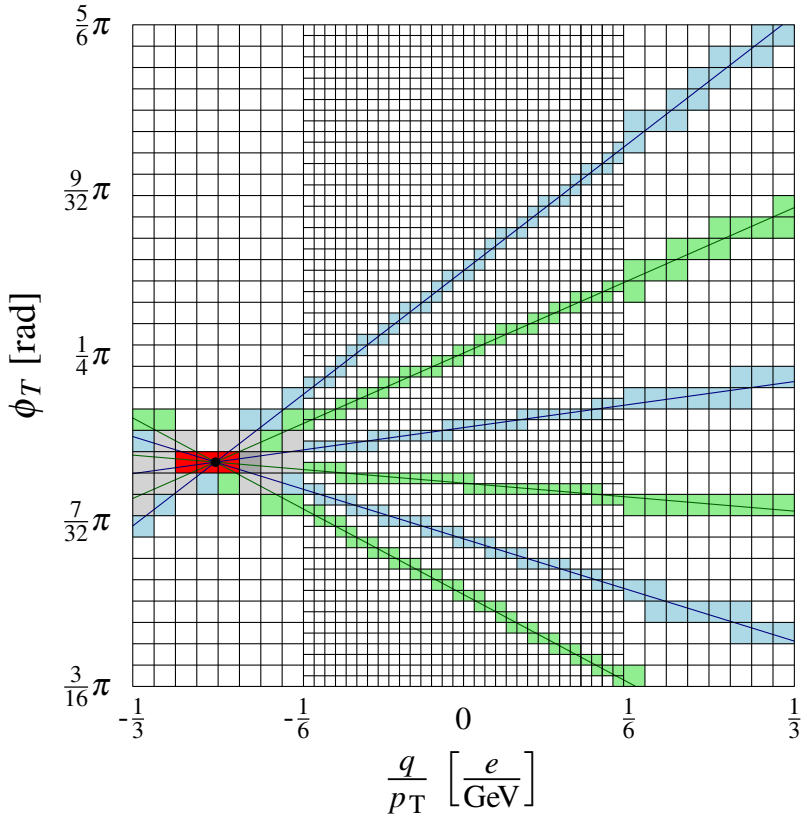

Abbildung 7.11 Darstellung eines Hough-Histogramms mit einer inhomogenen Granularität. Es ist befüllt mit den Testteilchen aus Abb. 5.8.

Pro Ereignis werden mit diesem Histogramm durchschnittlich nur 291 Spurkandidaten gefunden, welche im Mittel 7,2 Stubs enthalten, bei einer Effizienz von 96,8 %. Somit könnte man die Datenrate um 13% reduzieren, falls man bereit wäre 2 ‰ Effizienz zu opfern.

Alternative Histogrammformen

Eine störende Notwendigkeit in den q/p_{T}-Spalten ist der Zwischenspeicher, der die Stubs verdoppelt, die mit zwei Spuren desselben Transversalimpulses kompatibel sind. Problematisch dabei ist der Bedarf an einem halben BRAM, sowie die Ungewissheit, ob alle verdoppelten Stubs aufgrund der begrenzten Prozessdauer korrekt verarbeitet werden.

Daher führen wir nun drei Histogramme ein, in denen eine Gerade, mit begrenztem Gradienten, pro Spalte höchstens eine Zeile kreuzen kann, wodurch der Zwischenspeicher obsolet wird. Die Zellen in diesen Histogrammen haben eine hexagonale, eine rauten- oder eine mauerwerkartige Form und sind in Abb. 7.12 abgebildet.

Im Gegenzug benötigen diese Histogramme mehr Spalten, wobei die Spalten nur einen halben BRAM anstelle eines ganzen benötigen. Später werden wir feststellen, dass es die BRAM-Ressourcen sind, die die Anzahl der Histogramme begrenzen, die sich auf einem MP7 implementieren lassen. Daher nehmen wir nun an, dass sich auf einem FPGA genauso viele alternative Histogramme mit doppelt so vielen Spalten implementieren lassen, wie ursprüngliche Histogramme.

Im Vergleich zu den anderen Formen sind die Rauten besonders günstig, da diese nur ein ϕ_T in der Spaltenmitte berechnen müssen. Das Mauerwerk benötigt die Berechnung von ϕ_T an den beiden Spaltenrändern, so wie das normale Histogramm. Die Hexagone sind ein wenig teurer. Diese benötigen drei Berechnung, an den Rändern und in der Mitte, sowie weitere Logik, welche in wenige CLBs passen sollte, um die korrekte Zeilennummer zu berechnen.

Tab. 7.1 zeigt die Effizienz und Qualität, gemessen an der Anzahl gefundener Spurkandidaten, für diverse Konfigurationen. Dabei dienten die relevanten Stubs aus Tab. 5.2 als Eingangsdaten.

Der Vergleich der verschiedenen Histogrammarten bei ähnlicher Zellenanzahl zeigt, dass das hexagonale Histogramm die beste Datenreduktion, also beste Qualität der Spursuche, erreicht. Das mauerwerkartige und rautenförmige Histogramm erzielen zwar die besten Effizienzen, allerdings zu Lasten der Qualität.

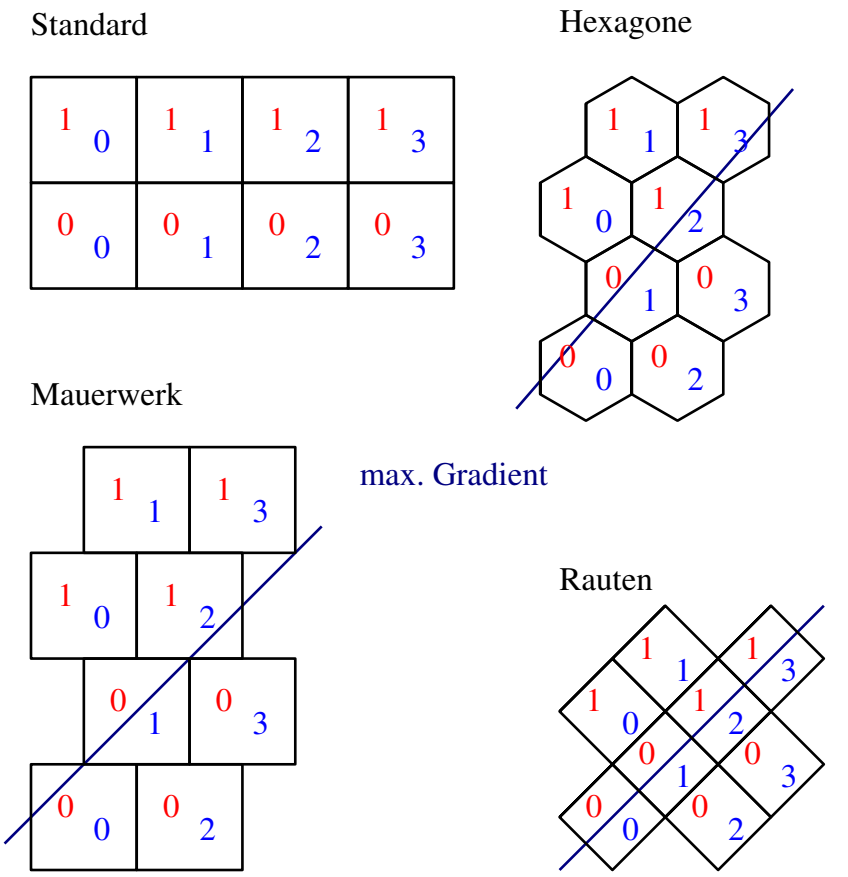

Abbildung 7.12 Alternative Histogrammformen. Alle Histogramme weisen zwei Zeilen und vier Spalten auf. Die Zeilennummern sind in der linken oberen Ecke jeder Zelle in Rot eingetragen und die Spaltennummer in der rechten unteren Ecke in Blau. Außerdem ist jeweils eine Gerade mit dem maximalen Gradienten eingezeichnet, bei dem nur eine Zeile pro Spalte mit der Geraden kompatibel ist.

Tabelle 7.1 Resultate von diversen Konfigurationen des Hough-Histogramms. Die Anzahl an Kandidaten entsprechen der mittleren Anzahl pro Ereignis.

Form	Spalten	Zeilen	Zellen	# Kandidaten	Effizienz in %
Standard	32	64	2048	332	97.2
Rauten	64	62	3968	288	96.9
	34	32	1088	456	97.5
	46	44	2024	358	97.3
Hexagone	64	42	2688	226	96.7
	50	32	1600	268	97.1
	56	36	2016	248	96.9
Mauerwerk	64	30	1920	360	97.6
	66	32	2112	343	97.3

7.2 Lineare Regression

Beginnen wir mit der Implementierung der Gleichungen 5.29 bis 5.31, welche für eine gegebene Menge an Stubs die ideale Trajektorie mit der kleinsten Fehlerquadratsumme bestimmen. Diesen funktionalen Block nennen wir Regressionskern. Die Gleichungen sind zwar simpel, enthalten aber eine Division, welche grundsätzlich schwierig zu implementieren ist, da die dedizierten Rechenwerke auf unserem FPGA nicht dividieren können. Daher ist der Grundgedanke, die Division durch eine Multiplikation mit dem nachgeschlagenen Kehrwert des Nenners zu ersetzen.

Grundsätzlich versuchen wir Rundungsfehler in den Zwischenberechnungen zu vermeiden, damit diese sich nicht fortpflanzen können. Allerdings führt das schnell zu breiten Variablen, wobei mit der Breite die Anzahl Bits, mit der die Variable dargestellt wird, gemeint ist. Die Breite eines nicht gerundeten Produktes ist durch die Summe der Breiten der jeweiligen Faktoren gegeben und die Breite einer nicht gerundeten Summe ist durch die Breite des breitesten Summanden plus eins gegeben.

Da ein DSP-Slice nur einen 25×18 Bit Multiplizierer enthält, werden für Multiplikationen von breiteren Zahlen mehrere Slices benötigt. Wesentlich schlimmer verhält es sich mit der Tiefe des Nachschlagewerkes und somit der Menge an BRAM-Einheiten, da diese exponentiell mit der Breite des Nenners wächst. Damit nur minimale Ressourcen vonnöten sind, versuchen wir alle Variablen mit möglichst wenigen Bits darzustellen. Dazu transformieren wir die Koordinaten der Stubs.

Betrachten wir zuerst die r-ϕ-Ebene. Analog zur Hough-Transformation translatieren wir r um T und ersetzen zusätzlich ϕ durch das Residuum von der gemessen Position des Stubs zur vorhergesagten Position, beruhend auf den Spurparametern, die mit der Hough-Transformation gefundenen wurden

$$\phi \quad \rightarrow \quad \phi_S - \left(\phi_T + r_T \cdot \frac{q}{p_\mathrm{T}} \right) . \tag{7.11}$$

Dadurch erhalten wir

$$\Delta \frac{q}{p_\mathrm{T}} \quad = \quad \frac{n\overline{r_T\,\phi} - \overline{r_T} \cdot \overline{\phi}}{n\overline{r_T\,r_T} - \overline{r_T} \cdot \overline{r_T}} , \tag{7.12}$$

$$\Delta \phi_T \quad = \quad \frac{\overline{\phi} \cdot \overline{r_T\,r_T} - \overline{r_T} \cdot \overline{r_T\,\phi}}{n\overline{r_T\,r_T} - \overline{r_T} \cdot \overline{r_T}} , \tag{7.13}$$

wobei wir wieder die in Gl. 5.27 eingeführte Kurzschreibweise verwenden. Die Ergebnisse $\Delta q/p_\mathrm{T}$ und $\Delta\phi_T$ stellen Korrekturen zu den gefundenen Werten der Hough-Transformation dar, welche noch auf diese addiert werden müssen, um die endgültigen Parameter zu erhalten.

In der r-z-Ebene translatieren wir r um t

$$r \quad \rightarrow \quad r_t \quad = \quad r - t , \tag{7.14}$$

wobei $t = 50\,\mathrm{cm}$ gewählt wurde für [17]. Dadurch transformiert sich Gl. 5.15 zu

$$z \quad = \quad z_t + r_t \cdot \cot\theta , \tag{7.15}$$

wobei z_t der z-Koordinate der Trajektorie bei dem Radius t entspricht. Die Unterteilung der r-z-Ebene in 18 Sektoren entspricht einer Unterteilung von z_t in 18 Regionen, welche nicht

äquidistant gewählt sind für [17]. Dies ist unschön und wir entfernen uns ein wenig von [17], behalten die Anzahl von 18 Sektoren bei, unterteilen z_t aber in gleichgroße Stücke. Nun bietet sich eine Translation von z um s, der z_t - Koordinate der Sektormitte, an

$$z \rightarrow z_s = z - s \,. \tag{7.16}$$

Dadurch transformiert sich Gl. 7.15 zu

$$z_s = z_t + r_t \cdot \cot\theta \,, \tag{7.17}$$

wobei z_t ab jetzt der z-Koordinate der Trajektorie bei dem Radius t relativ zur Sektormitte entspricht. Für die Lineare Regression erhalten wir

$$\Delta \cot\theta = \frac{n\overline{r_t z_s} - \overline{r_t}\cdot\overline{z_s}}{n\overline{r_t r_t} - \overline{r_t}\cdot r_t} \,, \tag{7.18}$$

$$z_t = \frac{\overline{z_s}\cdot\overline{r_t r_t} - \overline{r_t}\cdot\overline{r_t z_s}}{n\overline{r_t r_t} - \overline{r_t}\cdot\overline{r_t}} \,, \tag{7.19}$$

wobei $\Delta\cot\theta$ eine Korrektur zum $\cot\theta$ des Sektors entspricht.

7.2.1 Datenformate

Wir wählen für ϕ dieselbe Basis wie für ϕ_S. Konstruktionsbedingt ist der Wertebereich von ϕ doppelt so groß wie die Basis von ϕ_S beziehungsweise um sechs Zweierpotenzen kleiner als der Wertebereich. Dadurch werden nur 8 Bits benötigt, um ϕ darzustellen, während für ϕ_S 14 Bits benötigt werden.

Die Basen der Parameterkorrekturen $\Delta q/p_{\mathrm{T}}$ und $\Delta\phi_T$ haben wir um fünf Zweierpotenzen kleiner als die jeweiligen Basen der Spurparameter von der Hough-Transformation gewählt. Die Breite ist in beiden Fällen durch 6 Bit gegeben.

