UNIVERSITÉ DE STRASBOURG

UNIVERSITÈ DE STRASBOURG

Ecole Doctorale de Physique et Chimie Physique
Université de Strasbourg

ÈCOLE DOCTORALE DE PHYSIQUE ET CHIMIE-PHYSIQUE
Institut Pluridisciplinaire Hubert CURIEN (IPHC)

THÈSE présentée par:

Qingyuan LIU

soutenue le: 19 Septembre 2017

pour obtenir le grade de: **Docteur de l'Université de Strasbourg**
Discipline/Spécialité : Physique des particules èlèmentaires

Development of a CMOS Pixel Based Inner Tracker For the BES-III Experiment at BEPC-II

Directeur de thèse:	Marc WINTER	Directeur de recherche, IPHC-UMR 7178, Strasbourg
	Meng WANG	Professeur, Shandong University, China
Rapporteur externe:	Eric KAJFASZ	Directeur de recherche, CPPM-UMR 7346, Marseille
	Imad LAKTINEH	Professeur, IPNL-UMR 5822, Lyon
Examinateur:	Sandrine COURTIN	Professeur, IPHC-UMR 7178, Strasbourg

Acknowledgements

This thesis would not have been possible without the inspiration and support of many people. First and foremost, I would like to thank my two supervisors: Prof. Meng WANG and Dr. Marc WINTER for their encouragement and guidance throughout my doctoral studies. I am also grateful to Dr. Christine HU-GUO for her kind help in this co-advisor program between University of Strasbourg and Shandong University.

I would like to show my sincere gratitude to Prof. Jérôme BAUDOT. His introduction to beam tests and the fast tracking estimation tools led me to the right path of the thesis topic. I would like to thank Luis Alejandro PEREZ PEREZ, Mathieu GOFFE and Gilles CLAUS for beam test studies. It is also a pleasure to thank Michal SZELEZNIAK and Franck AGNESE for their guidance in CPS probing test.

I am heartily thankful to Dr. Eric KAJFASZ and Dr. Imad LAKTINEH for their insightful comments and advices as the thesis rapporteurs, as well as Dr. Sandrine COURTIN for being the examiner. I would also like to thank the other colleagues in the two Universities: Dr. Isabelle RIPP-BAUDOT, Prof. Auguste BESSON, Prof. Xueyao ZHANG, Prof. Cunfeng FENG and Prof. Qun WU for their help and intellectual discussions.

I would like to express my gratitude to China Scholarship Council (CSC) for the financial support of my joint doctoral study in France.

Finally, my appreciations goes out to my family and the friends in the lab. Thanks for their sympathetic ear and helpful conversations.

Contents

List of Figures

List of Tables

Abstract

The thesis addresses the upcoming particle physics programme foreseen at the Beijing Electron Positron Collider II (BEPCII), for which the Beijing Electron Spectrometer III (BESIII) is being upgraded. The work presented focused on the upgrade of the central tracker of the experiment by exploring the possibilities offered by a new approach using CMOS Pixel Sensors (CPS) instead of the present gas drift chamber.

The MDC inner chamber of BESIII suffers from ageing effects which lead to performance degrading in terms of detection efficiency, momentum resolution and spatial resolution. From the analysis in 2013, the efficiency of the MDC inner cell could be around 70% and its spatial resolution degraded from $\sim 120\mu m$ to $\sim 300\mu m$ in the worst case. This ageing effect would prevent BESIII from collecting the foreseen data sample and decrease the precision of the measurements. The collaboration decided therefore to upgrade the original inner chamber with a newly designed one or with a tracker using more advanced technology. The project of a Silicon Pixel Tracker (SPT) was proposed in this perspective.

Two issues in SPT development have been studied: a low material budget of a long ladder and the detecting geometry optimization. CPS prototypes using large pixel pitches and an improved rolling-shutter readout architecture have been tested and studied to validate the sensor performance of a spatial resolution $\sim 10\,\mu\mathrm{m}$, an integration time $\sim 20.8\,\mu\mathrm{s}$ and a power consumption around $80\,\mathrm{mW/cm^2}$. Therefore, the materials from cooling system are limited to guarantee the low material budget of the ladder. For tracker optimization, a general method and the related software based on tracking algorithms, numerical geometry scans and Geant4 Monte-Carlo simulations, has been developed and applied for BESIII SPT. The total number (3) of ladders and the tracker layout with a uniform spacing have been determined for the optimal momentum measurement for a sensor spatial resolution $\sim$10 µm and for a ladder material budget $\sim 0.36\% X_0$. To achieve a higher tracking efficiency, a double-sided ladder may be considered to replace the innermost one-sided ladder with acceptable tracking performance degradation for low momentum trajectories.

Keywords: pixel detector, tracker optimization, CPS test, tracking algorithm

Résumé en Français

R.1 Introduction

La thèse se place dans le contexte du programme de physique prévu auprès du collisionneur BEPCII (*Beijing Electron Positron Collider II*), pour lequel l'expérience BESIII (*Beijing Electron Spectrometer III*) doit subir une jouvence. Mon travail concerne l'amélioration des performances de la trajectométrie centrale de l'expérience, en explorant une nouvelle approche utilisant les capteurs à pixels CMOS (CPS) en remplacement partiel de la chambre à dérive actuelle.

BEPCII réalise des collisions $e^+ + e^-$ dont l'énergie dans le centre de masse atteint $E_{cm} \sim 2{-}4.6\,\mathrm{GeV}$ et avec une luminosité instantanée de $1 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Les deux anneaux de stockage et le point d'interaction conçu de la machine sont représentés sur la figure 1.1. Le programme de BESIII, le seul détecteur auprès de la machine BEPCII, inclut la spectroscopie des hadrons légers, la physique du quark charmé et du charmonium, et la physique du lepton tau. La

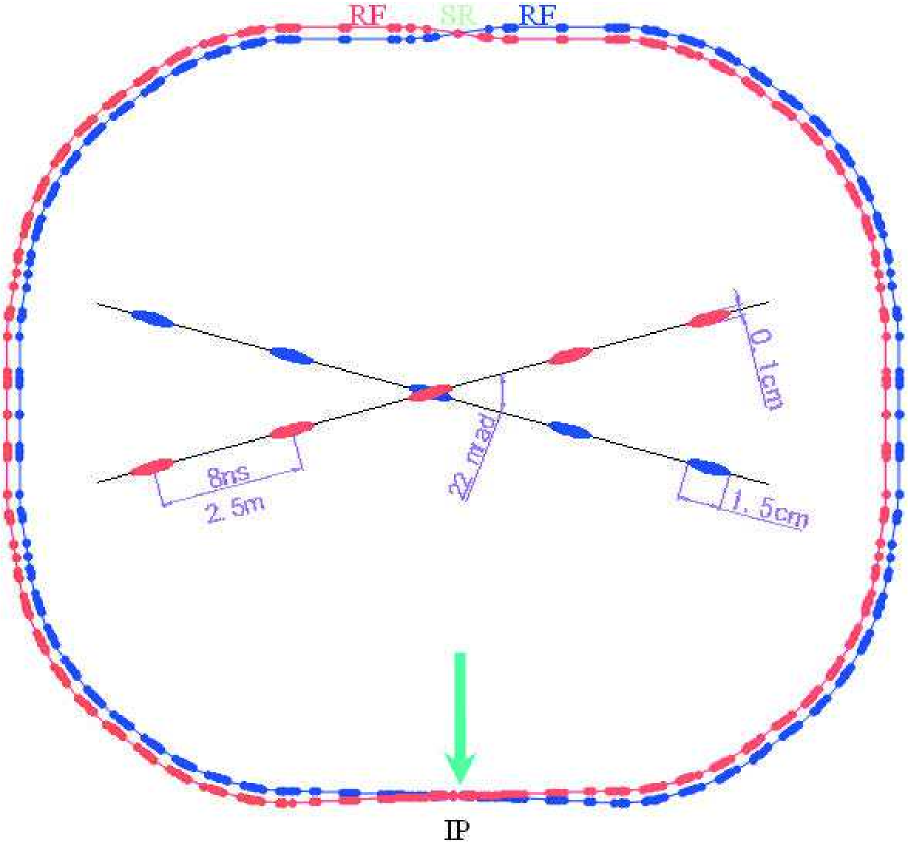


Figure 1: Les deux anneaux de stockage et le point d'interaction conçu de BEPCII

figure 1.3 illustre les sous-systèmes de détection de BESIII. L'élément central de l'expérience est le trajectomètre actuel, qui exploite un détecteur gazeux de type chambre à dérive multi-couche (*multilayer drift chamber* ou MDC). La MDC permet de localiser les points d'interaction des particules le long de leur trajectoire avec une résolution $\lesssim 130\mu m$, qui conduit à une résolution sur l'impulsion d'environ 0.5%@ 1 GeV/c.

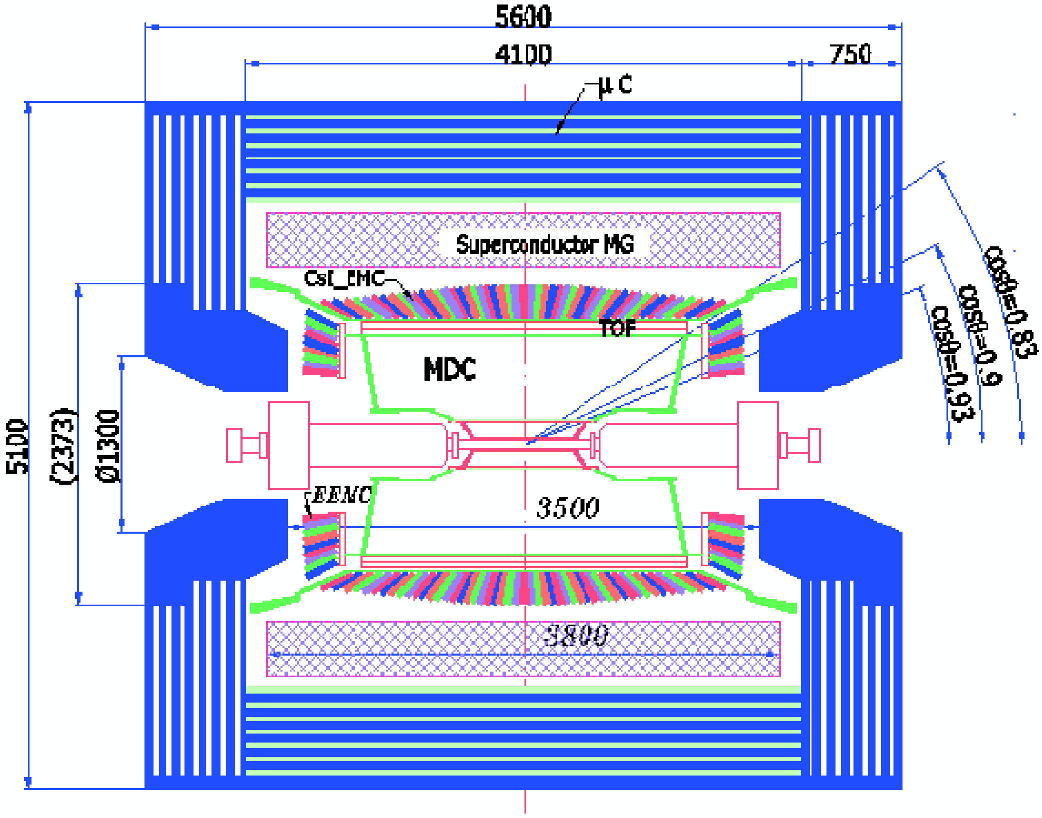


Figure 2: Les sous-systèmes de détection sur l'expérience BESIII[1]

R.2 Travaux effectués

Motivation

La partie interne de la MDC souffre d'effets liés au vieillissement du détecteur, ce qui diminue les performances en termes d'efficacité de détection et de résolution. Une analyse conduite en 2013 a montré que ces effets pouvaient dégrader l'efficacité jusque vers 70% et la résolution spatiale vers 300μm dans le pire des cas. La dégradation des performances est illustrée dans la figure 1.9 et la figure 1.10. Dans ces conditions, le vieillissement empêcherait BESSIII d'accumuler une quantité suffisante de données pour atteindre la précision statistique visée sur les mesures du programme de physique. C'est pourquoi la collaboration envisage soit le renouvellement des cellules internes de la chambre MDC soit leur

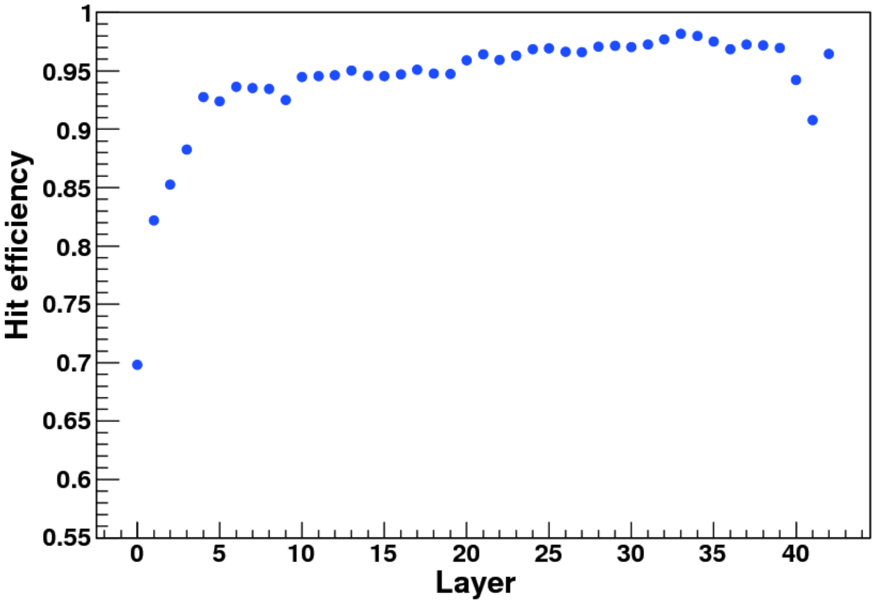


Figure 3: L'efficacité de détection de fil unique en fonction de l'indice de couche du MDC [2].

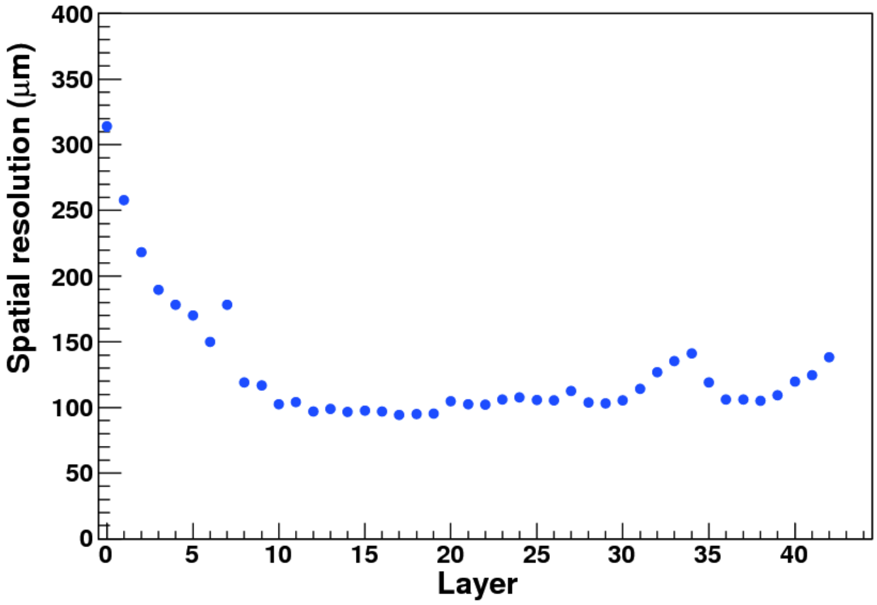


Figure 4: La résolution spatiale dans le plan R-ϕ de chaque couche en fonction de l'indice de couche du MDC [2].

remplacement par un trajectomètre exploitant une technologie plus avancée. Le projet de *Silicon Pixel Tracker* (SPT) a été proposé dans cette perspective.

Problématiques et méthodologies

Le développement du SPT soulève deux problématiques principales afin de garantir la résolution voulue sur l'impulsion : obtenir des éléments de détection (dit échelles) avec un budget de matière suffisamment faible et optimiser leur disposition géométrique. La technologie des capteurs à pixel CMOS (CPS) semble bien adaptée pour répondre à ces contraintes. L'expérience STAR [3] a validé le concept d'un détecteur composé de tels capteurs. En utilisant les circuits MIMOSA-28 développés à l'IPHC pour construire des échelles de détection refroidies par flux d'air, un budget de matière de $\sim 0.37\% X_0$ [*] a été atteint pour une résolution spatiale intrinsèque de $\sim 3.5\,\mu\text{m}$. Cependant, l'acceptance géométrique du SPT de BESIII requiert des échelles 2,5 fois plus longues que celles développées dans STAR tout en conservant un budget de matière similaire. Ceci n'est possible qu'avec des CPS dissipant une puissance thermique plus faible pour conserver le refroidissement à air, qui garanti l'absence de matériaux additionnels pour le refroidissement.

Le développement de CPS de très faible consommation de puissance est en cours au sein de l'expérience ALICE auprès du *Large Hadron Collider* selon deux stratégies. La première consiste à modifier entièrement la stratégie de lecture des pixels et a conduit au capteur ALPIDE. La seconde s'appuie sur l'architecture de MIMOSA-28 mais avec une taille de pixel agrandie et vise la création du capteur MISTRAL. Une partie de cette thèse s'intéresse à la validation des prototypes de CPS à grands pixels à travers des tests réalisés avec un faisceau de particules chargés.

La seconde partie de ce travail vise l'optimisation de la géométrie du trajectomètre. Une méthode générale a été développée et implémentée sous forme logicielle. Elle s'appuie sur une simulation Monte-Carlo portée par l'outil GEANT4, des algorithmes de reconstruction de trajectoire et la possibilité de tester de nombreuses géométries différentes. La méthode est ensuite appliquée au cas du SPT de BESSIII.

---

[*] La longueur de radiation X_0 (en $\text{g}\,\text{cm}^{-1}$) représente l'épaisseur moyenne, pour un matériau donné, après laquelle un électron de haute-énergie ne conserve qu'une fraction 1/2 de son énergie initiale après radiations par bremsstrahlung [4]

R.2.1 Caractérisation de prototypes de capteurs

Le développement de l'architecture pour MISTRAL requiert la caractérisation de différents prototypes avec des faisceaux de particules à des énergies au delà du GeV. La géométrie simplifiée de ces tests autorise la mesure de la résolution spatiale et de l'efficacité de détection d'un capteur positionné perpendiculairement au faisceau. Dans un premier temps, un dispositif comprenant 4 capteurs MIMOSA-28 accolés dans le même plan est utilisé pour comprendre et mettre au point l'algorithme d'alignement, c'est à dire la détermination de la position des capteurs par rapport au faisceau avec une résolution inférieure au micromètre.

Par la suite, le circuit FSBB (*Full Size Building Block*) est testé sur le faisceau. FSBB comprend des pixels plus grands, $22.0\times32.5\,\mu m^2$ contre $20.7\times20.7\mu m^2$ pour MIMOSA-28, il est fabriqué dans le process TowerJazz 0.18 µm plus évolué et représente un tier de la surface totale du capteur final MISTRAL. L'analyse des données montre qu'il est possible de trouver des paramètres nominaux de fonctionnement du circuit pour obtenir une efficacité de détection supérieur à 99%, un taux de fausses détections inférieur à 10^{-5}/pixel/lecture et une résolution spatiale $\sim 4.5\,\mu m$. Ces résultats sont illustrés dans la figure 3.10. Les performances du capteur sont étudiées en dehors du point nominal de fonctionnement, notamment avec des courants de polarisation plus faibles pour l'amplificateur présent dans chaque pixel et pour le comparateur déterminant la présence d'un signal ou non. Dans ces conditions de courant faibles, la résolution spatiale n'est pas affectée, l'efficacité s'améliore légèrement et le

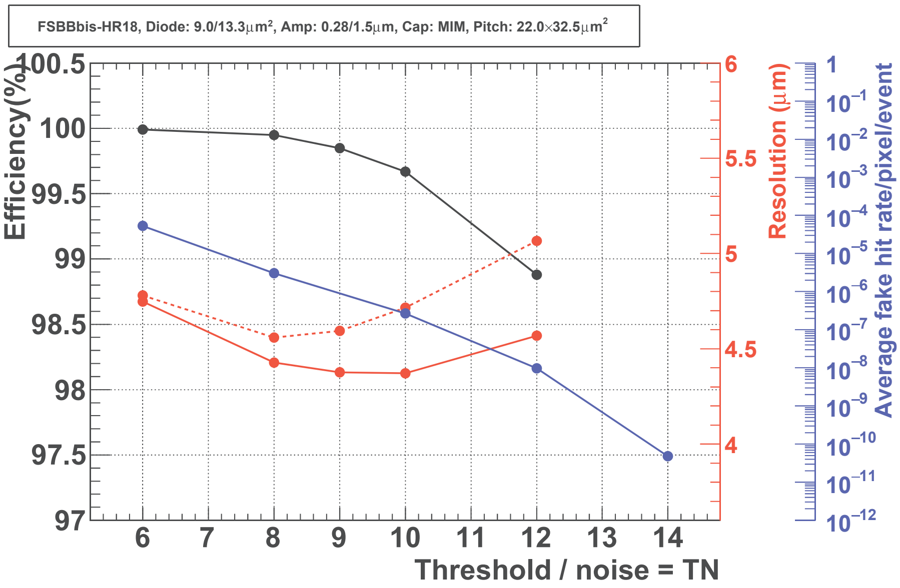


Figure 5: L'efficacité de détection (courbe noire), le taux de fausses détections (courbe bleue) et le résolution spatiale (courbe rouge) le long des côtés courts (solide) et longs (en pointillés) des pixels en fonction du seuil de discriminateur (en multiples de bruit thermique) du capteur FSBBbis-HR18 avec des pixels 22.0×32.5 µm^2.

taux de fausses détections augmente d'un facteur 2 à 4, demeurant toutefois dans les limites tolérables. L'étude valide ainsi la possibilité de diminuer la dissipation de puissance du circuit (courant plus faible), sans dégrader les performances, ce qui présente l'intérêt de limiter la puissance du système de refroidissement et donc du budget de matière.

R.2.2 Outil générique d'optimisation de la géométrie d'un trajectomètre

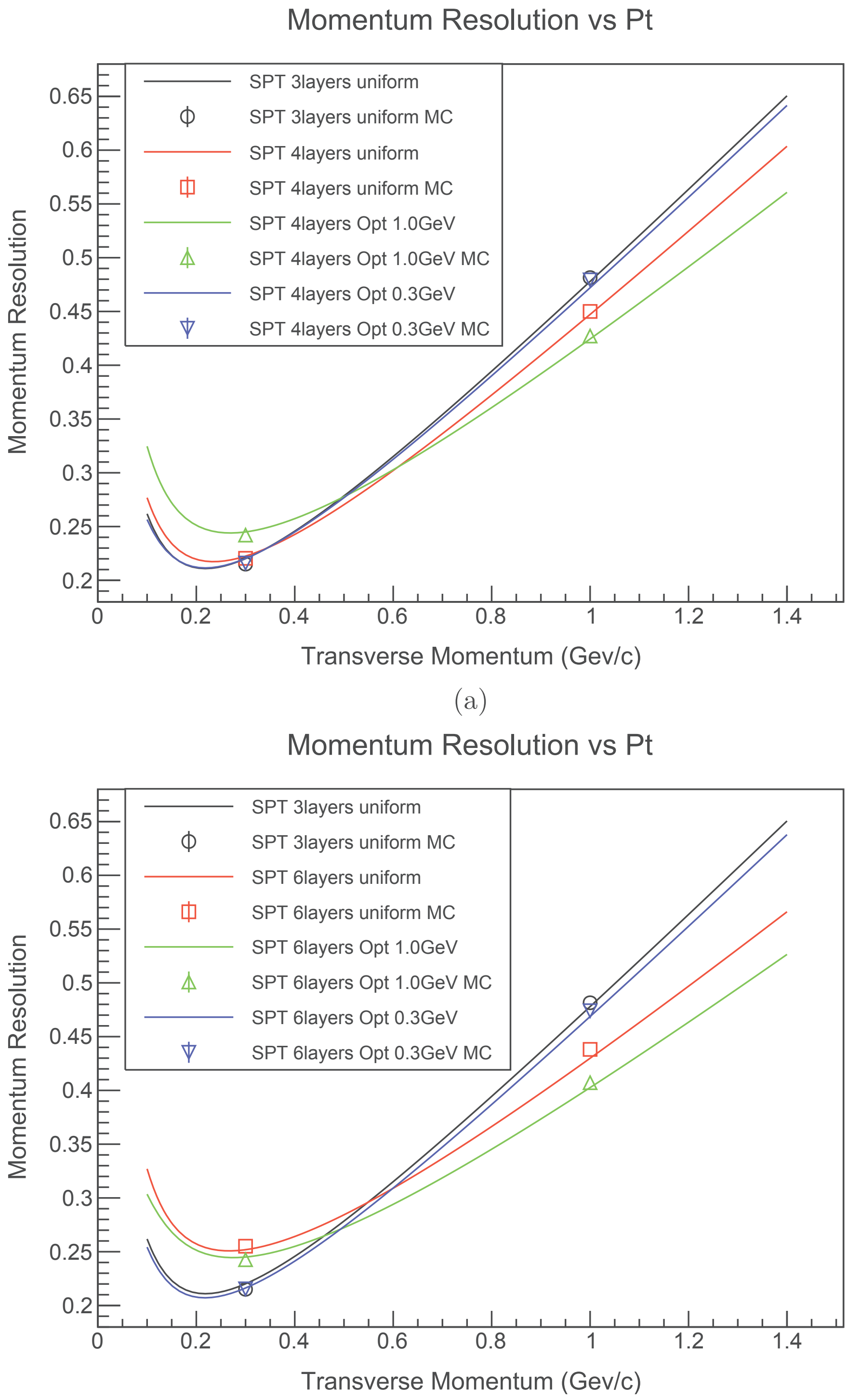
Le problème étudié est celui de la configuration d'un trajectomètre interne auprès d'un collisionneur. Il s'agit d'un ensemble de couches cylindriques concentriques autour du point de collision des faisceaux. La contrainte externe est le rayon maximal du dispositif et les paramètres libres comprennent le nombre de couches et leurs distances relatives. L'optimisation porte sur la résolution obtenue au final sur la mesure de l'impulsion des particules et l'extrapolation de leur trajectoire à proximité du point de collision.

La formule analytique de Gluckstern [5] est utilisée pour obtenir la résolution en fonction des paramètres de la géométrie testée. Une attention particulière est portée à la gestion de la diffusion multiple, qui dépend du budget de matière de chaque couche de détection. Un ensemble de géométries, couvrant l'espace des paramètres possibles, sont évaluées itérativement par un logiciel créer spécifiquement. Les résultats de cette méthode rapide sont validés par comparaison, pour quelques points, à une simulation complète et beaucoup plus lente exploitant une simulation Monte-Carlo utilisant la plateforme GEANT4 et un algorithme de reconstruction de trajectoire prenant en compte la diffusion multiple. La figure 5.5 illustre les résultats obtenus pour différentes configurations géométriques possibles pour l'expérience BESSIII. Les géométries connexes (3 à 6 couches), optimisées pour des trajectoires de 0.3 GeV et de 1.0 GeV avec un angle polaire $\theta = 0$, sont donnés dans le tableau 5.5.

N	For P_T	Step (mm)	Layer positions (mm)
3	all	any	72.58, 86.04, 99.50
4	0.3	0.45	72.58, 72.58, 85.59, 99.50
4	1.0	0.45	72.58, 84.25, 87.83, 99.50
5	0.3	0.67	72.58, 72.58, 78.64, 90.75, 99.50
5	1.0	0.67	72.58, 72.58, 83.35, 87.39, 99.50
6	0.3	0.90	72.58, 72.58, 79.76, 92.32, 99.50, 99.50
6	1.0	0.90	72.58, 72.58, 84.25, 87.83, 99.50, 99.50

Table 1: Géométries optimisées pour SPT détecteur avec une résolution spatiale $\sim 10\,\mu\mathrm{m}$.

Dans la mesure où la pluspart des particules détectées ont une impulsion < 0.5 GeV, il apparaît que la configuration à trois couches avec un espacement uniforme doit être favorisée.



(a)

(b)

Figure 6: Résolution relative sur l'impulsion mesurée en fonction de l'impulsion pour les trajectoires des pions courbées par le champ magnétique. Pour chaque géométrie, les lignes correspondent aux calculs alors que les points avec barres d'erreur sont issus des simulations Monte-Carlo.

R.2.3 Intérêt d’ échelles de détection double-face

Le groupe de Strasbourg a déjà démontré la possibilité de construire un élément de détection à partir d’un unique support mécanique et équipé de capteurs sur chaque face. Il s’agit des échelles double-face PLUME (*Pixelated Ladder with Ultra-Low Material Embedding*) [6] qui offrent un budget de matière le plus faible possible. Les points de mesure rapprochés (la distance entre les deux faces est de 2 mm) permettent par ailleurs d’obtenir une information supplémentaire à une simple position, puisqu’il est possible d’estimer la direction locale de la trajectoire. Les échelles PLUME présentent donc un intérêt particulier pour les couches internes du SPT de BESIII, pour lequel le budget de matière et la résolution sur les paramètres des trajectoires revêtent une importance considérable.

La configuration par défaut du SPT comprend trois couches simples, soit 3-plans, chacune avec un budget de matière équivalent à $\sim 0.36\% X_0/$. La configuration alternative remplace la couche la plus interne par une échelle double-face dont le budget de matière atteint $\sim 0.51\% X_0$, soit 4-plans.

La méthode discutée dans la section précédente est utilisée pour évaluer les performances des deux configurations. Comme élément de comparaison supplémentaire, une pseudo-efficacité de reconstruction des trajectoires est définie comme suit : la trace est considérée comme correctement reconstruite si trois vrais points sont associés et, dans le cas de la géométrie 4-plans, une mauvaise association sur les 4 points est tolérée. Les conclusions suivantes sont obtenues. La géométrie 3-plans permet une meilleure résolution sur l'extrapolation des trajectoires vers le point d'interaction pour les impulsions inférieures à 1 GeV/c, cependant sa pseudo-efficacité est plus faible de 16% par rapport à la configuration 4-plans pour les impulsions inférieures à 0.1 GeV/c et un niveau de bruit avoisinant 100 $\text{mm}^{-2}\,\text{s}^{-1}$ sur le premier plan de détection.

R.3 Conclusion et Perspectives

Cette thèse a étudié les performances de capteurs pixellisés et l'optimisation de la géométrie d'un nouveau trajectomètre interne multi-couche pour l'expérience BESIII. Les bons résultats initiaux obtenus avec le prototype FSBB, comportant des pixels de grande taille, ont été confirmés par la caractérisation d'un circuit ultérieur, MIMOSA22THRb [7]. Il est ainsi validé, qu'un pixel de $39.0\times50.8\,\mu\text{m}^2$ permet d'atteindre une résolution spatiale intrinsèque de l'ordre de 10 µm avec une efficacité de détection proche de 100%. Cette confirmation ouvre la voie pour le développement d'échelle de grande longueur, supérieure au mètre, dont BESSIII a besoin.

Un logiciel générique proposant une stratégie d'optimisation pour un trajectomètre pixellisé a été mis au point et appliqué au cas particulier du SPT de BESSIII. Il a permis de montrer qu'une configuration à trois couches simple-face optimise la résolution sur les trajectoire alors qu'une option incluant une couche double-face atteint une efficacité de reconstruction supérieure.

Les travaux réalisés permettent de conclure plus généralement sur les dispositifs de trajectométrie, notamment pour un futur collisionneur électron-positon en Chine (CEPC). Ils indiquent la possibilité de remplacer les dispositifs de détection gazeux usuellement employés à grand rayon mais limitant le taux d'interaction observables, par des dispositifs exploitant entièrement la technologie des capteurs à pixels CMOS. L'outil logiciel donne également la possibilité dans les études d'optimisation de considérer à la fois la résolution sur l'impulsion et celle sur l'extrapolation des trajectoires, qui sont deux quantités majeures pour l'identification robuste des états finals produits dans les collisions.

Introduction

In the field of τ-charm physics, the BEijing Spectrometer III (BESIII) experiment at Beijing Electron Positron Collider II (BEPCII) is remarkable for its fruitful results including a series of discoveries like four-quark bound states and precise measurements for R parameter, the τ mass and various decay modes of J/ψ, ψ' and $\psi(3770)$. One of the dominating sub-detector to bring the physics results is the multilayer drift chamber (MDC) which is the main tracker constructed with a spatial resolution better than $130\mu m$ on average and a high transfer momentum resolution around $0.5\%@1GeV/c$. Nevertheless, it suffers from ageing effects which lead to performance degrading in terms of detection efficiency, momentum resolution and spatial resolution. From the analysis in 2013, the efficiency of the MDC inner cell could be around 70% and its spatial resolution degraded from $\sim 120\mu m$ to $\sim 300\mu m$ in the worst case. This ageing effect would prevent BESIII to take as much data as its designed and would decrease the precision of measurements. The collaboration decided therefore to upgrade the original inner chamber with a newly designed one or with a tracker using more advanced technology. The project of a Silicon Pixel Tracker (SPT) was proposed in this perspective.

CMOS Pixel Sensor (CPS) also named as Monolithic Active Pixel Sensor (MAPS), first developed at IPHC, has offered an opportunity for nuclear and particle physics experiments to address the physics studies requiring accurate spatial resolution and low material budget. Therefore, it is possible now to open up a research frontier to replace a gas chamber with a CPS based silicon tracker. The CPS technology has been fully validated by the STAR experiment at the Relativistic Heavy Ion Collider (RHIC) in Brookhaven National Laboratory. New generations of CPS for particle physics applications are being researched and developed under the promotion of ALICE experiment at the large hadron collider (LHC) for the upgrade of its vertex detector. On the other hand, a more radiation tolerant technique of High voltage CMOS pixel sensor is favoured by detectors like

A Toroidal LHC ApparatuS (ATLAS) and the Compact Muon Solenoid (CMS), both of which were constructed for high luminosity p-p collisions. Benefiting from the rapid expansion of CPS technology in high energy physics experiments, this thesis could address the development of BESIII inner tracker using CPS sensors to replace the original gas chamber.

The required spatial resolution for BESIII is less demanding than for vertex detectors like the Heavy Flavor Tracker (HFT) in STAR, because the generated D mesons are almost at rest and there is no production of a jet from hadronizations. Therefore, the spatial resolution of CPS sensors that are developed for vertex detectors exceeds the requirement of a tracker and leads to a waste of power consumption and data transmission bandwidth. The CPS with larger pixel size is then favoured by BESIII. As the main specification of a tracker, the momentum resolution of the inner tracker is chosen as the figure of merit to guide the geometry optimisation. In this thesis, several sensors with larger pixel pitches have been tested to validate the latest CPS technology and the general considerations of detector design problems in tracking theory have been discussed and summarized. With the answers of the mathematical optimisation problem for tracking, the potential of CPS applications in BESIII will be explored by evaluating and optimising several tracker designs.

The chapters composing the thesis are organised as below:

- In Chapter 1, Beijing spectrometer III (BESIII) experiment at Beijing Electron Positron Collider II (BEPCII) will be introduced. The physics goals and detector design of BESIII will be presented and the ageing effect of the multilayer drift chamber will be emphasised since it is the motivation for this thesis.
- In Chapter 2, CMOS pixel sensor technology and performance will be presented to demonstrate the possibility of CPS to meet the requirements of BESIII inner tracker and vertex detectors in future colliders. The mature rolling-shutter design architecture and the latest asynchronous readout strategy will be illustrated with the design of state-of-the-art MIMOSA-28 chip and the design of the frontier ALPIDE chip respectively. Sensor test facilities and their physics objectives will also be introduced.
- In Chapter 3, general considerations of sensor design optimisation for BESIII will be presented. To verify the latest CPS technology, the data taken from the beam test of single arm large area telescope (SALAT) and the test about full size building block

(FSBB) sensors will be analysed. Considering the published results of MIMOSA-22THRb, the sensor performance with different pixel sizes and CMOS processes will be compared to show that the new TowerJezz 0.18 µm technology could make CPS work with larger pixel size, lower power consumption and higher readout speed than MIMOSA-28 in AMS 0.35 µm process, which offers more options for CPS designs.

- In Chapter 4, the two widely used tracking algorithms, generalized least squares (GLS) method and Kalman filtering/smoothing technique will be shown with derivations and examples. The implications of tracker parameters like the number of layers, the intrinsic resolution of each layer and the length of lever arm, will be illustrated using parabola model and a simple tracker geometry with uniform spacing. A Kalman filter implementation for beam trajectories will also be presented. Concerning track reconstruction efficiency, the method used in STAR experiment will be introduced for the related calculations.
- In Chapter 5, the general strategy to optimize barrel silicon pixelated trackers will be given using tracks in the bending plane, and the geometry optimization for vertexing will also be presented using telescope systems. In addition, the income of a double-sided ladder like the Pixel Ladders with Ultra-low Material Embedding (PLUME) on tracking performance, will also be discussed. The R&D is based on the algorithms described in Chapter 4.
- In Chapter 6, simulation tools and the relevant results will be presented. The standalone Geant4-based simulation software will be presented. It is developed for silicon tracker studies and the validation of the optimization strategy in Chapter 5. The full Monte-Carlo simulations processed in BESIII offline software system, will also be illustrated to show the overall performance of the silicon pixel tracker intensified with the outer gas chamber of BESIII.
- In the last chapter, the results of this thesis will be summarized with the main conclusions, and the perspectives of CPS based vertex detector at Circular Electron Positron Collider (CEPC) will be foreseen.