In der r-z-Ebene wird der Wertebereich von z_t durch die maximale Pseudorapidität von $2{,}4$ und der Zahl der Sektoren beschränkt.

Wir stellen z_t mit 11 Bit dar und wählen

$$\text{Basis}(z_t) \;=\; \frac{2 \cdot t \cdot \sinh(2,4)}{18} \cdot 2^{-11} \;=\; 0,015\,\text{cm}\,. \tag{7.20}$$

Der Wertebereich der Parameterkorrektur $\Delta \cot\theta$ hängt von denselben Parametern ab, wie der Wertebereich von z_t und zusätzlich von der Unsicherheit der Ausdehnung der primären Wechselwirkungszone, welche [17] mit ± 15 cm angibt. Wir stellen $\Delta \cot\theta$ auch mit 11 Bit dar und wählen

$$\text{Basis}(\Delta \cot\theta) \;=\; 2 \cdot \frac{\sinh(2,4)}{18} \cdot \frac{15\,\text{cm}}{t} \cdot 2^{-11} \;=\; 5,90 \cdot 10^{-4}\,. \tag{7.21}$$

Somit haben wir optimale Basen konstruiert, sodass jede darstellbare Zahl auch im Wertebereich liegt. Die tatsächliche Präzision ist nur von der Wahl der jeweiligen Breite abhängig, welche momentan noch willkürlich erscheint und später begründet wird.

Die Basen von z_s und r_t wählen wir so, dass sich daraus ein minimaler Aufwand für den FPGA ergibt. Dazu wählen wir

$$\text{Basis}(z_s) \;=\; \text{Basis}(z_t) \cdot 2^2 \;=\; 0,116\,\text{cm}\,, \tag{7.22}$$

$$\text{Basis}(r_t) \;=\; \frac{\text{Basis}(z_t)}{\text{Basis}(\Delta \cot\theta)} \cdot 2^{-8} \;=\; 0,13\,\text{cm}\,. \tag{7.23}$$

Die Zweierpotenzen wurden so gewählt, dass die Basen zwischen der halben und vollen Pixellänge liegen. Für r_t und z_s werden je 10 Bit benötigt. Da man natürlich nicht r_t und r_T in den Stubdaten unterbringen möchte, würden wir r_t aus r_T berechnen und kleine Rundungsfehler in Kauf nehmen.

7.2.2 Regressionskern

Beginnen wir mit Gl. 7.12 und 7.13. Wir setzen voraus, dass sich in einem Spurkandidaten höchstens ein Stub auf einer Detektorlage befindet beziehungsweise im Vorfeld für die jeweiligen Detektorlagen ein virtueller Stub generiert wurde. Darüber hinaus seien die Variablen ϕ und r_t aus den Stubdaten der gefundenen Spurkandidaten berechnet worden.

Diese Operationen sowie den Regressionskern gestalten wir so, dass sie einen Stub pro Takt prozessieren. Dabei verwenden wir dieselbe Taktrate wie bei der Hough-Transformation. Abb. 7.13 skizziert die Implementierung des Regressionskerns.

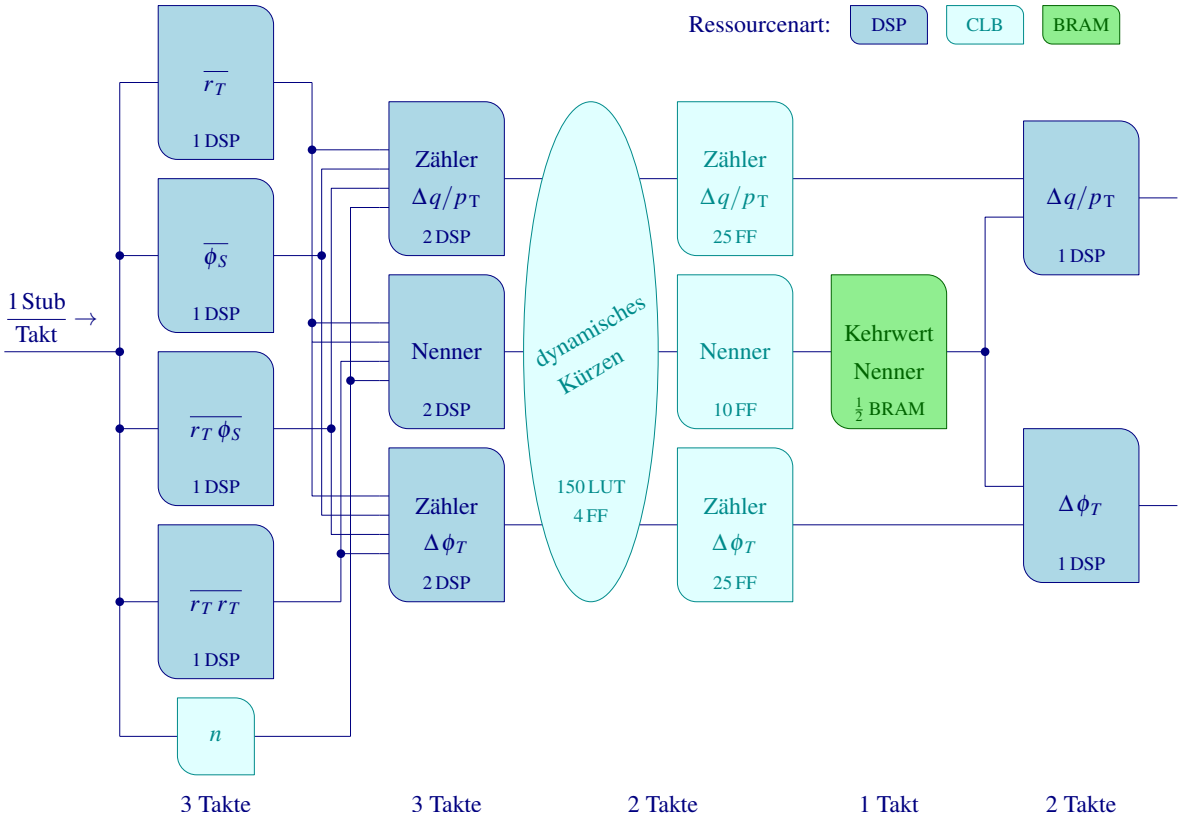

Abbildung 7.13 Blockschaltbild des Regressionskerns. Der Ressourcenbedarf und die Laufzeit der einzelnen Prozessschritte sind dargestellt.

Um die Zähler und den Nenner zu berechnen, benötigen wir die Summen $\overline{r_T}$, $\overline{\phi}$, $\overline{r_T\,\phi}$ und $\overline{r_T\,r_T}$, welche wir jeweils mit einem DSP-Slice generieren, und n, welche wir mit einem 3-Bit-Zähler realisieren. Dieser Zähler benötigt nur einen Bruchteil eines CLB-Slices. Die beiden Zähler und der Nenner bestehen jeweils aus einer Differenz von zwei Produkten, deren Faktoren durch je eine der eben genannten Summen gegeben ist. Für jedes Produkt verwenden wir ein DSP-Slice, wobei jedes zweite Slice zusätzlich die Differenz berechnet.

Beruhend auf Simulationsdaten haben wir die numerischen Maximalwerte der endgültigen Summen bestimmt. Dadurch konnten wir die Produkte ohne Rundungsfehler mit nur einem DSP-Slice ausführen. Die numerischen Maximalwerte der endgültigen Differenzen, also der beiden Zähler und Nenner, haben wir auch bestimmt. Hier ist es allerdings so, dass wir für den Zähler von $\Delta\phi_T$ 30 Bits benötigen, während der Nenner und der andere Zähler weniger als 25 Bits benötigen.

Da wir später den Zähler mit dem Kehrwert des Nenners multiplizieren wollen, kürzen wir den Zähler von $\Delta\phi_T$ um 5 Bits und erzeugen so einen Rundungsfehler. Dadurch wird allerdings später nur ein DSP-Slice benötigt. Die maximale Breite des Nenners beträgt 24 Bit, was viel zu groß ist, um diesen Wert als Adresse eines Speichers zu verwenden, da man einen riesigen Speicher $\mathcal{O}(100\,\mathrm{Mb})$ benötigen würde. Damit man nur den kleinsten verfügbaren Speicher benötigt, einen halben BRAM, müsste man die Breite auf 10 Bit reduzieren.

Dazu verwenden wir eine Methode, die wir dynamisches Kürzen nennen. Die Idee ist, den Zähler und den Nenner um so viele Zweierpotenzen zu kürzen wie nötig, damit der numerische Wert des gekürzten Nenners kleiner als $2^{10} - 1$ ist.

Dies geschieht in zwei Schritten. Im ersten Schritt wird ermittelt, um wie viele Zweierpotenzen gekürzt werden muss. Diese Zahl liegt zwischen 0 und 14. Der zweite Schritt besteht im Ausführen der Schiebeoperationen für beide Zähler und den Nenner um eben dieser Anzahl an Stellen. Im Gegensatz zum statischen Schieben, welches kostenfrei für den FPGA ist, werden hierfür allerdings Ressourcen und ein Takt benötigt. Dabei überlassen wir die genaue Umsetzung dem Synthesewerkzeug. Das Syntheseresultat benötigt dafür nur 150 LUTs, was uns zufriedenstellt.

Der gekürzte Nenner dient nun als Adresse für einen halben BRAM. In diesem liegen 1024 Kehrwerte gespeichert, welche mit 18 Bit dargestellt werden. Der letzte Schritt besteht in der Multiplikation des nachgeschlagenen gekürzten Nenners mit den gekürzten Zählern.

Die Berechnung der Spurparameter in der *r*-*z*-Ebene, siehe Gl. 7.18 und 7.19, funktionieren komplett analog. Die einzigen nennenswerten Unterschiede sind, dass für den maximalen numerischen Wert des ungekürzten Nenners 23 Bit und für den ungekürzten Zähler von z_t 29 Bit benötigt werden.

Die gesamte Laufzeit dieser Implementierung ist durch die Anzahl der einzulesenden Stubs plus 11 gegeben. In Anbetracht der Taktfrequenz von 240 MHz und einer maximalen Stubanzahl von 16 pro Spurkandidat beträgt die Laufzeit höchstens 112,5 ns.

7.2.3 Validierung der Division

Der Ressourcenbedarf und die Laufzeit fallen für den Regressionskern gering aus, allerdings stellt sich die Frage, ob die Präzision ausreicht, um als Fit zu dienen. Um das herauszufinden, haben wir mit der Simulationssoftware die Ergebnisse einerseits mit einer numerischen Präzision der Fließkommadarstellung und andererseits mit den genannten Präzisionen berechnet und miteinander verglichen.

Als Daten dienen uns die Stubs der Teilchen für Qualitätsstudien aus Tab. 5.2, wobei wir wieder, wie für die Bestimmung der Detektorauflösung, die den Teilchen zugeordneten Stubs bereinigen. Abb. 7.14 bis 7.17 zeigen die Auflösungen der vier Spurparameter. Hierbei ist die Auflösung eines Parameters über die mittlere absolute Abweichung von dem rekonstruierten Parameter zum simulierten Parameter definiert.

Für die Parameter in der r-ϕ-Ebene können wir keine Einbußen der Spurparameterauflösung aufgrund der ganzzahligen Implementierung des Regressionskerns samt Division feststellen. In der r-z-Ebene fallen diese sehr gering aus. Die Breiten der Spurparameter wurden so groß gewählt und damit die Basen so klein, dass keine Einbußen der Spurparameterauflösung sichtbar sind.

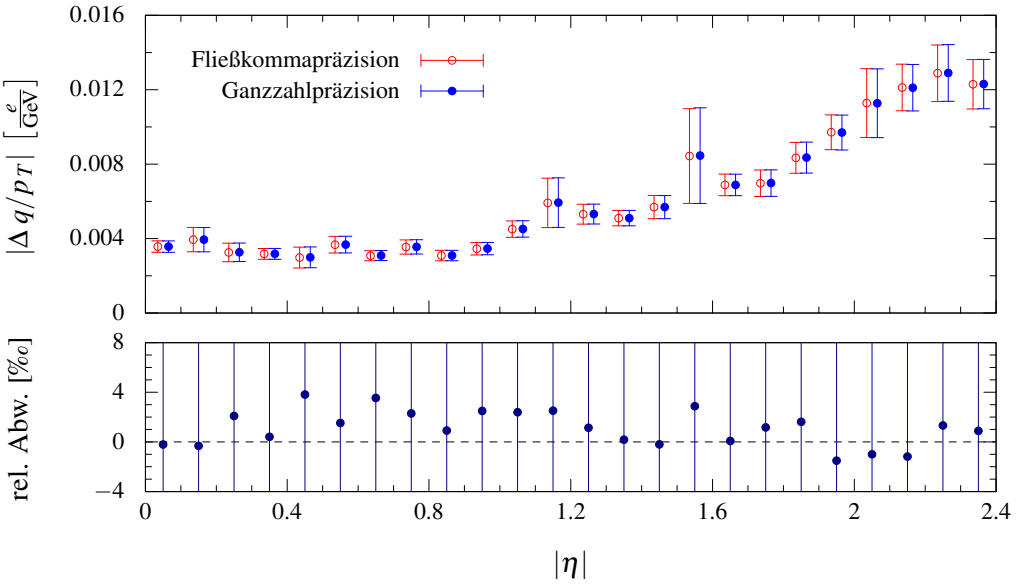

Abbildung 7.14 q/p_{T} - Auflösung des Regressionskerns, aufgetragen über $|\eta|$. In Rot ist sind die Ergebnisse beruhend auf Fließkommazahlen und in Blau die Ergebnisse beruhend auf den im Text beschriebenen Ganzzahlen eingezeichnet. Der untere Abschnitt zeigt die relative Abweichung von der Ganzzahlpräzision zur Fließkommapräzision.