Chapter 1

Detectors and physics in the BES-III experiment at BEPC-II

In this chapter, the physics motivation and the experiment facilities including the collider BEPCII and the detector BESIII will be presented. The requirements on the tracker system is given by the momentum distribution of the final states in important physics channels. Due to the higher background and increased physical events coming from the improved peak luminosity, the ageing effect of BESIII gas chamber becomes more serious, therefore the influence of this effect will be emphasized in this chapter to show the urgency of the Research and Development (R&D) of a new inner tracker.

1.1 Physic of BESIII experiment

As the standard model has been well tested and confirmed by numerous experiments in the electroweak sector, new physics are searched mainly through two approaches: One is the searching in higher energy region like the experiments ATLAS and CMS at LHC; the other is to observe with precise measurements and high luminosity in those colliders like the B-factory KEKB and the charm-factory BEPCII. On the other hand, the quantum chromodynamics (QCD) of the strong sector in standard model has also been verified precisely in the case of very high energy transfer, but its predictions are limited for low energy physics due to the loss of applicability of the perturbation theory. Therefore, the understanding of the strong interaction in low energy region is at the frontier of particle physics. As a hopeful non-perturbative approach, the lattice QCD and its improved calculation accuracy could bring more predictions including the properties of glueballs and hybrid states. However, the reliability of the calculations still require the calibration from precise experiments. The hadrons and their decay products in tau-charm energy region for which BEPCII/BESIII is built, play a decisive role in calculations of lattice QCD. In these two types of precise measurements, the BESIII collaboration has already made great contributions since 2009. The main physics topics in BESIII are introduced in the following sections.

1.1.1 Studies of electroweak interactions

The well constructed electroweak theory could be tested using the measurements of related branching fractions with high precision and the efficient particle reconstructions in BESIII.

1.1.1.1 Precise measurements of CKM matrix elements

Since the observed quarks in weak interactions are not their mass eigenstates, the CKM matrix which is proposed by Cabibbo [8] and generalized by Kabayashi and Maskawa [9], is required to describe the relations between the eigenstates in weak interactions and their mass eigenstates in strong interactions. By convention, CKM matrix is applied on the mass eigenstates of down-type quarks (d, s, and b):

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}, \tag{1.1.1}$$

where only 4 independent parameters (three mixing angles and one complex phase for CP violation) are to be determined because of the property of unitary matrix and the 5 absorbed phases from the 6 quark fields. In BESIII, the magnitudes of $|V_{cd}|$ and $|V_{cs}|$ could be deduced from the measured branching fractions of the semileptonic decays of D mesons including: $D^0 \to \pi^- e^+ \nu_e$, $D^0 \to \pi^- \mu^+ \nu_\mu$, $D^0 \to K^- e^+ \nu_e$, $D^0 \to K^- \mu^+ \nu_\mu$, $D^+ \to \bar{K}^0 e^+ \nu_e$, $D^+ \to \bar{K}^0 \mu^+ \nu_\mu$, etc. The relations [10, 11, 12] of semileptonic decay width for $D^0 \to K^-(\pi^-) e^+ \nu_e$, the form factor $|f_+^{K(\pi)}(0)|$ and the CKM elements $|V_{cs(d)}|$

are given below:

$$\Gamma(D^0 \to K^- e^+ \nu_e) = \frac{B(D^0 \to K^- e^+ \nu_e)}{\tau_{D^0}} = 1.53|V_{cs}|^2 |f_+^K(0)|^2 \times 10^{11} s^{-1}, \tag{1.1.2}$$

$$\Gamma(D^0 \to \pi^- \ e^+ \nu_e) = \frac{B(D^0 \to \pi^- \ e^+ \nu_e)}{\tau_{D^0}} = 3.01|V_{cd}|^2 |f_+^\pi(0)|^2 \times 10^{11} s^{-1}. \tag{1.1.3}$$

Since the form factors could be calculated by lattice QCD, the magnitudes of the related CKM elements could be extracted from Equation 1.1.3 and Equation 1.1.2. The systematic error of this method is given by

$$\frac{\Delta V_{cs(d)}}{V_{cs(d)}} = \sqrt{\left(\frac{\Delta B}{2B}\right)^2 + \left(\frac{\Delta \tau_{D^0}}{2\tau_{D^0}}\right)^2 + \left(\frac{\Delta f_+^{K(\pi)}(0)}{f_+^{K(\pi)}(0)}\right)^2}. \tag{1.1.4}$$

On the other hand, the form factors could also be extracted using the measured branching fractions, the current $|V_{cs(d)}|$ values and D^0 lifetime τ_{D^0} in PDG [4] with the system error:

$$\frac{\Delta f_+^{K(\pi)}(0)}{f_+^{K(\pi)}(0)} = \sqrt{\left(\frac{\Delta B}{2B}\right)^2 + \left(\frac{\Delta \tau_{D^0}}{2\tau_{D^0}}\right)^2 + \left(\frac{\Delta V_{cs(d)}}{V_{cs(d)}}\right)^2}. \tag{1.1.5}$$

Combining Equation 1.1.3 and Equation 1.1.2, the ratio $|V_{cd}|/|V_{cs}|$ could be determined from the two branching fractions and the lattice QCD ratio $|f_+^\pi(0)|/|f_+^K(0)|$ with a higher precision, as both the measurement error and the calculation uncertainty are significantly reduced.

1.1.1.2 Precision tests of lepton universality

As an assumption in standard model, the lepton universality has been verified using leptonic or semileptonic decays of τ and measurements of the branching fraction ratio of μ and τ decays with electron production. For example, tau-muon universality could be tested using the equation below:

$$\frac{B(\tau^- \to e^- \bar{\nu}_e \nu_\tau)}{B(\mu^- \to e^- \bar{\nu}_e \nu_\mu)} = \frac{\tau_\tau}{\tau_\mu} \left(\frac{m_\tau}{m_\mu}\right)^5, \tag{1.1.6}$$

and the muon-electron universality leads to

$$\Gamma(\tau^- \to e^- \bar{\nu}_e \nu_\tau) = \Gamma(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau). \tag{1.1.7}$$

Compared with the electron positron colliders working in a higher energy range, BEPCII could control the background by running under the production threshold of charmed mesons. Since tau pair production could reach the maximum cross-section rapidly after the production threshold, high statistics of pure tau pairs are guaranteed. In addition, the tau data in BESIII samples of ψ', ψ'' and $D_s^+ D_s^-$ are cleaner than the data in $B-$factories [13]. Consequently, the branching fractions of the dominating decay channels of τ could be measured with very low system uncertainties. Considering the precise calculations on the cross-section of tau pair production near the threshold energy, τ mass measurement accuracy could also be improved significantly.

1.1.2 Studies of strong interactions

As the perturbation theory has limitations in tau-charm energy region, parameter calculations in QCD are with large uncertainties. The data taken in BESIII could help theorists to develop their theories or methods like Lattice QCD. Main topics are summarized below:

- Using related calculation results, parameters like the mass of charm quark and the coupling constant α_s for strong interaction could be measured. For instance, R value scan, tau decay and J/ψ inclusive radiative decay could be used to study α_s.

- The hadron spectrum study in tau-charm energy range could be studied with a high precision. As the best process to study glue balls, the radiative decays of J/ψ could be analysed using partial wave analysis to search for the possible candidates and even to study their properties. In addition, hybrid states and the normal mesons could be studied by a series of processes including the radiative or hadronic decay of J/ψ, decay of $\psi(2S)$ and the other charmonium states, the non-leptonic decays of D or D_s, the semi-leptonic decay of τ, etc.

- Charmonium states and their decays could be used to study and develop QCD theories. Some input parameters of the potential model and non-relativistic QCD (NRQCD) could be determined by BESIII measurements. As an important topic, the spin-singlet P-wave state $h_c(^1P_1)$ could be studied in BESIII with better statistics and lower background compared with B-factories. Another famous problem in the study of charmonium dynamics is the $\rho\pi$ puzzle: The ratio of the branching fraction of some J/ψ decays to that of ψ' decays with the same final states, deviates from the perturbative-QCD predictions severely. As almost all the decay channels from the related charmonium states could be measured precisely in BESIII, A better comprehension of this puzzle could be foreseen.

1.1.3 Searches for new physics

As a high luminosity experiment in the precision frontier, rare decays and non standard model processes could be measured to search for new physics. The weak decays with D/D_s production and lepton family number violating decay modes of J/ψ samples, the flavor-changing neutral current (FCNC) decays of D meson samples and CP violating effect in $D_0\bar{D}_0$ mixing and D_s/D decays are interesting processes in BESIII. They are used for the precision test of standard model and the limit constrains of possible new mechanisms.

1.2 Beijing Electron Positron Collider II

As the successor of Beijing electron positron collider, BEPCII has reached its designed peak luminosity $1 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at beam energy $2\times1.89\,\mathrm{GeV}$ in April 2016. The machine consists of three components: the linear accelerator, the beam transportation line and the two storage rings. Electrons and positrons could be accelerated to 1.89 GeV or even 2.5 GeV in the updated 202 m-long linear accelerator [14]. After the beam injection, the transportation line will transport the e^- and e^+ beams into their storage rings respectively. As shown in Figure 1.1, the two 237.5 m-long rings are used to store 93 bunches of electrons and the same number of bunches of positrons respectively and guide them to collide with a bunch spacing time around 8 ns. The bunch interval is therefore around 2.4 m and the crossing angle is $\sim 22\,\mathrm{mrad}$ in the horizontal direction. The main design parameters of the machine are summarized in Table 1.1

At the interaction point, beam pipe is designed with minimal materials to reduce its interactions with the particles produced in the beam collisions. The cross-section view

Parameters	Designs
Centre of mass Energy (GeV)	$2 - 4.6$
Circumference (m)	237.5
Number of rings	2
Radio-frequency(RF) $f_{rf}(MHz)$	499.8
Peak luminosity ($cm^{-2}s^{-1}$) at $2 \times 1.89\ GeV$	$\sim 10^{33}$
Number of bunches	2×93
Beam current (A)	2×0.91
Bunch spacing (m/ns)	2.4/8
Bunch length σ_z (cm)	1.5
Bunch width σ_x (μm)	~ 380
Bunch height σ_y (μm)	~ 5.7
Relative energy spread	5×10^{-4}
Crossing angle (mrad)	± 11

Table 1.1: Main design parameters of BEPCII [1]

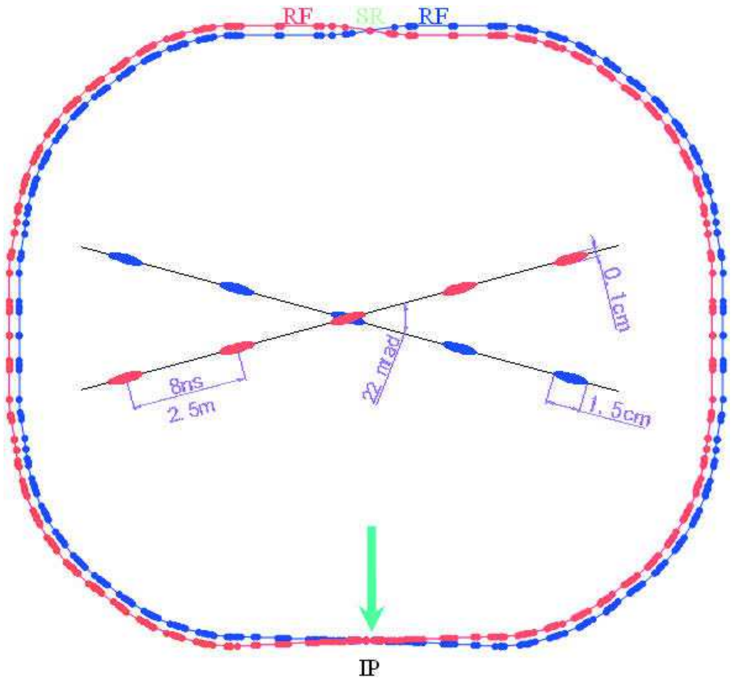


Figure 1.1: The storage rings and the designed interaction point of BEPCII

of the interaction region is shown in Figure 1.2 with the beam pipe, the MDC inner end plates and the inner tube. The beryllium central pipe welded on the copper extension on both sides, is 29.6 cm long with an inner radius of 31.5 mm. To reduce beam related backgrounds like synchrotron photons and to withstand the heat from the radiation on a solid pipe, the central region is designed using 14.6 µm gold coated on the inner surface of the 0.8 mm beryllium wall, a 0.8 mm channel full of pure mineral oil for cooling and a 0.6 mm thick beryllium outer wall. The radiation length for a straight trajectory from

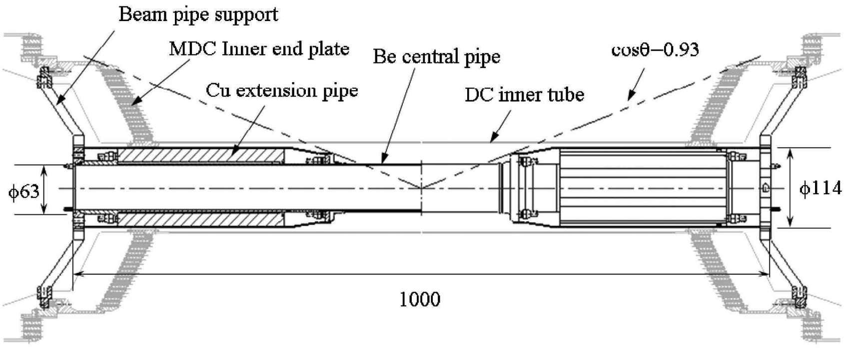


Figure 1.2: The cross-section view of the interaction region of BESIII

the primary vertex in the bending plane is $1.04\%X_0$.

1.3 The current design of BESIII

The BEijing Spectrometer III (BES-III) shown in Figure 1.3 is the only detector to explore particle physics at BEPC-II. It contains the multilayer drift chamber (MDC), time-of flight (TOF) detector, the CsI(Tl) electromagnetic calorimeter (EMC), the superconducting solenoid magnet (SSM) supporting 1 T magnetic field and the muon identifier (MU). The polar angle coverage of BESIII is $21° < \theta < 159°$. In one calendar year, the expected data samples are summarized in Table 1.2. And the main parameters and performance of BES-III are listed in Table 1.3. Brief introductions of all the sub-detectors are given in the following sub-sections.

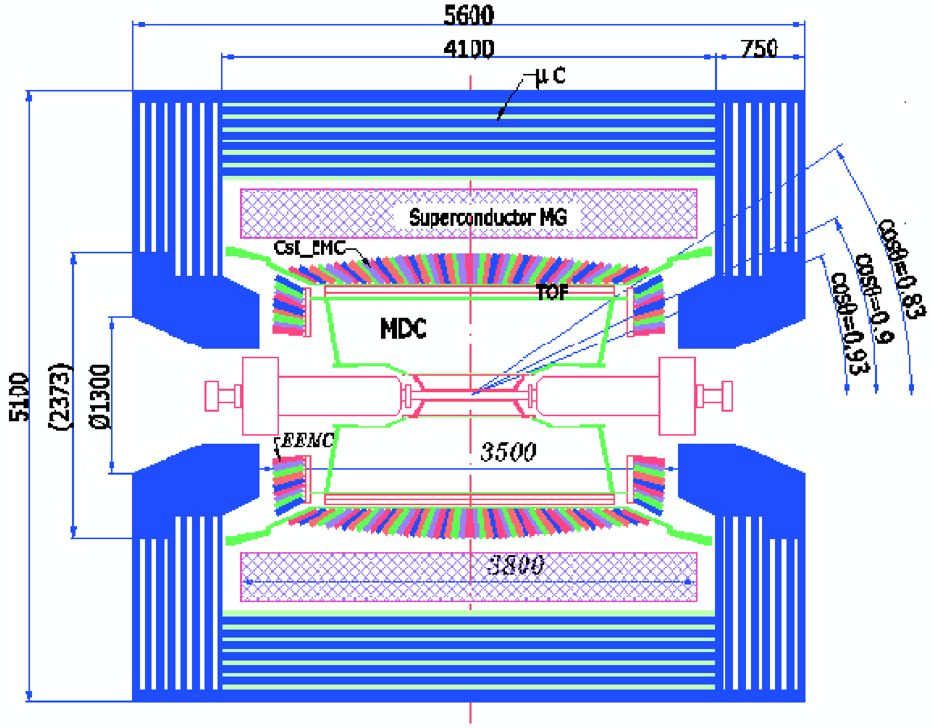


Figure 1.3: Schematic diagram of the BESIII detector[1]

states	Energy (GeV)	Peak luminosity ($10^{33}\ cm^{-2}s^{-1}$)	Cross section (nb)	Events/year
J/ψ	3.097	0.6	3400	1.0×10^{10}
$\psi(2S)$	3.686	1.0	640	3.0×10^{9}
$\tau^+\tau^-$	3.670	1.0	2.4	1.2×10^{7}
$D^0\bar{D}^0$	3.770	1.0	3.6	1.8×10^{6}
D^+D^-	3.770	1.0	2.8	1.4×10^{6}
$D_s^+D_s^-$	4.030	0.6	0.32	1.0×10^{6}
$D_s^+D_s^-$	4.170	0.6	1.0	2.0×10^{6}

Table 1.2: Predicted data samples in BESIII for one year's running [1, 15]

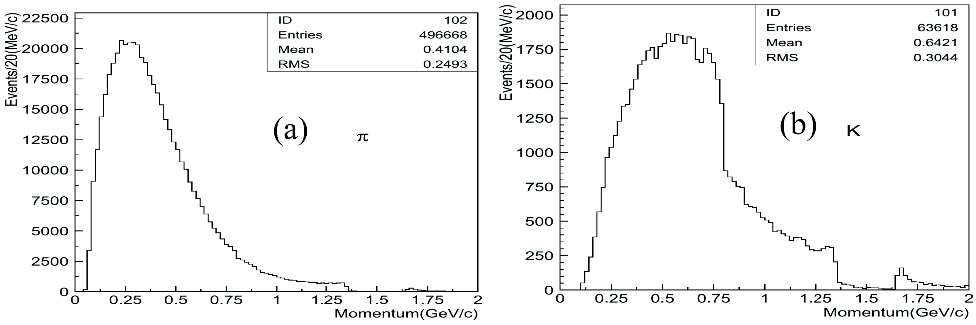
1.3.1 Multilayer drift chamber

The multilayer drift chamber (MDC) is designed to reconstruct the trajectories and to measure the momentum of the charged final states created by e^+e^- collisions in BEPCII , with a high precision to fulfil the physics goals aiming at precise measurements and new physics in $\tau - c$ energy range. Meanwhile, the measurements of charges ionized by the flying particles could be used in particle identification by checking their dE/dx spectrum. The momentum distributions of the charged final states in J/ψ and D^0 decays are shown in Figure 1.4 and Figure 1.5 respectively. They indicate that most of the charged particles have a momentum lower than 1.5 GeV and the averaged momentum of pions(kaons) is around 0.4(0.65) GeV. In addition, the electrons coming from D^0 decay have the mean momentum around 0.3 GeV and the most probable momentum lower than 0.25 GeV. Therefore, the tracker system in BESIII is extremely sensitive to material effects like the multiple scattering for all charged particles, and the energy loss of electrons and positrons. These effects raise the requirements on material budget of the inner tracker system. A gaseous tracker is thus favoured by BESIII, and the transverse momentum resolution of MDC is determined to be better than $\sigma_{p_t}/p_t = 0.32\% p_t \oplus 0.37\%/\beta$ [15].

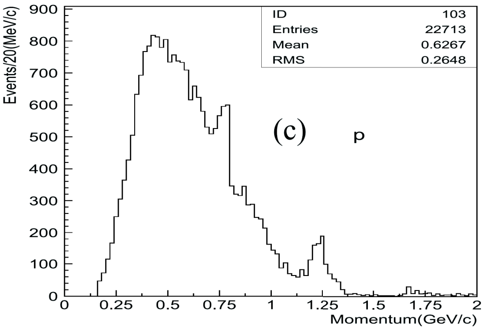
In the radial range from 59 mm to 810 mm, the MDC is designed with two parts: the inner chamber which consists of the inner carbon-fibre wall ($0.45\% X_0$) and 8 detecting layers, and the outer chamber which starts at 182.5 mm and contains 35 detecting layers. To reduce multiple scattering, there is no separation wall between the two chambers.

Subsystem parameters	Performance
MDC	
Single wire $\sigma_{r\phi}$(µm)	130
σ_P/P (1 GeV/c) (%)	0.5
$\sigma(dE/dx)$ (%)	6
EMC	
σ_E/E (1 GeV/c) (%)	2.5
Position resolution (1 GeV/c) (cm)	0.6
TOF	
σ_T (ps)	
Barrel	100
End cap	110
MU	
No. of layers (barrel/end cap)	9/8
Cut-off momentum (MeV/c)	0.4
SSM magnet field (T)	1.0
$\Delta\Omega/4\pi$	93%

Table 1.3: Detector parameters and performance of BESIII [1]

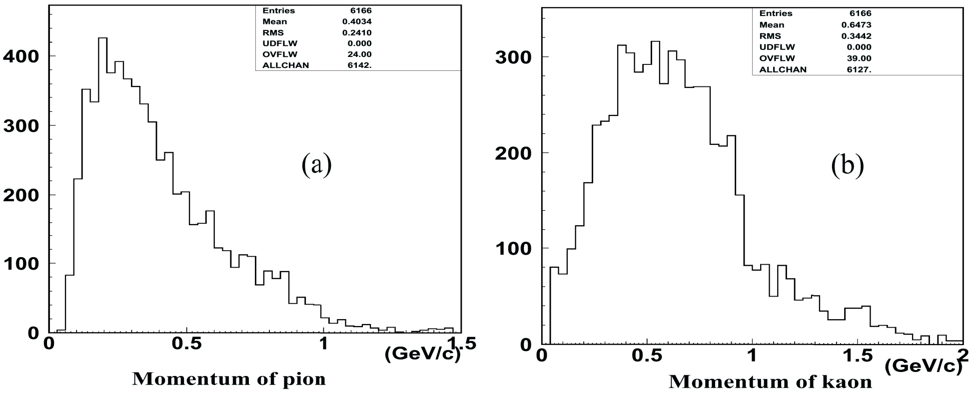


(a) Momentum distribution of pion (b) Momentum distribution of kaon

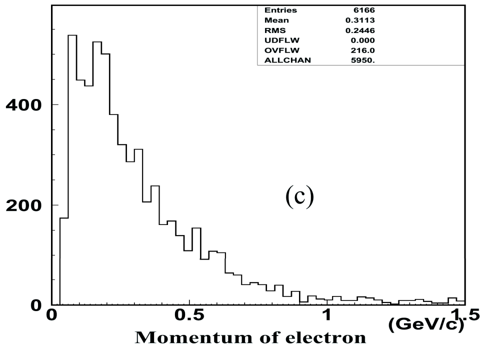


(c) Momentum distribution of proton

Figure 1.4: Momentum distributions of charged particels produced from J/ψ decays (BESII data)[16].



(a) Momentum distribution of pion (b) Momentum distribution of kaon



(c) Momentum distribution of electron

Figure 1.5: Momentum distributions of charged particels produced from D^0 decays (BESII data)[16].

The working gas is a helium based mixture He/C_3H_8 (60/40) with a radiation length around 550 m. This leads to an averaged radiation length X_0 ~218 m assuming that the materials of wires are distributed uniformly in the chamber [16]. As the result of MDC configurations, the single wire spatial resolution is better than 130 µm in $r - \phi$ plane in average to guarantee the transverse momentum resolution $\sim$ 0.5% for heavy charged particles at 1 GeV/c, and the position resolution at the impact point (IP) is ~2 mm in the beam direction. Moreover, the dE/dx resolution for particles in the bending plane is 6% better than the design goal and a 98% hit reconstruction efficiency in a cell is also achieved [17].

1.3.2 Time of flight system

The time of flight (TOF) counters between MDC and EMC are used to measure the flight time of the charged particles, which will contribute on particle identification, cosmic ray rejection and the level 1 trigger. The main components to make up the TOF incorporate plastic scintillators and photomultiplier tubes. To cover almost the same solid angle as that of MDC, there are two sub-systems: the barrel TOF with time resolution around 100 ps to cover the angle $|cos(\theta)| < 0.83$ and the end cap TOF with time resolution ~110 ps covering the range $0.85 < |cos(\theta)| < 0.95$.

1.3.3 CsI(Tl) electromagnetic calorimeter

The CsI(Tl) crystal based electromagnetic calorimeter (EMC) could measure the energy and position of photons and electrons precisely in the energy range from 20 MeV up to 2 GeV just outside of the TOF system. And a good e/π discrimination is guaranteed for the range above 200 MeV. Therefore, EMC could also offer the trigger information for neutral particles. The material budget of the crystals is about $15X_0$ which is a length of 28 cm. The designed energy resolution and position resolution are $\sigma_E/E = 2.5\%\sqrt{E}$ and $\sigma \leq 6\,mm/\sqrt{E}$ at 1 GeV respectively.

1.3.4 Muon identifier

Muon identifier (MU) is the outermost sub-detector used for identification and measuring of muons with momenta above 0.4 GeV/c. The resistive plate counters (RPC) are used for muon hit reconstruction, and steel plates are placed between two counters to absorb the other charged particles. The spatial resolution in ϕ direction is better than 1.2 cm, and the detecting efficiency is greater than 0.95%.

1.4 Ageing effect of MDC in BES-III

The deterioration problem of the inner chamber and its influence to physics studies will be presented in this section. Gas chambers have the ageing effect because of gas polymer condensing on electrodes and charge accumulation on these insulating layers, which lead to gain decrease and a lower signal to noise ratio. In addition, the ageing effect on cathodes named as Malter effect [18] could prevent positive ions to absorb electrons, which may result in a very high electric field to emit electrons. Once the electrons drift to the anodes, avalanches near the sensing wires will generate more positive ions to enhance this effect, thus the continuous self-discharge will make the chamber extremely noisy or even dead due to the high currents. To solve the cathode ageing problem, water vapour could be injected to the gas mixture to absorb the electrons from Malter effect, but it will result in gain decreases. Although using higher operating voltages on wires could increase the gain, the ageing effect will be accelerated. Consequently, the tracking efficiency and spatial resolution of a gas chamber will deteriorate under high irradiations.

The ageing effect in MDC of BESIII has been studied by literature [19]. As the hit rate in the outer chamber is much lower, the influence of ageing is not serious. However, the gain decrease in the inner chamber (the first 8 layers) is dramatic. The hit rates of single wires obtained from random triggered events at the beginning of each run are shown in Figure 1.6 to illustrate the crucial radiation environment of the inner chamber compared with the outer one. To avoid the influence on temperature, the data used for this analysis are taken on similar dates in each year. The changes of hit rate in different years are mainly due to the optimisation of machine running parameters. The gain decrease is studied using Bhabha events and accumulated charges calculated from the cell currents in related years [19]. The results of relative gains with respect to the gain in 2009 are shown in Figure 1.7 and Figure 1.8, where the gains of all the 8 layers in the inner chamber decrease dramatically. In the worst case, the degradation of the first layer has a factor about 29% during the 5 years operation. Finally, the implications for tracking performance could be seen from the decreased hit efficiency and spatial resolution in the

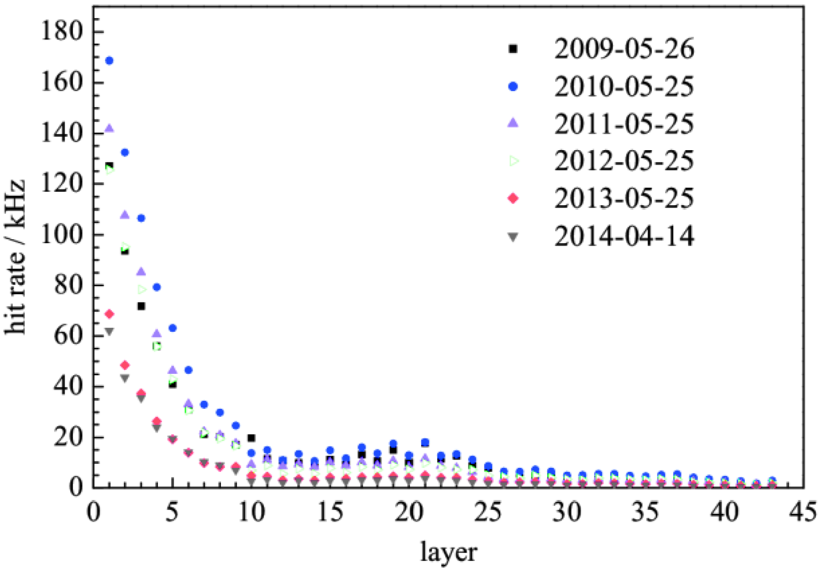


Figure 1.6: Single wire hit rate in May of each year as a function of MDC layer index [19].

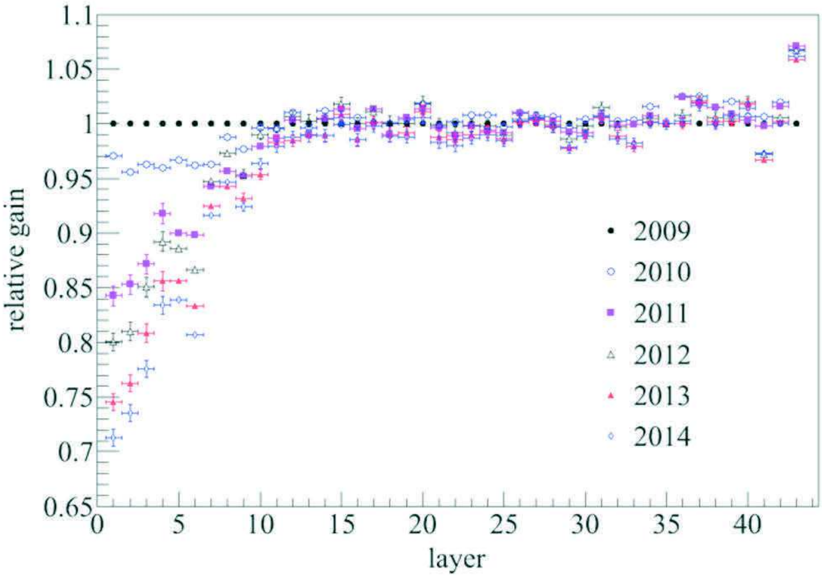


Figure 1.7: Gain decrease of cells in each year obtained from Bhabha events as a function of layer indices [19].

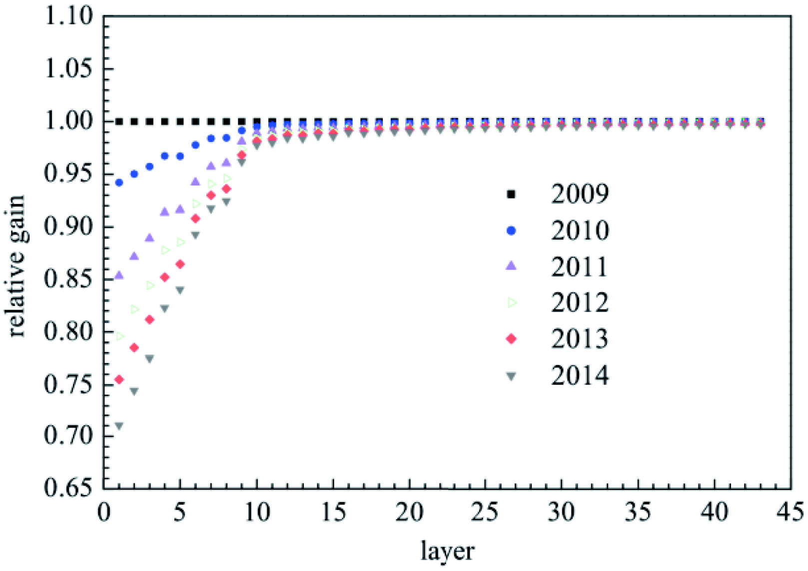


Figure 1.8: Gain decrease of cells in each year according to accumulated charges and ageing ratio as a function of layer indices [19].

analysis using the data taken in 2012. The hit efficiency of wires in each layer is shown in Figure 1.9. According to the beam test results in 2007, the hit efficiency of the first 8

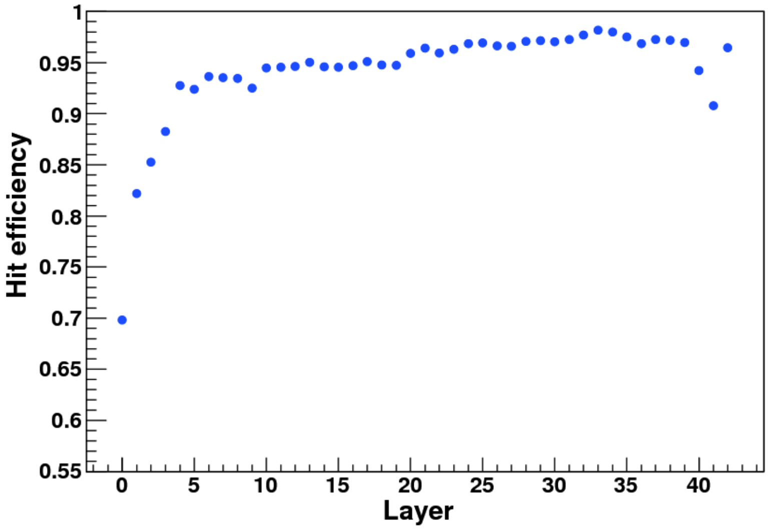


Figure 1.9: Single wire hit efficiency as a function of MDC layer index (analysis in 2012) [2].

layers was better than 98% [17]. The efficiency of the first 4 layers declined seriously. At the first layer, hit efficiency even decreased almost 29% [19]. Compared with the tested spatial resolution $< 130\,\mathrm{mm}$ for 1 GeV muon in 2009 [19], the spatial resolution is also affected greatly for layers in the inner chamber as shown in Figure 1.10.

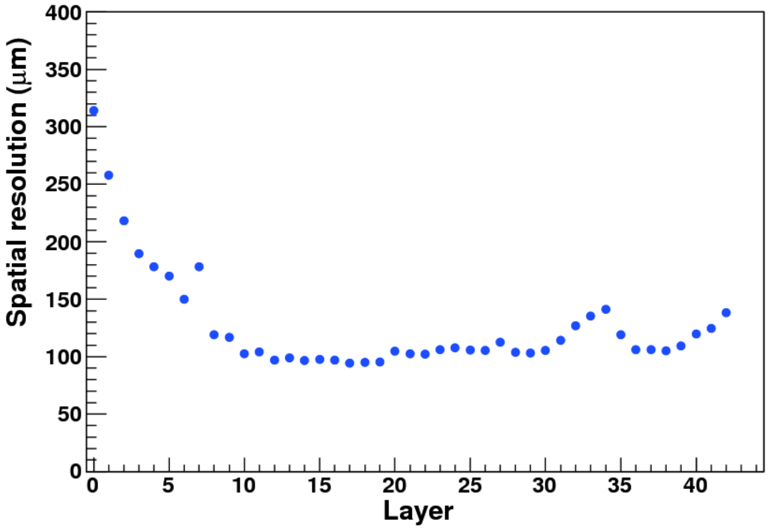


Figure 1.10: Spatial resolution in R-ϕ plane of each layer as a function of MDC layer index (analysis in 2012) [2].

1.5 Summary and conclusion

In this chapter, the BESIII experiment was introduced with the ageing effects of MDC, the motivation of this thesis. As the only collider and detector focusing on the physics in τ-charm region currently, the BESIII experiment at BEPCII is remarkable for its performance and fruitful results. Nevertheless, the ageing effect of the inner chamber of the main tracker system has led to performance degrading in detection efficiency, spatial resolution and thus momentum resolution of the tracker. Since both data acquisition and data quality are influenced seriously, it is quite urgent to develop a new generation of inner tracker for BESIII to face the coming challenges in some similar physics studies of B-factories. Because a gas chamber suffering strong radiations cannot get free from ageing effects, the study on using CMOS Pixel sensors to construct an inner tracker is profound and promising.

Chapter 2

CMOS Pixel Sensor

Complementary metal-oxide-semiconductor (CMOS) refers to the family of processes to implement CMOS circuits and the design style of integrated circuits itself. In CMOS design, both p-channel metal-oxide-semiconductor field effect transistor (MOSFET) and n-channel MOSFET are used to realize its low static power consumption, fast response in both high-to-low and low-to-high output transitions, noise immunity and high integration density. On the strength of CMOS technology, the pixel sensing circuits, the analogue shaping amplifier, the analogue to digital converter (ADC) and even the zero suppression circuits for fast data transmission could be implemented into one chip, which is the reason why CPS was also named as monolithic active pixel sensor (MAPS) at the beginning. The performance of CPS has been fully validated by the project of the Detector R&D towards the International Linear Collider founded by EU (EUDET) using MIMOSA-26 chips and the Heavy Flavor Tracker (HFT) containing two layers of MIMOSA-28 chips with the spatial resolution $\sim$3.5 µm in STAR experiment. Figure 2.1 shows the Distance of the Closest Approach (DCA) resolution which could be achieved in 200 GeV Au-Au data taken at RHIC in 2014 [20]. With a preliminary alignment, the pointing resolution of 750 MeV pions is $\leq$60 µm in both x-y plane and z direction. Figure 2.2 shows the background reduction performance of the HFT in the $D^0\bar{D}^0$ invariant mass signal [21].

This chapter will start from the particle interactions with matter to illustrate the detecting principles of CPS and then introduce the mature designs for charged particle detection. The specification of CPS and the related test methods are also presented.

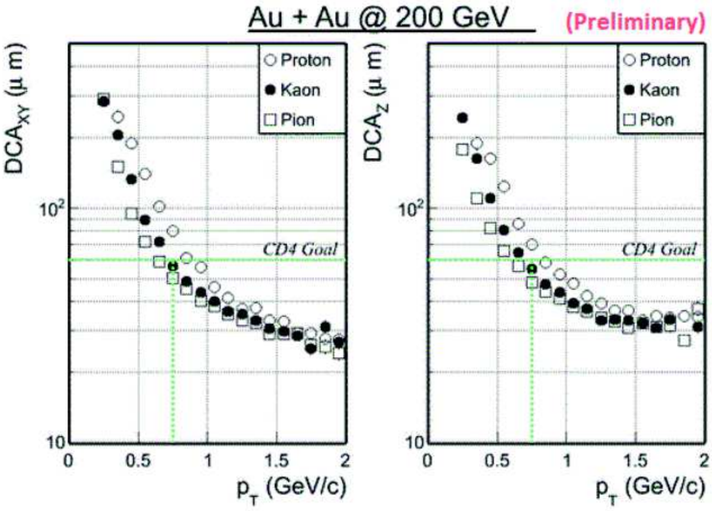


Figure 2.1: DCA resolution for TPC tracks with one hit from the Intermediate Silicon Tracker (IST) and two hits from the PiXel Layers (PXL) as a function of transverse momentum in STAR 2014 data [20].

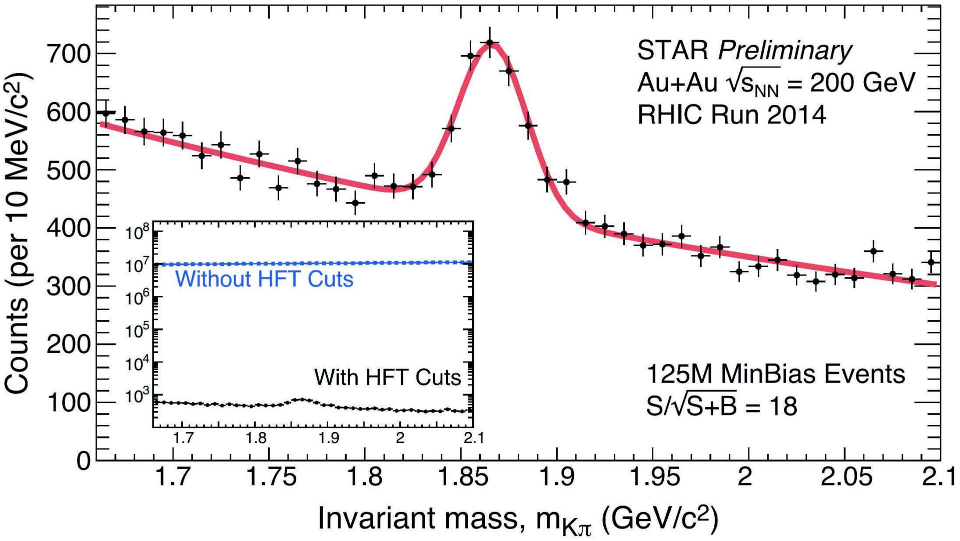


Figure 2.2: $D^0\bar{D}^0$ invariant mass signal in STAR 2014 Au-Au data [21].

2.1 Particle interactions with matter and detecting principles of CPS

The interactions between the passing particles and the medium are essential processes of particle detection and measurements. On the other hand, the material effects like the energy loss and the multiple Coulomb scattering bring uncertainties on the reconstructions of hits and tracks in a detector. Therefore, they are introduced before the explanation of CPS operation principles and the tracker optimization based on tracking algorithms.

2.1.1 Electronic energy loss of charged particles

The mean energy loss due to ionization and excitation of atoms along the flying path of charged heavy particles in the medium can be described by the Bethe equation:

$$\left\langle -\frac{dE}{dx} \right\rangle_{electronic} = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{2.1.1}$$

where x is the mass per unit area,
K is a constant $\sim 0.307\,075\,\mathrm{MeV\,mol^{-1}\,cm^2}$,
z is the charge number of the charged particle,
Z and A are the atomic number and atomic mass (in $\mathrm{g\,mol^{-1}}$) of the medium respectively,
m_e is the rest mass of an electron,
W_{max} is the maximum energy transfer of the incident particle in a single collision,
I is the mean excitation energy of the absorber,
$\delta(\beta\gamma)$ is the density effect correction.
Equation 2.1.1 is accurate with a few percent in the range $0.1 \leq \beta\gamma \leq 1000$ for materials with moderate Z and incident particles like muons, pions and protons [4]. For ions, additional corrections for finite target size and higher order photon coupling are required [4, 22]. If additional low energy corrections are applied into the Bethe equation properly, the accuracy could be in 1% down to $\beta \approx 0.05$ which is $\sim 1\,\mathrm{MeV}$ for protons [4, 23, 24, 25, 26]. For $\beta\gamma > 1000$, radiative energy loss has to be considered.

The maximum energy transfer W_{max} for a projectile with mass M is given below [4]:

$$W_{max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2} \tag{2.1.2}$$

The mean excitation energy I for different materials could be found in literatures [23, 24, 25, 26, 27] and the online tables of PDG website. The last term in Equation 2.1.1, the density effect correction, could be calculated as follows:

$$\delta(\beta\gamma) = \begin{cases} 2(\ln 10)x - \bar{C}, & if\ x \geq x_1; \\ 2(\ln 10)x - \bar{C} + a(x_1 - x)^k, & if\ x_0 \leq x < x_1; \\ 0, & if\ x < x_0(\text{nonconductors}); \\ \delta_0 10^{2(x-x_0)}, & if\ x < x_0(\text{conductors}) \end{cases} \tag{2.1.3}$$

in which $\bar{C} = 2\ln(I/\hbar\omega_p) + 1$ with the plasma energy $\hbar\omega_p = \sqrt{\rho < Z/A >} \times 28.816\,\text{eV}$. The relevant parameters of some materials of interest in this thesis are listed in Table 2.1. Some coefficients like a, k, x_0 and x_1 for $\delta(\beta\gamma)$ calculations can be estimated analytically by equations described in References [28, 27] in case of the materials without related measurements.

Z	A [g/mol]	ρ [g/cm^3]	I [eV]	a	$k = m_s$	x_0	x_1	$\bar{C}$	δ_0
4(Be)	9.01218	1.848	63.7	0.80392	2.4339	0.0592	1.6922	2.7847	0.14
6(C)	12.0107	1.700	78.0	0.20762	2.9532	0.0480	2.5387	3.1550	0.14
14(Si)	28.0855	2.329	173.0	0.14921	3.2546	0.2015	2.8716	4.4355	0.14
29(Cu)	63.5460	8.960	322.0	0.14339	2.9044	-0.0254	3.2792	4.4190	0.08

Table 2.1: The parameters of some materials [27]

The mean rate of energy loss for restricted energy transfer $T \leq W_{cut} \leq W_{max}$ is also summarized in Reference [4]. This rate depends on the value of W_{cut}. In addition, the mean loss rate is influenced dramatically by the rare knock-on electrons which take a large amount of energy away from the incident particle. Therefore, the mean energy loss is only useful for thick absorbers.

To estimate energy loss in materials with moderate thickness[*] x (in g cm^{-2}), one should use the most probable energy loss given below [29, 30, 31]:

$$\Delta_p = \xi \left[\ln \frac{2m_e c^2 \beta^2 \gamma^2}{I} + \ln \frac{\xi}{I} + 0.2000 - \beta^2 - \delta(\beta\gamma) \right] \tag{2.1.4}$$

---

[*] It refers to $\kappa \leq 0.05 - 0.1$, where the constant $\kappa \equiv \xi / W_{max}$ is used in Vavilov's equation.

where $\xi = (K/2)\,\langle Z/A\rangle\, z^2(x/\beta^2)$ MeV. The related energy loss probability distribution is named as straggling function which has been studied by several physicists using different methods summarized in literatures [31, 32]. The Landau-Vavilov distribution works well for moderate thickness, and it leads to the Δ_p above and a full width at half maximum (FWHM) $w_L \approx 4\xi$. However, the Δ_p given in Equation 2.1.4 is overestimated while w_L is underestimated significantly for very thin absorbers. In the case $\kappa < 0.6$, the straggling functions given by the convolution method coincides with experimental data [31]. Figure 2.3 presents the straggling functions for 500 MeV pions penetrating different thickness in silicon.

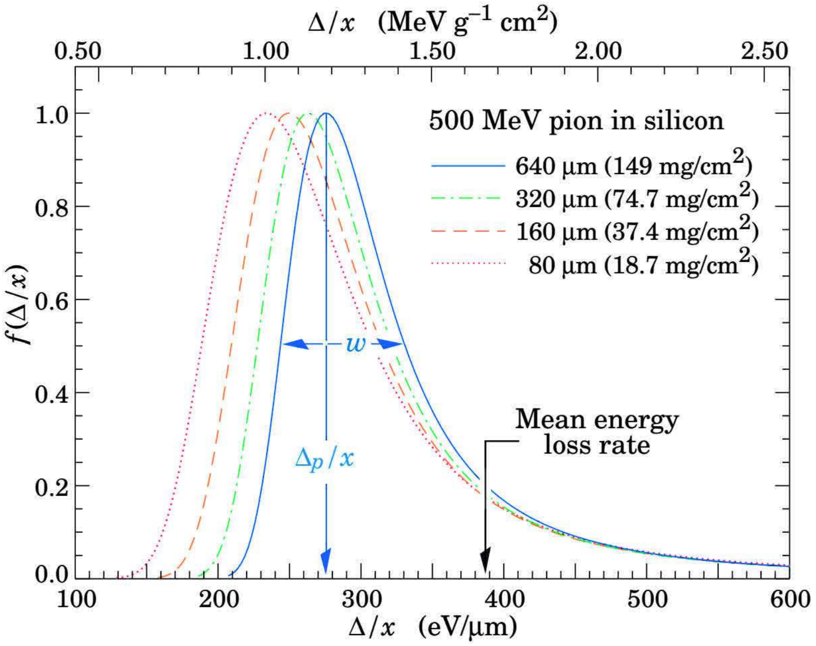


Figure 2.3: Straggling functions for 500 MeV pions penetrating different thickness in silicon. The distributions are normalized to have equal maximum [4].

The mean rates of electronic energy loss for electrons and positrons include different corrections from Møller cross section (in t/u-channel) and Bhabha equation (in s/t-channel) respectively. Although the form is different, the mean electronic energy loss of e^+/e^- is similar to that of charged heavy particles [4]. However, the radiative energy loss of e^+/e^- is much higher for the interested range of $\beta\gamma$ due to the small mass m_e.

2.1.2 Radiative energy loss of e^+/e^-

Bremsstrahlung is the main source of radiative energy loss for e^+/e^-. The relations and conclusions described here are for outer or external Bremsstrahlung in which the incident particle has small scattering angles. The differential cross section with respect to photon energy is described by

$$\frac{d\sigma}{dk} = 4\alpha r_e^2 \frac{1}{k}\left[f(Z, \frac{k}{E})\right] \tag{2.1.5}$$

where r_e is the classical electron radius, $f(Z, \frac{k}{E})$ is a function of the atomic number Z and the ratio of the photon energy k to the energy E of the incident particle. Function $f(Z, \frac{k}{E})$ is related to screening and its form for the complete screening is summarized in literature [33]. Substituting $y = k/E$ into Equation 2.1.5 and considering all possible energy transfer fractions, the energy loss rate due to Bremsstrahlung reads:

$$\begin{aligned} -\frac{dE}{dx} &= \frac{N_A}{A}\int_0^E k\frac{d\sigma}{dk}dk \\ &= 4\alpha E\frac{N_A}{A}\left(\frac{1}{4\pi\epsilon_0}\frac{e^2}{m_e c^2}\right)^2\int_0^1 f(Z, y)dy, \end{aligned} \tag{2.1.6}$$

which indicates that $-dE/dx$ is proportional to E/m^2 of the incident particle.* Therefore, it is natural to define the radiation length X_0 over which the electron has $1/e$ of its initial energy left in average due to Bremsstrahlung, such that the energy loss rate could be written as $-dE/dx = E/X_0$. The equation to calculate X_0 is given in Reference [33]. The tabulated values for some materials are also available:

$$X_0 = 4\alpha r_e^2\frac{N_A}{A}\left\{Z^2\left[L_{rad} - f(Z)\right] + ZL'_{rad}\right\}, \tag{2.1.7}$$

in which the two parameters L_{rad} and L'_{rad} are given in Table 2.2. The Coulomb correction $f(Z)$ used in literature [33] is shown below:

$$f(z) \approx 1.202z - 1.0369z^2 + 1.008z^3/(1+z), \tag{2.1.8}$$

where $z = (\alpha Z)^2$.

The radiative energy loss of e^-/e^+ is greater than the ionization loss when $E_e > 40\,\text{MeV}$

---

*It is also valid for charged heavy particles like muons

Z	L_{rad}	L'_{rad}
1	5.31	6.144
2	4.79	5.621
3	4.74	5.805
4	4.71	5.924
≥ 5	$\ln(184.15Z^{-1/3})$	$\ln(1194Z^{-2/3})$

Table 2.2: The parameters for X_0 calculation [4, 33]

in silicon. The critical energy E_c is thus defined as the energy of the incident electron or positron at which the loss rates due to Bremsstrahlung and ionization are equal. But the alternative definition from Rossi [34] is more accurate to describe the transverse development of electromagnetic cascades. In this case, E_c is the energy at which the ionization loss of one radiation length is equal to the electron energy. The formulae to calculate Rossi's E_c [34] are shown below:

$$E_c \approx \begin{cases} \dfrac{610\,\text{MeV}}{Z + 1.24}, & \text{for solids;} \\ \\ \dfrac{710\,\text{MeV}}{Z + 0.92}, & \text{for gases.} \end{cases} \tag{2.1.9}$$

In addition, electronuclear interactions will dominate e^-/e^+ energy loss for extremely high energy ($E > 10^9$ TeV) [4, 35].

2.1.3 Radiative energy loss of muons

The radiative energy loss of muons have been well studied and summarized in literatures [4, 27]. Considering the electronic stopping power described by the Bethe equation, the total loss rate is described by [36]:

$$\langle -dE/dx \rangle = a(E) + b(E)E \tag{2.1.10}$$

where a is the electronic energy loss and b includes the contributions from direct e^+e^- pair production, bremsstrahlung and muon photonuclear interactions [27]:

$$b \equiv b_{pair} + b_{bremsstrahlung} + b_{nuclear} \tag{2.1.11}$$

The b-values of muons in iron are shown in Figure 2.4.

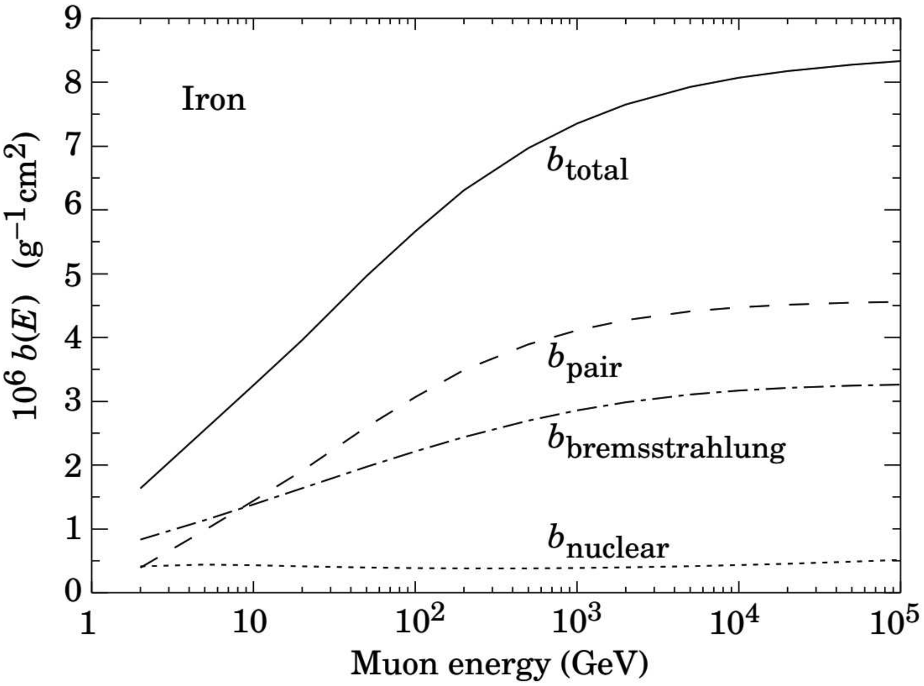


Figure 2.4: b-values as a function of muon energy in iron [27].

The critical energy for muons $E_{\mu c}$ is defined by $E_{\mu c} \equiv a/b$, such that energy losses due to ionization and the radiative processes above are equal. The approximated equations to calculate $E_{\mu c}$ for elements except hydrogen are given by [4]:

$$E_{\mu c} \approx \begin{cases} \dfrac{5700\,\text{GeV}}{(Z+1.47)^{0.838}}, & \text{for solids;} \\[2ex] \dfrac{7980\,\text{GeV}}{(Z+2.03)^{0.879}}, & \text{for gases.} \end{cases} \tag{2.1.12}$$

More accurate values could be found in PDG's website[*]. The muon critical energies of iron and silicon are 347 GeV and 582 GeV respectively.

2.1.4 Energy loss of photons

Photons mainly loss energies through photoelctric effect, Compton scattering and pair production. In addition, The Rayleigh scattering (coherent/elastic scattering) also contribute photon attenuation in low energy region. The rest beam intensity I of mono-energetic photons with an initial intensity I_0 penetrating a mass thickness x (in $\mathrm{g\,cm^{-2}}$) is given by [37]:

$$I = I_0 \exp\left[-(\mu/\rho)x\right] \tag{2.1.13}$$

where μ is the attenuation coefficient and ρ is the density of the absorber. The mass attenuation coefficient μ_m is then defined by:

$$\mu_m \equiv \mu/\rho = \sigma_{tot} N_A/A, \tag{2.1.14}$$

in which σ_{tot} is the total cross section of photon-matter interactions. The photo mass attenuation length is given by $\lambda = 1/\mu_m$. Figure 2.5 from UCSD's website[†] shows the mass attenuation coefficients as a function of photon energy for the predominant processes in silicon. The related data is also available in NIST's database[‡]. The photoelectric effect dominates the attenuation below 50 keV and the pair production is predominant when $E_{photon} > 15\,\mathrm{MeV}$. In the energy region of our interest, the photonuclear interactions could be neglected, but they will dominate photon attenuation for extremely high energy in the order of 10^{20} eV [4].

2.1.5 Small-angle multiple scattering

Particles passing through a medium will change their flying directions with an angle θ_{space} due to multiple Coulomb scatterings. The related theories have been reviewed by several literatures [4, 38, 39]. Only the small angle scatterings which could be described using Gaussian distributions, are introduced here. The distribution of the projected angle θ_{plane}

---

[*] `http://pdg.lbl.gov/2016/AtomicNuclearProperties`

[†] `http://www-ferp.ucsd.edu/LIB/PROPS/PHOTON/`

[‡] `https://www.nist.gov/pml/xcom-photon-cross-sections-database`

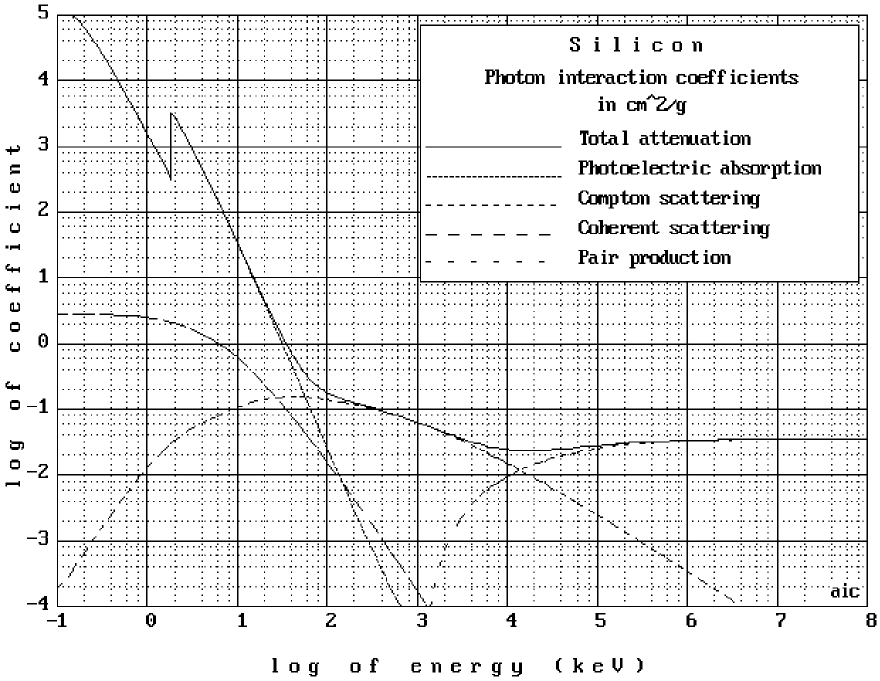


Figure 2.5: Photon attenuation coefficients as a function of photon energy in silicon.

of θ_{space} on a plane is approximated to a Gaussian for its central 98%. The approximated standard deviation θ_0 for single charged projectile could be described by [40, 41]:

$$\theta_0 = \sigma_{\theta_{plane}} \approx \frac{13.6MeV}{\beta pc}\sqrt{\frac{X}{X_0}}\left[1 + 0.088\log_{10}\left(\frac{X}{X_0}\right)\right] \tag{2.1.15}$$

where β is the speed of the flying particle divided by the light speed c, p is the momentum of this projectile, X_0 is the radiation length and X is the equivalent mass length of the path in the penetrated materials. Compared with the Gaussian fitting to the central 98% of the scattering angle distribution given by Molière theory[38, 42, 43], the accuracy of Equation 2.1.15 is better than 11% for all atomic number Z and for $0.001 < X/X_0 < 100$ when $\beta = 1$ [41]. Equation 2.1.15 is referenced by Particle Data Group (PDG) with a modified form in which a constant z indicating the charge of the projectile particle is multiplied [4]. However, article [41] illustrated that the consideration for β and z dependence in the lowest order should result in an equation shown below:

$$\theta_0 \approx \frac{13.6MeV}{pc}\frac{z}{\beta}\sqrt{\frac{X}{X_0}}\left[1 + 0.088\log_{10}\left(\frac{Xz^2}{X_0\beta^2}\right)\right] \tag{2.1.16}$$

This form works quite well for small atomic number Z but with pour accuracy for large Z and small X because of a worse estimation on the influence of β. Fortunately, for a silicon detector with $Z = 14$, it is safe to use equation 2.1.15 and equation 2.1.16 when $\beta \approx 1$. By checking the related plots in reference [41], one can find that the accuracy for singly charged particle and for $0.001 < X/X_0 < 1$ in a silicon scatterer is close to 2%.

In addition to the deflection angle θ_{plane}, the other related quantities including the lateral displacement y and the chord angle ψ are shown in Figure 2.6. The useful standard

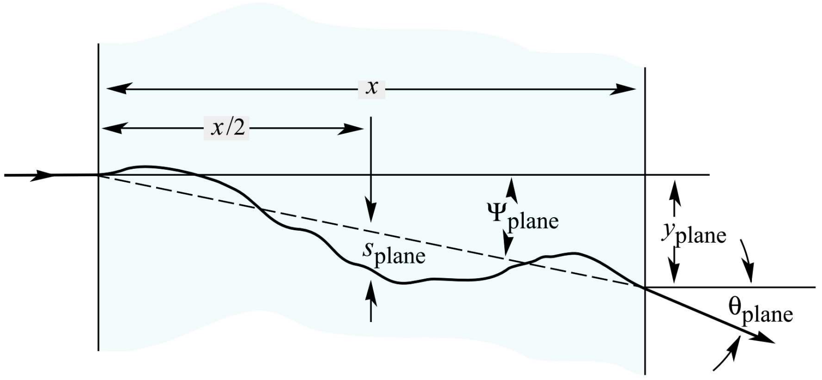


Figure 2.6: A scattered track in the paper plane at depth x with the deflected angle θ_{plane}, the chord angle ψ_{plane} and a displacement y_{plane} [4].

deviations and $y - \theta$ covariance are given below:

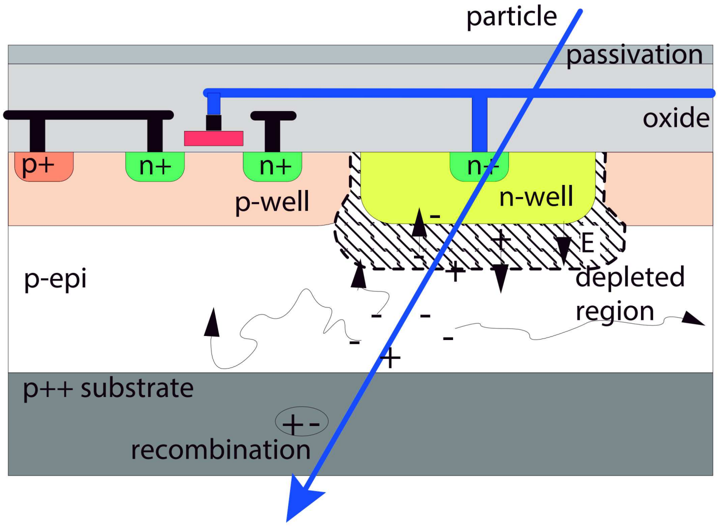
$$\sigma_{\psi_{plane}} = \frac{1}{\sqrt{3}}\theta_0, \quad \sigma_{y_{plane}} = \frac{1}{\sqrt{3}}x\theta_0, \quad Cov(y_{plane}, \theta_{plane}) = \frac{1}{2}x\theta_0^2 \tag{2.1.17}$$

2.1.6 CPS detection principles

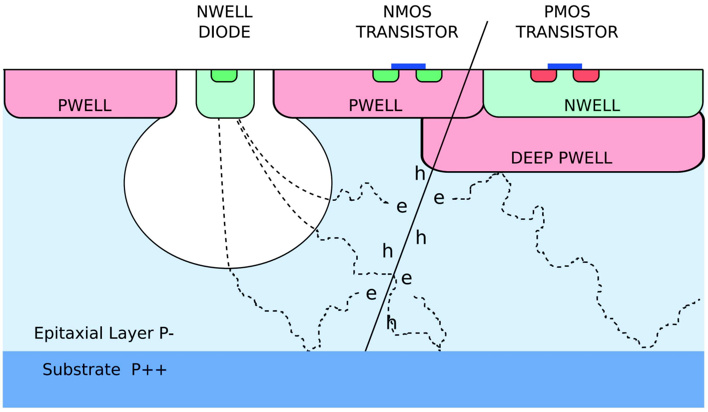
The average energy to generate an electron-hole pair in silicon is only $\sim 3.68\,\text{eV}$ for charged particles [31]. Therefore, several electron-hole pairs could be produced even for one collision of the incident particle. The minimum ionization energy loss rate is $\sim 3.876\,\text{MeV}\,\text{cm}^{-1}$, such that about 105 electron-hole pairs, on average, will be produced for a relativistic muon passing through 1 µm silicon. The most probable yield is smaller, around 60 pair/µm[*] for thin absorbers (∼30 µm) [31, 44]. The CPS is designed to collect these charges using an n-well in the p-epitaxial layer shown in Figure 2.7. The produced electrons diffuse in the epitaxial layer thermally and drift in the depleted region of the

---

[*] It is around 50 pair/µm for 10 µm silicon.



(a) AMS 0.35 µm Opto-CMOS (C35O) process with 15-20 µm epitaxial layer.



(b) TowerJazz 0.18 µm CMOS Image Sensor (CIS) process with high resistivity epitaxial layer and deep p-well [44].

Figure 2.7: CPS schematic diagrams using different CMOS processes.

n-well diode (a reverse-biased p-n junction). They are also reflected by the higher p-doped substrate and p-wells. Consequently, most of the charges are collected by the diode to generate the signal of a hit from the incident particle.

The AMS 0.35 µm Opto-CMOS process (C35O) shown in Figure 2.7a was used to develop the state-of-the-art mimosa-28. PMOS transistors (planted on an n-well) cannot be used in the detecting region due to the charge collection competition with the n-well diode. Hence, only simple circuits could be implemented. In addition, the resistivity of the epitaxial layer in this process is lower, which lead to a thin depleted region in the collecting diode. As the requirements of CPS on readout speed and power consumption are getting more stringent, TowerJazz 0.18 µm CMOS Image Sensor (CIS) process shown in Figure 2.7b, is introduced in CPS design. The deep p-well prevents the PMOS structure from absorbing electrons, so more complex circuits could be realized in CPS sensing region. The high-resistivity ($> 1\,\mathrm{k\Omega\,cm}$) epitaxial layer makes larger depleted region possible to improve the signal to noise ratio (SNR).

2.1.7 CPS noises

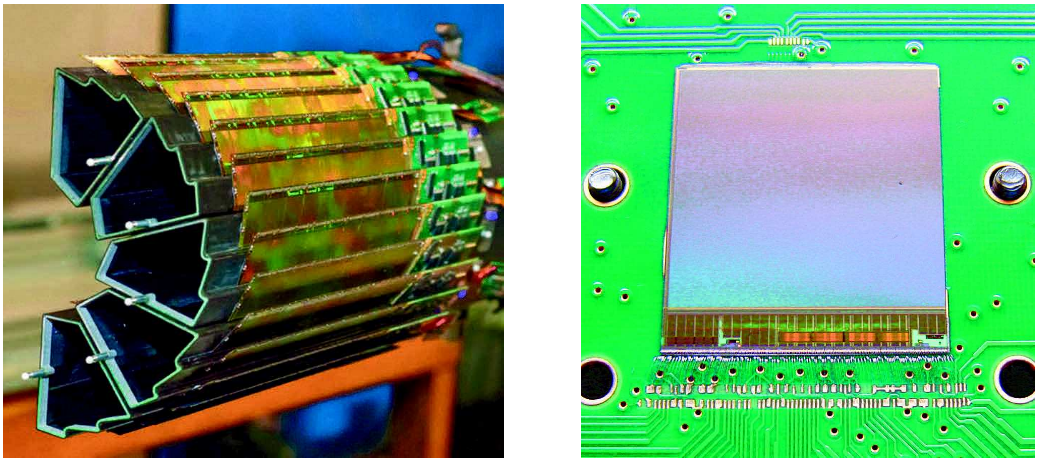
The typical signal in one pixel is only a few hundreds of electrons. To achieve a high SNR, circuits should be optimized to reduce the related noises. In the pixel array, the pedestals of the outputs of the pixels are different. Fixed pattern noise (FPN) is thus defined as the standard deviation of its distribution. In the readout circuits of one pixel unit, the temporal noise (TN) is mainly a sum of the kTC noise (thermal noise) during the reset phase of the collecting diode, the short noise caused by the leakage fluctuation of the diode, and the flicker noise ($1/f$ noise) of the transistors. The thermal noise of the collecting electrode is a few tens of equivalent noise charges (ENC) and the short noise is only a few ENC for an integration time of 100 µs in room temperature for non-irradiated sensors. Most of the temporal noise could be eliminated using correlated double sampling (CDS) method, except the random telegraph signal (RTS) noise with a step-like output signature in small transistors. The measurements of these noises are described in Section 2.4.1.

2.2 CPS chip design of MIMOSA series

Minimum Ionizing particle MOS Active pixel sensor (MIMOSA) series are developed by PICSEL (Physics with Integrated Cmos Sensors and ELectron machines) group at IPHC. Many accomplishments including the EUDET telescope using mimosa-26, shown in Figure 2.8, and HFT of STAR experiment using mimosa-28, have been achieved. The pixel layers of HFT and the top view of Mimosa28 are shown in Figure 2.9. The main characters of mimosa sensors are listed in Table 2.3, which includes the process technology, thickness of the epitaxial layer, pixel pitch size, number of metal layers, number of pixels with digital/analogue readout and the application description. The main specifications of the state-of-the-art MIMOSA-28 are summarized in Table 2.4.

Figure 2.8: Photo of EUDET telescope using Mimosa26 sensors

The typical design strategies of MIMOSA sensors are introduced in the following sections. Basic circuits for signal generation in each pixel and the readout architectures in chip level are presented to show the optimized solutions for particle detection.



(a) Photo of PXL ladders (b) Top view of MIMOSA-28

Figure 2.9: Pixel layers (PXL) of HFT in STAR experiment

Chip	Process	Epi. (µm)	Pitch (µm)	N_{metal}	pixel	Description
M22AHR	AMS0.35 HiRes	10/15/20	18.4 20.7	4	73k digital 4.6k analog	Prototype for EUDET
M26	AMS0.35	14	18.4	4	660k digital	EUDET chip
M26AHR	AMS0.35 HiRes	10/15/20	18.4	4	660k digital	
M28 Ultimate	AMS0.35 HiRes	15/20	20.7	4	890k digital	STAR chip

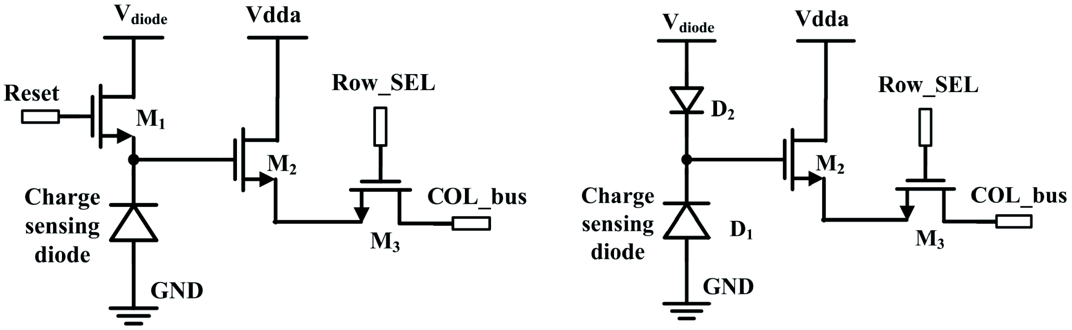
Table 2.3: List of chosen MIMOSA chips

Parameter	Performance/Description
Process	AMS 0.35 µm (C35O)
Pixel pitch	20.7 µm × 20.7 µm
Spatial resolution	$\gtrsim$ 3.6 µm
Pixel matrix	928 (rows) × 960 (columns)
Read architecture	Rolling-shutter (column parallel)
Epitaxial layer	20 µm, high resistivity($> 400\,\Omega \cdot$ cm)
Sensor thickness	50 µm
Integration time	$\sim$ 185.6 µs
Output bandwidth	2 × 160 Mbits/s LVDS
Power consumption	$\sim$ 160 mW/cm^2 at 3.3 V
Radiation tolerance	$>$ 0.15 MRad & 3×10^{12} n$_{\mathrm{eq}}$/cm^2

Table 2.4: Specifications of MIMOSA-28

2.2.1 Readout architectures in Pixel level

The objective of the readout circuits in one pixel is to collect the ionized charges (electrons) generated from an incident particle and to generate the signal for subsequent processes. Two basic architectures using reverse-biased diodes are shown in Figure 2.10. The three transistor (3T) design in Figure 2.10a, includes a switch M_1 to reset the state of the sensing diode, a source follower M_2 to output voltage signals and the row selection switch

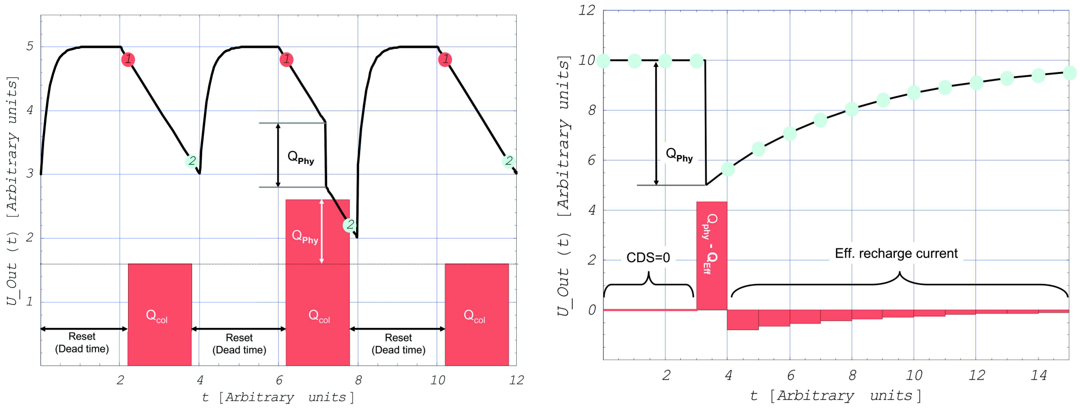


(a) Three transistors (b) Self-biased diode

Figure 2.10: Signal encoding of the pixel readout circuits [45]

M_3. The collected charges lead to a voltage drop on the sensing diode, and then the transistor M_2 makes the signal readable for the following circuits. Figure 2.10b shows the Self-bias diode (SB) structure with one forward biasing diode to replace the reset switch M_1 in the 3T architecture. Using a long recharge constant and small time window for signal reading, the SB structure could work without the dead time caused by the reset process in the 3T-pixel.

As the potential on the sensing diode cannot be the same after a reset or a recharging, the correlated double sampling (CDS) is mandatory for both of the pixel designs. Furthermore, CDS can eliminate most of the temporal noises (e.g., κTC noise of the diode) in the circuits. The physics signals of the two pixel structures are shown in Figure 2.11. The CDS circuit, which consists of a capacitor and a clamping switch, can also be implemented into the pixel structure after a pre-amplifier as shown in Figure 2.12 for a SB-pixel. The first sample is read in the reading phase, then the second sample is taken in a calibration phase after the clamping process. The two samples will be sent to a discriminator or an analogue-to-digital converter (ADC) for digitization.



(a) Three transistors

(b) Self-biased diode

Figure 2.11: Signal encoding of different pixel structures. The CDS values (Subtracting the first sample from the second one) are illustrated with red histograms.