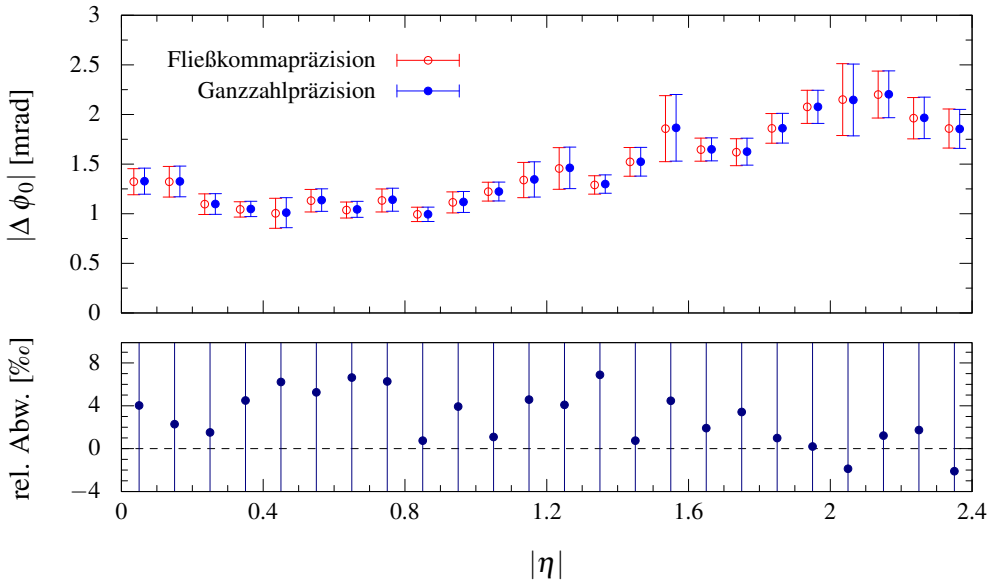

Abbildung 7.15 ϕ_0 - Auflösung des Regressionskerns analog zu Abb. 7.14.

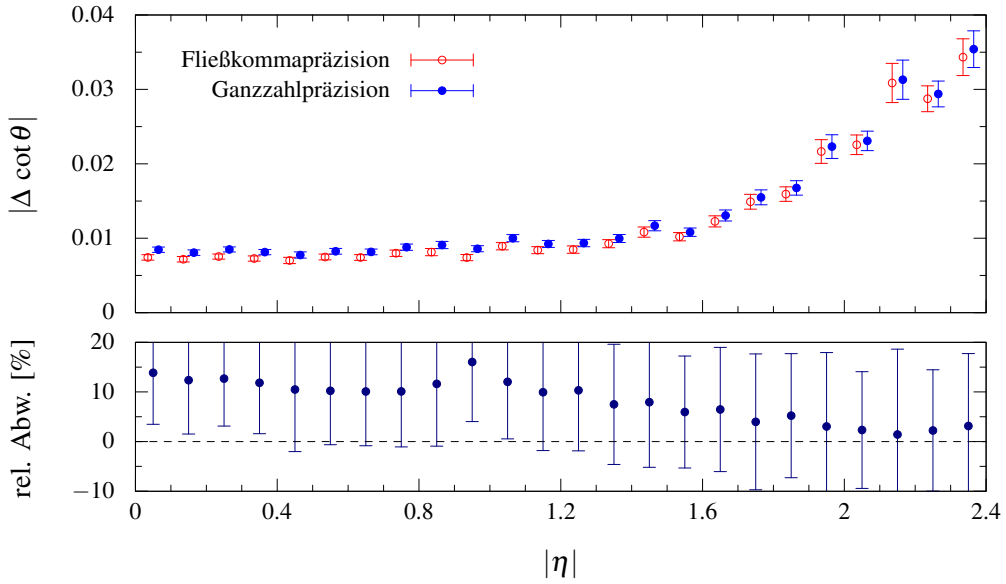

Abbildung 7.16 $\cot\theta$ - Auflösung des Regressionskerns analog zu Abb. 7.14.

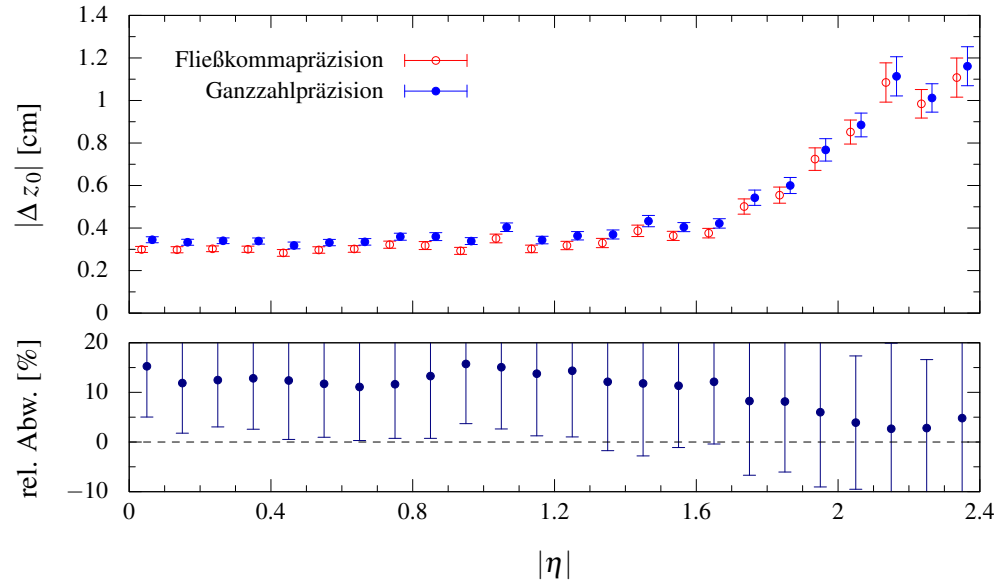

Abbildung 7.17 z_0 - Auflösung des Regressionskerns analog zu Abb. 7.14.

7.2.4 Globale Lineare Regression

Die globale Lineare Regression, siehe Abschnitt 5.4, ist für den Spurtrigger besonders interessant, da der Rechenaufwand linear mit der Stubanzahl skaliert. Dadurch kann das System so gebaut werden, dass die vollständige Verarbeitung der Spurkandidaten von der Hough-Transformation innerhalb einer konstanten Laufzeit gewährleistet ist.

Da die Hough-Transformation nur Kandidaten mit bis zu 16 Stubs erzeugen kann und eine rekonstruierbare Spur nach Tab. 5.2 mindestens vier Stubs von verschiedenen Detektorlagen aufweist, müssen höchstens 12 Stubs entfernt werden. Daher verwenden wir 12 Iterationen, wobei jede Iteration aus einer Berechnung der Spurparameter, einer Berechnung der Residuen, der möglichen Identifikation eines Störstubs und dessen Entfernung besteht. Die Berechnung der ungewichteten Teilresiduen χ_ϕ und χ_z sind sehr simpel aufgrund der linearen Bewegungsgleichungen. Sie sind gegeben durch

$$\chi_\phi = \left|\phi_S - \left(\Delta\phi_T + r_T \cdot \Delta q/p_\mathrm{T}\right)\right| , \tag{7.24}$$

$$\chi_z = \left|z_s - \left(z_t + r_t \cdot \Delta \cot\theta\right)\right| . \tag{7.25}$$

Das Residuum χ eines Stubs berechnen wir mit

$$\chi = \frac{\chi_\phi}{w_\phi} + \frac{\chi_z}{w_z} , \tag{7.26}$$

wobei w_ϕ und w_z Gewichte darstellen. Um die Division in der Berechnung des Residuums zu umgehen, wählen wir konstante Gewichte, da wir dann mit den im Vorfeld berechneten Kehrwerten multiplizieren können. In der r-z-Ebene wählen wir als Gewicht 0,7 mm für Stubs von PS-Modulen und 2,5 cm für Stubs von 2S-Modulen, also jeweils die halbe Sensorlänge. Für Stubs der Endkappen wird das Residuum zusätzlich mit den $\cot\Theta$ des Sektors gewichtet, da die Streifen und Pixel in den Endkappen in $\hat{r}$-Richtung zeigen. In der r-ϕ-Eben haben wir 1 mrad als Gewicht gewählt. Dieser Wert ist in Anbetracht des großen Hebelarms des äußeren Spurdetektors und der kleinen Streifen- beziehungsweise Pixelabstände von 90µm und 100 µm sehr groß und kompensiert somit die Unsicherheit der Teilchenflugbahn aufgrund von Vielfachstreuungen der Teilchen mit dem Detektor und der Ungewissheit des Impaktparameters.

Da wir an präzisen Spurparametern interessiert sind, benötigen wir mindestens zwei Stubs von PS-Modulen aus verschiedenen Detektorlagen. Spurkandidaten, die diese Bedingung nicht erfüllen, werden nicht erfolgreich rekonstruiert.

Da [17] eine Spur nur als erfolgreich rekonstruiert betrachtet, wenn sie sich mit einem Teilchen assoziieren lässt und ausschließlich Stubs enthält, die mit diesem Teilchen assoziiert sind, entfernen wir so lange Stubs, bis nur noch vier Stubs von verschiedenen Detektorlagen übrig sind, wobei davon mindesten zwei Stubs von PS-Modulen stammen. Dazu entfernen wir grundsätzlich mit jeder Iteration den Stub mit dem größten Residuum. Stubs, dessen Entfernung dazu führt, dass nur noch Stubs von drei Detektorlagen oder nur noch Stubs von PS-Modulen und einer Detektorlage vorhanden sind, werden bei der Bestimmung des Stubs mit dem größten Residuum nicht berücksichtigt.

So erhalten wir spätestens in der zwölften Iteration einen Spurkandidaten, der nur noch vier Stubs besitzt. Ist das größte Residuum dieser vier Stubs kleiner als zwei, bildet die Lineare Regression einen Spurkandidaten aus diesen vier Stubs und den letzten berechneten Spurparametern.

7.2.5 Resultate

Wir haben die globale Lineare Regression bisher nur in Software implementiert, bis auf den Regressionskern, der von uns auch in Hardware implementiert wurde, da nur die Implementierbarkeit dieses Teiles der Linearen Regression fragwürdig war. Als Eingangsdaten für die Lineare Regression dienen uns die mit der Hough-Transformation gefundenen Spurkandidaten, siehe Abschnitt 7.1.4.

Wir definieren die Effizienz durch das Verhältnis der erfolgreich rekonstruierten Spuren zu den zu rekonstruierenden Spuren. Die zu rekonstruierenden Spuren sind durch die Teilchen für Qualitätsstudien aus Tab. 5.2 gegeben. Eine Spur gilt als erfolgreich rekonstruiert, wenn sie sich mit einem Teilchen assoziieren lässt und ausschließlich Stubs enthält, die mit diesem Teilchen assoziiert sind. Damit ein Spurkandidat sich mit einem Teilchen assoziieren lässt, muss es mindestens vier Stubs von verschiedenen Detektorlagen mit diesem gemeinsam haben.

Die Auflösung der Spurparameter bestimmen wir mit den erfolgreich rekonstruierten Spurkandidaten, welche sich mit den Teilchen für Qualitätsstudien assoziieren lassen. Dazu bilden wir die mittleren absoluten Abweichungen zwischen den jeweiligen rekonstruierten Spurparametern zu den generierten.

Damit wir die Leistung der globalen Linearen Regression einschätzen können, vergleichen wir die Resultate mit den Resultaten eines Kalman-Filters [19], auf den später nochmal eingegangen wird und uns jetzt als Referenz dient. Tab. 7.2 zeigt die Effizienz und die Anzahl der gebildeten Spurkandidaten für die Hough-Transformation und für die anschließende Lineare Regression beziehungsweise für den anschließenden Kalman-Filter. Abb. 7.18 bis 7.21 zeigen die Auflösungen der vier Spurparameter für die globale Lineare Regression und den Kalman-Filter.

Tabelle 7.2 Die mittlere Effizienz und die mittlere Spurkandidatenanzahl pro Ereignis für die Hough-Transformation (HT), die globalen Linearen Regression (LR) und den Kalman-Filter (KF), welcher als Referenz dient. Als Eingangsdaten für die HT dienen die relevanten Stubs aus Tab. 5.2. LR und der KF verwenden die mit der HT gefundenen Spurkandidaten als Eingangsdaten. Die Resultate wurden mit der Simulationssoftware gewonnen.

Schritt	Effizienz [%]	# Kandidaten
HT	97,2	334
KF	95,1	164
LR	93,9	181

Der Kalman-Filter ist effizienter und generiert weniger Spurkandidaten, weist also eine kleiner Fehlidentifikationsrate auf als die globale Lineare Regression. Die Auflösungen der Spurparameter sind weitestgehend vergleichbar. Allerdings ist die Lineare Regression wesentlich simpler und möglicherweise im Spurtrigger schneller, billiger und zuverlässiger als der Kalman-Filter. Um hier eine Entscheidung zu treffen, wäre es wichtig zu wissen, ob ein Effizienzverlust von 1,2 % tragbar ist.

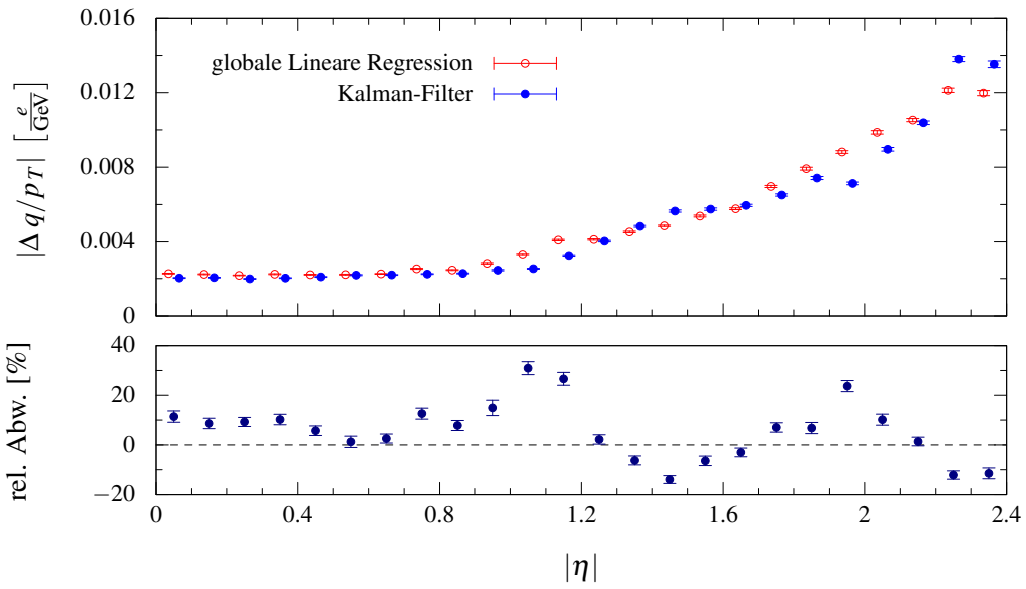

Abbildung 7.18 q/p_{T} - Auflösung der globalen Linearen Regression in Rot und in Blau die des Kalman-Filters aufgetragen über $|\eta|$. Der untere Abschnitt zeigt die relative Abweichung von der globalen Linearen Regression zum Kalman-Filter.