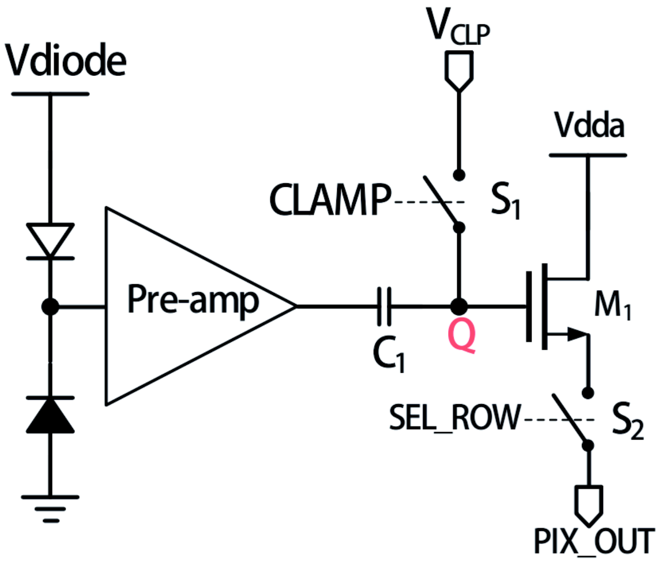


Figure 2.12: CDS circuits design [45]

2.2.2 Readout architectures in Chip level

2.2.2.1 The rolling-shutter readout

Rolling-shutter is a widely used method to read a pixel array in the CMOS imaging sensor of a modern digital camera. Figure 2.13 shows the schematic diagram of this readout architecture. Pixel signals are read row by row and then compared by the discriminators integrated in the pixel circuits or in a digital region at the end of pixel columns.

This column parallel readout is an optimal solution for full frame reading with a low power consumption and a relative high integration speed. In addition, the circuits in one pixel could be simple, such that designs using only NMOS transistors (like the pixels in MIMOSA-26 and MIMOSA-28) are acceptable. Using this architecture and proper pixel designs, applications on photon detecting or beam monitoring could be foreseen. However, the pixel occupancy in a vertex detector is below than one percent. The fired pixels are minor, therefore the asynchronous pixel readout used in ALPIDE sensors could have a higher time resolution (faster integration) and a lower power consumption. Because of the same reason, rolling-shutter architecture requires zero suppression and sparse readout circuits to reduce data transmission bandwidth.

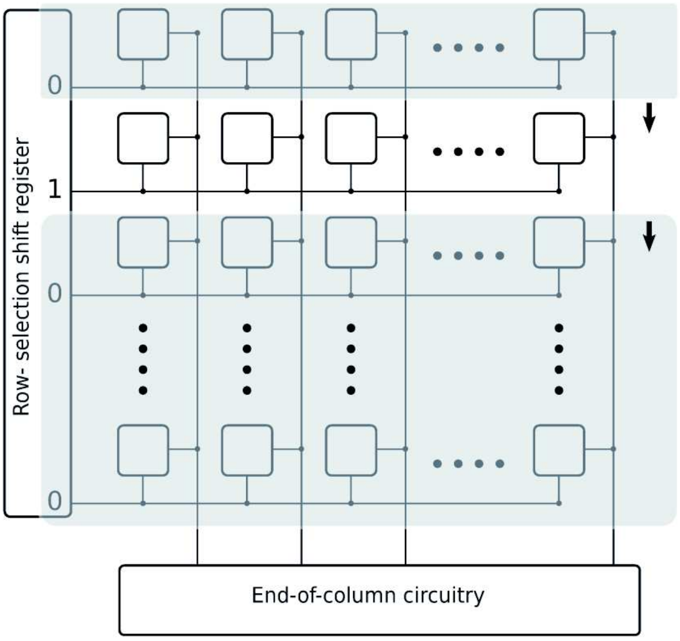


Figure 2.13: Schematic diagram of rolling-shutter readout [44]

2.2.2.2 Zero suppression and sparse readout

The fired pixels due to physics signals are of a small number in one frame, such that a sparse readout for the rolling-shutter architecture is necessary to increase data transmission speed. One solution is to use a zero suppression (SUZE) architecture shown in the block diagram of Figure 2.14 of MIMOSA-28, to record only the information of fired pixels. SUZE includes three parts: the sparse data scan (SDS), the multiplexer and the memory management. First, the 960 columns are processed in 15 banks in parallel for

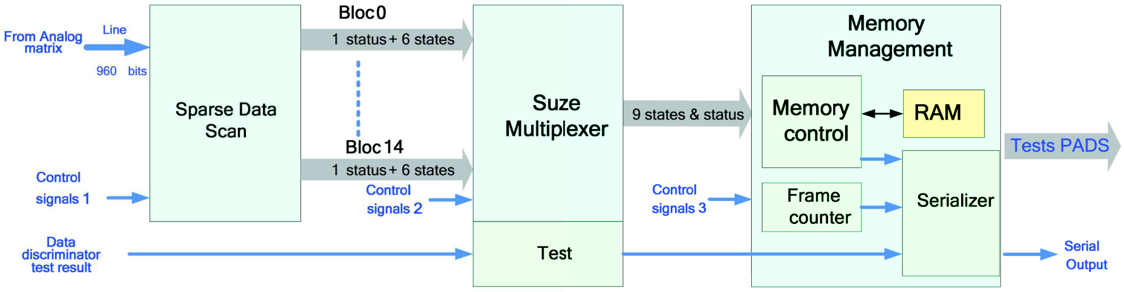


Figure 2.14: zero suppression logic of MIMOSA-28 [46]

sparse data scan. Each bank could handle 64 pixels and output a maximum of 6 states which consists of the address of the first fired pixel and the number of successive fired pixels, and one status indicating the number of states. Second, the data of the 15 banks are extracted out to generate a maximum of 9 states and one status indicating the total number of states and their common line address. Finally, the data with only fired pixels are buffered by the memory management module for readout. The algorithm and implementation details can be found in literature [47] and the user manual of Ultimate 2 (MIMOSA-28).

2.2.3 Upgraded designs

The general basic circuits and architectures used in MIMOSA series have been introduced. Some details like radiation hardness designs and noise reduction circuits are skipped. In addition, prototypes including MIMOSA-32 and MIMOSA-34 to explore the potential of TowerJazz 0.18 µm technology have been developed and studied. Modifications like in-pixel discriminators have been implemented and studied using AROM-0. New features like a rolling-shutter to read two rows (two discriminators for each column) simultaneously and the large pixel pitch ($> 36\,\mu\text{m}$) have also been validated to reduce power consumption and integration time by prototypes like FSBB-M0 and MIMOSA-22THRb for the R&D of the sensor MISTRAL-O.

2.3 CPS Chip design of ALPIDE series

ALPIDE (ALice PIxel DEtector) chips are designed using TowerJazz 0.18 µm CIS process for Alice vertex detector upgrade with a very fast readout speed and a low power consumption, which benefits from its in-pixel discriminator optimisation, asynchronous hit-driven readout architecture and the physics signal nature that only a small fraction of pixels are fired by incident particles in the integration period. The prototype pALPIDE-1 with a pixel pitch $28\,\mu\text{m} \times 28\,\mu\text{m}$ has a spatial resolution around 5 µm, a time resolution $\gtrsim 4\,\mu\text{s}$ [*] and a power consumption $\lesssim 40\,\text{mW}\,\text{cm}^{-2}$ [48]. This architecture is decided to be the baseline design for ITS upgrade of ALICE experiment.

Figure 2.15 shows the schematic diagram of the Address-Encoder and Reset-Decoder

---

[*] The peaking time of the amplifier is around 2 µs which allows data to be discriminated and strobbed in some 2 µs[48].

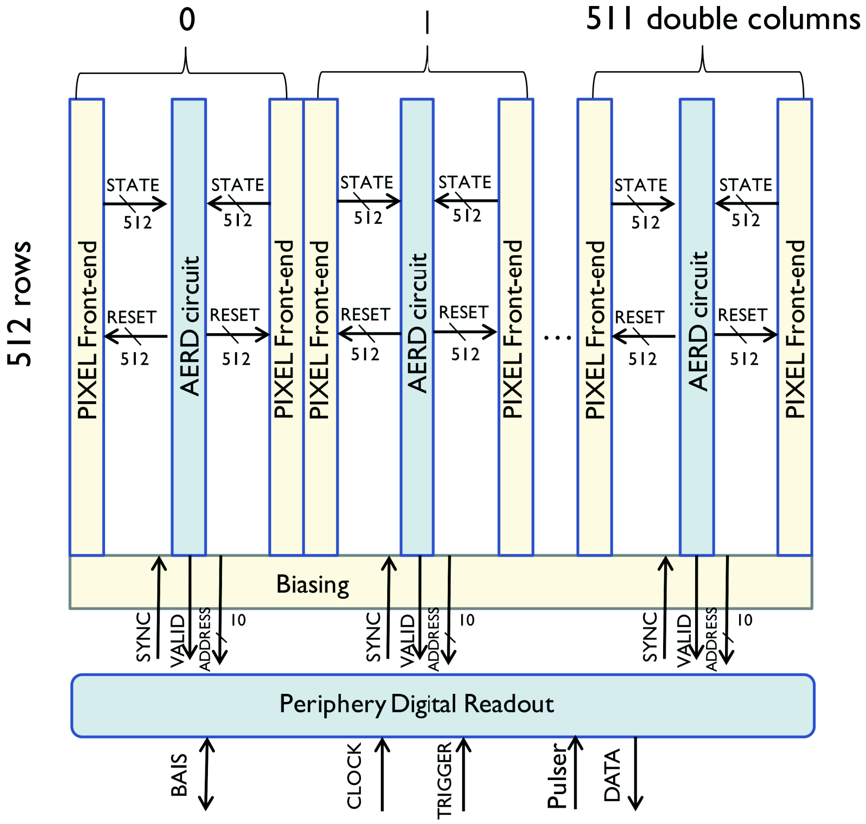


Figure 2.15: The AERD asynchronous readout structure of pALPIDEfs chip[49].

(AERD) readout structure used in ALPIDE sensors. The in-pixel circuit (Pixel front-end) is optimised to work with a low power consumption and a fast readout speed. A current comparator with a bias of only 20 nA is used for discrimination. It generates a pulse with a peaking time around 2 µs and a total shaping time around 4 µs when the pixel is fired by the impinging particle. This data will be latched as a STATE once a strobe is applied for this fired pixel. The AERD circuits for each two columns propagate the information and related digital signals of a fired pixel. The address of the pixel with the highest priority is read and reset by the periphery circuits first. Finally, data are compressed and assembled into events for subsequent transmissions. The design details and optimization discussions are introduced in reference [49].

2.4 Sensor characterisation and test methods

The basic CPS designs have been illustrated in previous text, and sensor specifications with the related test methods will be presented in this section. The specifications could be studied in two types of tests: the laboratory test and a beam test using charged particles.

During the development phase of a sensor, features like charge collection efficiency (CCE), charge to voltage factor (CVF), signal to noise ratio (SNR), temporal noise (TN) of a pixel and the fixed pattern noise (to describe pedestal dispersion) of a pixel array are of interest. These parameters could be tested using X-ray, Beta-ray or laser sources in a lab.

A beam test is used to study CPS performance like detecting efficiency, intrinsic spatial resolution and fake hit rate, which are input parameters for a tracker design. Accelerators generating high energy electron/proton/pion beams and a telescope system are needed.

In addition, the radiation tolerance of the sensor could be evaluated by the measurements of all the specifications after a given radiation dose.

For a final CPS design, the probe test, which is a fast lab test system using probing card designed with sensor reloading feature, is used for CPS quality control in the production phase.

2.4.1 Laboratory calibrations

In the laboratory test, both analogue and digital data are acquired for analysis. Even digital sensors contain some test pixels with analogue outputs. The test set-up of MIMOSA-18[*] is shown in Figure 2.16. The proximity board is used to place a CPS sensor with suited pre-amplifiers. The auxiliary board is then connected by flat cables to offer the voltage supply for the front-end and to buffer digital or analogue signals for long-distance data transmissions using network cables. The third VME board, named as imager USB board, provides clocks and reset signals, processes the analogue-digital conversion using a 12-bits ADC, and packs data with USB protocol for their transmissions to a PC.

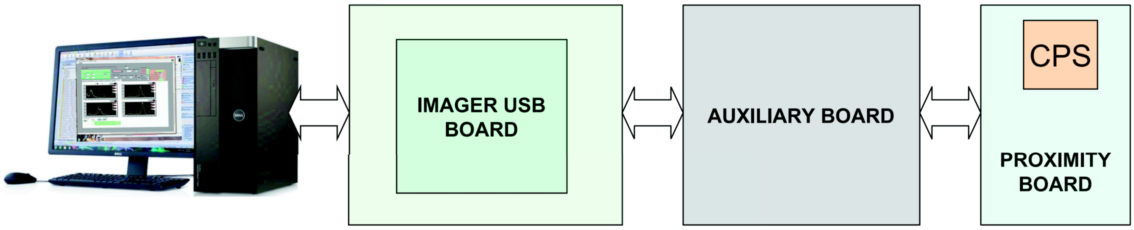


Figure 2.16: The block diagram of the lab test set-up for MIMOSA-18.

In laboratory tests, Iron-55 is a widely used X-ray source to study charge collection performance. The main emissions of ^{55}Fe decays include 5.19 keV Auger electrons ($\sim$ 60%), x-rays with energy of 5.9 keV ($\sim$ 24.4%) and 6.5 keV photons ($\sim$ 2.85%). These Auger electrons are absorbed on the surface of silicon, and the attenuation lengths of the two photon components are about 29 µm and 38 µm respectively.[†] The electron-hole pair yield of a 5.9 keV photon in silicon is about 1600 [50], which is used to calibrate the gain of the pixel readout circuits.

Figure 2.17 shows the charge collection distributions of a seed[‡] pixel in solid red, a cluster of 4 pixels with the highest signals in dashed blue and a cluster of 25 pixels centred at the seed in black. The large collection peaks are from the diffused charges whose origins are photon electron conversions in the epitaxial layers, while the small calibration peaks come from rare events in which the interactions occur inside the depleted region of a sensing diode. As almost all the charges are collected by one pixel for events of the

---

[*] MIMOSA-18 is a sensor with analogue outputs.

[†] http://henke.lbl.gov/optical_constants/atten2.html

[‡] A seed pixel is the one with the highest signal among its neighbours.

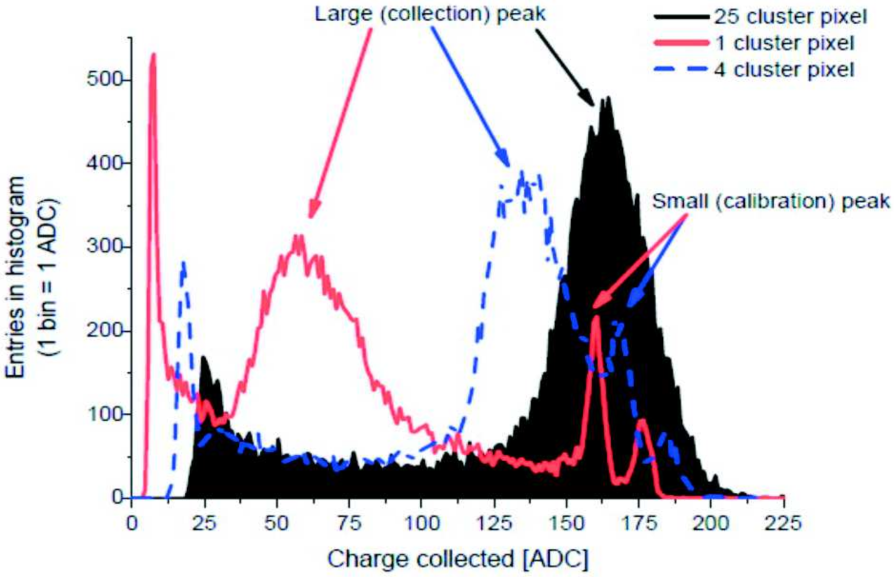


Figure 2.17: Charge collection distributions of a CPS [45]. The solid red curve is from the charges collected by the seed pixel. The dashed blue one is from a cluster of four pixels with the highest outputs. The black distribution is from the data of a cluster containing 25 pixels with the seed pixel in the centre. Bias of the blue calibration peaks is due to pixel temporal noises and the cluster reconstruction algorithm.

calibration peaks, CVF is then given by the ratio of 1600 to the ADC value of the first calibration peak which refers to the 5.9 keV photons. In the solid red curve, the ratio of the most probable ADC value at the collection peak to that at the first calibration peak, gives the CCE of the seed pixel. Comparing the positions of the collection peaks, one can conclude that a seed pixel collects $\sim 37\%$ of the total charges, a 4-pixel cluster collects $\sim 80\%$ and a 5×5 cluster collects almost 100%.

The temporal noise (TN) of a pixel and the fixed pattern noise (FPN) of the array could be extracted out directly from the analogue outputs. Nevertheless, it is tricky in case of digital sensors. A threshold scan of noise runs* is usually processed to generate S-curves (transfer functions) as shown in Figure 2.18. Both pixel noises and discriminator noises are included in each curve. The combined noise distribution of the pixel and its related discriminator circuits can be extracted out by a fit using a cumulative Gaussian or the error function. The width of this distribution is taken as the temporal noise of the pixel, and the mean value is defined as its pedestal. Temporal noise distribution in sensor level is then defined by the temporal noises of all pixels, and a threshold distribution is

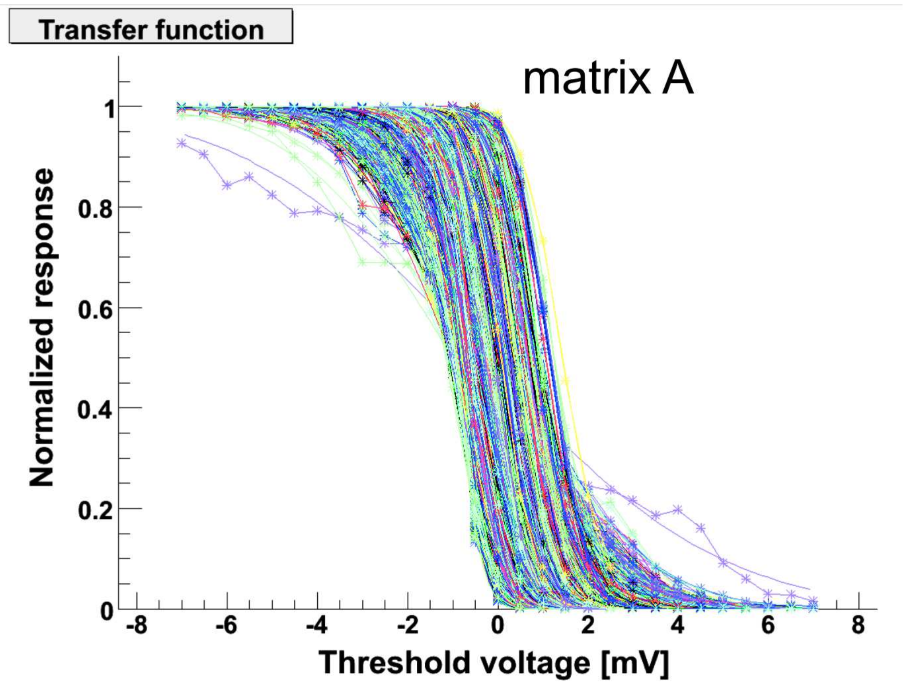
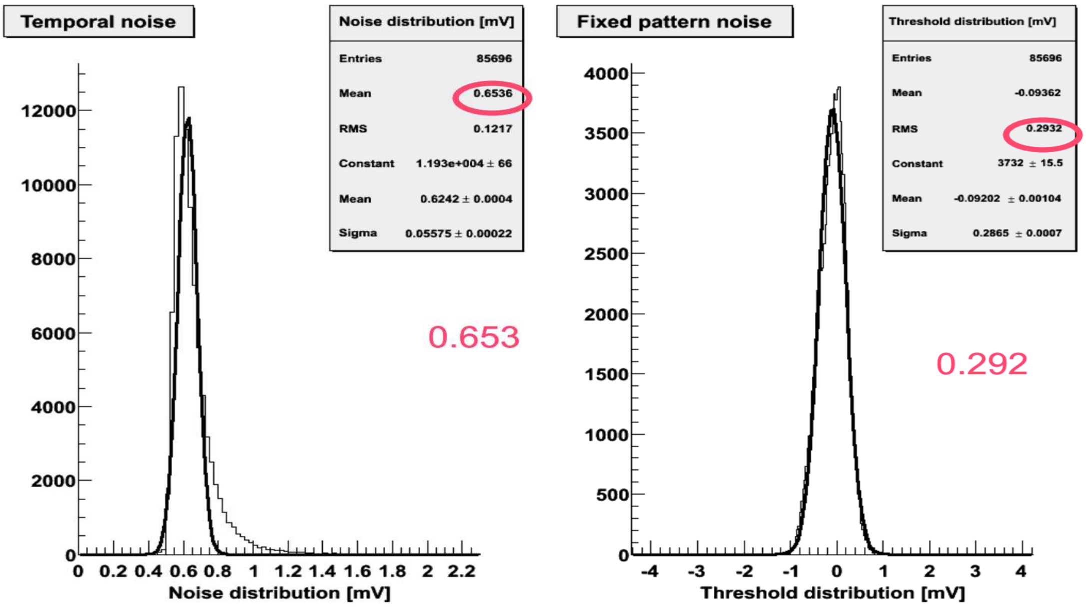


Figure 2.18: Transfer functions of the pixels in matrix A of a FSBBbis chip.

---

*The sensors are shielded from particle sources to present their intrinsic noise level.

defined using all the pedestal values. Figure 2.19 shows the related noise distributions of the matrix A in a FSBBbis* sensor. Finally the mean value of the sensor's temporal noise distribution is used as TN of the sensor while the standard deviation of the threshold distribution in Figure 2.19b defines the FPN.



(a) Temporal noises of all pixels

(b) Pedestal dispersion of the matrix

Figure 2.19: Temporal noise distribution and threshold distribution of 85696 pixels in the matrix A (416 rows × 208 columns) of a FSBBbis sensor. 4 noisy rows (row 412 to row 415) at the end, are removed from the threshold scan.

SNR could be studied in the lab test using electron sources or in a beam test, but radioactive sources of X-rays and gamma-rays are not favoured, as the signal strength from them is different with that from the fast heavy particles generated in particle physics experiments.

2.4.2 Beam test

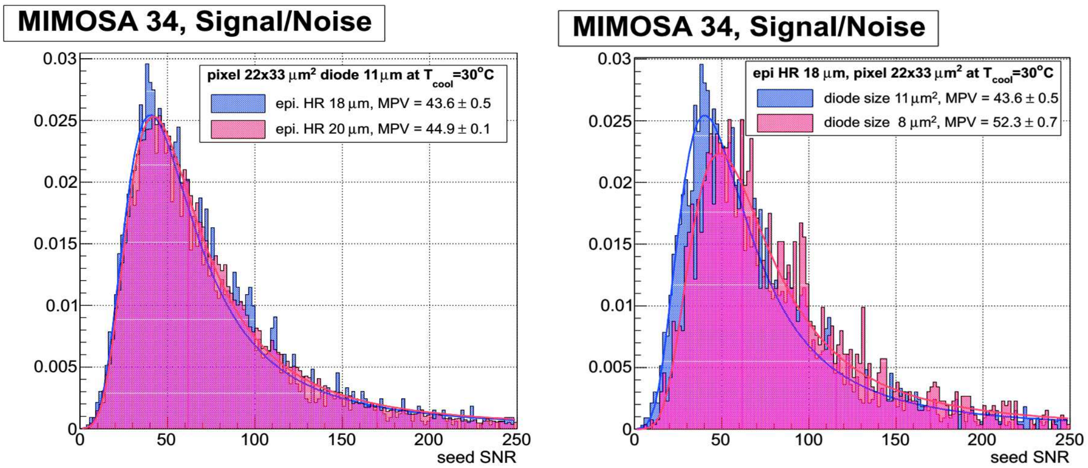
The beam test facility used for CPS studies is the telescope system like the EUDET telescope shown in Figure 2.8. It consists of several reference CPS sensors with known performance, a device under test (DUT), their mechanical support and cooling system, a

---

*Full Size Building Block (FSBB) is a chip designed as a block of a large chip containing three blocks. FSBBbis designed with a pixel pitch around 22×32.5 µm^2 using TowerJazz 0.18 µm process.

trigger system using scintillators and Photomultiplier Tubes (PMT), and a data acquisition system.

The SNR of a CPS could be measured without tracking. Figure 2.20 shows the results of MIMOSA-34 chips to optimize the epitaxial thickness and diode size. Consequently, a high resistivity epitaxial layer of 20 µm and a diode size (dominating parameter) around 8 μm^2 are favoured with the MPV of SNR $\sim$ 20% better.



(a) SNR of MIMOSA-34 with different epitaxial thickness.

(b) SNR of MIMOSA-34 with different diode size.

Figure 2.20: Signal-to-noise ratio (SNR) of seed pixels in different MIMOSA-34 chips [51].

The intrinsic spatial resolution and detecting efficiency are the two properties which rely on beam tests. Efficiency is defined by the number of the associated hits on the DUT divided by the total number of tracks reconstructed from the reference sensors. The spatial resolution of the DUT, σ_{DUT}, has to be extracted from the residual* distributions. That is $\sigma_{DUT} = \sqrt{\sigma^2_{residual} - \sigma^2_{tel}}$, where σ_{tel} is the track uncertainty at the DUT position. For a beam of 120 GeV pions, multiple scattering of materials could be neglected. But it has to be considered into σ_{tel} for beams like 4 GeV electrons. The resolution of a binary output sensor could be estimated using $\sigma \approx \text{pitch}/\sqrt{12}$, but charge sharing in adjacent pixels improves the spatial resolution when charge of gravity (COG) algorithm is used for hit reconstruction. This effect is shown in Figure 2.21. In addition, noise runs after the warm-up data taking are used to obtain the fake hit rate of a sensor. Since the

---

[*] A residual is the difference between the pointed track position and the associated hit position.

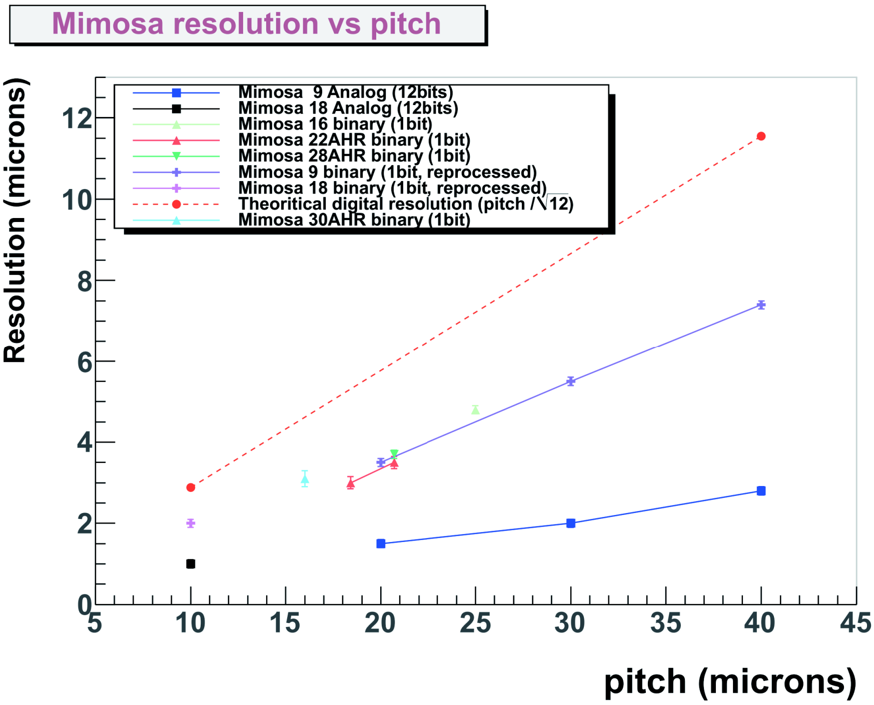


Figure 2.21: MIMOSA resolution VS pixel pitch [51]. The theoretical resolution (pitch/$\sqrt{12}$) is shown with the dashed line.

beam is shut down, the DUT works in triggerless mode. The main results of a beam test could be summarized in a three-color plot as shown in Figure 2.22. Moreover, the running parameter, the discriminator threshold, is also determined from the plots. With an efficiency better than 99% and a fake hit rate below 10^{-5}/pixel/frame, the thresholds between 5 times and 8 times of the TN, are preferred.

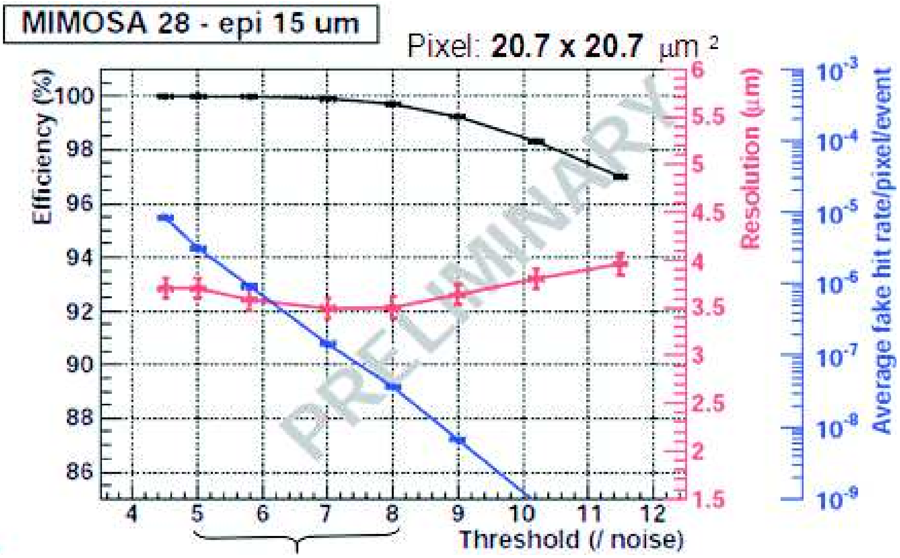


Figure 2.22: Efficiency (black), resolution (red) and fake hit rate (blue) as a function of threshold for MIMOSA-28.

2.4.3 Probe test

Probe test is used for sensor quality monitoring in the production phase. A dedicated probe card is developed to contact the sensors on a wafer. With the DAQ system and photon/radiation sources, the bench test is processed to check the quality of the sensors.

2.5 Summary and conclusion

Within this chapter, we introduced CPS principles, basic design strategies and sensor characterisations with the related test methods. Based on these understanding and the presented performance of MIMOSA sensors, the potential of their usage in high energy physics experiments like BESIII, could be foreseen.

For BESIII inner tracker, the requirement on spatial resolution is not as strict as that of STAR and Alice, such that a larger pixel pitch could be helpful to reduce power consumption and to increase readout speed. Since MIMOSA-28 is a well validated chip in STAR experiment and most of its specifications are good enough for BESIII, the group decided to build a prototype of BESIII inner silicon tracker with MIMOSA-28 chips. As for the real size detector, sensor optimization and layout designs will be discussed in the following chapters based on the performance of MIMOSA-28.

Chapter 3

Sensor design optimization for BESIII

During the development of CPS, the goal is to make very thin sensors with a high detecting efficiency, a fast readout speed, a low power consumption and a good spatial resolution which means a small pixel pitch. Under the given technical conditions, sensor design is a task to find the balance of all these parameters. For instance, faster readout speed and smaller pixel size lead to a higher power consumption. This trade-off is exactly what happened in sensor optimisation for BESIII. To replace a very long gas chamber compared with the silicon trackers or vertex detectors used in high energy physics experiments, the first worry is about the material budget, as the spatial resolution of CPS sensors usually exceeds the requirement. For BESIII trigger rate, the readout speed of MIMOSA-28 is good enough for its pixel occupancy. Therefore, the dominating factor in CPS design, is to reduce power consumption of the sensor to avoid additional cooling materials for the long ladders. A natural solution is to enlarge the pixel pitch, thus the number of channels to read the pixel array in the rolling-shutter architecture will be reduced to save power. However, a large pixel size may bring a higher noise level, a lower detecting efficiency and a higher pixel occupancy. This leads to the designs with newly balanced parameters. Meanwhile, the TowerJezz 0.18 µm technology enables more complex circuit implementations to help to improve all the sensor specifications. The new CPS sensors with different pixel sizes and improved readout designs have been developed recently. In this chapter, the general sensor design considerations will be applied in BESIII and the related beam test analysis were processed to study the performance of MIMOSA-28 (with a pixel pitch, $20.7\times20.7\,\mu m^2$) and FSBB/FSBBbis ($22\times32.5\,\mu m^2$). In conjunction with the test results of MIMOSA-22THRb ($39\times50.8\,\mu m^2$), the analysis results could give the design constraints on condition of the current technologies.

3.1 Detector requirements and CPS design considerations

The requirements of the silicon pixel tracker (SPT) in BESIII are listed in Table 3.1. Most of the requirements have already been fulfilled by MIMOSA-28. Since the momentum resolution is mainly decided by the length of a tracker lever arm, it is not a problem to reach the goal $\sigma_{P_T}/P_T \sim 0.5\%$ at 1 GeV conditioned on the given limits of material budget. Related studies will be presented in the following chapters. The potential risk exists in the sturdiness and cooling of a long ladder. As mentioned at the beginning of this chapter, the key point of CPS optimization for BESIII is its power consumption. A lower requirement on cooling could be fulfilled with air cooling and even a flow with lower speed, which may reduce the related vibration and thus save materials for mechanic support.

Parameters	BESIII inner tracker (Requirements)	STAR PXL (Achieved)
Particle rate	$\sim 10^4\,\mathrm{Hz/cm^2}$	$\sim 10^6\,\mathrm{Hz/cm^2}$
Spatial resolution	$\sigma_{r\phi} \sim 100\,\mu m$, $\sigma_Z \sim 400\,\mu m$	$\gtrsim 3.6\,\mu m$
σ_{p_T}/P_T	$\sim 0.5\%$ @ 1 GeV (with MDC)	N/A
Efficiency	$\sim 98\%$	$\sim 99.9\%$
Material budget	$< 0.5\%\ X_0$/layer (1.5% in total)	$\sim 0.37\%\ X_0$/layer
Radiation tolerance		
TID	~ 8.4 kRad/year	150 kRad
NIEL	$\sim 3\times10^{11}\,\mathrm{N_{eq}/cm^2}$/year	$3\times10^{12}\,\mathrm{N_{eq}/cm^2}$
Maximum Ladder length	~ 52 cm (bilateral support)	~ 20 cm (unilateral)

Table 3.1: The requirements of BESIII inner tracker and the related performance achieved by STAR Pixel Layer (PXL) using MIMOSA-28.

The new generation of CPS for particle physics applications is being studied under the promotion of ALICE experiment. Lower power consumption CPS could be achieved through two strategies: a larger pixel pitch (MISTRAL sensor) and a data driven readout architecture (pALPIDE sensor). The TowerJazz 0.18 µm process used for CPS development, offers new features like high resistivity epitaxial layer, deep p-well and more metal-layers to optimize sensor design for lower power consumption and faster readout at the same time. The pALPIDE-1 ($28\times28\,\mu m^2$) prototype consumes only 1/4 the power of MIMOSA-28 and is able to cope with a hit rate at least 10 times higher [48]. It offers the new potential of CPS in a data-driven readout using TowerJazz 0.18 µm. On the other hand, Our study focuses on the optimization of the pixel pitch. The special situation of this study is that the spatial resolution of CPS dose not need to be so small. Therefore, a larger pixel pitch and a rectangular shape for fewer columns (fewer discriminators) could be used to reduce power consumption. Together with the R&D of MISTRAL sensors, this thesis includes the validation of the large pixel CPS prototypes and the related beam test analysis.

3.2 Beam tests for CPS towards a large pixel pitch

The pixel pitch sizes of MIMOSA-28, FSBBbis and MIMOSA-22THRb are in ascending order. The last two sensors are technique exploring prototypes developed for ITS upgrade of ALICE experiment. FSBBbis includes corrections and improvements for its predecessor FSBB developed for validation of TowerJazz 0.18 µm process and a double-row rolling-shutter. MIMOSA-22THRb focuses on the performance study of large pixels with a pitch of $39\times50.8\,\mu m^2$ or $36\times62.5\,\mu m^2$.

The beam test results of MIMOSA-22THRb have been published in literature [7]. We participated and analysed the beam tests of MIMOSA-28 based Single Arm Large Area Telescope (SALAT) modules and FSBBbis sensors. The final goal is to validate the performance of a sensor with a spatial resolution $\sim 10\,\mu m$ and a modified rolling-shutter which reads two rows at a time.

Moreover, each test addresses dedicated objectives. The analysis of a SALAT module, a structure with 4 MIMOSA-28 sensors to have a larger detecting area, explored the ladder alignment method and studied the tracking performance of SALAT modules. The FSBBbis beam test was also used for optimisation of two running parameters I_{pix} and I_{dis} for sensor configurations.

3.2.1 SALAT beam Test

SALAT is a telescope system with a large detecting area (4 times of a MIMOSA-28). Because of financial reasons, SALAT works with a single arm, i.e., the whole system is set up on one side of a DUT.

3.2.1.1 Test configurations

In the beam test, the modules were tested first using 4 MIMOSA-28 sensors (working like a small telescope) to repeat the validation tests for the 4 sensors in each module. Simultaneously, relative positions of each sensor with respect to the module centre are recorded for the subsequent ladder[*] alignment. The telescope layout used for this study is illustrated in Figure 3.1. The front views of the PCB board of the modules are also shown in 3.1a. Two MIMOSA-28 are on the front, and the other two are glued on the Mylar foil of the back side. Techniques like UV curing gluing and wire bonding are used for SALAT fabrication. Once the information of all the modules are recorded, SALAT is used as the telescope to detect standalone MIMOSA-28 sensors in a configuration shown in Figure 3.2.