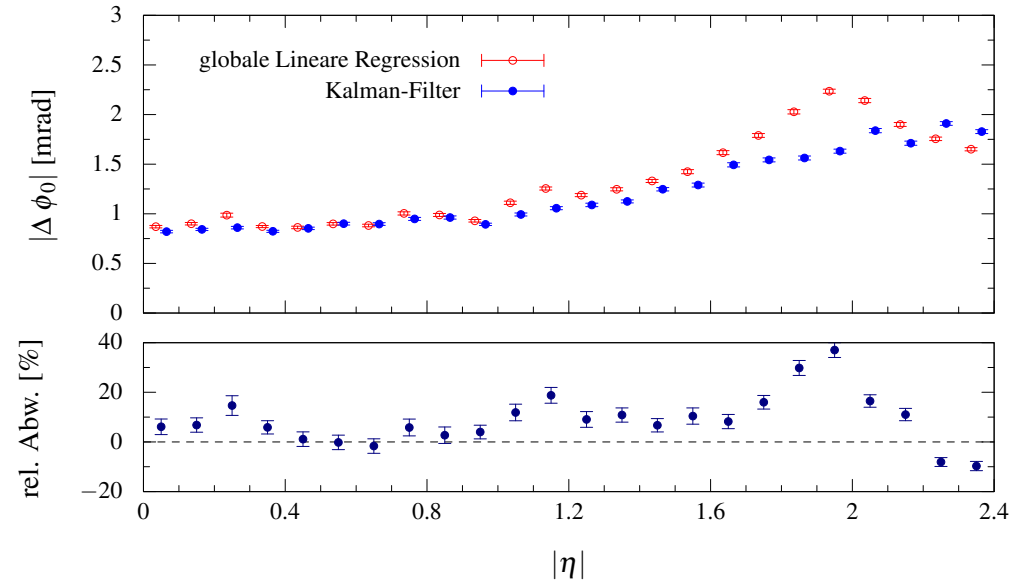

Abbildung 7.19 ϕ_0 - Auflösung der globalen Linearen Regression analog zu Abb. 7.18.

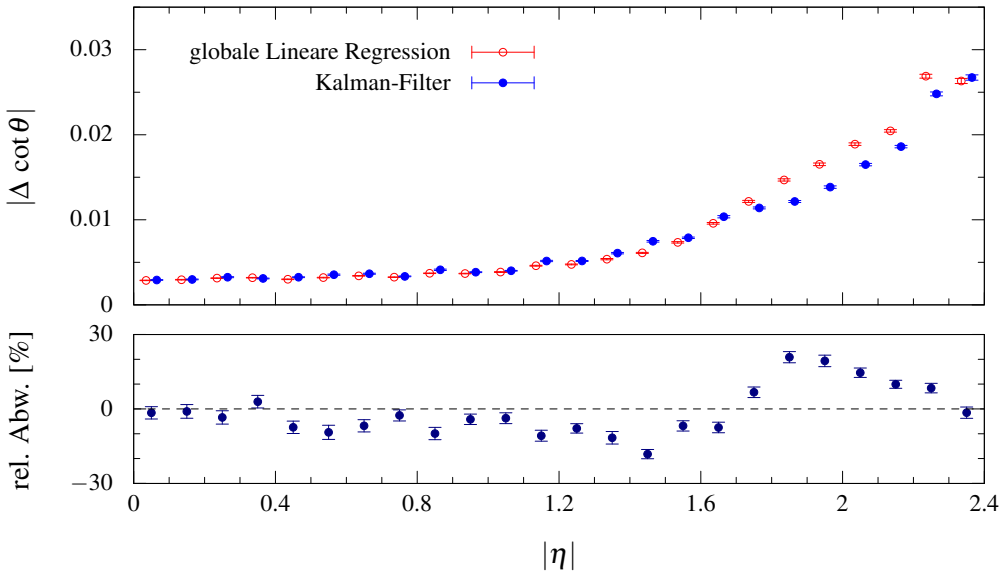

Abbildung 7.20 $\cot\theta$ - Auflösung der globalen Linearen Regression analog zu Abb. 7.18.

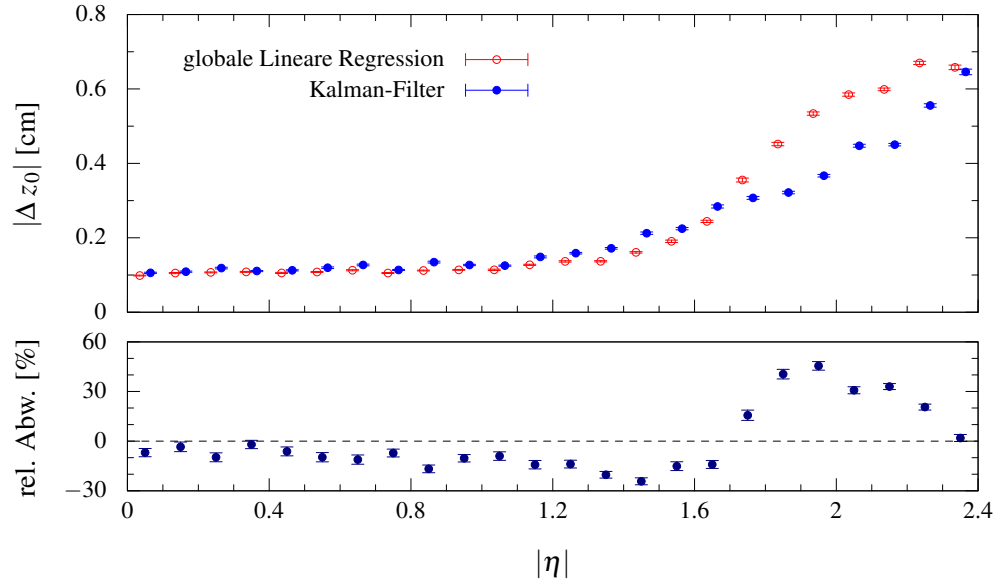

Abbildung 7.21 z_0 - Auflösung der globalen Linearen Regression analog zu Abb. 7.18.

Kapitel 8

Demonstratorarchitektur

Es gibt eine Vielzahl an Möglichkeiten, einen Spurtrigger aufzubauen. Angefangen bei der Wahl der Algorithmen, der Wahl der Elektronik, auf der die Algorithmen implementiert werden sollen und die Wahl, wie die Algorithmen auf die Elektronik abgebildet werden. Innerhalb von CMS haben sich drei Ansätze etabliert, welche TMTT-, AM-, und Tracklet-Ansatz genannt werden.

Der TMTT-Ansatz verwendet unsere Implementierung der Hough-Transformation und wird noch im Detail betrachtet. Der AM-Ansatz benutzt sowohl FPGAs als auch ASICs und sucht nach Spuren mittels frei wählbarer Schablonen. Dies ähnelt sehr der Hough-Transformation, wodurch sich diese beiden Ansätze gut vergleichen lassen. Der Tracklet-Ansatz verwendet nur FPGAs und beruht auf einer klassischen propagierenden Spursuche.

Um die Realisierbarkeit der jeweiligen Ansätze zu untersuchen, wurde für jeden Ansatz ein Demonstrator gebaut. Dieser soll ein Achtel des äußeren Spurdetektors (d.h. einen 45° ϕ-Sektor), auch Detektoroktant genannt, abdecken.

Demonstriert wird weder das Data-, Trigger- and Control-System (DTC), noch der Teil des Spurtriggers, welcher die rekonstruierten Spuren verwendet, sondern nur die Spurrekonstruktion an sich.

Die Anbindung des Detektors an die DTCs ist für den Demonstrator nicht genau festgelegt. Allerdings ist diese derart eingeschränkt, sodass der Detektoroktant an nicht mehr als 36 DTCs angeschlossen wird. Die maximale Bandbreite, die der Demonstrator verarbeiten muss, beträgt somit 21,6 Tb/s.

8.1 Systemarchitektur

Der TMTT-Ansatz verwendet die sogenannte „time multiplexed architecture“ [29]. Der Grundgedanke dieser Architektur besteht darin, dass man alle Stubs eines Ereignisses zu einem FPGA transportiert, welcher die Spurrekonstruktion durchführt.

Der reine Transfer aller Stubs eines Ereignisses zu einem FPGA wird man zwar rein technologisch, aufgrund der begrenzten Bandbreite eines FPGAs, nicht für den gesamten Detektor umsetzen können, allerdings ist dies für einen Oktanten möglich. Auch aufgrund der limitierten Ressourcen wird man die Rekonstruktion von Spuren aus den Stubs eines Oktanten nicht auf nur einem FPGA durchführen können. Daher ersetzt man den einen FPGA durch eine Kette von FPGAs, wobei der erste FPGA dieser Kette alle Stubs eines Oktanten empfängt und der letzte FPGA der Kette die rekonstruierten Spuren sendet. Diese Kette von FPGAs fasst man als logische Einheit auf und nennt sie den Track-Finder-Prozessor (TFP).

Ein Vorteil dieser Architektur besteht darin, dass man die TFPs derartig an die DTCs anschließen kann, dass keine Kommunikation zwischen den TFPs notwendig ist, was den gesamten Prozessablauf vereinfacht und die Komplexität der einzelnen TFPs reduziert. Ein weiterer Vorteil besteht darin, dass nur ein TFP benötigt wird, um die Funktion des gesamten Systems zu demonstrieren. Um diese Verbindung zu ermöglichen, werden Prozessoktanten definiert, welche auch 45° in ϕ abdecken, aber zu den Detektoroktanten um die Hälfte verschoben sind.

Ein Prozessoktant besteht aus 36 TFPs, wobei jedes TFP ein anderes Ereignis prozessiert und mit den 72 DTCs der beiden Detektoroktanten verbunden ist. Somit ist auch jedes DTC eines Detektoroktanten mit den 72 TFPs der beiden Prozessoktanten verbunden. Abb. 8.1 zeigt diesen Aufbau.

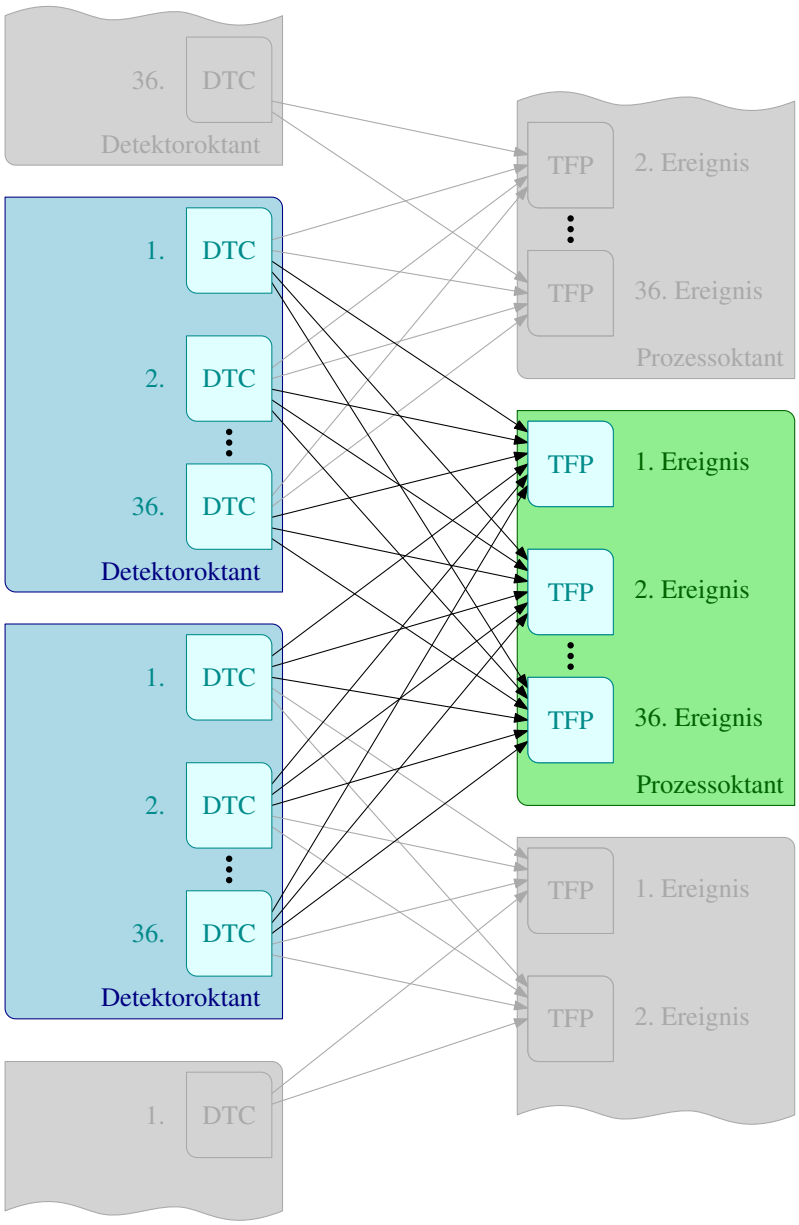

Abbildung 8.1 Systemarchitektur des TMTT-Ansatzes. Der Übersicht wegen wurden nur 3 der 36 DTCs in einem Detektoroktanten sowie nur 3 der 36 TFPs in einem Prozessoktanten abgebildet. Jeder Pfeil entspricht einem Lichtwellenleiter.

Die Bandbreite von einem DTC zu einem TFP beträgt 8 Gb/s und zu einem Prozessoktanten $36 \times 8\,\mathrm{Gb/s} = 288\,\mathrm{Gb/s}$. Diese Bandbreite sollte ausreichend sein, stellt aber möglicherweise

einen Flaschenhals dar. Da die genaue Verkabelung der DTCs an den Detektor nicht spezifiziert ist und der DTC nicht Bestandteil des Demonstrators ist, wird dieser potentielle Flaschenhals durch den Demonstrator nicht ersichtlich.

Der erste Prozessschritt des DTCs ist die Konvertierung der Stubdaten der Sensormodule in ein globales Koordinatensystem. Um ein Stub in diesem Koordinatensystem adäquat zu beschreiben, werden 48 Bit benötigt [19]. Danach kann ein Stub dem einen oder anderen Prozessoktant zugeordnet werden. Falls ein Stub mit beiden Prozessoktanten konsistent ist, wird er dupliziert und beiden zugeordnet.

Dieser Fall tritt aufgrund der Krümmung einer Spur mit einem Transversalimpuls von 3 GeV dann ein, wenn ein Stub mit geladenen Teilchen aus beiden Detektoroktanten vereinbar wäre. Durch das Ausnutzen der stubeigenen Transversalimpulsinformation lässt sich die Zahl der zu duplizierenden Stubs reduzieren. Die gesamte Laufzeit für das Extrahieren, Formatieren und das örtliche Zuordnen der Stubs im DTC wird auf 250 ns geschätzt. Die Datenübertragung vom DTC zum TFP beträgt ungefähr weitere 150 ns.

Der TFP ist in vier logische Einheiten unterteilt:

- Geometric-Processor (GP): Vorbearbeitungsstufe. Der GP unterteilt den Oktanten in weitere feinere Sektoren in η und ϕ. Dies erleichtert die anschließende Spursuche und erhöht die Parallelität und somit die Prozessbandbreite.
- Hough Transform (HT): Hochgradig parallelisierte Spurfindung. Die HT bildet einen Satz von Stubs, die auf einem Kreis in der r-ϕ Ebene liegen. Dies reduziert massiv die Kombinatorik und liefert grobe Spurparameter.
- Kalman Filter (KF): Spurbereinigung und präziser Fit. Der KF entfernt falsche Kandidaten und erhöht Spurparameterauflösung.
- Duplicate Removal (DR): Entfernung von Duplikaten. Das DR benutzt präzise Spurinformation, um die Duplikate zu entfernen, welche von der Hough Transformation kreiert wurden.

Die Firmwareblöcke und deren Verbindung sind in Abb. 8.2 dargestellt.

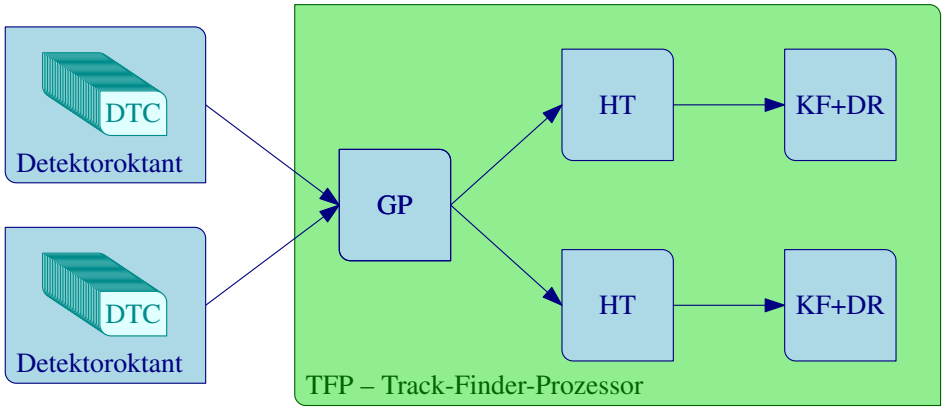

Abbildung 8.2 Blockschaltbild des Demonstrators. Jeder blauer Block entspricht einer MP7-Karte. Adaptiert von [19].

8.2 Geometric-Processor (GP)

Der GP wird auf einer MP7-Karte realisiert. Diese MP7-Karte ist hypothetisch mit einem eingehenden Lichtwellenleiter an einer DTC-Karte angeschlossen. Der GP verarbeitet die DTC-Stubs eines Prozessoktanten, welche mit 48 Bit dargestellt werden. Das Stubformat wird auf 64 Bit erweitert, wodurch die Hough-Transformation erleichtert wird. Des Weiteren werden die Stubs den Sektoren aus Abb. 7.2 und 7.3 zugeordnet.

Die Firmware besteht aus einem mathematischen Block, welcher die korrekten Sektoren der Stubs aus den globalen Koordinaten berechnet, gefolgt von einem mehrschichtigen Verteilungsblock. Die Stubs, die zu einem Sektor gehören, werden zu einem dedizierten ausgehenden Lichtwellenleiterpaar geführt. Ein Lichtwellenleiterpaar überträgt einen Stub pro Takt mit einer Taktfrequenz von 240 MHz.

Die Abb. 7.3 und 7.2 zeigen, wie der GP einen Oktanten in 36 Sektoren unterteilt und verdeutlicht die Überlappregionen. Die r-ϕ Ebene wird zweigeteilt und die r-z Ebene wird 18 Mal unterteilt. Die relativ schmalen η-Sektoren führen zwar dazu, dass sich die Überlappregionen in der r-z-Ebene überlappen und somit manche Stubs vervielfachen, bringen aber den Vorteil, dass die anschließend in der r-ϕ Ebene gefundenen Spuren in der r-z Ebene in guter Näherung auf einer Geraden liegen.

Ein Sektor wird über Bereiche in den schon erwähnten ϕ_T und z_t Koordinaten, siehe Gl. 7.2 und 7.15, definiert. Diese Werte wurden weitestgehend so gewählt, dass die Anzahl der Stubs, die mit mehreren Sektoren konsistent sind und daher vervielfacht werden, minimal wird. In der r-ϕ-Ebene wurde allerdings ein größerer Abstand gewählt, um Ressourcen für die Hough-Transformation einzusparen, welche einen Abstand von 65 cm, dem radialen Mittelpunkt des Detektors, präferiert. Die 50 cm entsprechen dem mittleren Abstand aller relevanten Stubs aus Tab. 5.2.

Die Bereiche in ϕ_T und z_t sind exklusiv, das heißt, dass sich zwei benachbarte Sektoren nicht überlappen. In der r-ϕ Ebene sind die Bereiche gleich groß, während die Größe der Bereiche in der r-z so optimiert sind, dass die resultierenden Stubraten möglichst gleichmäßig verteilt sind.

Der GP muss Stubs den Sektoren zuordnen, deren Bereiche in ϕ_T und z_t mit allen möglichen Spuren zwischen der primären Wechselwirkungszone und dem Stub vereinbar sind. Dies ist nötig aufgrund der Krümmung der Trajektorie von geladenen Teilchen im Magnetfeld, eingeschränkt durch eine konfigurierbare Transversalimpulsschwelle (typischerweise 3 GeV) oder aufgrund der Ausdehnung der primären Wechselwirkungszone in Strahlrichtung. Deren halbe Länge ist auch ein konfigurierbarer Parameter für den GP, welcher typischerweise bei 15 cm liegt.

8.3 Hough-Transformation (HT)

Die Hough-Transformation eines TFPs besteht aus 36 Hough-Histogrammen, welche auf zwei MP7-Karten verteilt sind. Eine einzelne MP7-Karte hat zwar eine ausreichende optische Anbindung, um 36 Hough-Histogramme anzuschließen, allerdings reichen die restlichen Ressourcen eines FPGAs nicht aus. Hierbei stellen sich die BRAM-Einheiten als die begrenzende Ressource heraus. Jede MP7-Karte beinhaltet daher 18 Hough-Histogramme und benötigt nur 36 der 72 Hochgeschwindigkeitsreceiver.

Bei der Bildung der Ausgangspakete ergibt sich ein Problem. Die Hough-Transformation reduziert zwar im Schnitt die Stubanzahl um eine Größenordnung, aber nicht immer. In Jets

kann sich die Stubanzahl sogar vergrößern. Um auch in diesem Fall dennoch alle Stubs der gefundenen Spurkandidaten auszulesen, werden die Stubs der 18 Histogramme auf einem FPGA vermengt und über alle 72 Hochgeschwindigkeitstransmitter versendet.

8.4 Kalman-Filter (KF)

Um die mit der Hough-Transformation gefundenen Spuren zu fitten wurde für den Demonstrator der Kalman-Filter gewählt. Der Filter startet mit geschätzten Spurparametern und deren Unsicherheit, auch Zustand genannt. Die Stubs werden iterativ genutzt, um den Zustand nach dem Kalman-Formalismus [30] zu aktualisieren und die Unsicherheit der Spurparameter zu reduzieren.

Typischerweise weisen über die Hälfte der gefundenen Spurkandidaten, die einem Teilchen zugeordnet werden können, zumindest einen Stub auf, welcher nicht mit dem Teilchen assoziiert ist. Besonders wenn diese Stubs weit von der Trajektorie entfernt sind, würden diese das Fit-Ergebnis stark verfälschen. Daher ist die Entfernung solcher Stubs sinnvoll.

Darüber hinaus können die Hälfte aller Spurkandidaten keinen Teilchen zugeordnet werden. Das Entfernen solcher Kandidaten ist erwünscht, allerdings nur mit minimalen Verlusten an Effizienz. Der Kalman-Filter ist in der Lage während des Iterierens inkorrekte Stubs zu erkennen und zu entfernen, um die bestmöglichen Spurparameter zu erhalten.

Realisiert wird der Kalman-Filter auf zwei MP7-Karten. Je eine Karte empfängt die gefundenen Spuren von einer HT-Karte.

8.5 Duplicate-Removal (DR)

Über die Hälfte der mit dem Kalman-Filter gefitteten Spurkandidaten sind unerwünschte Spuren, die von der Hough-Transformation mehrfach gefunden wurden. Die Aufgabe dieses Blocks ist deren Entfernung.

Der Algorithmus beruht auf dem Verständnis, wie die Hough-Transformation Duplikate erzeugt. In Abb. 8.3 wird das Beispiel aus Abb. 7.1, in dem aus sechs Stubs drei Kandidaten geformt werden, vergrößert dargestellt. Die entsprechenden Zellen sind grün und blau markiert. Da diese Kandidaten dieselben Stubs beinhalten, werden sie dieselben gefitteten Spurparameter erhalten. Diese präzisen Spurparameter sollten in der blauen Zelle liegen, wo sich die zu Geraden transformierten Stubkoordinaten schneiden.

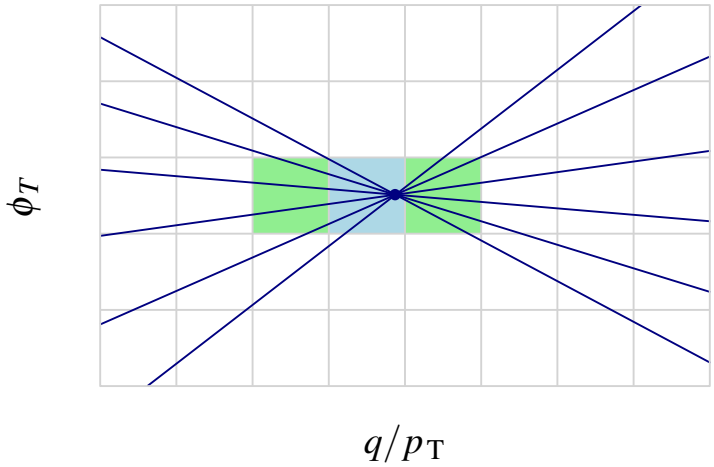

Abbildung 8.3 Veranschaulichung der Bildung von Duplikaten in der Hough-Transformation. Zu sehen ist ein Ausschnitt des Hough-Histogramms aus Abb. 7.1. Die Zellen, die Spurkandidaten darstellen, sind farblich markiert. Die grünen Zellen stellen Duplikate dar.

Duplikate erkennt man daher daran, dass die präzisen gefitteten Spurparameter nicht mit den groben gefundenen Spurparametern der $(\phi_T, q/p_\mathrm{T})$-Matrix vereinbar sind. Dies ist eine äußerst simple Methode, da man nur einzelne Kandidaten betrachten und keine Paare von Kandidaten vergleichen muss, um zu überprüfen, ob diese identisch sind.

Allerdings birgt dieser Algorithmus ein kleines Problem, denn er verliert ein paar Prozent Effizienz aufgrund von Auflösungseffekten. Daher wird der Algorithmus erweitert, um diesen Effizienzverlust zu reduzieren. Dazu werden die Duplikate, deren grobe gefundenen Spurparameter mit keinen präzisen gefitteten Spurparametern eines anderen Kandidaten vereinbar sind, doch nicht als Duplikate identifiziert.

Dies stellt zwar eine Vorkomplizierung dar, diese lässt sich aber effizient mit wenigen Ressourcen auf einem FPGA realisieren. Daher passt dieser Algorithmus sogar auf dieselben FPGAs, auf denen der Kalman-Filter realisiert ist.

8.6 Demonstrator

Um den TMTT-Ansatz für einen Detektoroktanten in echter Hardware zu implementieren und die Qualität der Spurrekonstruktion innerhalb der Laufzeitgrenzen zu validieren, wurde ein Demonstrator aufgebaut. Als Eingangsdaten des Demonstrators dienen die relevanten Stubs aus Tab. 5.2, welche auf der in Abb. 4.5 dargestellten Detektorgeometrie beruhen.

Dieses Teilsystem entspricht einem TFP, welcher konstruiert wurde, um das Konzept mit heutiger Technologie zu demonstrieren. Da die TFPs vollkommen unabhängig voneinander operieren, kann der gesamte Detektor abgebildet werden, obwohl das Teilsystem nur ein Achtel des Detektors in ϕ abdeckt. Dazu werden die Stubs der acht Oktanten sequentiell abgearbeitet.

Der Demonstrator befindet sich in der Tracker Integration Facility (TIF) am CERN und besteht aus einem Schroff dual MicroTCA Crate [31]. Dieses ist bestückt mit einem kommerziell erhältlichen NAT MicroTCA Carrier Hub (MCH) für eine Gigabit-Ethernet-Kommunikation über die Backplane und einer CMS-spezifischen Karte namens AMC13 [32], welche unter anderem einen gemeinsamen Takt für die restlichen Karten im Crate bereitstellt. Die TFP-Algorithmen sind auf einem Satz von fünf MP7-Karten implementiert.