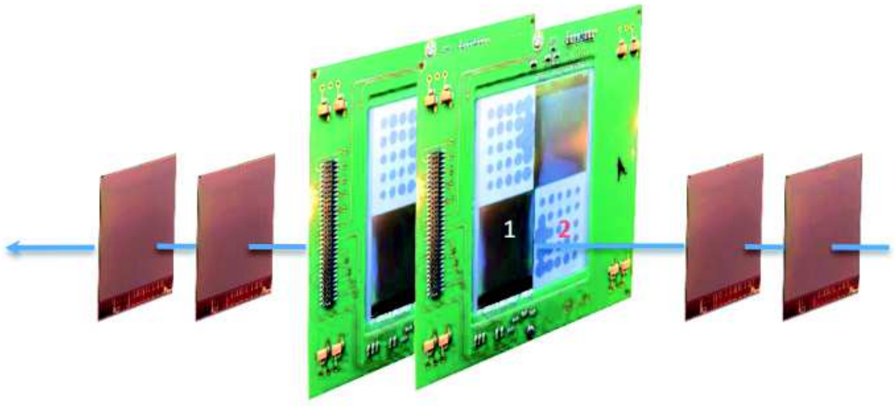
3.2.1.2 Data analysis

The analysis started with the normal alignment method (sensor-by-sensor) to align all the sensors penetrated by the same tracks. Each SALAT module has to be aligned four times for its four sensors. The results of this analysis are chosen to be the reference for the outcomes of the ladder alignment. A dedicated hit map of a SALAT module is shown in Figure 3.3, in which the gaps between the sensors could be seen and measured. Figure 3.4a and Figure 3.4b show the measured results of sensor gaps using microscope photo and beam test analysis respectively. Because sensors are glued in diagonal on each surface (front/back), the gap between the two sensors on the same side (left/right) cannot be avoided. For two sensors from different surfaces, the gap may be replaced by an overlap like the green and yellow part shown in figure 3.4b.

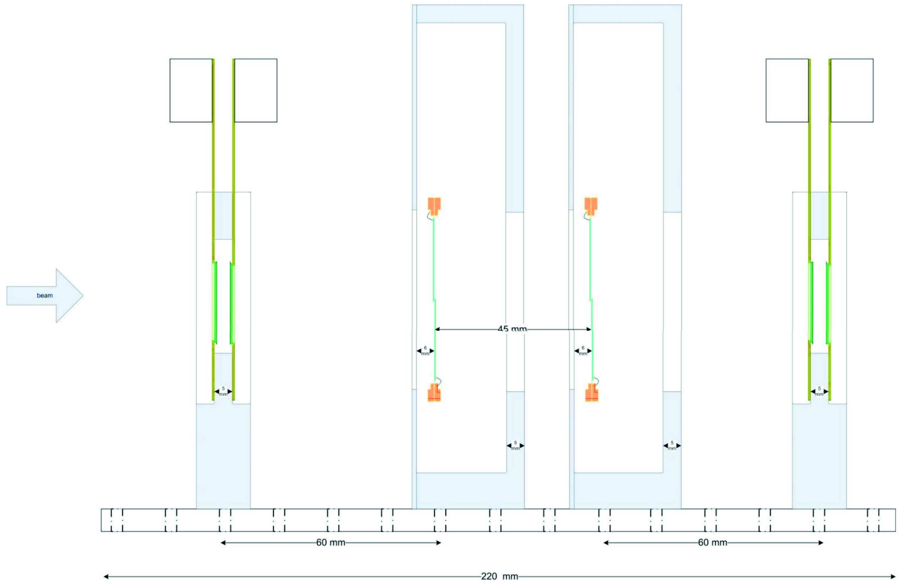
The second analysis is related to the new ladder alignment method. The idea is to rotate or translate all sensors belonging to the same ladder together. Based on the relative positions of each sensor with respect to the ladder centre, the new position for each sensor

---

[*] One module is treated like a ladder with 4 sensors



(a) Schematic diagram with photos.



(b) The configuration diagram. The distance between two modules is 45 mm, and that of the two standalone MIMOSA-28 on each side is only 5 mm.

Figure 3.1: The configuration for SALAT module test.

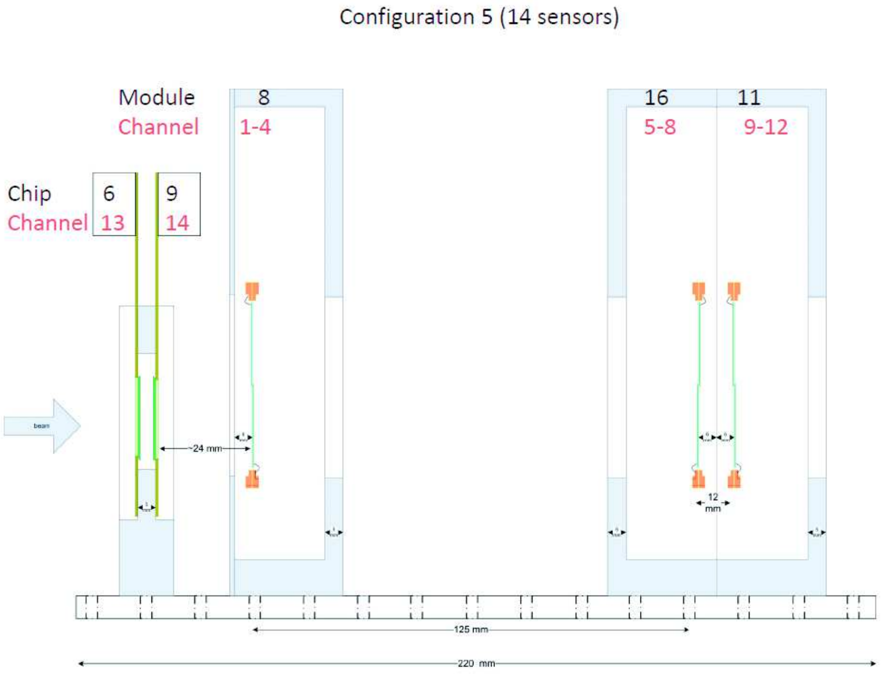


Figure 3.2: One of the configurations for MIMOSA-28 test using a SALAT system with 3 modules.

could be calculated directly. In this way, the alignment could be processed more efficiently. We tried this new algorithm using the position information taken at the beam test of the configuration in Figure 3.1b. As the calculations of relative positions propagate errors, the alignment of ladders is more difficult than the normal sensor-by-sensor alignment. We checked the alignment performance using SALAT modules as references and single MIMOSA-28 as DUT to correct possible mismatches in the same configuration.[*] The procedures are given together with the related results as follows:

1. Using modules as references to align the first single mimosa-28 upstream, the monitoring plots are given in Figure 3.5. There was a misalignment region in the scattered plot of ΔU as a function of U (The plot on the second row and the first column), the local coordinate of the sensor in horizontal direction (along the short end of the sensor). A biased region of around 1 mm was found. By checking the geometry, the reason is found at the edges of two sensors which were penetrated by the same

---

[*] The previous analysis of this configuration is to use 4 single sensors as the telescope to record the relative positions of sensors in each SALAT module. Here, the roles of modules and single sensors are exchanged.

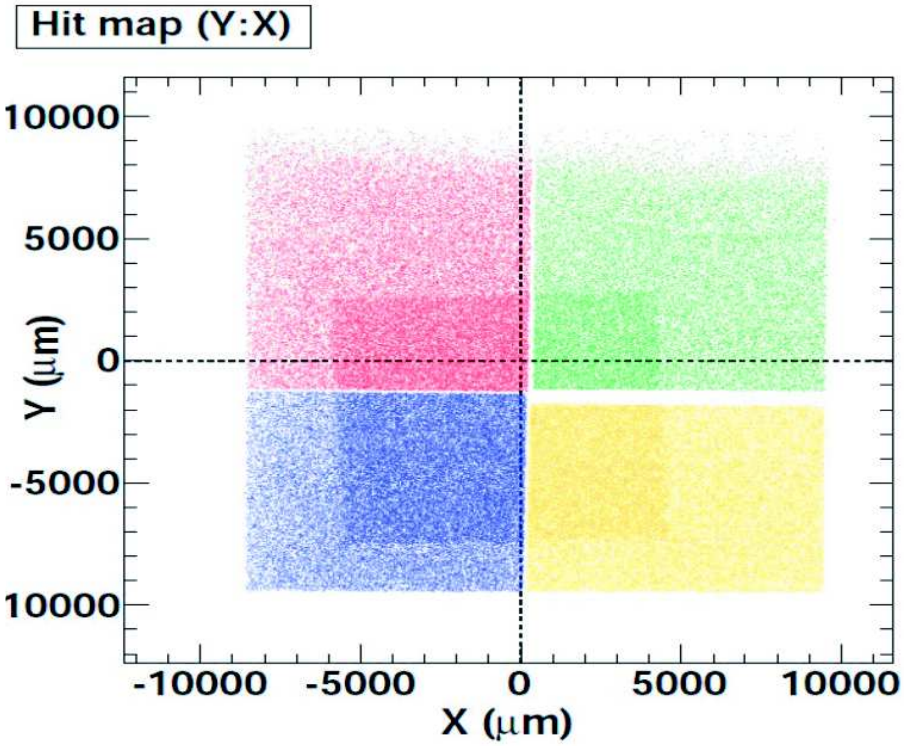
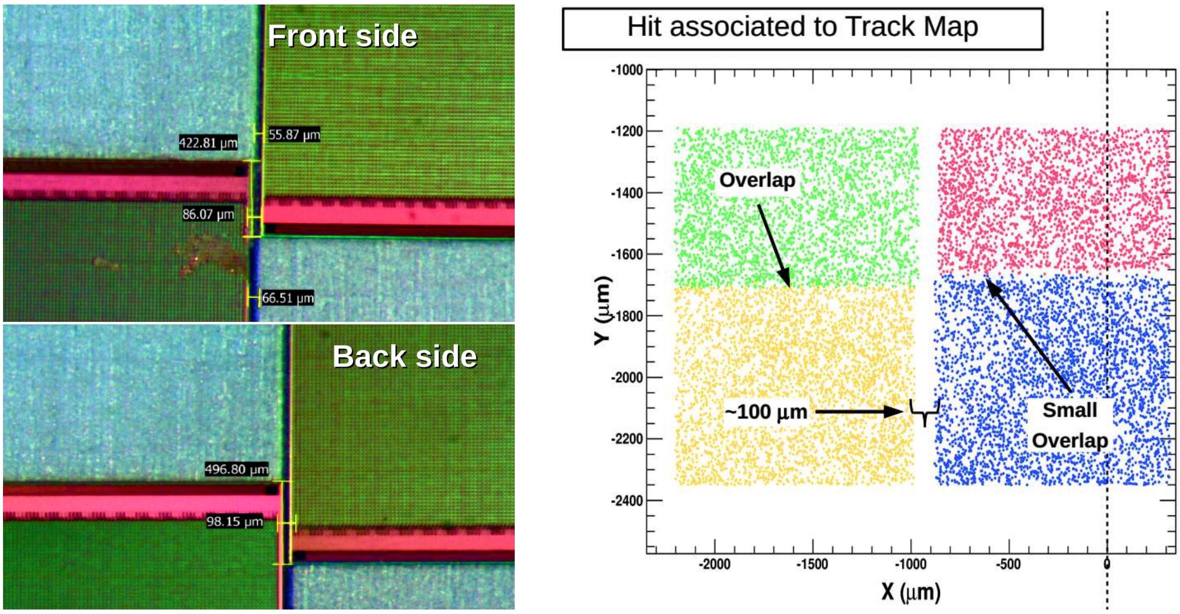


Figure 3.3: Hit map of a SALAT module. Four different colours are used four the hits from 4 MIMOSA-28 sensors inside one module. The distorted rectangular of higher hit density is the shadow of the scintillator trigger.



(a) Measurements of SALAT module 8 (b) Centre focused hit map of module 8

Figure 3.4: The configuration for SALAT module test

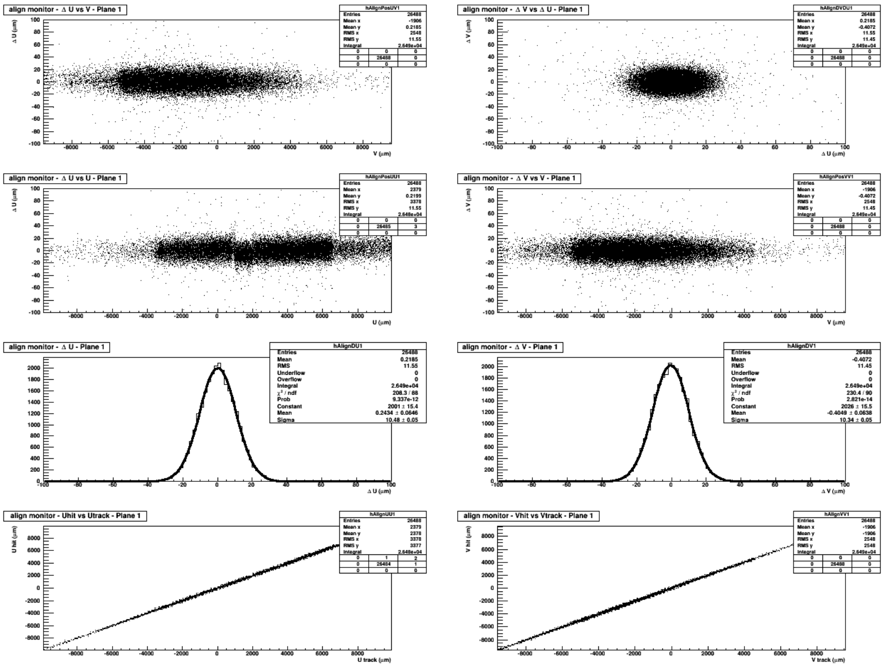


Figure 3.5: Alignment monitoring plots for a DUT (MIMOSA-28) using SALAT modules as references. The misalignment is due to the SALAT system.

tracks, on different quarters of two modules (e.g., the up-left sensor of module 1 and the up-right sensor of module 2). As one module needs four independent alignments for each quarter (or each sensor), these two sensors with mismatches belong to different simple alignment processes. Usually, the alignment algorithm suffers minimization problems for at least one of them.

2. The roles of SALAT modules and single MIMOSA-28 chips were switched back to the original ones, i.e., single sensors are used as references. More tracks were processed to improve the alignment of questionable sensors.

3. Using modules as references again to align single sensors. Figure 3.6 shows the monitoring plots with module misalignment corrected.

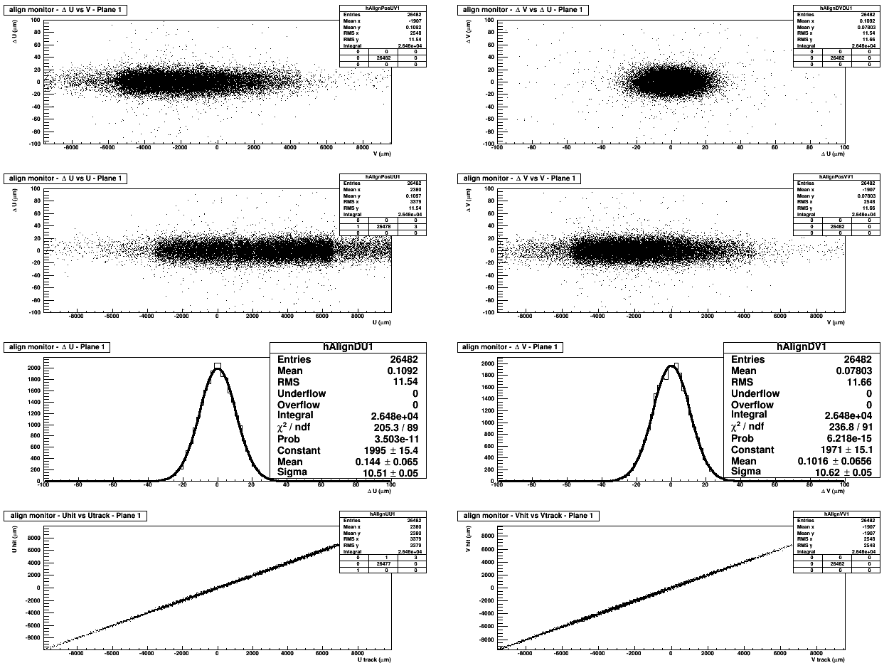


Figure 3.6: Alignment monitoring plots for a DUT (MIMOSA-28) using SALAT modules as references in the case that the SALAT modules are well aligned.

4. The final procedure is to check the performance of the new ladder alignment algorithm directly in the current configuration before other applications. First, the original roles are used to re-align modules using the new ladder alignment method. Second, the roles of DUT and references are exchanged to process the alignment

for the first single chip upstream. The result is shown in Figure 3.7. There is no obvious mismatch like that in Figure 3.5, but the scattered plot of ΔU VS U shows a deformation-like result, which is not true for that sensor. It means that sensors in the modules are misaligned slightly after using the ladder alignment implementation.

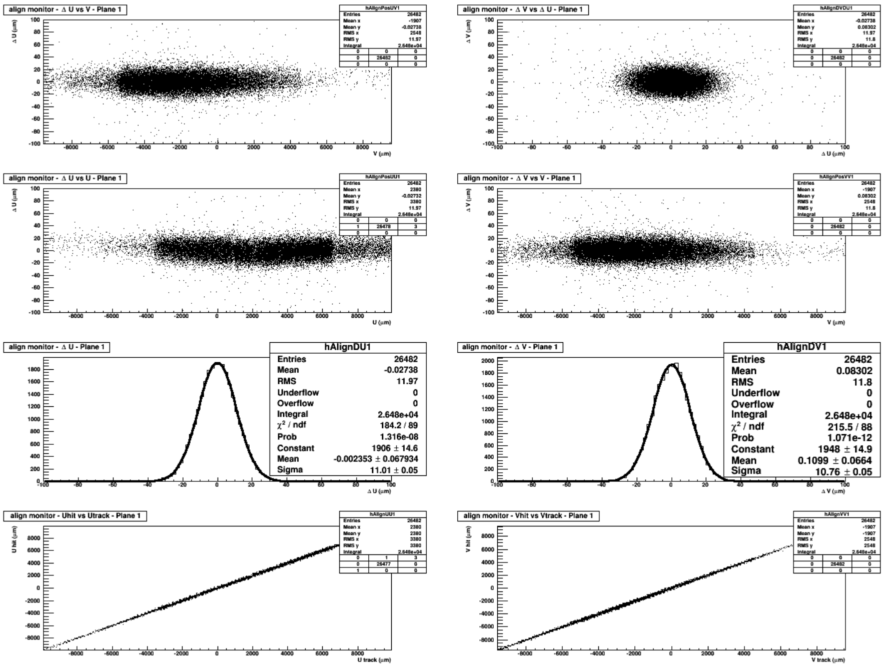


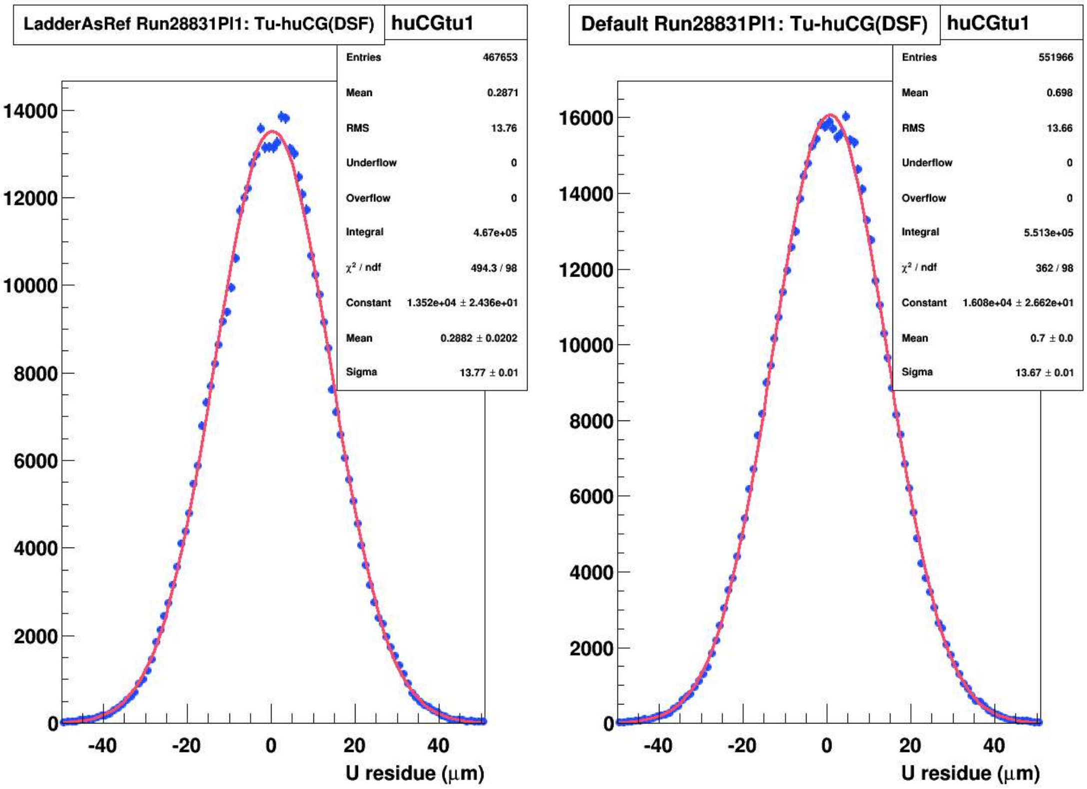
Figure 3.7: Alignment monitoring plots for a DUT (MIMOSA-28) using SALAT modules as references. SALAT modules are well aligned, but using ladder alignment method leads to a deformation-like result in the scatter diagram of ΔU VS U.

The procedures above have to be processed for all the modules, then the configuration in Figure 3.2, could be used for DUT studies with a SALAT telescope. The residual distribution in U direction for the first chip upstream is shown in Figure 3.8a. Compared with the distribution in Figure 3.8b, using a normal sensor-by-sensor alignment, fewer tracks ($\sim 85\%$) were reconstructed with a similar uncertainty. Figure 3.9 compares the impact positions of tracks for different pixel multiplicities[*] using different alignment algorithms. The patterns are similar, but plots from ladder alignment show a lower statistics.

The conclusions for the ladder alignment algorithm is given below:

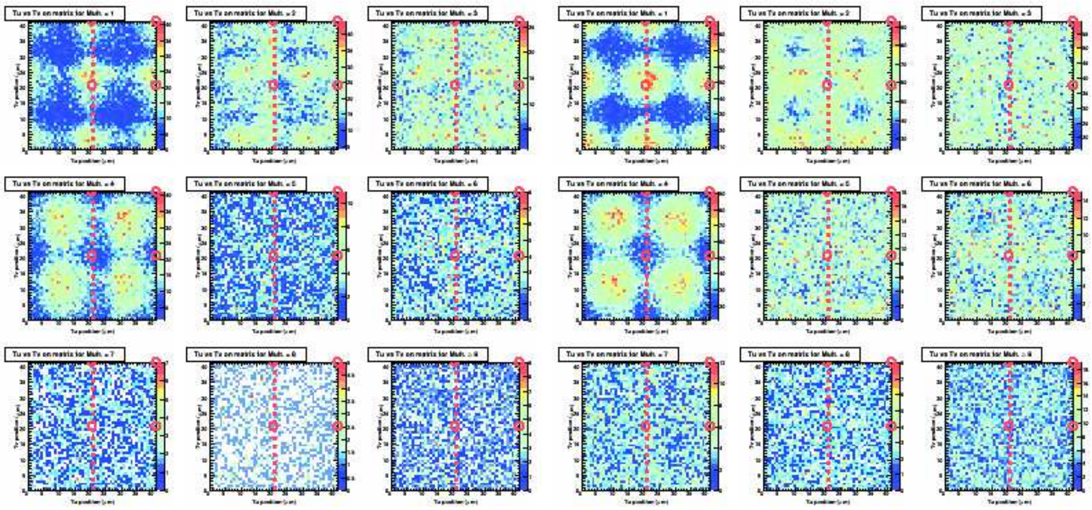
---

[*] Pixel multiplicity is the number of fired pixels for a reconstructed hit.



(a) Using ladder alignment algorithm (b) Using sensor alignment algorithm

Figure 3.8: Residual distributions in U direction



(a) Using ladder alignment algorithm (b) Using sensor alignment algorithm

Figure 3.9: Track associated hits for different pixel multiplicities. The plots from top to bottom and left to right correspond to pixel multiplicities from 1 to 9. High density is shown with higher temprature and the red circles refer to the positions of adjacent collecting diodes.

1. An unbiased performance requires a more precise alignment of sensors, especially for those belonging to different quarters of different modules with overlaps.
2. After applying ladder alignment algorithm, fewer tracks are reconstructed from the SALAT telescope. The reason is due to a slightly worse alignment performance. As the relative positions of sensors are fixed, the degrees of freedom for each module are reduced.
3. The performance on DUT studies including sensor tracking efficiency and spatial resolution is similar with that using the normal alignment procedure.

3.2.2 FSBBbis Test

FSBBbis consists of a slightly enlarged pixel pitch and a double-row rolling-shutter for a faster readout using TowerJazz 0.18 µm process. The design is optimized for the outer layer of ALICE ITS. For BESIII, the integration time of FSBBbis (41.6 µs) is beyond the required. However, a larger pixel pitch accompanied with a similar readout architecture as that in MIMOSA-28 will lead to a higher pixel occupancy. This double-row readout offers new options for reducing integration time with acceptable power consumptions. On the other hand, the enlarged pixel size (in row direction) is coincident with the pixel optimization routine in our study. The influence of pixel pitch is thus illustrated step by step using FSBBbis and MIMOSA-22THRb. Moreover, the beam test analysis of this sensor is also used for tuning two power consumption related parameters: the bias current I_{pix} of the pixel amplifier and the discriminator bias current I_{dis}. Using the same telescope configuration, the irradiated sensors were also studied to assess the radiation tolerance of FSBBbis.

The test results of one FSBBbis sensor with an epitaxial layer thickness $\sim 18\,\mu\text{m}$, are summarised in the three-colour plot shown in Figure 3.10. For the sensor requirements

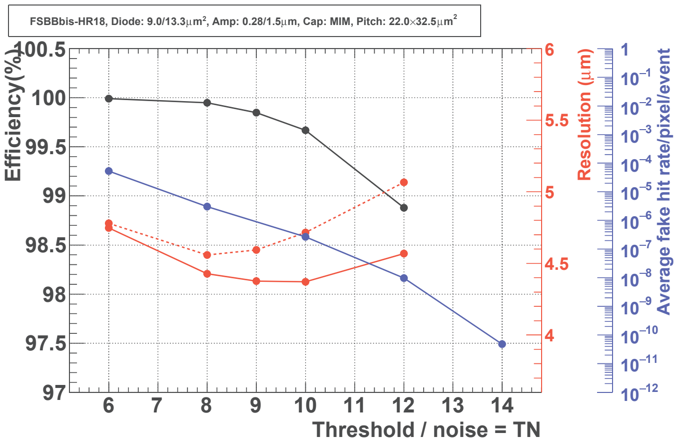
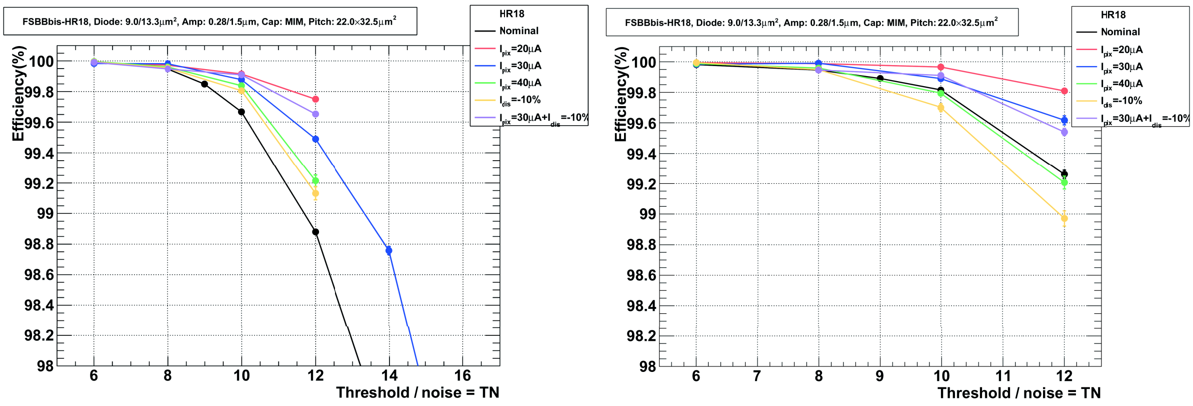


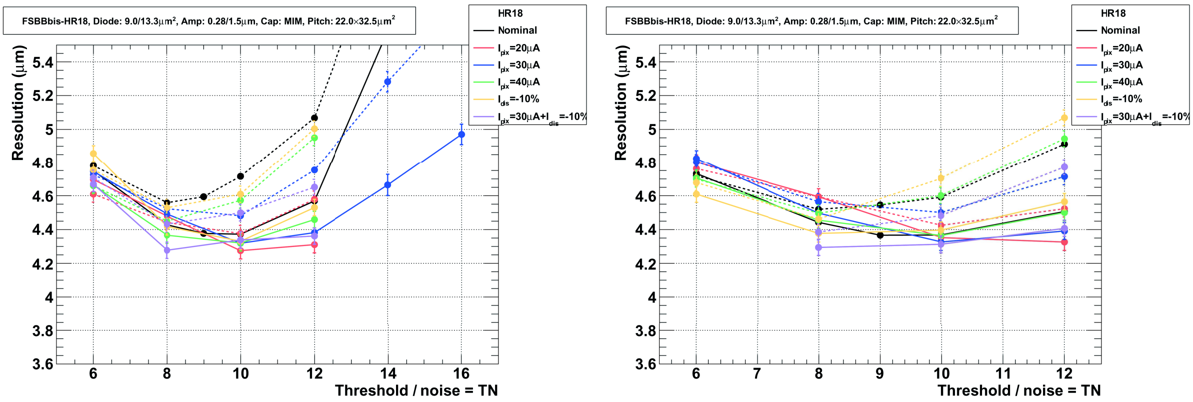
Figure 3.10: Three colour plot of a FSBBbis-HR18 sensor with the pitch around $22.0\times32.5\,\mu\text{m}^2$. The solid black line, solid blue line, solid red line and dashed red line are used to indicate the detecting efficiency, average fake hit rate, spatial resolution in U direction (along the short side) and in V coordinate (along the long side) as a function of threshold respectively.

of an efficiency $> 99\%$ and a fake hit rate $< 10^{-5}$/pixel/frame, the thresholds between 7.5 and 11.5 times the temporal noise are suitable. The intrinsic spatial resolution in this range is better than $5\,\mu\text{m}$ in both U and V directions. One could also find a resolution degrading for the elongated pixel size in V when higher thresholds are applied.

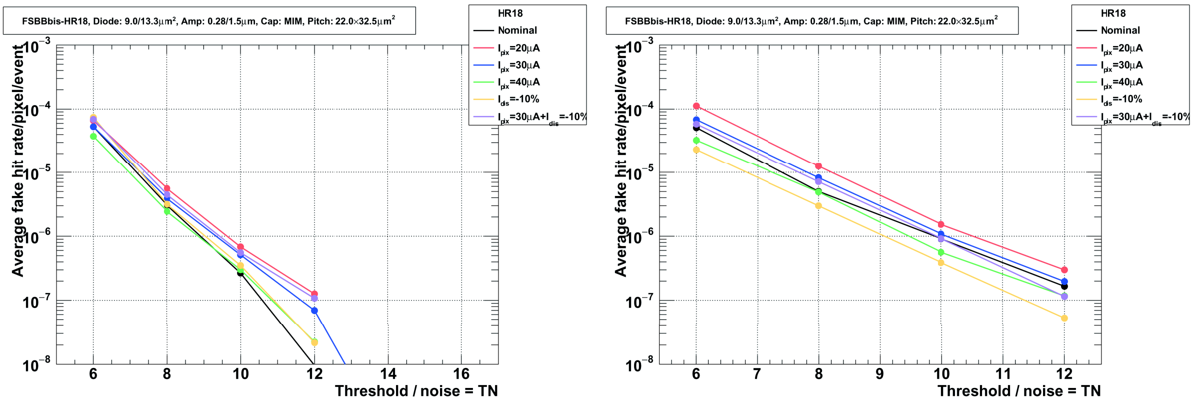
In the analysis for running parameter optimization, Figure 3.11 was shown to study the influence on detecting efficiencies. Two sensors without irradiation were tested. For



(a) Efficiency of sensor 1 (b) Efficiency of sensor 2



(c) Resolution of sensor 1 (d) Resolution of sensor 2



(e) Fake hit rate of sensor 1 (f) Fake hit rate of sensor 2

Figure 3.11: Efficiency, resolution and fake hit rate as a function of threshold for different I_{pix} and I_{dis}. The black curves are for the nominal configuration with the highest currents. The red, blue, green, yellow and pink colours are used for I_{pix} =20 µA, I_{pix} =30 µA, I_{pix} =40 µA, 90% of the nominal I_{dis} and I_{pix} =30 µA together with 90% of the nominal I_{dis} respectively. Dashed lines in the resolution curves are for the V direction (long side) while solid lines are used for U direction (short side). Plots on the left are the test results of sensor 1, and these on the right are for the second FSBBis sensor.

the first sensor, the nominal configuration (with the highest currents settings for both I_{pix} and I_{dis}) has the worst efficiency and resolution, and the improvements due to lower currents could be found with a small cost in fake hit rate. The conclusions are only partly consistent with these of the second sensor as the configuration of I_{pix} =40 µA or $I_{dis} = -10\%$ led to a worse efficiency and resolution. The reason of this different behaviour on sensors is not clear. In case of a low current I_{pix} =20 µA, the two sensors have the best efficiency, increased fake hit rate with a factor of 2 to 4, and a similar spatial resolution. Since this fake hit rate is still acceptable, decreasing I_{pix} and I_{dis} is of interest to mitigate the power dissipation, which in turn helps to reduce the material budget.

3.3 Summary and conclusion

In this chapter, the beam test results of SALAT modules and FSBBbis sensors have been presented. Considering the published performance of MIMOSA-22THRb sensors shown in Figure 3.12, the resolutions of sensors with different pixel pitches and on condition of

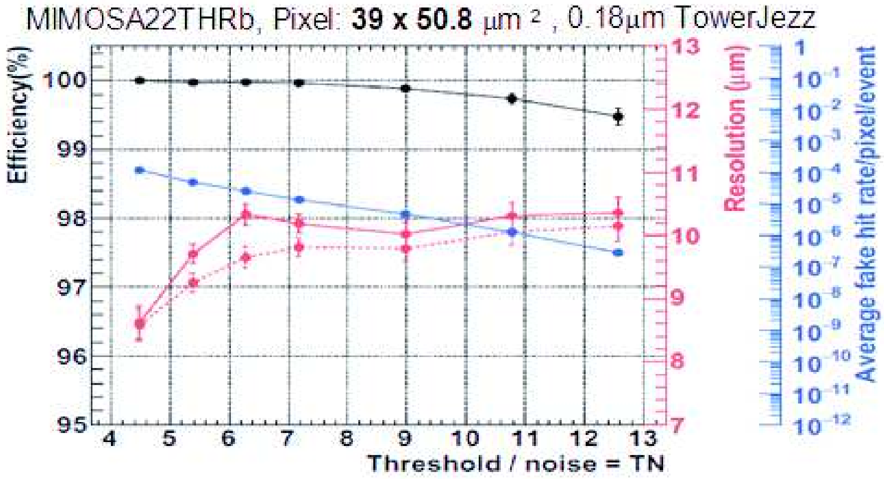


Figure 3.12: Efficiency (black), resolution (red) and fake hit rate (blue) as a function of threshold for MIMOSA-22THRb [7].

a good detecting efficiency (> 99%) and a low fake hit rate ($< 10^{-5}$/pixel/frame) are summarized in Table 3.2. The resolution of the new TowerJazz 0.18 µm process, degrades with the increase of pixel pitches. This trend is consistent with the results of sensors using AMS 0.35 µm process, which are summarized in Figure 2.21 in Chapter 2, and the

CPS	Pixel pitch (U×V [µm])	resolution [µm]
MIMOSA-28	20.7×20.7	~ 3.5
FSBBbis	22.0×32.5	~ 4.4
MIMOSA-22THRb	39.0×50.8	~ 10

Table 3.2: The requirements of BESIII inner tracker and the related performance achieved by STAR Pixel Layer (PXL) using MIMOSA-28.

10 µm resolution of a large pitch is a new record for CPS sensors. It means that the SNR and related pixel circuits of such a large pixel are good and functional for particle detection. Using the double-row rolling-shutter of FSBBbis and a large pixel pitch like that of the MIMOSA-22THRb, the MISTRAL-M0 sensor designed for ITS outer layer has the potential to run with a power consumption around 80 mW/cm^2 and an integration time about 20.8 µs. By reducing readout speed, even more power could be saved for the CPS designs focusing on BESIII requirements. Moreover, the pALPIDE sensors could work with a power consumption $\lesssim$ 40 mW cm^{-2} [48] using a pixel pitch 28 µm × 28 µm. Combined with a large pixel pitch for a resolution around 10 µm, ALPIDE architecture could reach a new limit for low power dissipation. The maximum ladder length for BESIII SPT is 2.6 times as long as that of the STAR PXL, a power saving factor of 4 is therefore helpful to loosen the requirements on cooling, which guarantees a low material budget.

Chapter 4

Tracking Algorithms

To reconstruct the trajectory of a charged particle, two types of algorithms are needed. First, all the hits belonging to one track will be collected by a track finder whose main task is to solve pattern recognition problems and then to find the track with its parameters. Algorithms like conformal mapping, Hough transform and cellular animation are popular methods in track finding. After the track parameters are given, a track fitting algorithm will be used to optimise the parameters and to output their covariance matrix. The two widely used methods for fitting are the generalized least squares (GLS) and the Kalman filter (KF), both of which are originally used to solve linear problems and have been extended in the non-linear case. Moreover, Kalman filter could be used for track finding, which benefits from its recursive strategy. Therefore, it will also be used to estimate tracking efficiencies. The probability that a real hit is associated with the reconstructed track will also be calculated, which is related to the tracking efficiency and purity of a detector. The following sections about the least chi-square and Kalman filter will present the main ideas and the related mathematical derivations. Tracking performance in terms of impact parameter resolution, momentum resolution and tracking efficiency of a given tracker design, is then estimated using a tracking algorithm.

4.1 The generalized least squares fitting

The generalized least squares algorithm is a widely used statistical technique to solve linear regression problems. In charged particle tracking, straight tracks could be handled naturally while the helix trajectories could be linearised or simplified using parabola approximation in $r-\phi$ plane. In this section, the generalized linear square fitting algorithm will be introduced.

4.1.1 Variables and notations

The lower-case letters stand for variables. They could be unknown true values, the known measured variables and the known user decided independent variables. For example, in the 2D track fitting equation

$$\eta = \theta_1 + \theta_2 x + \theta_3 x^2.$$

Letter x stands for user-decided variable which is the position of a telescope plane and is supposed to be accurate. θ_1, θ_2 and θ_3 are the parameters we need to figure out, and η is the true value, the real hit position. When η is measured, we could get the measurement variable y which has a mean value η with an uncertainty σ. To reconstruct a track is to evaluate the track parameters and their errors from the track fitting equation, such that measurements should be done in a series of detecting planes located at $x_1, x_2, \ldots, x_n$. Finally, we get N equations:

$$\begin{aligned} \eta_1 &= \theta_1 + \theta_2 x_1 + \theta_3 x_1^2 \\ &\vdots \\ \eta_j &= \theta_1 + \theta_2 x_j + \theta_3 x_j^2 \\ &\vdots \\ \eta_n &= \theta_1 + \theta_2 x_n + \theta_3 x_n^2 \end{aligned} \tag{4.1.1}$$

And the notation y_j is used as the related measurement of the true value η_j.

The bold lower-case letters designate vectors which are $m \times 1$ matrices with only one column, while upper-case boldface is used for $m \times n$ matrix. Therefore, the fitting problem in Equations 4.1.1 could be simplified as

$$\boldsymbol{y} = \boldsymbol{\eta} + noise = \boldsymbol{A\theta} + noise \tag{4.1.2}$$

where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_j \\ \vdots \\ \eta_N \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ & \vdots & \\ 1 & x_j & x_j^2 \\ & \vdots & \\ 1 & x_N & x_N^2 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

In the general case, η is a linear function with M parameters:

$$\eta_j = f_j(\theta) = \theta_1 a_{j1} + \theta_2 a_{j2} + \cdots + \theta_M a_{jM}, \tag{4.1.3}$$

in which j stands for the j'th measurement. The relevant matrices are changed as follows:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1M} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \cdots & a_{ji} & \cdots & a_{jM} \\ \vdots & & \vdots & & \vdots \\ a_{N1} & \cdots & a_{Ni} & \cdots & a_{NM} \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_i \\ \vdots \\ \theta_M \end{bmatrix},$$

The covariance matrix of y could be a diagonal one when the measurements are independent to each other. Nevertheless, it's not the case in particle tracking in which multiple scattering should be considered. We have a general covariance matrix:

$$\boldsymbol{V_y} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1i}^2 & \cdots & \sigma_{1N}^2 \\ \vdots & & \vdots & & \vdots \\ \sigma_{j1}^2 & \cdots & \sigma_{ji}^2 & \cdots & \sigma_{jN}^2 \\ \vdots & & \vdots & & \vdots \\ \sigma_{N1}^2 & \cdots & \sigma_{Ni^2} & \cdots & \sigma_{NN}^2 \end{bmatrix}$$

It can be proved that this matrix is symmetric, i.e., $\sigma_{ij} = \sigma_{ji}$. The related discussion could be found in the following section involving multiple scattering.

4.1.2 Fitting tracks without multiple scattering

As multiple scattering raises the measurement correlations to complicate the covariance matrix of the observed values, it is easier and more intuitive to start without considering this material effect. In this case, the covariance matrix for $\boldsymbol{y}$ is a diagonal matrix. Therefore, we write the chi-square as follows:

$$\chi^2 = \sum_{j=1}^{N} \frac{(y_j - \eta_j)^2}{\sigma_j^2} = \sum_{j=1}^{N} \frac{[y_j - f_j(\theta_1, \theta_2, \cdots, \theta_M)]^2}{\sigma_j^2} \tag{4.1.4}$$

In the matrix notation, equation 4.1.4 reads:

$$\chi^2 = (\boldsymbol{y} - \boldsymbol{A\theta})^T \boldsymbol{V}_{\boldsymbol{y}}^{-\mathbf{1}} (\boldsymbol{y} - \boldsymbol{A\theta}) \tag{4.1.5}$$

This equation also works for the general case considering multiple scattering effects, since $\boldsymbol{V_y}$ is symmetric.

4.1.2.1 Parameter estimation

The best estimation is to minimize this chi-square which means that the first-order partial derivative of each parameter should be zero, hence we can write:

$$\frac{\partial \chi^2}{\partial \theta_l} = \sum_{j=1}^{N} \left\{ \frac{2\,[y_j - f_j(\theta_1, \theta_2, \cdots, \theta_M)]}{\sigma_j^2} \cdot (-\frac{\partial f_j}{\partial \theta_l}) \right\} = 0$$

By substituting f_j with equation 4.1.3 and eliminating the constant -2, we find:

$$\sum_{j=1}^{N} \left\{ \frac{[y_j - (\theta_1 a_{j1} + \cdots + \theta_l a_{jl} + \cdots + \theta_M a_{jM})]}{\sigma_j^2} \cdot a_{jl} \right\} = 0, \tag{4.1.6}$$

such that we get M linear equations with M parameters and therefore all the estimates $\hat{\theta}_l$ could be solved.

In the matrix notation, the partial derivative could be written more simply:

$$\begin{aligned}\frac{\partial\chi^2}{\partial\boldsymbol{\theta}} &= (\boldsymbol{y}-\boldsymbol{A\theta})^T\boldsymbol{V_y^{-1}}(-\boldsymbol{A}) + (\boldsymbol{y}-\boldsymbol{A\theta})^T\boldsymbol{V_y^{-1T}}(-\boldsymbol{A}) \\ &= -2(\boldsymbol{y}-\boldsymbol{A\theta})^T\boldsymbol{V_y^{-1}A} \\ &= -2\boldsymbol{A}^T\boldsymbol{V_y^{-1}}(\boldsymbol{y}-\boldsymbol{A\theta}) \\ &= 0\end{aligned}$$

And the set of linear equation 4.1.6 for all the parameters $(\theta_1, \theta_2, \cdots, \theta_M)$ reads:

$$\boldsymbol{A}^T\boldsymbol{V_y^{-1}}(\boldsymbol{y}-\boldsymbol{A\theta}) = 0$$

In this way, the estimation for $\boldsymbol{\theta}$ could be solved with one formula:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1}\boldsymbol{A}^T\boldsymbol{V_y^{-1}y} \tag{4.1.7}$$

4.1.2.2 Covariance of parameters

In this section, only matrix notation is used to show the derivation of the covariance of the estimated parameters. First of all, the definition of covariance for y_i and y_j is given below:

$$\boldsymbol{V_y}_{ij} = \langle (y_i - \overline{y_i})(y_j - \overline{y_j}) \rangle \equiv \langle \delta y_i \delta y_j \rangle\,.$$

Therefore, the covariance matrix of $\boldsymbol{y}$ could be written by:

$$\boldsymbol{V_y} = \left\langle \boldsymbol{\delta y \delta y^T} \right\rangle$$

Since matrix $\boldsymbol{A}$ and $\boldsymbol{V_y}$ are constant matrices and an inverse matrix of a symmetric matrix is also symmetric, we can derive the covariance matrix of $\hat{\boldsymbol{\theta}}$ as follows:

$$\begin{aligned}\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} &= (\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1}\boldsymbol{A}^T\boldsymbol{V_y^{-1}}\left\langle \boldsymbol{\delta y \delta y^T} \right\rangle \boldsymbol{V_y^{-1}A}(\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1} \\ &= (\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1}\boldsymbol{A}^T\boldsymbol{V_y^{-1}V_yV_y^{-1}A}(\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1} \\ &= (\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1}(\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})(\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1} \\ &= (\boldsymbol{A}^T\boldsymbol{V_y^{-1}A})^{-1}\end{aligned} \tag{4.1.8}$$

Although this section started with a simple diagonal $\boldsymbol{V_y}$ without considering multiple scattering, all the equations with matrix notations are also valid for a complex $\boldsymbol{V_y}$ containing the correlation information of the measurements.

4.1.2.3 An Example in N uniform layers detector

Assuming a particle flying in x-y plane under a magnetic field along z-axis is measured by N detector layers placed uniformly along x-axis, the trajectory is an arc of a circle. Since circle function is not linear, the parabola approximation could be a good choice for high momentum particles. The track equation reads:

$$y = a + bx + cx^2 + noise$$

where a, b and c are parameters to be figured out. The data we have is the measured value y_j(j ranges from 1 to N) at the relevant layer position x_j. To simplify the calculation, the spatial resolution of each layer is σ and $x_j = \frac{jL}{N}$ in which L is the position of the last layer. We can write the parameter matrix to be estimated and those constant matrices given by data and the detector geometry as follows

$$\boldsymbol{\theta} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ & \vdots & \\ 1 & x_j & x_j^2 \\ & \vdots & \\ 1 & x_N & x_N^2 \end{bmatrix}, \quad \boldsymbol{V_y^{-1}} = \begin{bmatrix} 1/\sigma^2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1/\sigma^2 \end{bmatrix}$$

The parameter covariance matrix could be solved according to equation 4.1.8:

$$\begin{aligned}
\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} &= (\boldsymbol{A}^T\boldsymbol{V}_{\boldsymbol{y}}^{-1}\boldsymbol{A})^{-1} \\
&= \left(\begin{bmatrix} 1 & x_1 & x_1^2 \\ & \vdots & \\ 1 & x_j & x_j^2 \\ & \vdots & \\ 1 & x_N & x_N^2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} 1/\sigma_1^2 & x_1/\sigma_1^2 & x_1^2/\sigma_1^2 \\ & \vdots & \\ 1/\sigma_j^2 & x_j/\sigma_j^2 & x_j^2/\sigma_j^2 \\ & \vdots & \\ 1/\sigma_N^2 & x_N/\sigma_N^2 & x_N^2/\sigma_N^2 \end{bmatrix}\right)^{-1} \\
&= \begin{bmatrix} \sum\limits_{j=1}^N 1/\sigma_j^2 & \sum\limits_{j=1}^N x_j/\sigma_j^2 & \sum\limits_{j=1}^N x_j^2/\sigma_j^2 \\ \sum\limits_{j=1}^N x_j/\sigma_j^2 & \sum\limits_{j=1}^N x_j^2/\sigma_j^2 & \sum\limits_{j=1}^N x_j^3/\sigma_j^2 \\ \sum\limits_{j=1}^N x_j^2/\sigma_j^2 & \sum\limits_{j=1}^N x_j^3/\sigma_j^2 & \sum\limits_{j=1}^N x_j^4/\sigma_j^2 \end{bmatrix}^{-1}
\end{aligned} \tag{4.1.9}$$

Since $\sum x_j^k$ includes the sum of k'th power of integers, the accurate expression of the matrix above could be expanded by equations A.1.1–A.1.5 in Appendix A.1. If the number of layers are very large, we could get a simplified matrix. Considering the equation below:

$$\begin{aligned}
\lim_{N\to\infty}\sum_{j=1}^{N}\left(\frac{jL}{N}\right)^k &= \int_\delta^1 x^k dx \times NL^k \\
&= \frac{NL^k}{k+1}
\end{aligned}$$

where $\delta \to 0$, the covariance matrix reads:

$$\begin{aligned}
\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} &= \frac{\sigma^2}{N}\begin{bmatrix} 1 & \frac{1}{2}L & \frac{1}{3}L^2 \\ \frac{1}{2}L & \frac{1}{3}L^2 & \frac{1}{4}L^3 \\ \frac{1}{3}L^2 & \frac{1}{4}L^3 & \frac{1}{5}L^4 \end{bmatrix}^{-1} \\
&= \frac{\sigma^2}{N}\begin{bmatrix} 9 & -36/L & 30/L^2 \\ -36/L & 192/L^2 & -180/L^3 \\ 30/L^2 & -180/L^3 & 180/L^4 \end{bmatrix}
\end{aligned}$$

Finally, the track parameters could be calculated by Equation 4.1.7 as shown below

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= (\boldsymbol{A}^T \boldsymbol{V}_{\boldsymbol{y}}^{-\mathbf{1}} \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{V}_{\boldsymbol{y}}^{-\mathbf{1}} \boldsymbol{y} \\
&= \boldsymbol{V}_{\hat{\boldsymbol{\theta}}} \boldsymbol{A}^T \boldsymbol{V}_{\boldsymbol{y}}^{-\mathbf{1}} \boldsymbol{y} \\
&= \frac{\sigma^2}{N} \begin{bmatrix} 9 & -36/L & 30/L^2 \\ -36/L & 192/L^2 & -180/L^3 \\ 30/L^2 & -180/L^3 & 180/L^4 \end{bmatrix} \frac{1}{\sigma^2} \begin{bmatrix} \sum y_j \\ \sum x_j y_j \\ \sum x_j^2 y_j \end{bmatrix} \\
&= \frac{1}{N} \begin{bmatrix} 9 & -36/L & 30/L^2 \\ -36/L & 192/L^2 & -180/L^3 \\ 30/L^2 & -180/L^3 & 180/L^4 \end{bmatrix} \begin{bmatrix} \sum y_j \\ \sum x_j y_j \\ \sum x_j^2 y_j \end{bmatrix}
\end{aligned}
$$

4.1.3 Chi-square fitting with Multiple scattering

There are mainly two methods to consider multiple scattering into least squares fitting. The first algorithm takes the multiple scattering as a part of the measurement error. In this case, one needs to rewrite the covariance matrix for $\boldsymbol{y}$, and it is not a diagonal matrix any more. The other method is to estimate all the kink angles for each scattering plane as described in article [52].

4.1.3.1 Method 1: Include multiple scattering error into the measurement covariance matrix

Figure 4.1 demonstrates a four-layer detector measuring a track with the original direction along x-axis. The layers are placed at the positions from r_1 to $r4$, and are perpendicular to x-axis. As the scattering angle α is small, the material contribution of the i'th layer (as a scatterer) to the fourth layer is $\Delta y_{4\_i} \approx \alpha_i(r_4 - r_i)$ with the index i ranging from 1 to 3. Therefore, the total multiple scattering shift on the j'th layer is the sum of the contributions from all the previous layers:

$$
\Delta y_j = \begin{cases} \sum\limits_{l=1}^{j-1} \alpha_l (r_j - r_l), & j > 1 \\ 0, & j = 1 \end{cases} \tag{4.1.10}
$$

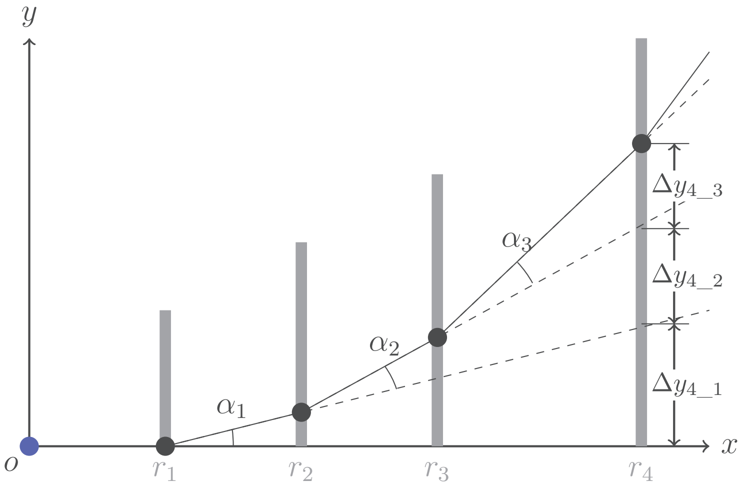


Figure 4.1: Multiple scattering in several layers of a detector

An ideal threshold function could be defined by:

$$T(z) = \begin{cases} 1, & z > 0 \\ 0, & z \leq 0 \end{cases}$$

With this function, Equation 4.1.10 in an N-layer detector could be rewritten as:

$$\Delta y_j = \sum_{l=1}^{N} \alpha_l (r_j - r_l) T(r_j - r_l) \tag{4.1.11}$$

Since the average scattering angle is 0, we have $\overline{\Delta y} = 0$. And the deflection angle at each layer is independent with each other. It means that multiple scattering will not bias our measurements on y and only contribute to the covariance matrix with the element shown below:

$$\begin{aligned} \boldsymbol{V}_{\boldsymbol{y}_{ij}} &= \langle \Delta y_i \Delta y_j \rangle + \sigma_i^2 \delta_{ij} \\ &= \sum_{l=1}^{N} (r_i - r_l) T(r_i - r_l)(r_j - r_l) T(r_j - r_l) \left\langle \delta\alpha_l \delta\alpha_l \right\rangle \\ &\quad + \sigma_i^2 \delta_{ij} \end{aligned} \tag{4.1.12}$$

where the second term σ_i is the intrinsic single point resolution on the i'th layer which only exists in the diagonal elements and $\langle \delta\alpha_l \delta\alpha_l \rangle$ is the variance of small multiple scattering angle with the approximated standard deviation for single charged projectile described by [40, 41]

$$\sqrt{\langle \delta\alpha_l \delta\alpha_l \rangle} = \theta_0 \approx \frac{13.6MeV}{\beta pc} \sqrt{\frac{X}{X_0}} \left[1 + 0.088 \log_{10} \left(\frac{X}{X_0} \right) \right] \tag{4.1.13}$$

or by a formula considering β and z dependence in the lowest order:

$$\sigma_{MS} \approx \frac{13.6MeV}{pc} \frac{z}{\beta} \sqrt{\frac{X}{X_0}} \left[1 + 0.088 \log_{10} \left(\frac{Xz^2}{X_0 \beta^2} \right) \right] \tag{4.1.14}$$

Equation 4.1.12 contains the sum from $l = 1$ to $l = N$, but the term with r_l larger than r_i or r_j is 0. Therefore, if the layer index $i \leq j$, it could be simplified as

$$\boldsymbol{V}_{\boldsymbol{y}ij} = \sum_{l=1}^{i} (r_i - r_l)(r_j - r_l) \langle \delta\alpha_l \delta\alpha_l \rangle + \sigma_i^2 \delta_{ij} \tag{4.1.15}$$

With this symmetric covariance matrix $\boldsymbol{V_y}$, the estimations of track parameters $\hat{\boldsymbol{\theta}}$ and the covariance matrix $\boldsymbol{V}_{\hat{\boldsymbol{\theta}}}$ could be calculated using equation 4.1.7 and equation 4.1.8 respectively.

This method using the same equations and procedures as the non scattering case is easy to use in theory. Nevertheless, its disadvantage is too much time consumption to inverse $\boldsymbol{V_y}$ when the number of detection layers is large.

4.1.3.2 Method 2: Fitting all kink angles

The kink angles or deflection angles of multiple scattering could be evaluated at the same time as what we do for track parameters in a modified chi-square function. The key issue of this method is to evaluate some angles which are not measured. In the Gaussian-like approximation, these scattering angles have the mean value of zero and the standard derivative which could be described by equation 4.1.13 and equation 4.1.14, such that a term $\sum \alpha_l^2 / \sigma_{MS}^2$ could be added into the chi-square function given by [52]:

$$\chi^2 = \sum_{j=1}^{N} \frac{[y_j - f_j(\theta_1, \theta_2, \cdots, \theta_M) - \Delta y_j]^2}{\sigma_j^2} + \sum_{l=1}^{N} \frac{(\alpha_l - 0)^2}{\sigma_{\alpha_l}^2} \tag{4.1.16}$$

where, f_j is the linear function of the j'th measurement which is represented in equation 4.1.3, and Δy_j described by equation 4.1.11, is the displacement coming from multiple scattering.

4.2 Kalman Filter

In modern particle physics experiments, Kalman Filter is a widely used algorithm for track finding and fitting. The original Discrete Kalman Filter is a recursive solution to the discrete-time version of the Wiener problem in signal processing[53]. The theory description and derivations in this section mainly refer to literatures [53, 54, 55].

Let us start with simple measurements of a constant η for n times. Then we have:

$$y_i = \eta + w_i$$

where y_i is the i'th value we get and w_i stands for the i'th measurement error we can not know. In the easiest case, the best estimate of η is shown below:

$$\hat{\eta}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$$

This expression could be written in terms of $\hat{\eta}_{n-1}$ and y_n:

$$\begin{aligned}\hat{\eta}_n &= \frac{n-1}{n} \cdot \frac{1}{n-1}\sum_{i=1}^{n-1} y_i + \frac{y_n}{n} \\ &= \frac{n-1}{n}\hat{\eta}_{n-1} + \frac{y_n}{n} \\ &= \hat{\eta}_{n-1} + \frac{1}{n}(y_n - \hat{\eta}_{n-1}) \qquad (4.2.1)\end{aligned}$$

The trick is straightforward, but this derivation shows the hint towards the much more complex Kalman Filter whose key idea is to calculate the current estimator ($\hat{\eta}_n$) from the prior estimate ($\hat{\eta}_{n-1}$) and the difference between the current measurement (y_n) and the previous estimator ($\hat{\eta}_{n-1}$). In this simple case, a factor $1/n$ delivers the difference to $\hat{\eta}_n$ in the second term, which will be derived later and will be replaced by the Kalman gain.

4.2.1 The Discrete Kalman Filter

The random process which could be solved by Kalman filtering can be generally modelled by the system state equation

$$\boldsymbol{\theta_{k+1}} = \boldsymbol{F_k}\boldsymbol{\theta_k} + \boldsymbol{w_k} \tag{4.2.2}$$

and the measurement equation

$$\boldsymbol{y_k} = \boldsymbol{H_k}\boldsymbol{\theta_k} + \boldsymbol{v_k} \tag{4.2.3}$$

The related notation explanations are listed in Table 4.1

Notation	Dimension	Description
$\boldsymbol{\theta_k}$	$m \times 1$	the k'th state vector (at time t_k)
$\boldsymbol{F_k}$	$m \times m$	process propagator from state k to state $k+1$
$\boldsymbol{w_k}$	$m \times 1$	process noise vector assumed with a mean value of 0 and a known covariance
$\boldsymbol{Q_k}$	$m \times m$	the diagonal covariance matrix of $\boldsymbol{w_k}$
$\boldsymbol{y_k}$	$n \times 1$	the k'th measurement vector
$\boldsymbol{H_k}$	$n \times m$	ideal measurement matrix which connects the state vector with the measurements
$\boldsymbol{v_k}$	$n \times 1$	the measurement error independent from $\boldsymbol{w_k}$ and with a known covariance structure
$\boldsymbol{V_k}$	$n \times n$	the diagonal covariance matrix of $\boldsymbol{v_k}$

Table 4.1: Notation Elaboration

To solve the estimation problem above, Kalman filtering requires the prior (a priori) estimate $\boldsymbol{\hat{\theta}_k^-}$ for the current state vector and the related estimation covariance matrix:

$$\boldsymbol{C_k^-} = \left\langle \left(\boldsymbol{\theta_k} - \boldsymbol{\hat{\theta}_k^-}\right)\left(\boldsymbol{\theta_k} - \boldsymbol{\hat{\theta}_k^-}\right)^T \right\rangle \tag{4.2.4}$$

These prior informations are predicted before the state update. Since we have the measurement of current state $\boldsymbol{y_k}$, the estimation of the current state $\boldsymbol{\theta_k}$ could be improved by a similar technique used in equation 4.2.1:

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} + \boldsymbol{K}_{\boldsymbol{k}} \left(\boldsymbol{y}_{\boldsymbol{k}} - \boldsymbol{H}_k \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) \tag{4.2.5}$$

in which $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ is the updated (a posteriori) estimate and $\boldsymbol{K}_{\boldsymbol{k}}$ is the kalman gain[*] which will be determined later using minimum mean-square error (MMSE) method.

Before finding the form of $\boldsymbol{K}_{\boldsymbol{k}}$, the error covariance matrix of the current estimation have to be indicated:

$$\boldsymbol{C}_k = \left\langle \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}} \right) \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}} \right)^T \right\rangle \tag{4.2.6}$$

By considering that $\boldsymbol{v}_{\boldsymbol{k}}$ is independent from the error $(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-})$ and then substituting equation 4.2.3 and equation 4.2.5 into equation 4.2.6, we have

$$\begin{aligned}
\boldsymbol{C}_k = & \left\langle \left[\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} - \boldsymbol{K}_{\boldsymbol{k}} \left(\boldsymbol{H}_k \boldsymbol{\theta}_{\boldsymbol{k}} + \boldsymbol{v}_{\boldsymbol{k}} - \boldsymbol{H}_k \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) \right] \right. \\
& \left. \left[\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} - \boldsymbol{K}_{\boldsymbol{k}} \left(\boldsymbol{H}_k \boldsymbol{\theta}_{\boldsymbol{k}} + \boldsymbol{v}_{\boldsymbol{k}} - \boldsymbol{H}_k \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) \right]^T \right\rangle \\
= & \left\langle \left[\left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) - \boldsymbol{K}_k \boldsymbol{H}_k \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) - \boldsymbol{K}_k \boldsymbol{v}_{\boldsymbol{k}} \right] \right. \\
& \left. \left[\left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) - \boldsymbol{K}_k \boldsymbol{H}_k \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) - \boldsymbol{K}_k \boldsymbol{v}_{\boldsymbol{k}} \right]^T \right\rangle \\
= & \left\langle \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right)^T \right\rangle \\
& - \left\langle \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right)^T \right\rangle \boldsymbol{H}_{\boldsymbol{k}}{}^T \boldsymbol{K}_{\boldsymbol{k}}{}^T \\
& - \boldsymbol{K}_k \boldsymbol{H}_k \left\langle \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right)^T \right\rangle \\
& + \boldsymbol{K}_k \boldsymbol{H}_k \left\langle \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right) \left(\boldsymbol{\theta}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} \right)^T \right\rangle \boldsymbol{H}_{\boldsymbol{k}}{}^T \boldsymbol{K}_{\boldsymbol{k}}{}^T \\
& + \boldsymbol{K}_k \left\langle \boldsymbol{v}_k \boldsymbol{v}_{\boldsymbol{k}}{}^T \right\rangle \boldsymbol{K}_{\boldsymbol{k}}{}^T \\
= & \boldsymbol{C}_{\boldsymbol{k}}^{-} - \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_{\boldsymbol{k}}{}^T \boldsymbol{K}_{\boldsymbol{k}}{}^T - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \\
& + \boldsymbol{K}_k \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_{\boldsymbol{k}}{}^T + \boldsymbol{V}_k \right) \boldsymbol{K}_{\boldsymbol{k}}{}^T && (4.2.7) \\
= & \left(\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k \right) \boldsymbol{C}_{\boldsymbol{k}}^{-} \left(\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k \right)^T + \boldsymbol{K}_k \boldsymbol{V}_k \boldsymbol{K}_{\boldsymbol{k}}{}^T && (4.2.8)
\end{aligned}$$

The expression 4.2.7 and expression 4.2.8 are general for the updated error covariance matrix. They can be used for any type of blending factor $\boldsymbol{K}_{\boldsymbol{k}}$ in equation 4.2.5. The prior

---

[*] Depending on the derivation, this mixing factor $\boldsymbol{K}_{\boldsymbol{k}}$ could have different forms and values. In this thesis, only Kalman gains will be introduced.

covariance matrix $\boldsymbol{C_k^-}$, the measurement error $\boldsymbol{V_k}$ and the measurement matrix $\boldsymbol{H_k}$ are already known, such that $\boldsymbol{C_k}$ is only a function of $\boldsymbol{K_k}$. Sine the trace of matrix $\boldsymbol{C_k}$ is the sum of mean-square errors, we could minimize it by asking its derivative with respect to $\boldsymbol{K_k}$ equal to zero. By applying formula A.2.1 and formula A.2.2 in Appendix A.2 into $\boldsymbol{C_k}$ expansion 4.2.7 in which all the covariance matrices are symmetric, we have

$$\begin{aligned}\frac{d\left[tr(\boldsymbol{C_k})\right]}{d\boldsymbol{K_k}} &= -\boldsymbol{C_k^-}\boldsymbol{H_k}^T - \left(\boldsymbol{H_k}\boldsymbol{C_k^-}\right)^T + 2\boldsymbol{K_k}\left(\boldsymbol{H_k}\boldsymbol{C_k^-}\boldsymbol{H_k}^T + \boldsymbol{V_k}\right) \\ &= -2\boldsymbol{C_k^-}\boldsymbol{H_k}^T + 2\boldsymbol{K_k}\left(\boldsymbol{H_k}\boldsymbol{C_k^-}\boldsymbol{H_k}^T + \boldsymbol{V_k}\right) \qquad (4.2.9)\end{aligned}$$

To minimize the trace of $\boldsymbol{C_k}$, we require

$$\frac{d\left[tr(\boldsymbol{C_k})\right]}{d\boldsymbol{K_k}} = 0 \qquad (4.2.10)$$

Finally the Kalman gain reads

$$\boldsymbol{K_k} = \boldsymbol{C_k^-}\boldsymbol{H_k}^T\left(\boldsymbol{H_k}\boldsymbol{C_k^-}\boldsymbol{H_k}^T + \boldsymbol{V_k}\right)^{-1} \qquad (4.2.11)$$

Substituting equation 4.2.11 into equation 4.2.7, the updated covariance matrix could be written with the optimal gain (Kalman gain $\boldsymbol{K_k}$) as follows:

$$\begin{aligned}\boldsymbol{C_k} &= \boldsymbol{C_k^-} - \boldsymbol{C_k^-}\boldsymbol{H_k}^T\left(\boldsymbol{H_k}\boldsymbol{C_k^-}\boldsymbol{H_k}^T + \boldsymbol{V_k}\right)^{-1}\boldsymbol{H_k}\boldsymbol{C_k^-} \qquad (4.2.12) \\ &= \boldsymbol{C_k^-} - \boldsymbol{K_k}\left(\boldsymbol{H_k}\boldsymbol{C_k^-}\boldsymbol{H_k}^T + \boldsymbol{V_k}\right)\boldsymbol{K_k}^T \qquad (4.2.13) \\ &= \left(\boldsymbol{I} - \boldsymbol{K_k}\boldsymbol{H_k}\right)\boldsymbol{C_k^-} \qquad (4.2.14)\end{aligned}$$

Note that equations (4.2.12-4.2.14) are only valid for the optimal gain which is derived from MMSE method. All the equations to calculate $\boldsymbol{C_k}$ are identical, but some of them may be more efficient or easier to be processed in a given engineering condition. After this step, the current state and its error covariance matrix are figured out, such that the Kalman filter is ready to process the next state. Because the process noise $\boldsymbol{w_k}$ has zero mean and is independent to all the other noises $\boldsymbol{w_i}$ in which $i \neq k$, the prediction of the next state vector $\boldsymbol{\hat{\theta}_{k+1}^-}$ could be written as below

$$\boldsymbol{\hat{\theta}_{k+1}^-} = \boldsymbol{F_k}\boldsymbol{\hat{\theta}_k} \qquad (4.2.15)$$

The related error covariance matrix of the predicted state could be written as

$$\boldsymbol{C}_{k+1}^{-} = \left\langle \left(\boldsymbol{\theta}_{k+1} - \hat{\boldsymbol{\theta}}_{k+1}^{-}\right) \left(\boldsymbol{\theta}_{k+1} - \hat{\boldsymbol{\theta}}_{k+1}^{-}\right)^{T} \right\rangle \tag{4.2.16}$$

Considering that the process noise $\boldsymbol{w_k}$ and the measurement noise $\boldsymbol{v_k}$ have no correlation and substituting equation 4.2.2 and equation 4.2.15 into equation 4.2.16, we have

$$\begin{aligned}
\boldsymbol{C}_{k+1}^{-} &= \left\langle \left[\boldsymbol{F}_k \left(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^{-}\right) + \boldsymbol{w}_k\right] \left[\boldsymbol{F}_k \left(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^{-}\right) + \boldsymbol{w}_k\right]^{T} \right\rangle \\
&= \boldsymbol{F}_k \left\langle \left(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k\right) \left(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k\right)^{T} \right\rangle \boldsymbol{F}_k{}^{T} + \boldsymbol{Q}_k \\
&= \boldsymbol{F}_k \boldsymbol{C}_k \boldsymbol{F}_k{}^{T} + \boldsymbol{Q}_k
\end{aligned} \tag{4.2.17}$$

Once we have all the formulae, the Kalman filter recursive loop could be summarized in figure 4.2. One important trick which has to be emphasized is that the general expression 4.2.6 to update error covariance matrix $\boldsymbol{C_k}$ is computational superior that the one shown in figure 4.2, due to the fact that the general expression is naturally symmetric to avoid the possible round-off errors and to overcome some problems coming from suboptimal gains and incorrect $\boldsymbol{Q_k}$ or $\boldsymbol{V_k}$[53, 55]. Compared with the least squares algorithm, the inputs of Kalman filter have to include the prior estimation $\hat{\boldsymbol{\theta}}_{\boldsymbol{0}}^{-}$ of state vector and its covariance matrix $\boldsymbol{C}_{\boldsymbol{0}}^{-}$. This fact leads to the shortcoming of Kalman filter, that an improper prediction at the beginning could deteriorate the final results.

4.2.1.1 Alternative form of the Kalman Filter

The Kalman Filter loop shown in figure 4.2 can be manipulated with a variety of forms[53, 55, 56, 57]. When the inverses of matrices $\boldsymbol{C}_{\boldsymbol{k}}^{-}$, $\boldsymbol{C_k}$ and $\boldsymbol{V_k}$ exist or can be calculated, the following alternative form of the Kalman Filter illustrated in figure 4.3 could be convenient in computation. Different from the normal Kalman Filter procedure, the first goal is to compute $\boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{-1}}$ rather than the Kalman gain $\boldsymbol{K}_k$. Recall that the error covariance matrix is given by equation 4.2.12.

$$\boldsymbol{C}_k = \boldsymbol{C}_k^{-} - \boldsymbol{C}_k^{-} \boldsymbol{H}_k{}^{T} \left(\boldsymbol{H}_k \boldsymbol{C}_k^{-} \boldsymbol{H}_k{}^{T} + \boldsymbol{V}_k\right)^{-1} \boldsymbol{H}_k \boldsymbol{C}_k^{-} \tag{4.2.18}$$

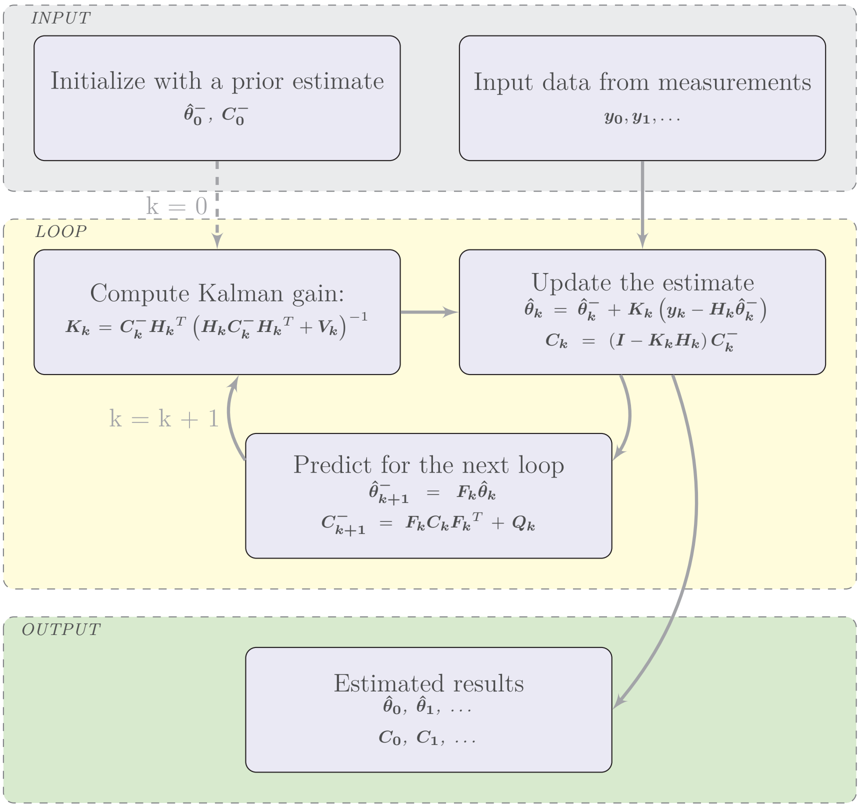


Figure 4.2: Kalman Filter Loop. The line refers to the data flow which contains all the results calculated by the previous step. The dashed line indicates that $\hat{\boldsymbol{\theta}}_0^-$ and $\boldsymbol{C}_0^-$ come from a model or a proper prediction rather than from the measurements, and these initialized values will be only used for the first recursion.