Die MP7-Karten sind mit einem Xilinx Virtex-7 XC7VX690T FPGA und 12 Avgo Technologies MiniPOD optischen Transceivern, welche an 72 Lichtwellenleiterpaare angeschlossen werden können, ausgestattet. Für den Demonstrator sind die Transceiver auf 10 Gb/s mit einer 8 b-zu-10 b Enkodierung eingestellt, sodass sich eine effektive Transferrate von 8 Gb/s pro Lichtwellenleiter ergibt.

Das MP7 beinhaltet eine gewisse Infrastruktur, welche die Transceiver, Serialisierung und Deserialisierung, Daten-Puffer, I/O-Formatierung, Karten- und Taktkonfiguration, sowie die

externe Kommunikation mittels Ethernet-Schnittstelle bereitstellt. Die Algorithmus-Firmware ist somit von diesen Aufgaben entkoppelt, wodurch ein System wie der Demonstrator leicht aus mehreren miteinander mittels Lichtwellenleitern verbundenen Karten zusammengebaut werden kann. Eine Einteilung des Demonstrators in dieser Weise, führt dazu, dass man die anfallende Firmware-Entwicklung einfach aufteilen kann. Darüber hinaus kann der Ressourcenbedarf und die Leistung des Gesamtsystems abgeschätzt werden, ohne von der derzeit zur Verfügung stehenden Technologie beschränkt zu sein.

Es werden insgesamt acht MP7-Karten benutzt, um die gesamte Verarbeitungskette abzubilden. Dabei dienen zwei Karten als Quelle. Diese beiden Karten repräsentieren die 36 DTCs eines Detektoroktanten und sind als ein großer Stub-Speicher implementiert. Sie können mit simulierten Stubs von bis zu 30 Ereignissen mittels IPBus [33] geladen werden. Jeder Lichtwellenleiter überträgt die vorformatierten 48-Bit-Stubs eines DTCs in den GP. Eine Karte dient als Senke, welche in der Lage ist, die gefundenen Spuren der 30 Ereignisse aufzunehmen und kann wieder mittels IPBus ausgelesen werden. Abb. 8.4 zeigt den Demonstrator.

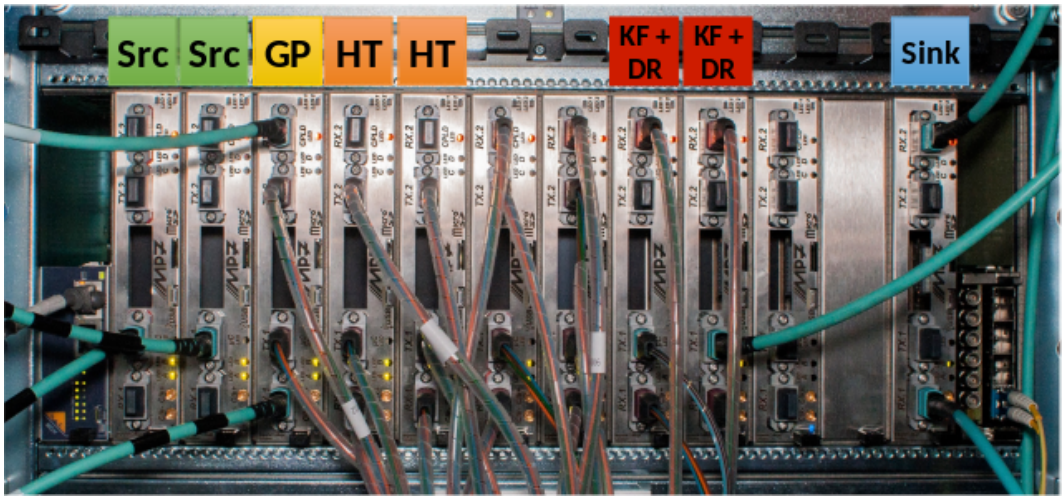

Abbildung 8.4 Photographie des Demonstrators [19].

Kapitel 9

Resultate mit dem Demonstrator

Mit der Simulations-Software von CMS (CMSSW) wurden Physik-Ereignisse (typischerweise $t\bar{t}$-Produktion) mit 200 Überlappereignissen simuliert und die Wechselwirkung der Teilchen mit dem Detektor modelliert. Um die Leistung des Demonstrators zu studieren, wurde Software entwickelt, welche Stubs von den simulierten Ereignissen in den Demonstrator einspeist und die Algorithmen des Demonstrators emuliert.

Dazu werden die Stubs in ein Textformat konvertiert und mittels IPBus übertragen. Die vom Demonstrator rekonstruierten Spuren werden mittels IPBus ausgelesen und analysiert. Dabei werden die mit dem Demonstrator gefundenen Spuren mit den Spuren verglichen, die mit der Software gefunden wurden.

Durch den Bau eines Demonstrators ist man gezwungen die komplette Verarbeitungskette vollständig fertigzustellen. Des Weiteren testet man, ob die Echtzeitanforderungen erfüllt werden und die Auswirkungen der begrenzten arithmetischen Präzision. Vor allem aber lässt sich durch den Demonstrator die Funktionsfähigkeit des Konzeptes beweisen.

9.1 Effizienz der Spurrekonstruktion

Die Effizienz wird relativ zu den generierten Teilchen der Selektion für Qualitätsstudien aus Tab. 5.2 gemessen. Eine Teilchenspur gilt als erfolgreich rekonstruiert, wenn mindestens ein Kandidat rekonstruiert wurde, welcher folgende Bedingungen erfüllt:

- der Kandidat teilt sich Stubs aus mindestens vier Detektorlagen mit dem Teilchen
- der Kandidat enthält keine Stubs, die nicht mit dem Teilchen assoziiert wurden

Kandidaten, die diese Bedingungen für kein rekonstruierbares Teilchen erfüllen, werden Fakes genannt. Falls diese Bedingungen für ein rekonstruierbares Teilchen von mehreren Kandidaten erfüllt werden, so werden diese Kandidaten bis auf einen als Duplikat bezeichnet.

Tab. 9.1 zeigt, wie sich einige Parameter der Spurrekonstruktion über die einzelnen Bearbeitungsschritte entwickeln. Dabei wurde die zweite Rekonstruktionsbedingung nur bei der Zeile für die vollständige Verarbeitungskette angewandt. Anhand der Tabelle kann man erkennen, dass die Hough-Transformation hoch effizient Spuren findet, sich unter den gefundenen Spurkandidaten aber auch viele Fakes und Duplikate befinden. Der Fit eliminiert die meisten Fakes und das Duplicate Removal entfernt die meisten Duplikate.

Tabelle 9.1 Resultate der einzelnen Prozessschritte des Demonstrators für $t\bar{t}$-Ereignisse mit 200 Überlappereignissen. Es werden die Durchschnittswerte gezeigt. [19]

Schritt	Effizienz [%]	# Kandidaten	# Fakes	# Duplikate
HT	97,1	331	139	126
KF	95,1	190	27	103
DR	94,4	79	16	3
gesamt	94,4	79	16	3

Die in Hardware gemessene Effizienz der Spurrekonstruktion stimmt zu 99,5 % mit der Simulationssoftware überein. Abb. 9.1 und 9.2 zeigen die Effizienz der Spurrekonstruktion über der Pseudorapidität beziehungsweise über dem Transversalimpuls der Teilchen, sowohl für die Hardware als auch für die Software.

Die Rekonstruktionseffizienz für Muonen überschreitet 97 %, während die Effizienz für Elektronen niedriger ist, wie in Abb. 9.3 und 9.4 zu sehen ist. Der Verlust in Effizienz für Elektronen ist zu erwarten und wird hauptsächlich durch Bremsstrahlung verursacht, welche dazu führt, dass die Trajektorie von einer perfekten Helix, wie sie von den Spurrekonstruktionsalgorithmen vorausgesetzt wird, abweicht.

Die Qualität von Spurkandidaten nimmt innerhalb von dichten Jets leicht ab. Das liegt an der erhöhten Stubdichte und der damit einhergehenden erhöhten Wahrscheinlichkeit ein Stub von einem anderen Teilchen zu finden, welches besser zum Kandidaten passt.

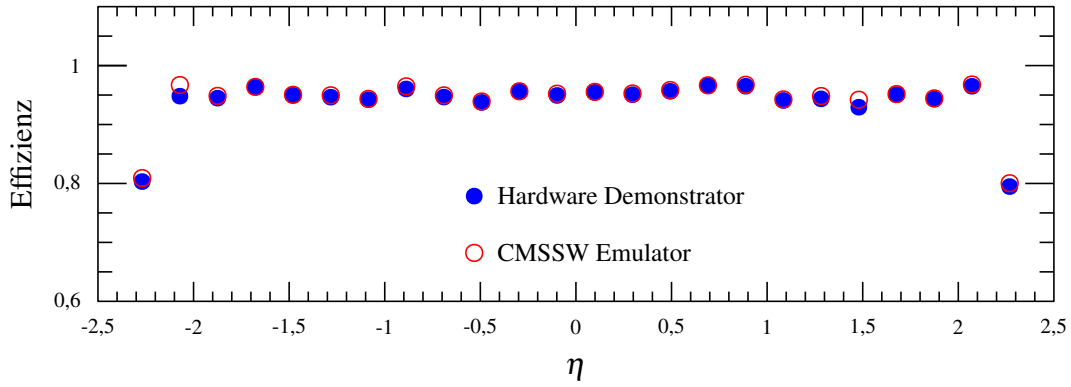

Abbildung 9.1 Effizienz der Spurrekonstruktion des Demonstrators aufgetragen über der Pseudorapidität. Messungen mit der Hardware sind in Blau und Messungen mit der Software sind in Rot eingetragen. Adaptiert von [19].

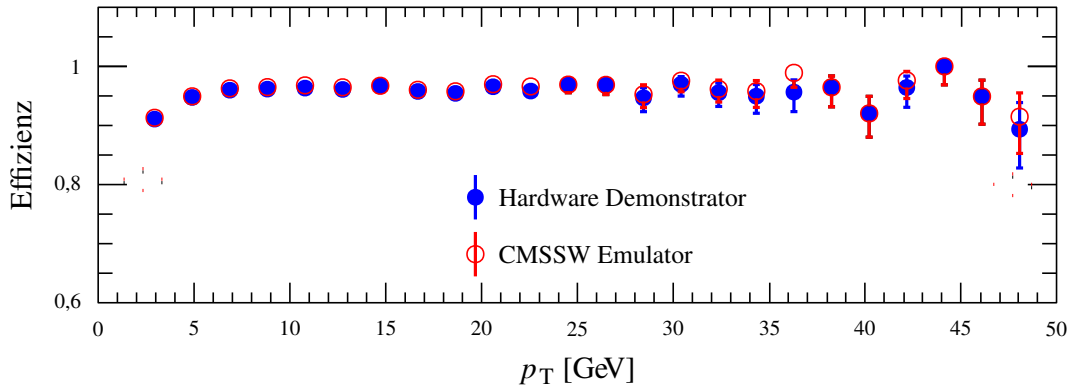

Abbildung 9.2 Effizienz der Spurrekonstruktion des Demonstrators aufgetragen über dem Transversalimpuls. Messungen mit der Hardware sind in Blau und Messungen mit der Software sind in Rot eingetragen. Adaptiert von [19].

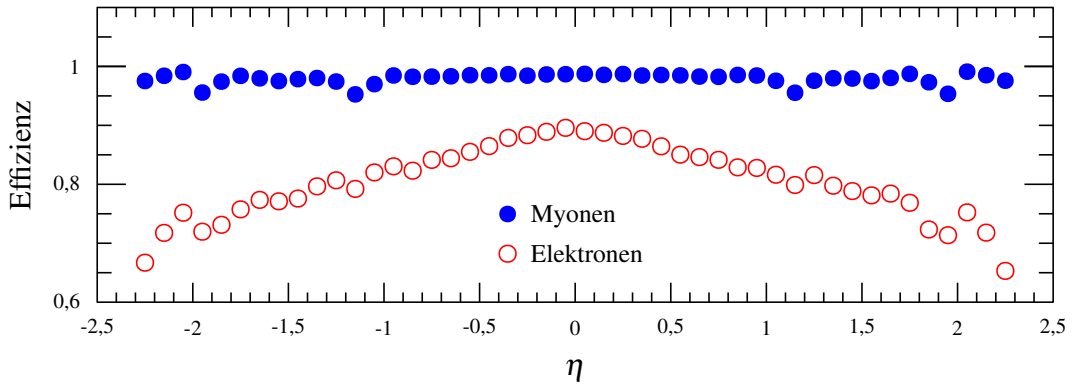

Abbildung 9.3 Effizienz der Spurrekonstruktion des Demonstrators für Myonen (blau) und Elektronen (rot), aufgetragen über der Pseudorapidität. Adaptiert von [19].

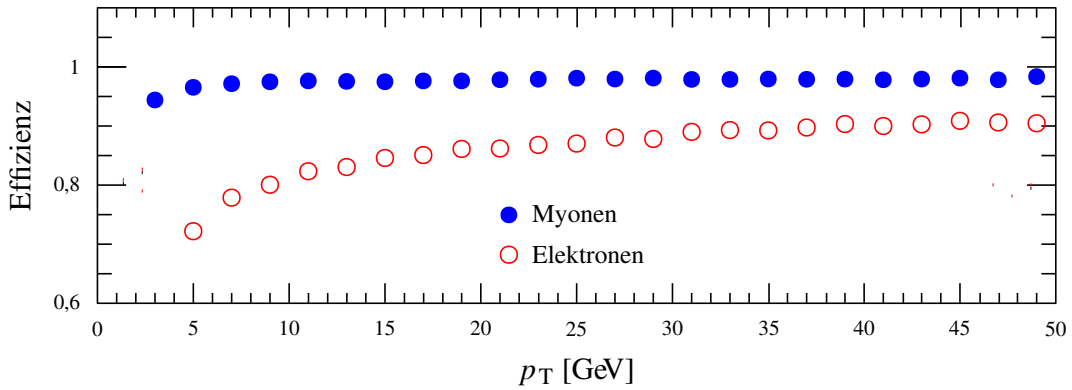

Abbildung 9.4 Effizienz der Spurrekonstruktion des Demonstrators für Myonen (blau) und Elektronen (rot), aufgetragen über dem Transversalimpuls. Adaptiert von [19].

9.2 Auflösung der Spurparameter

Abb. 9.5 bis 9.8 zeigen die Auflösungen der vier Spurparameter der in Software und Hardware rekonstruierten Spuren, welche mit den Teilchen aus der Selektion für Qualitätsstudien aus Tab. 5.2 assoziiert werden konnten.

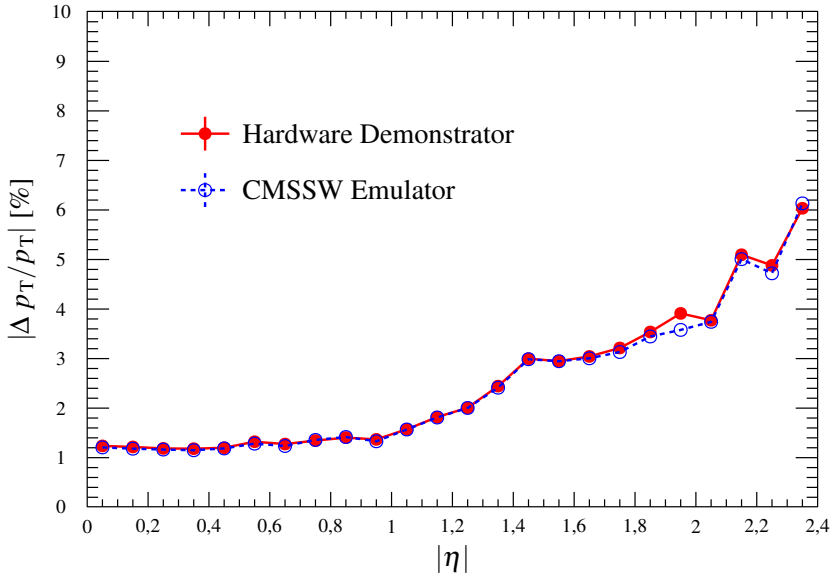

Abbildung 9.5 Relative p_T - Auflösung des Demonstrators aufgetragen über den Absolutwert der Pseudorapidität. Messungen mit der Hardware sind in Rot und Messungen mit der Software sind in Blau eingetragen. Adaptiert von [19].

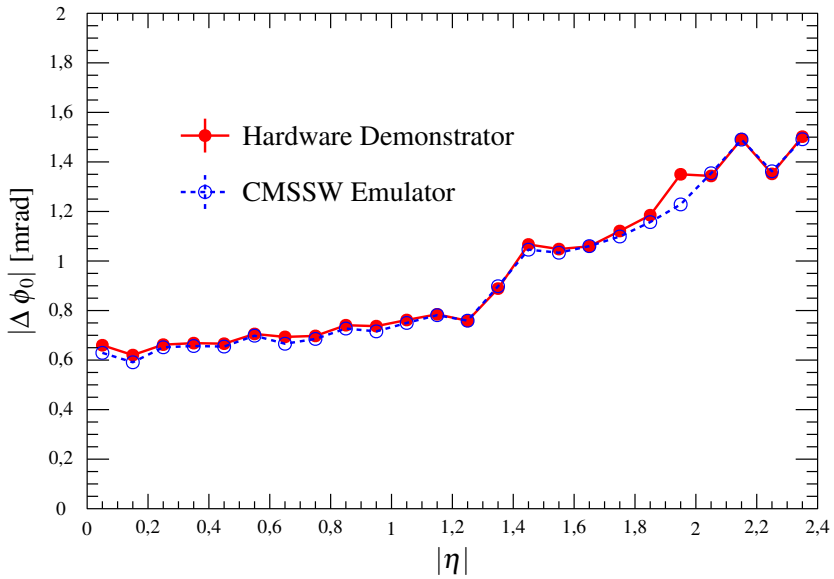

Abbildung 9.6 ϕ_0 - Auflösung des Demonstrators aufgetragen über den Absolutwert der Pseudorapidität. Messungen mit der Hardware sind in Rot und Messungen mit der Software sind in Blau eingetragen. Adaptiert von [19].

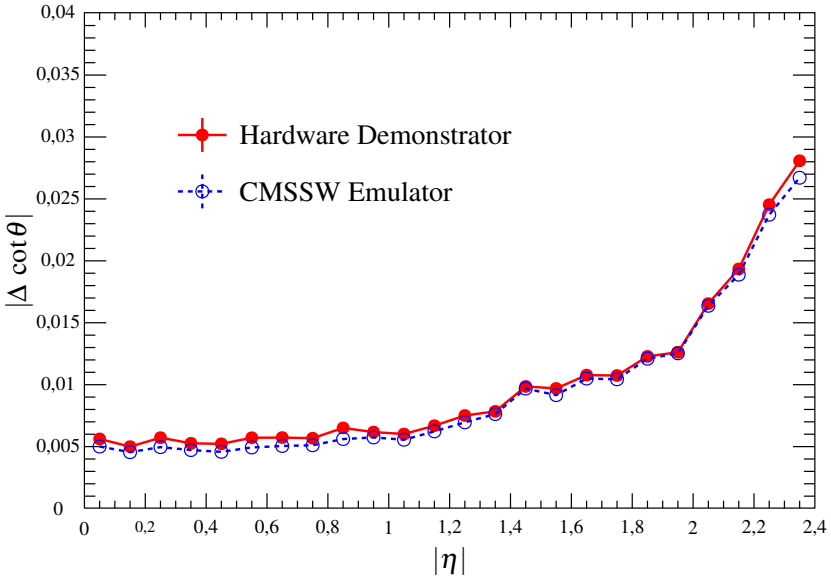

Abbildung 9.7 $\cot\theta$ - Auflösung des Demonstrators aufgetragen über den Absolutwert der Pseudorapidität. Messungen mit der Hardware sind in Rot und Messungen mit der Software sind in Blau eingetragen. Adaptiert von [19].

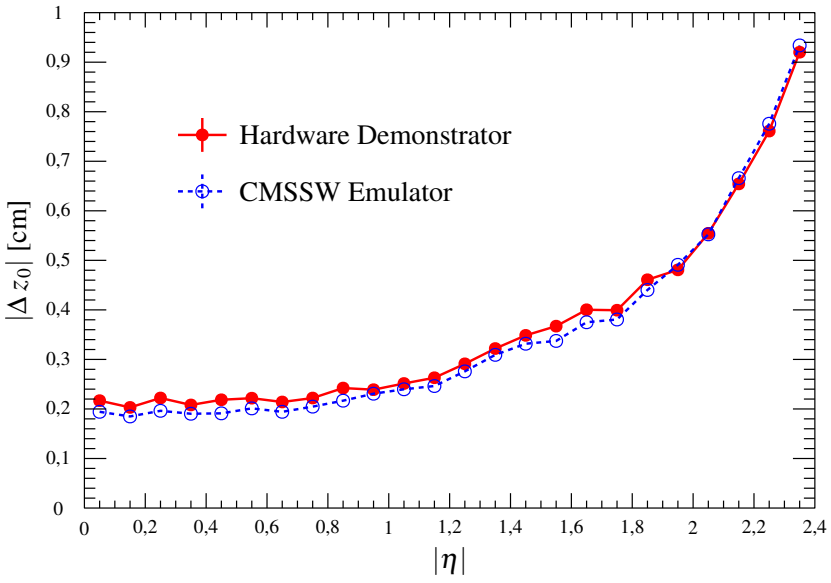

Abbildung 9.8 z_0 - Auflösung des Demonstrators aufgetragen über den Absolutwert der Pseudorapidität. Messungen mit der Hardware sind in Rot und Messungen mit der Software sind in Blau eingetragen. Adaptiert von [19].

Die Resultate von Emulator und Demonstrator stimmen gut über ein. Die verbleibenden Abweichung resultieren aus Teilen des Emulators, welche auf Fließkommapräzision beruhen.

Die Auflösungen nehmen mit größer werdender Pseudorapidität deutlich ab. Das liegt einmal an der größeren zurückgelegten Wegstrecke der Teilchen und der damit einhergehenden größeren Ablenkungen der Trajektorie durch beispielsweise Vielfachstreuung. Darüber hinaus wird der Hebelarm für die Messung der Parameter kleiner, da der radiale Bereich über den sich die Stubs verteilen kleiner wird.

Die Auflösung von p_T, ϕ_0 und $\cot\theta$ des Demonstrators sind vergleichbar mit den Auflösungen der Ereignisrekonstruktion, wie in Abb. 9.9 angedeutet [17]. Dabei steht der Ereignisrekonstruktion der volle Spurdetektor zur Verfügung und kann aufwändigerer Algorithmen nutzen, da weder Ressourcen noch Zeit begrenzt sind.

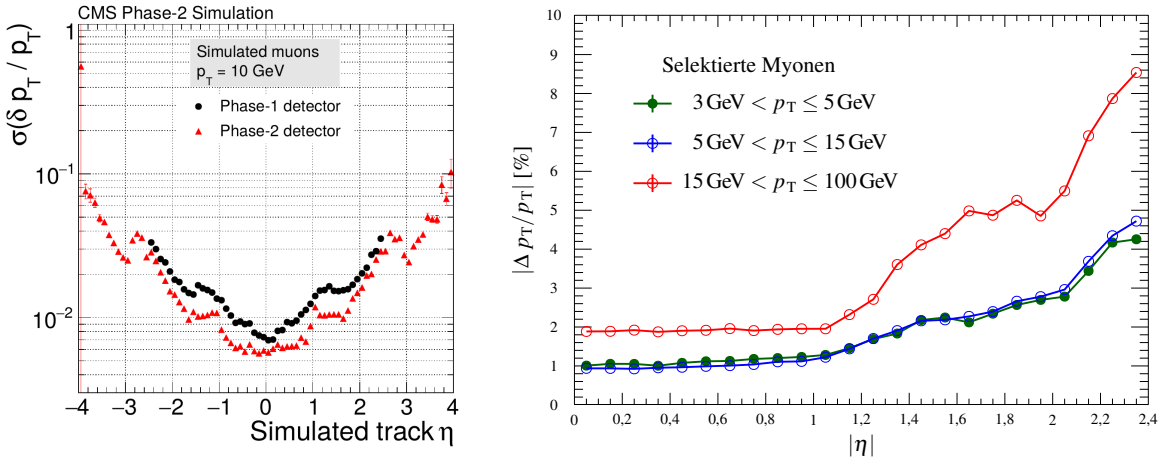

Abbildung 9.9 Relative p_T - Auflösungen aufgetragen über der Pseudorapidität. Das linke Bild zeigt die Auflösung für den aktuellen Spurdetektor der Phase-I in Schwarz und für den Spurdetektor der Phase-II in Rot. Simuliert wurden einzelne isolierte Myonen mit einem Transversalimpuls von 10 GeV. Adaptiert von [17]. Das rechte Bild zeigt die Auflösung für verschiedene Transversalimpulsbereiche für den Demonstrator. Als Eingangsdaten dienten die relevanten Stubs aus Tab. 5.2, wobei nur die Myonen aus der Selektion für Qualitätsstudien berücksichtigt wurden. Adaptiert von [19].

Die z_0 - Auflösung des Demonstrators könnte um einen Faktor zwei verbessert werden, wenn man die r und die z Koordinate jeweils mit zwei zusätzlichen Bits beschreiben würde. Dadurch würde der Auflösungsverlust aufgrund der Ganzzahlarithmetik vernachlässigbar werden [19]. Dennoch wäre die Auflösung um eine bis zwei Größenordnungen schlechter als die der Ereignisrekonstruktion, da der Pixeldetektor für eine sehr gute Messung von z_0 oder d_0 unerlässlich ist.

9.3 Laufzeiten

Die Laufzeiten wurden für die gesamte Verarbeitungskette sowie für die einzelnen Blöcke aus Abb. 8.2 gemessen. Diese sind in Tab. 9.2 eingetragen. Die Messungen beinhalten die Übertragungsverzögerungen sowie die Laufzeiten der Serialisierung und Deserialisierung (SERDES), welche die Hochgeschwindigkeitstransceiver benötigen.

Tabelle 9.2 Gemessene Laufzeiten des Demonstrators für ein Ereignis. Die Laufzeiten der einzelnen Prozessschritte wurden vom ersten einlaufenden Datenwort bis zum ersten auslaufenden Datenwort gemessen. Die Laufzeit des Gesamtsystems wurde vom ersten einlaufenden Stub bis zur letzten auslaufenden Spur gemessen.

Prozessschritt	Laufzeit [ns]
SERDES + Übertragungsstrecke Quelle zum GP	143
Geometric Prozessor	251
SERDES + Übertragungsstrecke GP zur HT	144
Hough Transform	1025
SERDES + Übertragungsstrecke HT zum KF+DR	129
Kalman Filter + Duplicate Removal	1658
SERDES + Übertragungsstrecke KF+DR zur Senke	129
Summe	3479
Gesamtsystem	3704

Die Laufzeit des Systems ist statisch und unabhängig von der Anzahl der Überlappereignisse oder der lokalen Stubdichten. Diese Größen beeinflussen nur die Effizienz und Qualität der Spurrekonstruktion. Die mit dem Demonstrator erreichte Laufzeit von 3,7 µs ist kurz genug für den Spurtrigger, da man bis zu 4 µs für die Spurrekonstruktion einräumt [17].

Kapitel 10

Schlussbemerkung

Das CMS-Experiment muss erheblich ausgebaut werden, um das vollständige Potenzial des Hochluminositätsbetriebes des LHC auszuschöpfen. Dabei ist der Spurtrigger, also die Spurrekonstruktion im äußeren Spurdetektor in der ersten Triggerstufe, ein vollkommen neuartiger und entscheidender Bestandteil dieses Ausbaus. Die Anforderungen an den Spurtrigger, die großen Datenrate, die kurze Laufzeit und die Komplexität der Rekonstruktion von Spuren sind derart herausfordernd, dass die Realisierbarkeit zum Jahr 2026 ungewiss ist.

Daher wurden drei Demonstratoren gebaut, um die Realisierbarkeit von drei unterschiedlichen Ansätzen der Spurrekonstruktion für die erste Triggerstufe des CMS-Detektors der Phase-II zu erproben. Der TMTT-Demonstrator implementiert auf einem Hardwaresystem meine Hough-Transformation, um grobe Spuren zu finden und einen Kalman-Filter, um diese zu bereinigen und die Spurparameter zu bestimmen.