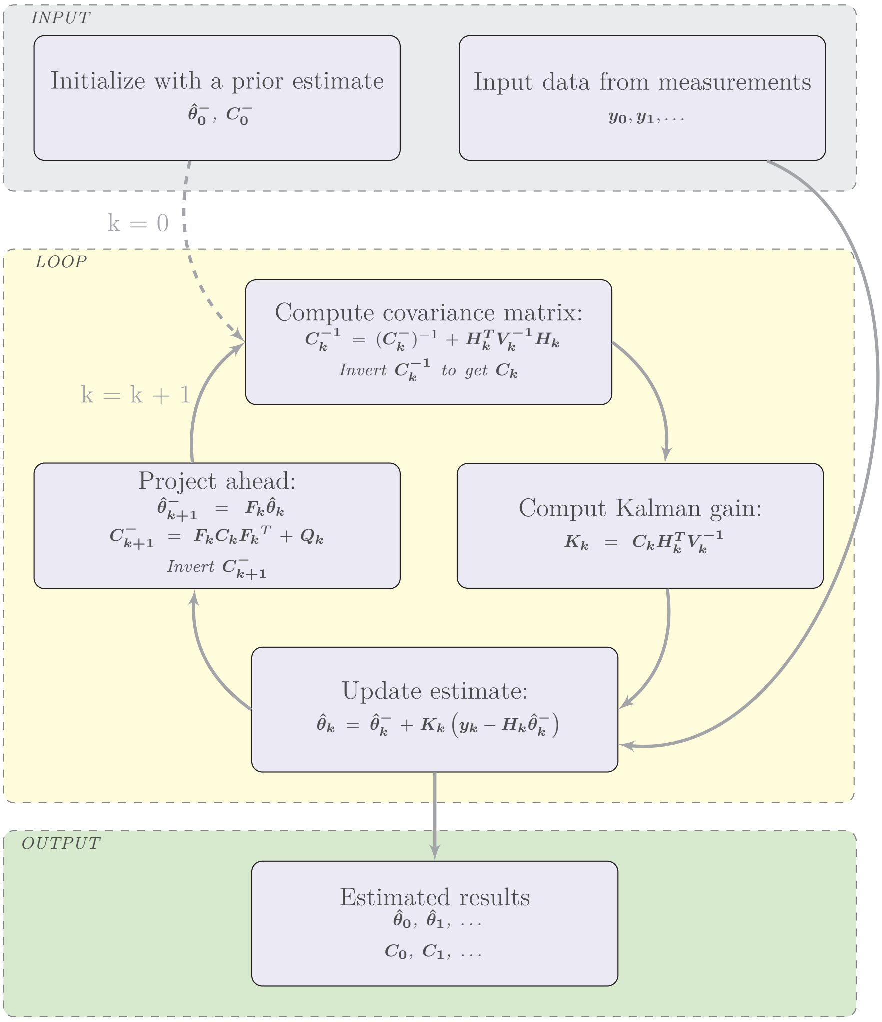


Figure 4.3: Alternative Kalman Filter Loop. The line refers to the data flow which contains all the results calculated by the previous step. The dashed line indicates that $\hat{\boldsymbol{\theta}}_0^-$ and $\boldsymbol{C}_0^-$ come from a model or a proper prediction rather than from the measurements, and these initialized values will be only used for the first recursion.

Using the matrix inversion lemma A.4.1 as shown in the justification of Equation A.4.8, it is straightforward to prove that:

$$\boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{-1}} = (\boldsymbol{C}_{\boldsymbol{k}}^{-})^{-1} + \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k \tag{4.2.19}$$

The following procedure shows that the product of the right sides of equation 4.2.18 and equation 4.2.19 reduces to the identity matrix:

$$\begin{aligned}
&\left[\boldsymbol{C}_{\boldsymbol{k}}^{-} - \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T + \boldsymbol{V}_k\right)^{-1} \boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-}\right] \left[(\boldsymbol{C}_{\boldsymbol{k}}^{-})^{-1} + \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k\right] \\
&= \boldsymbol{I} + \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k - \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T + \boldsymbol{V}_k\right)^{-1} \\
&\qquad \left[\boldsymbol{H}_k + \boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k\right] \\
&= \boldsymbol{I} + \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T + \boldsymbol{V}_k\right)^{-1} \left[\left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T + \boldsymbol{V}_k\right) \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k\right. \\
&\qquad \left. - \boldsymbol{H}_k - \boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k\right] \\
&= \boldsymbol{I}
\end{aligned}$$

Since $\boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{-1}}$ could be calculated before $\boldsymbol{K}_{\boldsymbol{k}}$, it is possible to derive the alternative expression for Kalman gain to reduce computing time. Inserting $\boldsymbol{C}_k \boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{-1}}$ and $\boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{V}_{\boldsymbol{k}}$ into equation 4.2.11, we have

$$\begin{aligned}
\boldsymbol{K}_k &= \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T + \boldsymbol{V}_k\right)^{-1} \\
&= \boldsymbol{C}_k \boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{V}_k \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T + \boldsymbol{V}_k\right)^{-1} \\
&= \boldsymbol{C}_k \boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} + \boldsymbol{I}\right)^{-1}
\end{aligned}$$

Substituting equation 4.2.19 into this $\boldsymbol{K}_{\boldsymbol{k}}$ expression above, we obtain

$$\begin{aligned}
\boldsymbol{K}_k &= \boldsymbol{C}_k \left[(\boldsymbol{C}_{\boldsymbol{k}}^{-})^{-1} + \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k\right] \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} + \boldsymbol{I}\right)^{-1} \\
&= \boldsymbol{C}_k \left(\boldsymbol{I} + \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-}\right) \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} + \boldsymbol{I}\right)^{-1} \\
&= \boldsymbol{C}_k \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \left(\boldsymbol{I} + \boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}}\right) \left(\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{k}}^{-} \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} + \boldsymbol{I}\right)^{-1} \\
&= \boldsymbol{C}_k \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}}
\end{aligned} \tag{4.2.20}$$

Once the Kalman gain is figured out, the state vector $\boldsymbol{\theta_k}$ could be updated by using Equation 4.2.5. An alternative form expressed with covariance matrix rather than the Kalman gain is also possible. Substituting Equation 4.2.19 and Equation 4.2.20 into Equation 4.2.5 leads to

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} &= \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} + \boldsymbol{C}_k \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \left(\boldsymbol{y_k} - \boldsymbol{H}_k \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-}\right) \\ &= \boldsymbol{C}_k \left[\left(\boldsymbol{C}_k^{-}\right)^{-1} \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} + \boldsymbol{H}_k{}^T \boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}} \boldsymbol{y_k}\right] \qquad (4.2.21)\end{aligned}$$

The main expressions in the alternative Kalman Filter have now been derived. In the following prediction step after Kalman gain computation and estimate updating, similar calculations will be processed as the normal Kalman Filter but one more task to invert the predicted covariance matrix $\boldsymbol{C}_{\boldsymbol{k+1}}^{-}$.

Compared with the normal Kalman Filter, this alternative could start with an infinite prior covariance matrix. On the other hand two $((m \times m))$ matrix inversions can not be avoided, such that when the state vector has a large dimension, the alternative Kalman filter will not be as fast as the regular Kalman procedure which contains the matrix inversions with the dimension $(n \times n)$ that is correlated to the measurement vector $\boldsymbol{y_k}$. In charged particle tracking under magnetic field, state vector has the dimension $m = 5$ while the measurement vector is in the order of $n = 3$.

4.2.2 Derivations using conditional probability density

The key ideal of Kalman filter is to minimise the mean-square error. Except the derivation of using the creative Equation 4.2.5 to update the state vector in the previous section, A kalman filter can also be derived from the point of view of conditional probability density, which introduces Equation 4.2.5 naturally.

The choice of the estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ should minimize the mean-square error in Equation 4.2.6 on condition of the given data samples. The Mean Square Error (MSE) is recalled and expanded as below:

$$\begin{aligned}
&\left\langle \left(\boldsymbol{\theta_k} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}\right)\left(\boldsymbol{\theta_k} - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}\right)^T \right\rangle \\
&= \left\langle \left(\boldsymbol{\theta_k}\boldsymbol{\theta_k}^T - \boldsymbol{\theta_k}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^T - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}\boldsymbol{\theta_k}^T + \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^T\right)\right\rangle \\
&= \left\langle \boldsymbol{\theta_k}\boldsymbol{\theta_k}^T \right\rangle - \langle \boldsymbol{\theta_k} \rangle \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^T - \hat{\boldsymbol{\theta}}_{\boldsymbol{k}} \left\langle \boldsymbol{\theta_k}^T \right\rangle + \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^T \\
&= \left\langle \boldsymbol{\theta_k}\boldsymbol{\theta_k}^T \right\rangle + \left(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \langle \boldsymbol{\theta_k} \rangle\right)\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \langle \boldsymbol{\theta_k} \rangle\right)^T - \langle \boldsymbol{\theta_k} \rangle \langle \boldsymbol{\theta_k} \rangle^T
\end{aligned} \tag{4.2.22}$$

Since only the middle term in the right side of Equation 4.2.22 depends on $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$, the MMSE method requires

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} = \left\langle \boldsymbol{\theta_k} | \boldsymbol{y}_{\boldsymbol{k}}^* \right\rangle \tag{4.2.23}$$

where $\boldsymbol{y}_{\boldsymbol{k}}^*$ stands for all the measurements up to the k'th state, that is the stream $\boldsymbol{y_0}$, $\boldsymbol{y_1}$, ..., $\boldsymbol{y_k}$. For a given data set, the state vector $\boldsymbol{\theta_k}$ in Equation 4.2.6 is implicitly on condition of $\boldsymbol{y}_{\boldsymbol{k}}^*$. Though Equation 4.2.23 is derived for the update step in Kalman filter, the same procedure could be applied in both the prediction and the smoothing.

Assuming all the noises and estimates have Gaussian distribution, The task to find the optimal estimate is now replaced by writing the explicit form of the probability density function of $\boldsymbol{\theta_k}$ conditioned on the measurements set $\boldsymbol{y}_{\boldsymbol{k}}^*$

$$p(\boldsymbol{\theta_k}|\boldsymbol{y}_{\boldsymbol{k}}^*) = \mathcal{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}, \boldsymbol{C_k}\right) \tag{4.2.24}$$

Kalman filter is started with the prior estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^-$ and its related covariance matrix $\boldsymbol{C}_{\boldsymbol{k}}^-$ which are predicted from the data stream $\boldsymbol{y}_{\boldsymbol{k-1}}^*$ when $k \geq 1$ or from some knowledges used for the initialization. The state density function on condition of $\boldsymbol{y}_{\boldsymbol{k-1}}^*$ reads

$$p(\boldsymbol{\theta_k}|\boldsymbol{y}_{\boldsymbol{k-1}}^*) = \mathcal{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^-, \boldsymbol{C}_{\boldsymbol{k}}^-\right) \tag{4.2.25}$$

The measurement model in Equation 4.2.3 connects the current state $\boldsymbol{\theta_k}$ and its measurement $\boldsymbol{y_k}$ and leads to the density function below:

$$p(\boldsymbol{y_k}|\boldsymbol{y}_{\boldsymbol{k-1}}^*) = \mathcal{N}\left(\boldsymbol{H_k}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^-, \boldsymbol{H_k}\boldsymbol{C}_{\boldsymbol{k}}^-\boldsymbol{H_k}^T + \boldsymbol{V_k}\right) \tag{4.2.26}$$

Since the measurement noise $\boldsymbol{v_k}$ is independent from the state $\boldsymbol{\theta_k}$, the joint Gaussian distribution of $\boldsymbol{y_k}$ and $\boldsymbol{\theta_k}$ conditioned on $\boldsymbol{y^*_{k-1}}$ could be written as follows

$$p(\boldsymbol{\theta_k}, \boldsymbol{y_k} | \boldsymbol{y^*_{k-1}}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\hat{\theta}^-_k} \\ \boldsymbol{H_k \hat{\theta}^-_k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{C^-_k} & \boldsymbol{C^-_k H^T_k} \\ \boldsymbol{H_k C^-_k} & \boldsymbol{H_k C^-_k H^T_k} + \boldsymbol{V_k} \end{bmatrix}\right) \tag{4.2.27}$$

By using lemma A.3.4 in Appendix A.3 into Equation 4.2.27, one can find the probability density function of $\boldsymbol{\theta_k}$ conditioned on $\boldsymbol{y^*_k}$

$$\begin{aligned} p(\boldsymbol{\theta_k} | \boldsymbol{y^*_k}) &= p(\boldsymbol{\theta_k} | \boldsymbol{y_k}, \boldsymbol{y^*_{k-1}}) \\ &= \mathcal{N}\Big(\boldsymbol{\hat{\theta}^-_k} + \boldsymbol{C^-_k H^T_k} \left(\boldsymbol{H_k C^-_k H^T_k} + \boldsymbol{V_k}\right)^{-1} \left(\boldsymbol{y_k} - \boldsymbol{H_k \hat{\theta}^-_k}\right), \\ &\qquad \boldsymbol{C^-_k} - \boldsymbol{C^-_k H^T_k} \left(\boldsymbol{H_k C^-_k H^T_k} + \boldsymbol{V_k}\right)^{-1} \boldsymbol{H_k C^-_k}\Big) \end{aligned} \tag{4.2.28}$$

Comparing the relevant terms of Equation 4.2.24 and Equation 4.2.28 leads to the update equations

$$\boldsymbol{\hat{\theta}_k} = \boldsymbol{\hat{\theta}^-_k} + \boldsymbol{C^-_k H^T_k} \left(\boldsymbol{H_k C^-_k H^T_k} + \boldsymbol{V_k}\right)^{-1} \left(\boldsymbol{y_k} - \boldsymbol{H_k \hat{\theta}^-_k}\right) \tag{4.2.29}$$

$$\boldsymbol{C_k} = \boldsymbol{C^-_k} - \boldsymbol{C^-_k H^T_k} \left(\boldsymbol{H_k C^-_k H^T_k} + \boldsymbol{V_k}\right)^{-1} \boldsymbol{H_k C^-_k} \tag{4.2.30}$$

Considering the definition of Kalman gain in Equation 4.2.11, the state and its covariance matrix update equation are exactly what we have derived in the previous section about the normal Kalman Filter.

4.2.3 The figure of merit in tracking

In generalized least squares fitting algorithm, track parameters are optimised by minimising the global chi-square function. Consequently, a track will always have a chi-square value to tell the goodness of fit, which is useful in practice to chose good tracks by setting a threshold. Kalman filter can construct a similar chi-square value.

After the filtering process, the residual $\boldsymbol{r_k}$ of the current measurement could be written:

$$\boldsymbol{r_k} \equiv \boldsymbol{y_k} - \boldsymbol{H_k \hat{\theta}_k} \tag{4.2.31}$$

Then by using state update Equation 4.2.5 and the alternative Kalman gain Equation

4.2.20 the covariance matrix $\boldsymbol{R_k}$ of the residual could be derived as follows:

$$\begin{aligned}
\boldsymbol{R_k} &= \left\langle \boldsymbol{r_k} \boldsymbol{r_k^T} \right\rangle \\
&= \left\langle \left(\boldsymbol{y_k} - \boldsymbol{H_k}\boldsymbol{\hat{\theta}_k}\right)\left(\boldsymbol{y_k} - \boldsymbol{H_k}\boldsymbol{\hat{\theta}_k}\right)^T \right\rangle \\
&= \boldsymbol{V_k} - \boldsymbol{V_k}\boldsymbol{K_k^T}\boldsymbol{H_k^T} - \boldsymbol{H_k}\boldsymbol{K_k}\boldsymbol{V_k} + \boldsymbol{H_k}\boldsymbol{C_k}\boldsymbol{H_k^T} \\
&= \boldsymbol{V_k} - \boldsymbol{H_k}\boldsymbol{C_k}\boldsymbol{H_k^T} && (4.2.32) \\
&= \left(\boldsymbol{I} - \boldsymbol{H_k}\boldsymbol{K_k}\right)\boldsymbol{V_k} && (4.2.33)
\end{aligned}$$

The chi-square increment in filtering is then given by

$$\chi_F^2 = \boldsymbol{r_k^T}\boldsymbol{R_k^{-1}}\boldsymbol{r_k} \tag{4.2.34}$$

The global chi-square until the current residual $\boldsymbol{r_k}$ reads:

$$\begin{aligned}
\chi_0^2 &= \boldsymbol{r_0^T}\boldsymbol{R_0^{-1}}\boldsymbol{r_0} \\
\chi_k^2 &= \chi_{k-1}^2 + \boldsymbol{r_k^T}\boldsymbol{R_k^{-1}}\boldsymbol{r_k}, \quad k \in \mathbb{Z}, k \geq 1
\end{aligned}$$

in which, χ_{k-1}^2 is the value calculated for the previous state when $k \geq 1$.

In track finding, one important goal it to get rid of the outliers from the track. Those outliers are some hits that do not belong to the reconstructed track. The global least squares fit could reject hits with large residuals only when multiple scattering is not dominating. Nevertheless, when multiple scattering could not be ignored, The kalman filter and smoother could overcome this disadvantage of the global fitting by using the chi-square of predictions[58]. Like Equation 4.2.31, the residual of prediction $\boldsymbol{r_k^-}$ is defined as

$$\boldsymbol{r_k^-} \equiv \boldsymbol{y_k} - \boldsymbol{H_k}\boldsymbol{\hat{\theta}_k^-} \tag{4.2.35}$$

As the current measurement $\boldsymbol{y_k}$ is independent on the predicted state $\boldsymbol{\hat{\theta}_k^-}$, it is straightforward to write the related covariance

$$\begin{aligned}\boldsymbol{R_k^-} &= \left\langle \left(\boldsymbol{r_k^-}\right)\left(\boldsymbol{r_k^-}\right)^T \right\rangle \\ &= \left\langle \left(\boldsymbol{y_k} - \boldsymbol{H}_k\boldsymbol{\hat{\theta}_k^-}\right)\left(\boldsymbol{y_k} - \boldsymbol{H}_k\boldsymbol{\hat{\theta}_k^-}\right)^T \right\rangle \\ &= \boldsymbol{V_k} + \boldsymbol{H_k}\boldsymbol{C_k^-}\boldsymbol{H_k^T} \qquad (4.2.36)\end{aligned}$$

Then the chi-square of the prediction reads

$$\chi_P^2 = (\boldsymbol{r_k^-})^T(\boldsymbol{R_k^-})^{-1}\boldsymbol{r_k^-} \qquad (4.2.37)$$

It is easy to show that the chi-square of prediction is equivalent to the filtered chi-square. Substituting the update Equation 4.2.5 into the residual of Equation 4.2.31 and choosing Equation 4.2.33 of the residual covariance, the chi-square of filtering yields

$$\begin{aligned}\chi_F^2 &= (\boldsymbol{r_k^-})^T\left(\boldsymbol{I} - \boldsymbol{H}_k\boldsymbol{K_k}\right)^T\boldsymbol{V_k^{-1}}\left(\boldsymbol{I} - \boldsymbol{H}_k\boldsymbol{K_k}\right)^{-1}\left(\boldsymbol{I} - \boldsymbol{H}_k\boldsymbol{K}_k\right)\boldsymbol{r_k^-} \\ &= (\boldsymbol{r_k^-})^T\left(\boldsymbol{I} - \boldsymbol{H}_k\boldsymbol{K_k}\right)^T\boldsymbol{V_k^{-1}}\boldsymbol{r_k^-}\end{aligned}$$

Using the Kalman gain in Equation 4.2.11 and considering the symmetry of all the covariance matrices lead to

$$\begin{aligned}\chi_F^2 &= (\boldsymbol{r_k^-})^T\left(\boldsymbol{V_k} + \boldsymbol{H}_k\boldsymbol{C_k^-}\boldsymbol{H_k^T}\right)^{-1}\boldsymbol{V}_k\boldsymbol{V_k^{-1}}\boldsymbol{r_k^-} \\ &= \boldsymbol{V_k} + \boldsymbol{H_k}\boldsymbol{C_k^-}\boldsymbol{H_k^T} \\ &= \chi_P^2\end{aligned}$$

It means that one can do the outlier elimination before updating the current state by using the predicted chi-square χ_P^2. For the hit or the measurement $\boldsymbol{y_k}$ belongs to the reconstructed track with a well known covariance matrix, χ_F^2 follows the χ^2 distribution whose degree of freedom is equal to the dimension of $\boldsymbol{y_k}$. Therefore, a hit could be rejected when its χ_P^2 is larger than a given cut value. Once a Kalman smoother is also applied, there is also a similar smoothed chi-square to take on this work and to offer a better performance to overcome noise hits. Nevertheless one has to refit the track if a hit is rejected in the smoothing process.

4.2.4 Smoothing

One may notice that a local recursive algorithm just takes the previous measurements to estimate the current state, such that only the final state is dealt with all the data we have. To improve the estimation of the specified state or all the states, a technique named smoothing is often used to combine all the information from the whole data samples to get the optimal estimates. There are three categories of problems in smoothing[56]

- *Fixed-interval smoothing* is used to deal with the measurements with fixed data interval or data span (i.e., the number of all states is fixed). In this off-line smoother, the optimal estimate of a certain state or all the states could be found.
- *Fixed-point smoothing* will only estimate a single fixed state with continuing updated data. This smoother could be used for on-line data smoothing and the first state could be chosen as the fixed-point to be optimised with the new measurements just acquired.
- *Fixed-lag smoothing* is to smooth the state which has a fixed distance (a fixed number of states away) from the current state. This smoother could also be used in on-line data flow. It is similar to the fixed-interval smoothing since it also incorporates a forward filtering sweep and a backward smoothing sweep with only a fixed number of states instead of all the previous states.

In the tracking problem of charged particles, track segments are often reconstructed to constitute a full trajectory, which means the fixed-interval smoother could take its advantages of finding the optimal estimates for all the hits inside a track and of combining the target segment to the neighbour segments from its two sides. Therefore, only the fixed-interval smoothing algorithm will be presented in this section. The details of the other two smoothers could be found in references [53, 55, 56, 59].

4.2.4.1 Expressions of the optimal smoother

The straightforward idea to use all the measurements to optimise the estimate of index k is to combine two filtered results $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ and $\hat{\boldsymbol{\theta}}^{-}_{\boldsymbol{k},\boldsymbol{b}}$ shown in Figure 4.4, in which $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ is the result of a forward filter started from the first measurement and $\hat{\boldsymbol{\theta}}^{-}_{\boldsymbol{k},\boldsymbol{b}}$ is a prediction, the outcome from the backward filter started at the last index T.

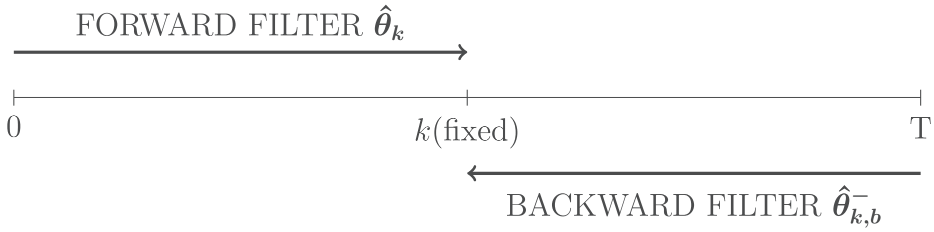


Figure 4.4: Smoothing based on the Forward and backward filters

To derive the necessary formulae of the backward filter, the required equations of the forward Kalman filter are recalled in Table 4.2. The backward filter loop is a recursion in

The model:

$$\boldsymbol{\theta_{k+1}} = \boldsymbol{F_k \theta_k} + \boldsymbol{w_k}$$
$$\boldsymbol{y_k} = \boldsymbol{H_k \theta_k} + \boldsymbol{v_k}$$

The update:

$$\boldsymbol{C_k^{-1}} = (\boldsymbol{C_k^-})^{-1} + \boldsymbol{H_k^T V_k^{-1} H_k}$$
$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} = \boldsymbol{C_k}\left[\left(\boldsymbol{C_k^-}\right)^{-1}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-} + \boldsymbol{H_k}^T\boldsymbol{V_k^{-1} y_k}\right]$$

The prediction:

$$\hat{\boldsymbol{\theta}}^{-}_{\boldsymbol{k+1}} = \boldsymbol{F_k}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$$
$$\boldsymbol{C^-_{k+1}} = \boldsymbol{F_k C_k F_k}^T + \boldsymbol{Q_k}$$

Table 4.2: Forward Kalman filter

reversed direction, such that the model will be changed as below:

$$\boldsymbol{\theta_{k,b}} = \boldsymbol{F_k}^{-1}\boldsymbol{\theta_{k+1,b}} - \boldsymbol{F_k}^{-1}\boldsymbol{w_k}$$

Since the covariance matrix could be updated by the prediction directly, its update formula shares the same form with the forward filter:

$$\boldsymbol{C_k^{-1}} = (\boldsymbol{C_k^-})^{-1} + \boldsymbol{H_k^T V_k^{-1} H_k} \tag{4.2.38}$$

$$\boldsymbol{C_{k,b}^{-1}} = (\boldsymbol{C_{k,b}^-})^{-1} + \boldsymbol{H_k^T V_k^{-1} H_k} \tag{4.2.39}$$

The prediction equation for the previous state and its covariance could be derived from the model with similar processes as the forward Kalman filter:

$$\boldsymbol{\hat{\theta}_{k,b}^-} = \boldsymbol{F_k}^{-1}\boldsymbol{\hat{\theta}_{k+1,b}} \tag{4.2.40}$$

$$\begin{aligned}\boldsymbol{C_{k,b}^-} &= \left\langle \left(\boldsymbol{\theta_{k,b}} - \boldsymbol{\hat{\theta}_{k,b}^-}\right)\left(\boldsymbol{\theta_{k,b}} - \boldsymbol{\hat{\theta}_{k,b}^-}\right)^T \right\rangle \\ &= \left\langle \boldsymbol{F_k}^{-1}\left[\left(\boldsymbol{\theta_{k+1,b}} - \boldsymbol{\hat{\theta}_{k+1,b}}\right) - \boldsymbol{w_k}\right]\right. \\ &\qquad \left.\left[\left(\boldsymbol{\theta_{k+1,b}} - \boldsymbol{\hat{\theta}_{k+1,b}}\right) - \boldsymbol{w_k}\right]^T \left(\boldsymbol{F_k}^{-1}\right)^T \right\rangle \\ &= \boldsymbol{F_k}^{-1}\left(\boldsymbol{C_{k+1,b}} + \boldsymbol{Q_k}\right)(\boldsymbol{F_k}^{-1})^T \end{aligned} \tag{4.2.41}$$

With the predicted covariance matrix above and the filtered covariance of Equation 4.2.39, the state vector in backward filtering could also be updated using Equation 4.2.21

$$\boldsymbol{\hat{\theta}_k} = \boldsymbol{C_{k,b}}\left[\left(\boldsymbol{C_{k,b}^-}\right)^{-1}\boldsymbol{\hat{\theta}_{k,b}^-} + \boldsymbol{H_k}^T\boldsymbol{V_k^{-1}}\boldsymbol{y_k}\right] \tag{4.2.42}$$

Finally, some important relations in backward filter is summarized in Table 4.3

The smoothed estimate we seek is a linear combination of $\boldsymbol{\hat{\theta}_k}$ and $\boldsymbol{\hat{\theta}_{k,b}^-}$. The reason to use prediction of the backward filter is on account of the fact that the data at index k is already taken into account in the forward filter. Using the same idea to construct Equation 4.2.5 in the state update step, we can write the smoothed state:

$$\boldsymbol{\hat{\theta}_k^s} = \boldsymbol{M}\boldsymbol{\hat{\theta}_k} + \boldsymbol{M'}\boldsymbol{\hat{\theta}_{k,b}^-} \tag{4.2.43}$$

The three estimates contains the true value $\boldsymbol{\theta_k}$ and the related estimation error $\boldsymbol{\tilde{\theta}_k^s}$, $\boldsymbol{\tilde{\theta}_k}$

The model:

$$\boldsymbol{\theta_{k,b}} = \boldsymbol{F_k}^{-1}\boldsymbol{\theta_{k+1,b}} - \boldsymbol{F_k}^{-1}\boldsymbol{w_k}$$

$$\boldsymbol{y_k} = \boldsymbol{H_k}\boldsymbol{\theta_k} + \boldsymbol{v_k}$$

The update:

$$\boldsymbol{C^{-1}_{k,b}} = (\boldsymbol{C^{-}_{k,b}})^{-1} + \boldsymbol{H^T_k}\boldsymbol{V^{-1}_k}\boldsymbol{H_k}$$

$$\boldsymbol{\hat{\theta}_k} = \boldsymbol{C_{k,b}}\left[\left(\boldsymbol{C^{-}_{k,b}}\right)^{-1}\boldsymbol{\hat{\theta}^{-}_{k,b}} + \boldsymbol{H_k}^T\boldsymbol{V^{-1}_k}\boldsymbol{y_k}\right]$$

The prediction:

$$\boldsymbol{\hat{\theta}^{-}_{k,b}} = \boldsymbol{F_k}^{-1}\boldsymbol{\hat{\theta}_{k+1,b}}$$

$$\boldsymbol{C^{-}_{k,b}} = \boldsymbol{F_k}^{-1}\left(\boldsymbol{C_{k+1,b}} + \boldsymbol{Q_k}\right)\left(\boldsymbol{F_k}^{-1}\right)^T$$

Table 4.3: Backward Kalman filter

and $\boldsymbol{\tilde{\theta}^{-}_{k,b}}$. Replacing the estimates with $\boldsymbol{\theta_k}$ and the relevant error, we have

$$\begin{aligned}\boldsymbol{\tilde{\theta}^{s}_k} &= \boldsymbol{M}(\boldsymbol{\tilde{\theta}_k} + \boldsymbol{\theta_k}) + \boldsymbol{M'}(\boldsymbol{\tilde{\theta}^{-}_{k,b}} + \boldsymbol{\tilde{\theta}_k}) - \boldsymbol{\theta_k} \\ &= \boldsymbol{M}\boldsymbol{\tilde{\theta}_k} + \boldsymbol{M'}\boldsymbol{\tilde{\theta}^{-}_{k,b}} + (\boldsymbol{M} + \boldsymbol{M'} - \boldsymbol{I})\boldsymbol{\theta_k}\end{aligned}$$

The unbiased estimation requires that the term containing $\boldsymbol{\theta_k}$ is zero. Therefore, it is straightforward to write the relation between $\boldsymbol{M}$ and $\boldsymbol{M'}$.

$$\boldsymbol{M'} = \boldsymbol{I} - \boldsymbol{M} \tag{4.2.44}$$

Substituting Equation 4.2.44 into Equation 4.2.43 yields

$$\boldsymbol{\hat{\theta}^{s}_k} = \boldsymbol{M}\boldsymbol{\hat{\theta}_k} + (\boldsymbol{I} - \boldsymbol{M})\boldsymbol{\hat{\theta}^{-}_{k,b}} \tag{4.2.45}$$

Therefore, the error covariance matrix of $\boldsymbol{\hat{\theta}^{s}_k}$ could be written in terms of the two independent matrices $\boldsymbol{C_k}$ and $\boldsymbol{C^-_{k,b}}$.

$$\begin{aligned}\boldsymbol{C^s_k} &= \left\langle \tilde{\boldsymbol{\theta}}^s_k \tilde{\boldsymbol{\theta}}^{s\,T}_k \right\rangle \\ &= \left\langle \left[\boldsymbol{M}\tilde{\boldsymbol{\theta}}_k + (\boldsymbol{I}-\boldsymbol{M})\tilde{\boldsymbol{\theta}}^-_{k,b}\right]\left[\boldsymbol{M}\tilde{\boldsymbol{\theta}}_k + (\boldsymbol{I}-\boldsymbol{M})\tilde{\boldsymbol{\theta}}^-_{k,b}\right]^T \right\rangle \\ &= \boldsymbol{M}\left\langle \tilde{\boldsymbol{\theta}}_k \tilde{\boldsymbol{\theta}}_k^T \right\rangle \boldsymbol{M}^T + (\boldsymbol{I}-\boldsymbol{M})\left\langle \tilde{\boldsymbol{\theta}}^-_{k,b}(\tilde{\boldsymbol{\theta}}^-_{k,b})^T \right\rangle (\boldsymbol{I}-\boldsymbol{M})^T \\ &= \boldsymbol{M}\boldsymbol{C_k}\boldsymbol{M}^T + (\boldsymbol{I}-\boldsymbol{M})\boldsymbol{C^-_{k,b}}(\boldsymbol{I}-\boldsymbol{M})^T \end{aligned} \tag{4.2.46}$$

Like the process to find the Kalman gain $\boldsymbol{K_k}$, we chose the matrix $\boldsymbol{M}$ to minimize the trace of $\boldsymbol{C^s_k}$ with the help of Equation A.2.2

$$\begin{aligned}\frac{d\left[tr(\boldsymbol{C^s_k})\right]}{d\boldsymbol{M}} &= 2\boldsymbol{M}\boldsymbol{C_k} + 2(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{C^-_{k,b}}(-\boldsymbol{I}) \\ &= 2\boldsymbol{M}\boldsymbol{C_k} + 2(\boldsymbol{M}-\boldsymbol{I})\boldsymbol{C^-_{k,b}} \\ &= 2\boldsymbol{M}(\boldsymbol{C_k}+\boldsymbol{C^-_{k,b}}) - 2\boldsymbol{C^-_{k,b}} \\ &= 0\end{aligned}$$

The form of matrix $\boldsymbol{M}$ is then written:

$$\boldsymbol{M} = \boldsymbol{C^-_{k,b}}(\boldsymbol{C_k}+\boldsymbol{C^-_{k,b}})^{-1} \tag{4.2.47}$$

The term $\boldsymbol{I}-\boldsymbol{M}$ is also given:

$$\begin{aligned}\boldsymbol{I}-\boldsymbol{M} &= \boldsymbol{I} - \boldsymbol{C^-_{k,b}}(\boldsymbol{C_k}+\boldsymbol{C^-_{k,b}})^{-1} \\ &= \boldsymbol{C_k}(\boldsymbol{C_k}+\boldsymbol{C^-_{k,b}})^{-1}\end{aligned} \tag{4.2.48}$$

It is convenient to start with the covariance matrix $\boldsymbol{C^s_k}$. Substituting Equation 4.2.47 and Equation 4.2.48 into Equation 4.2.46 leads to

$$\boldsymbol{C^s_k} = \boldsymbol{C^-_{k,b}}(\boldsymbol{C_k}+\boldsymbol{C^-_{k,b}})^{-1}\boldsymbol{C_k}(\boldsymbol{C_k}+\boldsymbol{C^-_{k,b}})^{-1}\boldsymbol{C^-_{k,b}} + (\text{swap } \boldsymbol{C^-_{k,b}} \text{ and } \boldsymbol{C_k})$$

To simplify the expression above, the matrix inversion lemma results in Equation A.4.3

and Equation A.4.5 are used in the following process:

$$\begin{aligned}
\boldsymbol{C}_k^s &= \boldsymbol{C}_{k,b}^-(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\left[\boldsymbol{C}_k{}^{-1} + (\boldsymbol{C}_{k,b}^-)^{-1}\right]^{-1} + \\
&\quad \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\left[\boldsymbol{C}_k{}^{-1} + (\boldsymbol{C}_{k,b}^-)^{-1}\right]^{-1} \\
&= \left[\boldsymbol{C}_k{}^{-1} + (\boldsymbol{C}_{k,b}^-)^{-1}\right]^{-1} && (4.2.49) \\
&= \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\boldsymbol{C}_{k,b}^- && (4.2.50) \\
&= \boldsymbol{C}_{k,b}^-(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\boldsymbol{C}_k && (4.2.51)
\end{aligned}$$

The other two identical expressions could be derived using Equation A.4.2 and Equation A.4.4.

$$\begin{aligned}
\boldsymbol{C}_k^s &= \boldsymbol{C}_k - \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\boldsymbol{C}_k && (4.2.52) \\
&= \boldsymbol{C}_{k,b}^- - \boldsymbol{C}_{k,b}^-(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\boldsymbol{C}_{k,b}^- && (4.2.53)
\end{aligned}$$

The smoothed covariance matrix $\boldsymbol{C}_k^s$ is calculated with all the information from the measurements and the theory, such that it should be better than any covariance estimated based on a part of the data. Inversion of the Equation 4.2.49 yields

$$(\boldsymbol{C}_k^s)^{-1} = \boldsymbol{C}_k{}^{-1} + (\boldsymbol{C}_{k,b}^-)^{-1}$$

The equation above shows $(\boldsymbol{C}_k^s)^{-1} \geq \boldsymbol{C}_k{}^{-1}$ and $(\boldsymbol{C}_k^s)^{-1} \geq (\boldsymbol{C}_{k,b}^-)^{-1}$. Consequently, the smoothed covariance matrix should satisfy $\boldsymbol{C}_k^s \leq \boldsymbol{C}_k$ and $\boldsymbol{C}_k^s \leq \boldsymbol{C}_{k,b}$.