Dieses Hardwaresystem hat erfolgreich gezeigt, dass die Spurrekonstruktion von geladenen Teilchen aus der primären Wechselwirkungszone mit einem Transversalimpuls über 3 GeV unter den herausfordernden Bedingungen des HL-LHC innerhalb von 4 µs mit derzeit zur Verfügung stehender Technologie möglich ist.

Dies wurde durch meine Implementierung der Hough-Transformation ermöglicht, welche bei einer Laufzeit von einer µs die Anzahl der Stubs um eine Größenordnung reduziert und somit die Komplexität der Ereignisse derart massiv reduziert, dass sogar ein komplexer Algorithmus wie der Kalman-Filter angewendet werden kann, um die Spurrekonstruktion abzuschließen.

Die Entwicklung des CMS-Spurtriggers ist mit meiner Arbeit noch nicht abgeschlossen, welche ich mehr als Pionierarbeit betrachte. Ich habe große moderne FPGAs erkundet, eine Spursuche sowie Kernkomponenten eines Spurfits auf diesen realisiert und die Stellschrauben, welche die Qualität und die Effizienz der Spurrekonstruktion bestimmen, herausgearbeitet.

Die wichtigsten Stellschrauben sind die Anzahl der Sektoren in der r-ϕ- und r-z-Ebene, die Granularität der Hough-Histogramme und die Komposition der Stubs von verschiedenen Detektorlagen, um einen Spurkandidaten zu identifizieren. Diese Größen definieren aber auch die benötigten Ressourcen und somit auch die benötigte Anzahl an FPGAs beziehungsweise die benötigte Anzahl an Karten, Crates und Racks, also auch die Größe des Systems und die Systemarchitektur.

Ohne die Kenntnis, welche Anforderungen die erste Triggerstufe an die Qualität und die Effizienz der Spurrekonstruktion stellt, lässt sich aber keine optimale Wahl für der Stellschrauben treffen. Somit lässt sich auch derzeit die endgültige Systemarchitektur nicht festlegen. Nichtsdestotrotz beginnt heute schon die Entwicklung der Karten für den Spurtrigger.

Als Konsequenz ist es bei der Hardwareentwicklung unerlässlich auf maximale Flexibilität zu achten. Diese Flexibilität hatte ich beim Bau des Demonstrators mit der MP7-Karte, welche sich dadurch auszeichnet, dass sich ein Großteil der Hochgeschwindigkeitsreceiver über Lichtwellenleiter unidirektional mit beliebigen MP7-Karten im System verbinden ließ. Diese Flexibilität sollten auch unbedingt die Karten für den Spurtrigger aufweisen.

Letztendlich meistert der TMTT-Demonstrator die gewaltigen Herausforderungen des Hochluminositätsbetriebes und stellt einen Meilenstein für den Spurtrigger, für das gesamte CMS-Experiment der Phase-II sowie für zukünftige Hochenergiephysikexperimente, dar.

Anhang A

Appendix

A.1 Native Auflösung

Die native Auflösung ist eine Abschätzung der Detektorauflösung, beruhend auf der Granularität der Sensoren. Dazu betrachten wir eine eindimensionale Detektorlage aus Sensoren mit einem gleichmäßigen Abstand und nehmen an, dass nur die Sensoren einen Teilchendurchgang detektieren, die diesem am nächsten sind. Die beste Wahl für die gemessene Position eines Teilchendurchgangs ist durch die Sensormitte des detektierenden Sensors gegeben. Die Wahrscheinlichkeitsdichte $f(x)$ für die Differenz x von der korrekten Position zur gemessenen Position des Teilchendurchgangs ist in Abb. A.1 gezeigt.

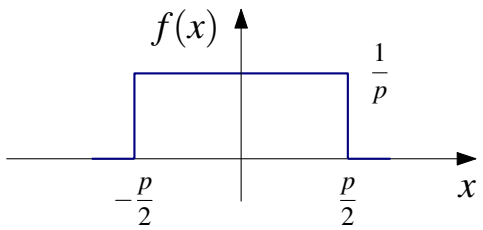

Abbildung A.1 Wahrscheinlichkeitsdichte der Differenz von der korrekten Position zur gemessenen Position des Teilchendurchgangs.

Die Varianz der Differenz von der korrekten Position zur gemessenen Position des Teilchendurchgangs ist gegeben durch

$$\text{Var} = \int_{-\infty}^{+\infty} \mathrm{d}x\, x^2 \cdot f(x) = \int_{-p/2}^{+p/2} \mathrm{d}x\, \frac{x^2}{p}\,, \tag{A.1}$$

$$= \frac{1}{3\,p} \cdot \left[\frac{p^3}{8} + \frac{p^3}{8}\right] = \frac{p^2}{12}\,. \tag{A.2}$$

Wir definieren die native Auflösung σ als die Standardabweichung der Differenz von der gemessenen Position des Teilchendurchgangs zu der korrekten Position und erhalten

$$\sigma = \sqrt{\text{Var}} = \frac{p}{\sqrt{12}}\,. \tag{A.3}$$

A.2 Lineare Regression

Für die Minimierung von Gl. 5.26 betrachten wir zunächst deren Ableitung, welche geben ist durch

$$\begin{pmatrix} \frac{\partial}{\partial c} \\ \frac{\partial}{\partial m} \end{pmatrix} \overline{r^2} = \begin{pmatrix} 2\bar{r} \\ 2\overline{xr} \end{pmatrix}\,. \tag{A.4}$$

Durch das Nullsetzen der Ableitungen zusammen mit

$$\bar{r} = m \cdot \bar{x} + c \cdot n - \bar{y}\,, \tag{A.5}$$

$$\overline{xr} = m \cdot \overline{xx} + c \cdot \bar{x} - \overline{xy}\,, \tag{A.6}$$

erhalten wir das folgende lineare Gleichungssystem

$$\begin{pmatrix} n & \bar{x} \\ \bar{x} & \overline{xx} \end{pmatrix} \cdot \begin{pmatrix} c \\ m \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \overline{xy} \end{pmatrix}\,, \tag{A.7}$$

welches wir mit dem Gaußschen Eliminationsverfahren in den nächsten drei Zeilen lösen.

$$\begin{aligned}
\begin{pmatrix} n & \bar{x} \\ 0 & n\overline{xx} - \bar{x}\cdot\bar{x} \end{pmatrix} \cdot \begin{pmatrix} c \\ m \end{pmatrix} &= \begin{pmatrix} \bar{y} \\ n\overline{xy} - \bar{x}\cdot\bar{y} \end{pmatrix}, \\
\begin{pmatrix} n\left(n\overline{xx} - \bar{x}\cdot\bar{x}\right) & 0 \\ 0 & n\overline{xx} - \bar{x}\cdot\bar{x} \end{pmatrix} \cdot \begin{pmatrix} c \\ m \end{pmatrix} &= \begin{pmatrix} \bar{y}\left(n\overline{xx} - \bar{x}\cdot\bar{x}\right) - \bar{x}\left(n\overline{xy} - \bar{x}\cdot\bar{y}\right) \\ n\overline{xy} - \bar{x}\cdot\bar{y} \end{pmatrix}, \\
\begin{pmatrix} n\overline{xx} - \bar{x}\cdot\bar{x} & 0 \\ 0 & n\overline{xx} - \bar{x}\cdot\bar{x} \end{pmatrix} \cdot \begin{pmatrix} c \\ m \end{pmatrix} &= \begin{pmatrix} \bar{y}\cdot\overline{xx} - \bar{x}\cdot\overline{xy} \\ n\overline{xy} - \bar{x}\cdot\bar{y} \end{pmatrix}. \qquad \text{(A.8)}
\end{aligned}$$

A.3 Zweierkomplement

Das Zweierkomplement stellt ganze Zahlen binär dar. Eine binäre Zahl besteht aus einer Anzahl Bits, welche eine gewisse Wertigkeit aufgrund ihrer Reihenfolge aufweisen. Wir nummerieren die Bits einer n-stelligen Zahl von 0 bis $n-1$ und bezeichnen den binären Wert des i-ten Bit mit b_i.

Das nullte Bit b_0 weist die niedrigste Wertigkeit auf und wir nennen es LSB (aus dem Englischen für *least significant bit*), während das $n-1$-te Bit b_{n-1} die höchste Wertigkeit aufweist und von uns MSB (aus dem Englischen für *most significant bit*) genannt wird.

Wenn das MSB den Wert '0' aufweist ist die Zahl positiv und der Dezimalwert d ist gegeben durch

$$d = \sum_{i=0}^{n-1} b_i \cdot 2^{n-i-1}. \qquad \text{(A.9)}$$

Weist das MSB den Wert '1' auf, ist die Zahl negativ und der Dezimalwert ist gegeben durch

$$d = -1 - \sum_{i=0}^{n-1} \bar{b}_i \cdot 2^{n-i-1}, \qquad \text{(A.10)}$$

wobei mit $\bar{b}$ die Negation des binären Wertes gemeint ist, welche eine '0' zu einer '1' und eine

'1' zu einer '0' transformiert. Um das Zweierkomplement zu veranschaulichen zeigt Tab. A.1 gewisse Binärzahlen und deren Dezimalwerte, wobei die Binärzahlen mit fallender Wertigkeit von links nach rechts geschrieben sind.

Tabelle A.1 Dezimalwerte aller 2-stelligen Binärzahlen für das Dualsystem und das Zweierkomplement.

Binärzahl	Dezimalwert	
	Dualsystem	Zweierkomplement
"00"	0	0
"01"	1	1
"10"	2	-2
"11"	3	-1

Literaturverzeichnis

[1] Amos Breskin and Rüdiger Voss. The CERN Large Hadron Collider: Accelerator and Experiments. (2009). URL https://cds.cern.ch/record/1244506

[2] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, (2012). doi: https://doi.org/10.1016/j.physletb.2012.08.020

[3] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, (2012). doi: https://doi.org/10.1016/j.physletb.2012.08.021

[4] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, (2008). doi: 10.1088/1748-0221/3/08/S08004

[5] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, (2008). doi: 10.1088/1748-0221/3/08/S08003

[6] G Apollinari et al. High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report. (2015). URL https://cds.cern.ch/record/2116337

[7] D Contardo et al. Technical Proposal for the Phase-II Upgrade of the CMS Detector. (2015). URL https://cds.cern.ch/record/2020886

[8] LEP design report. (1984). URL https://cds.cern.ch/record/102083

[9] LHC Guide. (2017). URL https://cds.cern.ch/record/2255762

[10] W.J. Stirling. persönliche Kommunikation

[11] The CMS Collaboration. Detector Drawings. (2012). URL `https://cds.cern.ch/record/1433717`

[12] Teilchen und Kerne : Eine Einführung in die physikalischen Konzepte, (2009). URL `http://dx.doi.org/10.1007/978-3-540-68080-2`

[13] The CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. (2006). URL `http://cds.cern.ch/record/922757`

[14] The HL-LHC project. LHC/HL-LHC Plan, (2015). URL `http://hilumilhc.web.cern.ch/about/hl-lhc-project`

[15] G Hall et al. 2-D PT module concept for the SLHC CMS tracker. *JINST*, (2010). URL `http://stacks.iop.org/1748-0221/5/i=07/a=C07012`

[16] M Pesaresi and G Hall. Simulating the performance of a p T tracking trigger for CMS. *JINST*, (2010). URL `http://stacks.iop.org/1748-0221/5/i=08/a=C08003`

[17] K Klein. The Phase-2 Upgrade of the CMS Tracker. (2017). URL `http://cds.cern.ch/record/2272264`

[18] J Butler et al. CMS Phase II Upgrade Scope Document. (2015). URL `https://cds.cern.ch/record/2055167`

[19] R. Aggleton et al. An FPGA Based Track Finder for the L1 Trigger of the CMS Experiment at the High Luminosity LHC. *JINST*, to be published.

[20] Paulo Rodrigues. The LpGBT Project, Status and Overview. ACES 2016 - Fifth Common ATLAS CMS Electronics Workshop for LHC Upgrades. (2016). URL `https://cds.cern.ch/record/2137809`

[21] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, (2003). doi: 10.1016/S0168-9002(03)01368-8

[22] Hough Paul V.C. Method and means for recognizing complex patterns. (1962). URL `http://www.google.de/patents/US3069654`

[23] *IEEE standard VHDL language reference manual*. (1994). ISBN 1-55937-376-8

[24] *7 Series FPGAs Configurable Logic Block User Guide*. Xilinx, Inc., (2016). URL `https://www.xilinx.com/support/documentation/user_guides/ug474_7Series_CLB.pdf`. Rev. 1.8

[25] *7 Series FPGAs Memory Resources User Guide*. Xilinx, Inc., (2016). URL `https://www.xilinx.com/support/documentation/user_guides/ug473_7Series_Memory_Resources.pdf`. Rev. 1.12

[26] *7 Series DSP48E1 User Guide*. Xilinx, Inc., (2016). URL `https://www.xilinx.com/support/documentation/user_guides/ug479_7Series_DSP48E1.pdf`. Rev. 1.9

[27] *7 Series FPGAs GTX/GTH Transceivers User Guide*. Xilinx, Inc., (2016). URL `https://www.xilinx.com/support/documentation/user_guides/ug476_7Series_Transceivers.pdf`. Rev. 1.12

[28] K Compton et al. The MP7 and CTP-6: multi-hundred Gbps processing boards for calorimeter trigger upgrades at CMS. *JINST*, (2012). URL `http://stacks.iop.org/1748-0221/7/i=12/a=C12024`

[29] G. Hall. A time-multiplexed track-trigger for the CMS HL-LHC upgrade. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, (2016). doi: http://doi.org/10.1016/j.nima.2015.09.075

[30] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, (1987). doi: http://dx.doi.org/10.1016/0168-9002(87)90887-4

[31] *Micro Telecommunications Computing Architecture base specification: Micro TCA*. PICMG, (2006). URL `https://cds.cern.ch/record/1159873`

[32] E Hazen et al. The AMC13XG: a new generation clock/timing/DAQ module for CMS MicroTCA. *JINST*, (2013). URL `http://stacks.iop.org/1748-0221/8/i=12/a=C12036`

[33] C. Ghabrous Larrea et al. IPbus: a flexible Ethernet-based control system for xTCA hardware. *JINST*, (2015). URL
`http://stacks.iop.org/1748-0221/10/i=02/a=C02019`