Once $\boldsymbol{C}_k^s$ is determined, the form of the smoothed state will be derived naturally. Substituting Equation 4.2.47 and Equation 4.2.48 into Equation 4.2.45, and using Equation 4.2.50 and 4.2.51 in the derivation process, the smoothed state reads

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_k^s &= \boldsymbol{C}_{k,b}^-(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\hat{\boldsymbol{\theta}}_k + \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\hat{\boldsymbol{\theta}}_{k,b}^- \\
&= \left[\boldsymbol{I} + \boldsymbol{C}_k(\boldsymbol{C}_{k,b}^-)^{-1}\right]^{-1}\hat{\boldsymbol{\theta}}_k + \boldsymbol{C}_k^s(\boldsymbol{C}_{k,b}^-)^{-1}\hat{\boldsymbol{\theta}}_{k,b}^- && (4.2.54)
\end{aligned}$$

Considering the matrix inversion lemma result A.4.6 into equation 4.2.54 leads to

$$\begin{aligned}\hat{\boldsymbol{\theta}}_k^s &= \left[\boldsymbol{I} - \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\right]\hat{\boldsymbol{\theta}}_k + \boldsymbol{C}_k^s(\boldsymbol{C}_{k,b}^-)^{-1}\hat{\boldsymbol{\theta}}_{k,b}^- \\ &= (\boldsymbol{I} - \boldsymbol{K}_k^s)\hat{\boldsymbol{\theta}}_k + \boldsymbol{C}_k^s(\boldsymbol{C}_{k,b}^-)^{-1}\hat{\boldsymbol{\theta}}_{k,b}^- \end{aligned} \tag{4.2.55}$$

in which $\boldsymbol{K}_k^s$, the smoother gain at index k is defined by

$$\boldsymbol{K}_k^s \equiv \boldsymbol{C}_k(\boldsymbol{C}_{k,b}^-)^{-1}\left[I + \boldsymbol{C}_k(\boldsymbol{C}_{k,b}^-)^{-1}\right]^{-1} \tag{4.2.56}$$

The other forms of the gain $\boldsymbol{K}_k^s$ could be found by using Equation 4.2.50

$$\begin{aligned}\boldsymbol{K}_k^s &= \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1} \\ &= \boldsymbol{C}_k^s(\boldsymbol{C}_{k,b}^-)^{-1}\end{aligned} \tag{4.2.57}$$

Therefore, a more simplified expression for the smoothed state is written as

$$\hat{\boldsymbol{\theta}}_k^s = (\boldsymbol{I} - \boldsymbol{K}_k^s)\hat{\boldsymbol{\theta}}_k + \boldsymbol{K}_k^s\hat{\boldsymbol{\theta}}_{k,b}^- \tag{4.2.58}$$

The two-filter smoothing method is a good choice when there is only one target state to be estimated. Nevertheless, a flexible tracking algorithm requires more smoothed states in one trajectory. The following fixed-interval smoother will be presented to bring the convenience of computation based on a forward Kalman filter and a backward smother.

4.2.4.2 RTS fixed-interval smoothing

The derivation of RTS (Rauch, Tung, and Striebel) smoother is published in literatures [60, 61]. The idea of this smoother is to process only one forward Kalman filter and a subsequent backward recursive smoothing sweep. Since all the variables required by the smoothing loop are already calculated in the filtering algorithm and all the states could be smoothed, this smoother is tremendously convenient and is therefore widely used in track fitting. Figure 4.5 shows the schematic diagram of RTS fixed-interval smoothing. At the end of the normal Kalman filter sweeping, the state vector and its error covariance matrix are already well estimated as all the measurements have included in the estimation process, so the backward smoothing loop will start to optimise the estimate at state $T-1$.

A popular notation in reference to what used to describe conditional probability is

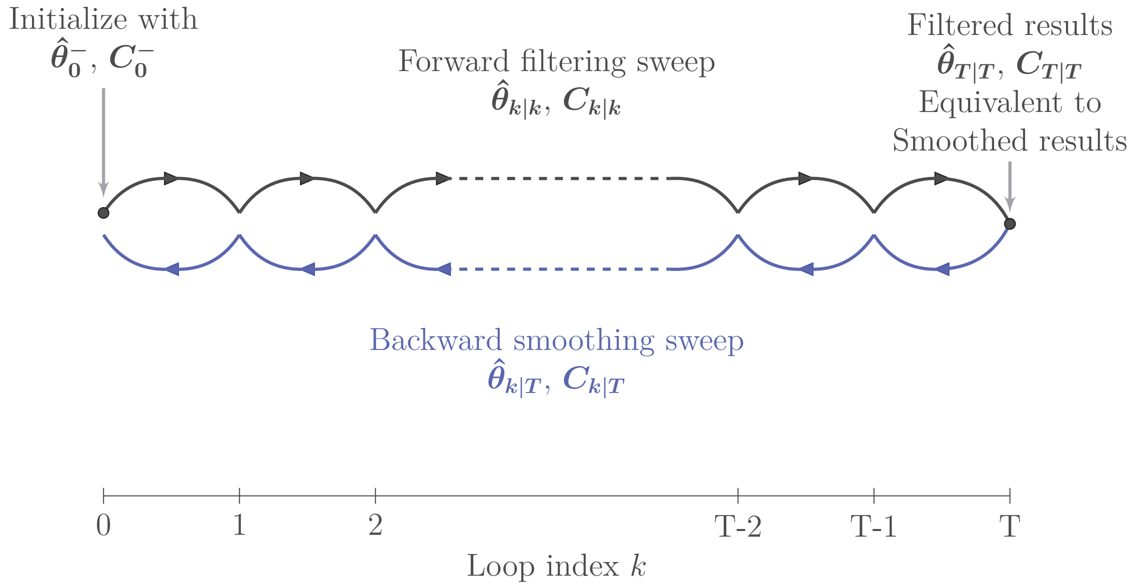


Figure 4.5: RTS fixed-interval Smoothing

used in Figure 4.5. $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}|\boldsymbol{T}}$ is the estimation at state k by using the measurements until the index T which means, in this case, all the data, so it is a smoothed state. $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}|\boldsymbol{k-1}}$ indicates a prediction at index k since the information from measurements between k and T is not taken into account, while $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}|\boldsymbol{k}}$ stands for the filtered state because of involving the data at the same index k. In smoothing, we prefer to use the measurements in the whole interval, so a smoothing state like $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}|\boldsymbol{k+1}}$ may be not interested in tracking. To avoid using $_{-1}$ in the initializer and to make the notation intuitive and straightforward, the superscript ^{s} and $^{-}$ are used instead of the conditional probability suffix. Table 4.4 shows the notation system preferred by this thesis and the corresponding notations in conditional probability style .

The recursive form of the estimated covariance matrix $\boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{s}}$ in RTS smoother is easy to derive from the expressions in forward-backward filtering smoother. To achieve the goal of only using the filtered variable and the previous smoothed covariance, Equation 4.2.52 is a good choice to start with.

$$\boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{s}} = \boldsymbol{C}_{\boldsymbol{k}} - \boldsymbol{C}_{\boldsymbol{k}}(\boldsymbol{C}_{\boldsymbol{k}} + \boldsymbol{C}_{\boldsymbol{k,b}}^{-})^{-1}\boldsymbol{C}_{\boldsymbol{k}}$$

The inverse term containing the backward filter element is to be replaced. Using Equation

This thesis	conditional style	Description
$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{-}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{k\|k-1}}$	the k'th state prediction
$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{k\|k}}$	the k'th state estimate of the filter
$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}^{\boldsymbol{s}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{k\|T}}$	the k'th state estimate of the smoother ($T > k$)
$\boldsymbol{C}_{\boldsymbol{k}}^{-}$	$\boldsymbol{C}_{\boldsymbol{k\|k-1}}$	covariance matrix prediction at k
$\boldsymbol{C}_{\boldsymbol{k}}$	$\boldsymbol{C}_{\boldsymbol{k\|k}}$	covariance matrix of the estimated state at k
$\boldsymbol{C}_{\boldsymbol{k}}^{\boldsymbol{s}}$	$\boldsymbol{C}_{\boldsymbol{k\|T}}$	covariance matrix of the smoothed state at k ($T > k$)

Table 4.4: Notation Elaboration in Kalman filter and smoother

4.2.41 and the Kalman filter $\boldsymbol{C}_{\boldsymbol{k+1}}^{-}$ update Equation 4.2.17, one can find

$$\begin{aligned}(\boldsymbol{C}_{\boldsymbol{k}} + \boldsymbol{C}_{\boldsymbol{k,b}}^{-})^{-1} &= \left[\boldsymbol{F}_{\boldsymbol{k}}{}^{-1}\left(\boldsymbol{F}_{\boldsymbol{k}}\boldsymbol{C}_{\boldsymbol{k}}\boldsymbol{F}_{\boldsymbol{k}}{}^{T} + \boldsymbol{C}_{\boldsymbol{k+1,b}} + \boldsymbol{Q}_{\boldsymbol{k}}\right)(\boldsymbol{F}_{\boldsymbol{k}}{}^{-1})^{T}\right]^{-1} \\ &= \left[\boldsymbol{F}_{\boldsymbol{k}}{}^{-1}\left(\boldsymbol{C}_{\boldsymbol{k+1}}^{-} + \boldsymbol{C}_{\boldsymbol{k+1,b}}\right)(\boldsymbol{F}_{\boldsymbol{k}}{}^{-1})^{T}\right]^{-1} \end{aligned} \tag{4.2.59}$$

where, $\boldsymbol{C}_{\boldsymbol{k+1,b}}$ could be expressed with the results from the forward filter and the previous smoothing. Substituting Equation 4.2.38 into Equation 4.2.39 to eliminate the term $\boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}}\boldsymbol{H}_{\boldsymbol{k}}$ and using Equation 4.2.49 yields

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{k+1,b}}{}^{-1} &= (\boldsymbol{C}_{\boldsymbol{k+1,b}}^{-})^{-1} + \boldsymbol{C}_{\boldsymbol{k+1}}{}^{-1} - (\boldsymbol{C}_{\boldsymbol{k+1}}^{-})^{-1} \\ &= \boldsymbol{C}_{\boldsymbol{k+1}}^{\boldsymbol{s}}{}^{-1} - (\boldsymbol{C}_{\boldsymbol{k+1}}^{-})^{-1} \end{aligned} \tag{4.2.60}$$

Substituting Equation 4.2.60 into Equation 4.2.59 and using matrix inverse lemma Equation A.4.7 lead to

$$\begin{aligned}&(\boldsymbol{C}_{\boldsymbol{k}} + \boldsymbol{C}_{\boldsymbol{k,b}}^{-})^{-1} \\ &= \boldsymbol{F}_{\boldsymbol{k}}{}^{T}\left\{\boldsymbol{C}_{\boldsymbol{k+1}}^{-} + \left[(\boldsymbol{C}_{\boldsymbol{k+1}}^{\boldsymbol{s}})^{-1} - (\boldsymbol{C}_{\boldsymbol{k+1}}^{-})^{-1}\right]^{-1}\right\}^{-1}\boldsymbol{F}_{\boldsymbol{k}} \\ &= \boldsymbol{F}_{\boldsymbol{k}}{}^{T}(\boldsymbol{C}_{\boldsymbol{k+1}}^{-})^{-1}(\boldsymbol{C}_{\boldsymbol{k+1}}^{-} - \boldsymbol{C}_{\boldsymbol{k+1}}^{\boldsymbol{s}})(\boldsymbol{C}_{\boldsymbol{k+1}}^{-})^{-1}\boldsymbol{F}_{\boldsymbol{k}} \end{aligned} \tag{4.2.61}$$

Finally, substituting Equation 4.2.61 into Equation 4.2.52 yields

$$\begin{aligned} \boldsymbol{C}_k^s &= \boldsymbol{C}_k - \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{C}_{k,b}^-)^{-1}\boldsymbol{C}_k \\ &= \boldsymbol{C}_k - \boldsymbol{A}_k(\boldsymbol{C}_{k+1}^- - \boldsymbol{C}_{k+1}^s)\boldsymbol{A}_k{}^T \\ &= \boldsymbol{C}_k + \boldsymbol{A}_k(\boldsymbol{C}_{k+1}^s - \boldsymbol{C}_{k+1}^-)\boldsymbol{A}_k{}^T \end{aligned} \tag{4.2.62}$$

in which, the smoothing gain $\boldsymbol{A}_k$ is defined by

$$\boldsymbol{A}_k \equiv \boldsymbol{C}_k\boldsymbol{F}_k{}^T(\boldsymbol{C}_{k+1}^-)^{-1} \tag{4.2.63}$$

$$= \boldsymbol{C}_k\boldsymbol{F}_k{}^T(\boldsymbol{F}_k\boldsymbol{C}_k\boldsymbol{F}_k{}^T + \boldsymbol{Q}_k)^{-1} \tag{4.2.64}$$

The update equation below of the smoothed state could be derived using equations in Section 4.2.4 of two-filter smoothing algorithm. Since its process is too complex, the derivation from the point of view of conditional probability will be introduced following the filter proof in Section 4.2.2.

$$\hat{\boldsymbol{\theta}}_k^s = \hat{\boldsymbol{\theta}}_k + \boldsymbol{A}_k(\hat{\boldsymbol{\theta}}_{k+1}^s - \hat{\boldsymbol{\theta}}_{k+1}^-) \tag{4.2.65}$$

Recalling Equation 4.2.24, we have the probability density of state $\boldsymbol{\theta}_k$ conditioned on the measurement data set $\boldsymbol{y}_k^*$:

$$p(\boldsymbol{\theta}_k|\boldsymbol{y}_k^*) = \mathcal{N}\left(\hat{\boldsymbol{\theta}}_k, \boldsymbol{C}_k\right) \tag{4.2.66}$$

Given the current state, the Kalman filter model leads to the probability density of the next state:

$$p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k) = \mathcal{N}\left(\boldsymbol{F}_k\hat{\boldsymbol{\theta}}_k, \boldsymbol{Q}_k\right) \tag{4.2.67}$$

Using lemma Equation A.3.2, the joint probability density of $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k+1}$ reads

$$p(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k+1}|\boldsymbol{y}_k^*) = \mathcal{N}\left(\begin{bmatrix} \hat{\boldsymbol{\theta}}_k \\ \boldsymbol{F}_k\hat{\boldsymbol{\theta}}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{C}_k & \boldsymbol{C}_k\boldsymbol{F}_k^T \\ \boldsymbol{F}_k\boldsymbol{C}_k & \boldsymbol{F}_k\boldsymbol{C}_k\boldsymbol{F}_k^T + \boldsymbol{Q}_k \end{bmatrix}\right) \tag{4.2.68}$$

Using lemma A.3.4 in Appendix A.3, $\boldsymbol{\theta}_k$ density conditioned on $\boldsymbol{\theta}_{k+1}$ and data set $\boldsymbol{y}_k^*$

could be written

$$p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1}, \boldsymbol{y}_k^*) = \mathcal{N}\left(\hat{\boldsymbol{\theta}}_k + \boldsymbol{C}_k\boldsymbol{F}_k^T\left(\boldsymbol{F}_k\boldsymbol{C}_k\boldsymbol{F}_k^T + \boldsymbol{Q}_k\right)^{-1}\left(\boldsymbol{\theta}_{k+1} - \boldsymbol{F}_k\hat{\boldsymbol{\theta}}_k\right), \right.$$
$$\left. \boldsymbol{C}_k - \boldsymbol{C}_k\boldsymbol{F}_k^T\left(\boldsymbol{F}_k\boldsymbol{C}_k\boldsymbol{F}_k^T + \boldsymbol{Q}_k\right)^{-1}\boldsymbol{F}_k\boldsymbol{C}_k\right) \tag{4.2.69}$$

With the given $\boldsymbol{\theta}_{k+1}$ and the measurements up to k, the current state $\boldsymbol{\theta}_k$ is independent of the future measurements. Therefore one can find:

$$p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1}, \boldsymbol{y}_T^*) = p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1}, \boldsymbol{y}_k^*)$$

Substituting Equation 4.2.64 into Equation 4.2.69 to use the auxiliary matrix $\boldsymbol{A}_k$ yields

$$p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1}, \boldsymbol{y}_T^*) = \mathcal{N}\left(\hat{\boldsymbol{\theta}}_k + \boldsymbol{A}_k\left(\boldsymbol{\theta}_{k+1} - \boldsymbol{F}_k\hat{\boldsymbol{\theta}}_k\right), \right.$$
$$\left. \boldsymbol{C}_k - \boldsymbol{A}_k\left(\boldsymbol{F}_k\boldsymbol{C}_k\boldsymbol{F}_k^T + \boldsymbol{Q}_k\right)\boldsymbol{A}_k{}^T\right) \tag{4.2.70}$$

Substituting the $\boldsymbol{C}_{k+1}^-$ identity 4.2.17 into Equation 4.2.70, one can get the simplified covariance term $\boldsymbol{C}_{k|\boldsymbol{\theta}_{k+1},\boldsymbol{y}_T^*}$ as follows

$$\begin{aligned} \boldsymbol{C}_{k|\boldsymbol{\theta}_{k+1},\boldsymbol{y}_T^*} &= \boldsymbol{C}_k - \boldsymbol{A}_k\left(\boldsymbol{F}_k\boldsymbol{C}_k\boldsymbol{F}_k^T + \boldsymbol{Q}_k\right)\boldsymbol{A}_k{}^T \\ &= \boldsymbol{C}_k - \boldsymbol{A}_k\boldsymbol{C}_{k+1}^-\boldsymbol{A}_k{}^T \end{aligned} \tag{4.2.71}$$

Assuming that the smoothed state $\hat{\boldsymbol{\theta}}_{k+1}^s$ and the related covariance $\boldsymbol{C}_{k+1}^s$ are known, the probability density of state $\boldsymbol{\theta}_{k+1}$ conditioned on all the data reads

$$p(\boldsymbol{\theta}_{k+1}|\boldsymbol{y}_T^*) = \mathcal{N}\left(\hat{\boldsymbol{\theta}}_{k+1}^s, \boldsymbol{C}_{k+1}^s\right)$$

Consequently, the joint Gaussian probability density of the next state and the current state conditioned on the whole data set could be written using lemma Equation A.3.2

$$\begin{aligned} &p(\boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k|\boldsymbol{y}_T^*) \\ &= \mathcal{N}\left(\begin{bmatrix} \hat{\boldsymbol{\theta}}_{k+1}^s \\ \hat{\boldsymbol{\theta}}_k + \boldsymbol{A}_k\left(\hat{\boldsymbol{\theta}}_{k+1}^s - \boldsymbol{F}_k\hat{\boldsymbol{\theta}}_k\right) \end{bmatrix}, \begin{bmatrix} \boldsymbol{C}_{k+1}^s & \boldsymbol{C}_{k+1}^s\boldsymbol{A}_k^T \\ \boldsymbol{A}_k\boldsymbol{C}_{k+1}^s & \boldsymbol{A}_k\boldsymbol{C}_{k+1}^s\boldsymbol{A}_k^T + \boldsymbol{C}_{k|\boldsymbol{\theta}_{k+1},\boldsymbol{y}_T^*} \end{bmatrix}\right) \end{aligned} \tag{4.2.72}$$

where the term $\boldsymbol{F_k}\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ is the prediction $\hat{\boldsymbol{\theta}}^{-}_{\boldsymbol{k+1}}$ of the next state. Using the marginal Gaussian Equation A.3.3 and substituting Equation 4.2.71 into Equation 4.2.72, one can finally have the probability density of $\boldsymbol{\theta_k}$ conditioned on all the data:

$$\begin{aligned} p(\boldsymbol{\theta}_k|\boldsymbol{y}^*_{\boldsymbol{T}}) &= \mathcal{N}\left(\hat{\boldsymbol{\theta}}^s_{\boldsymbol{k}}, \boldsymbol{C}^s_{\boldsymbol{k}}\right) \\ &= \mathcal{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} + \boldsymbol{A_k}\left(\hat{\boldsymbol{\theta}}^s_{\boldsymbol{k+1}} - \hat{\boldsymbol{\theta}}^-_{\boldsymbol{k+1}}\right),\right. \\ &\qquad \left.\boldsymbol{A_k}\boldsymbol{C}^s_{\boldsymbol{k+1}}\boldsymbol{A}^{\boldsymbol{T}}_{\boldsymbol{k}} + \boldsymbol{C_k} - \boldsymbol{A_k}\boldsymbol{C}^-_{\boldsymbol{k+1}}\boldsymbol{A_k}^T\right) \end{aligned} \tag{4.2.73}$$

Finally the smoothed state and its related covariance are given in Equation 4.2.73. All the equations in RTS smoothing is summarized in Table 4.5.

The initialization:

$$\hat{\boldsymbol{\theta}}^s_{\boldsymbol{T}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{T}}$$

$$\boldsymbol{C}^s_{\boldsymbol{T}} = \boldsymbol{C_T}$$

The stacked inputs($0 \le k \le T$):

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}, \boldsymbol{C_k}, \boldsymbol{F_k}, \hat{\boldsymbol{\theta}}^-_{\boldsymbol{k}}, \boldsymbol{C}^-_{\boldsymbol{k}}$$

The smoothing equations($0 \le k \le T-1$):

$$\boldsymbol{A_k} = \boldsymbol{C_k}\boldsymbol{F_k}^T(\boldsymbol{F_k}\boldsymbol{C_k}\boldsymbol{F_k}^T + \boldsymbol{Q_k})^{-1}$$

$$\hat{\boldsymbol{\theta}}^s_{\boldsymbol{k}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{k}} + \boldsymbol{A_k}\left(\hat{\boldsymbol{\theta}}^s_{\boldsymbol{k+1}} - \hat{\boldsymbol{\theta}}^-_{\boldsymbol{k+1}}\right)$$

$$\boldsymbol{C}^s_{\boldsymbol{k}} = \boldsymbol{C_k} + \boldsymbol{A_k}(\boldsymbol{C}^s_{\boldsymbol{k+1}} - \boldsymbol{C}^-_{\boldsymbol{k+1}})\boldsymbol{A_k}^T$$

Table 4.5: Rauch-Tung-Striebel fixed interval smoothing

4.2.5 Applications in tracking

In this section, the kalman filter based straight track fitting algorithm will be introduced. With proper treatments of coordinates rotation and translation, this tracking method could be used in beam test directly. But a proper initialization is required for a given beam and the telescope facility.

A line in space could be defined with 6 independent parameters, which are the coordinates of the origin $\vec{r_0} = (x_0, y_0, z_0)$, the unit vector of the direction of the line $\vec{v} = (a, b, c)$ and l, the distance between the point on the line and the chosen origin point. We have the constraint $\sqrt{a^2 + b^2 + c^2} = 1$, and the coordinates of a point (x, y, z) on this line read

$$\begin{cases} x = x_0 + l \cdot a \\ y = y_0 + l \cdot b \\ z = z_0 + l \cdot c \end{cases}$$

In beam test, telescope planes are almost perpendicular to particle trajectories and z axis is usually defined in the beam direction , such that we have $c \neq 0$. The projected lines in $x - z$ plane and $y - z$ plane could be written as:

$$x = \frac{a}{c} z + x_0 - \frac{a}{c} z_0 \tag{4.2.74}$$

$$y = \frac{b}{c} z + y_0 - \frac{b}{c} z_0 \tag{4.2.75}$$

Each of the projected line could be parametrized by the slope and the intercept. In $x - z$ plane, θ_x is defined as the angle from z-axis to the line direction in the range $-\pi/2 < \theta_x < \pi/2$, then $\tan\theta_x = a/c$ is the slop of the projected line. The state vector which chooses the intercept $x_0 - \tan\theta_x z_0$ leads to elements containing z_0 in the covariance matrix of multiple scattering noise[62]. The straightforward and convenient solution is to construct the state vector with the hit position and the slope as follows

$$\boldsymbol{\theta_k} = \begin{bmatrix} x_k \\ \tan\theta_{x_k} \\ y_k \\ \tan\theta_{y_k} \end{bmatrix} \tag{4.2.76}$$

In this parametrization, the track length l is a function of the two tangents and Δz. It is straightforward to get the relation below:

$$l = \frac{\Delta z}{c} = \Delta z \sqrt{1 + \tan^2\theta_x + \tan^2\theta_y} \tag{4.2.77}$$

in which, Δz could be the thickness of the scatterer and then l stands for the track length in the material, which could be used in Equation 4.1.14 to calculate multiple scattering error. After choosing the state vector, the measurement matrix could be written directly from Equation 4.2.3 as the measurements in beam test are only hit positions:

$$\boldsymbol{H_k} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

To get the transfer matrix $\boldsymbol{F_k}$, we need to rewrite Equation 4.2.74 into an recursive form. At k'th state, the point of the projected line in $x-z$ plane fulfils

$$x_k = \tan\theta_{x_k} z_k + x_0 - \tan\theta_{x_k} z_0,$$

At $(k+1)'$th state, the new point is a function of the previous one:

$$\begin{aligned} \tan\theta_{x_{k+1}} &= \tan\theta_{x_k} \\ x_{k+1} &= \tan\theta_{x_k} z_{k+1} + x_0 - \tan\theta_{x_k} z_0 \\ &= \tan\theta_{x_k}(z_{k+1} - z_k) + x_k \end{aligned}$$

Considering Equation 4.2.2, the transfer matrix reads

$$\boldsymbol{F_k} = \begin{bmatrix} 1 & \Delta z_k & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta z_k \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $\Delta z_k = z_{k+1} - z_k$.

The process noise considered in straight line fitting is mainly from multiple scattering. The $\boldsymbol{Q_k}$ matrix has been derived in several literatures[62, 63].

$$\boldsymbol{Q_k} = \begin{bmatrix} Q_{22}\frac{t^2}{3} & Q_{22}\frac{t}{2} & Q_{24}\frac{t^2}{3} & Q_{24}\frac{t}{2} \\ \cdots & Q_{22} & Q_{24}\frac{t}{2} & Q_{24} \\ \cdots & \cdots & Q_{44}\frac{t^2}{3} & Q_{44}\frac{t}{2} \\ \cdots & \cdots & \cdots & Q_{44} \end{bmatrix}$$

where t is the thickness of the current scatterer in $z-$axis and the key elements are given:

$$Q_{22} = \sigma_{MS}^2(1 + \tan^2\theta_x)(1 + \tan^2\theta_x + \tan^2\theta_y) \tag{4.2.78}$$

$$Q_{44} = \sigma_{MS}^2(1 + \tan^2\theta_y)(1 + \tan^2\theta_x + \tan^2\theta_y) \tag{4.2.79}$$

$$Q_{24} = \sigma_{MS}^2 \tan\theta_x \tan\theta_y(1 + \tan^2\theta_x + \tan^2\theta_y) \tag{4.2.80}$$

in which σ_{MS} is the multiple scattering angle deviation described by Equation 4.1.14 with corrected equivalent length $X = \rho t\sqrt{1 + \tan^2\theta_x + \tan^2\theta_y}$ and ρ is the density of the medium.

At each detection layer, measurements and the related uncertainties will be organized as the matrices below to feed the Kalman filter:

$$\boldsymbol{y_k} = \begin{bmatrix} m_x \\ m_y \end{bmatrix}, \quad \boldsymbol{V_k} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

where m_x and m_y are the measured values at z_k in X-axis and Y-axis respectively. In practice, the measurement covariance matrix may not be a diagonal one since the alignment will lead to coordinate rotations and translations.

The last and one of the most important preparation of kalman filter is the initialization. From the derivation in the viewpoint of conditional probability in Section 4.2.2, one has to make sure that the initialized state vector and covariance matrix are as close to their true values as possible.

4.3 Tracking efficiency estimation

Hit finding is usually the task of a track finding algorithm. Since tracking efficiency is one of the figure of merit of a tracker, one cannot avoid hit finding considerations in the tracker design. A straightforward idea to estimate tracking efficiency is to use an algorithm like Kalman filter that could be used both in track finding and track fitting for its recursive feature of track parameters estimation and their covariance matrix propagation at each local layer. Therefore, the associated hit could be found by picking the hit closest to the predicted track intersection on the layer in a given searching window. Assuming that the noise hits distribute uniformly on a detecting layer, it is possible to do hand calculations on the probability that a signal hit is associated with a track. The probability calculations in this section are summarized from the notes [64, 65, 66] about the heavy flavour tracker (HFT) design for Star (Solenoid Tracker at RHIC) experiment at the Relativistic Heavy Ion Collider (RHIC).

4.3.1 Pileup Probability

The pileup probability P_{pileup} is the probability that one or more noise hits inside an area a with a given surface hit density ρ could be found. Obviously, we can write the equation below:

$$P_{pileup} = 1 - P_{noNoise} \tag{4.3.1}$$

where $P_{noNoise}$ is the probability that there is no hit found in area a. It is easier to calculate $P_{noNoise}$ first. Assuming the small area a is included in a very larger area A, this probability for the total number of hits n could be written as:

$$P_{noNoise} = (1 - \frac{a}{A})^n \tag{4.3.2}$$

in which, $n = \rho \times A$. Applying binomial theorem to equation 4.3.2, one can find:

$$P_{noNoise} = \sum_{k=0}^{n} \left[\frac{n!}{k! \cdot (n-k)!} \cdot (-\frac{a}{A})^k \cdot 1^{n-k} \right] \tag{4.3.3}$$

Substituting $A = n/\rho$ into equation 4.3.3, we have

$$P_{noNoise} = \sum_{k=0}^{n} \left[\frac{n!}{k! \cdot (n-k)! \cdot n^k} \cdot (-a \cdot \rho)^k \right] \tag{4.3.4}$$

In case of large $n \to \infty$, equation 4.3.4 reads

$$\begin{aligned} P_{noNoise} &= \sum_{k=0}^{n} \left[\frac{(n-0)\cdots(n-k+1)}{k! \cdot n^k} \cdot (-a \cdot \rho)^k \right] \\ &= \sum_{k=0}^{n} \left[\frac{(1-\frac{0}{n})\cdots(1-\frac{k-1}{n})}{k!} \cdot (-a \cdot \rho)^k \right] \\ &\approx \sum_{k=0}^{\infty} \left[\frac{(-a \cdot \rho)^k}{k!} \right] \end{aligned} \tag{4.3.5}$$

As equation 4.3.5 is the Taylor series of $e^{-a\cdot\rho}$, we finally have

$$\begin{aligned} P_{pileup} &= 1 - P_{noNoise} \\ &= 1 - \sum_{k=0}^{n} \left[\frac{n!}{k! \cdot (n-k)! \cdot n^k} \cdot (-a \cdot \rho)^k \right] \\ &\approx 1 - e^{-a\cdot\rho} \end{aligned} \tag{4.3.6}$$

In this calculation, $P_{noNoise}$ is calculated from a binomial distribution. When n is larger than 20 and a/A is smaller than 0.05, it is a good approximation to use Poisson distribution instead of binomial probability. Since A is a very large area chosen by us, the conditions to use the Poisson distribution is always satisfied. The Poisson parameter is given below:

$$\begin{aligned} \lambda &= n \cdot \frac{a}{A} \\ &= \rho \cdot A \cdot \frac{a}{A} \\ &= a \cdot \rho \end{aligned} \tag{4.3.7}$$

Consequently, $P_{noNoise}$ could be derived quickly as shown in the following

$$\begin{aligned} P_{noNoise} &\approx Poiss(0, \lambda) \\ &= \frac{\lambda^0 \cdot e^{-\lambda}}{0!} \\ &= e^{-\lambda} \\ &= e^{-a\cdot\rho} \end{aligned} \tag{4.3.8}$$

4.3.2 Probability of finding the correct hit

In this section, the probability of finding a real hit ($P_{CorrectHit}$) will be calculated for both symmetric and asymmetric resolution case. Once a track is associated with a fake hit which is closer to the predicted track position than the real signal hit on a layer, the track may be guided away from its correct trajectory and finally becomes a ghost track.

Assuming that charged particles penetrate the same position (0,0) on the detecting layer to which the tracks are perpendicular, the signal hit density could be described by a two dimensional Gaussian distribution:

$$dP_{hit} = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} dxdy \tag{4.3.9}$$

in which σ_x^2 and σ_y^2 are the variances of the hit position in x and y respectively. These variances have already included the intrinsic spatial resolution of the sensor and the track uncertainty of the current hit at the measurement plane.

4.3.2.1 Symmetric Resolution

If both the spatial resolution of the sensor and the error of the track are symmetric, that is $\sigma_x = \sigma_y = \sigma$, equation 4.3.9 could be written only as a function of the distance r between a selected hit and the predicted track position:

$$\begin{aligned} dP_r &= \int_0^{2\pi} dP_{hit} \\ &= \int_0^{2\pi} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} dxdy \\ &= \int_0^{2\pi} \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\cdot\frac{r^2}{\sigma^2}} rdrd\phi \\ &= \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr \end{aligned} \tag{4.3.10}$$

To find a correct hit means there is no background hit closer to the predicted track position than the real hit. Considering equation 4.3.8, we can write:

$$\begin{aligned} P_{CorrectHit} &= \int_0^{\infty} P_{noNoise} \cdot dP_r \\ &= \int_0^{\infty} P_{noNoise} \cdot \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr \\ &= \int_0^{\infty} e^{-a\cdot\rho} \cdot e^{-\frac{r^2}{2\sigma^2}} \cdot \frac{r}{\sigma^2} dr \\ &= \int_0^{\infty} e^{-\pi\rho r^2} \cdot e^{-\frac{r^2}{2\sigma^2}} \cdot \frac{r}{\sigma^2} dr \\ &= \int_0^{\infty} e^{-\pi\rho r^2} \cdot e^{-\frac{r^2}{2\sigma^2}} d\left(\frac{r^2}{2\sigma^2}\right) \\ &= \int_0^{\infty} e^{-\pi\rho r^2} d\left(-e^{-\frac{r^2}{2\sigma^2}}\right) \end{aligned} \tag{4.3.11}$$

This integration could be solved easily with the hint below:

$$\begin{aligned} d(e^{-\pi\rho r^2}\cdot e^{-\frac{r^2}{2\sigma^2}}) &= e^{-\pi\rho r^2} d(e^{-\frac{r^2}{2\sigma^2}}) + e^{-\frac{r^2}{2\sigma^2}} d(e^{-\pi\rho r^2}) \\ &= e^{-\pi\rho r^2} d(e^{-\frac{r^2}{2\sigma^2}}) + e^{-\pi\rho r^2}\cdot e^{-\frac{r^2}{2\sigma^2}}\cdot(-2\pi\rho r)dr \\ &= e^{-\pi\rho r^2} d(e^{-\frac{r^2}{2\sigma^2}}) + e^{-\pi\rho r^2}\cdot e^{-\frac{r^2}{2\sigma^2}}\cdot(-2\pi\rho\sigma^2)d(\frac{r^2}{2\sigma^2}) \\ &= e^{-\pi\rho r^2} d(e^{-\frac{r^2}{2\sigma^2}}) + 2\pi\rho\sigma^2\cdot e^{-\pi\rho r^2} d(e^{-\frac{r^2}{2\sigma^2}}) \\ &= (1+2\pi\rho\sigma^2)e^{-\pi\rho r^2} d(e^{-\frac{r^2}{2\sigma^2}}) \end{aligned} \tag{4.3.12}$$

Finally we have:

$$P_{CorrectHit} = \frac{1}{1+2\pi\rho\sigma^2} \tag{4.3.13}$$

4.3.2.2 Asymmetric resolution

If the coordinate origin is the predicted track position, X axis is along the width of the sensor and Y axis is along its length, Asymmetric resolution in sensor surface means that the convolution of track error projected in X and sensor spatial resolution in X is different from what one can find in Y axis. In this case, equation 4.3.9 describes the hit density. Figure 4.6 shows a two-dimensional Gaussian distribution, with which one can find that the contour of equivalent possibility is an ellipse rather than a circle.

The associated hit should be the one in the smallest ellipse whose centre is the predicted track position. Therefore, if a fake hit is inside of the equivalent probability ellipse defined by a real hit, the reconstructed track would unfortunately become a ghost track. Assuming the fake hit is located at point (x_{fake}, y_{fake}) and the real hit is at (x, y), a ghost track which associates a fake hit have the inequality below:

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} > \frac{x_{fake}^2}{\sigma_x^2} + \frac{y_{fake}^2}{\sigma_y^2} \tag{4.3.14}$$

we change the sign *larger than* (>) into *equal* (=) to get the largest ellipse for those fake hits, such that the ellipse equation is written as:

$$\frac{x_{fake}^2}{\frac{\sigma_x^2\cdot y^2+\sigma_y^2\cdot x^2}{\sigma_y^2}} + \frac{y_{fake}^2}{\frac{\sigma_x^2\cdot y^2+\sigma_y^2\cdot x^2}{\sigma_x^2}} = 1 \tag{4.3.15}$$

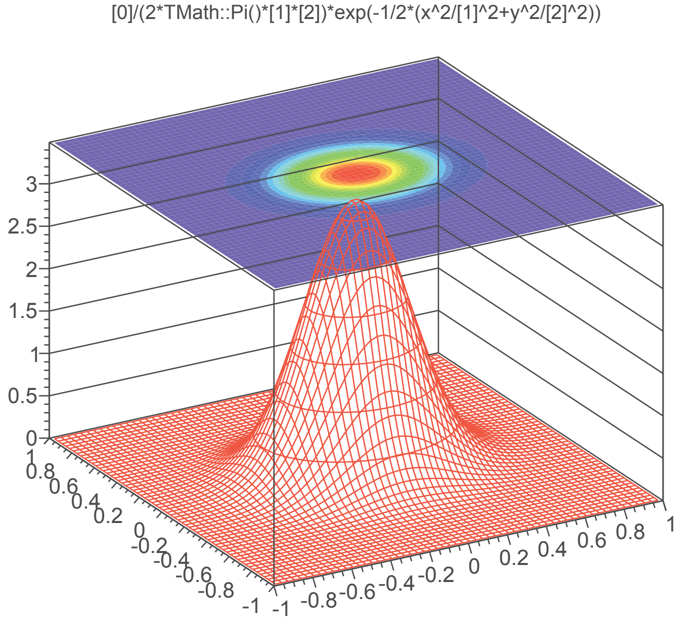


Figure 4.6: An example of a 2D Gaussian distribution.

The area a of this ellipse is what we need for the pileup calculation.

$$\begin{aligned} a &= \pi \cdot \frac{\sqrt{\sigma_x^2 \cdot y^2 + \sigma_y^2 \cdot x^2}}{\sigma_y} \cdot \frac{\sqrt{\sigma_x^2 \cdot y^2 + \sigma_y^2 \cdot x^2}}{\sigma_x} \\ &= \pi \cdot \frac{\sigma_x^2 \cdot y^2 + \sigma_y^2 \cdot x^2}{\sigma_x \cdot \sigma_y} \end{aligned} \tag{4.3.16}$$

Using equation 4.3.16, equation 4.3.8 and equation 4.3.9, we can find the probability to associate the correct hit with the track in the following derivations:

$$\begin{aligned}
P_{CorrectHit} &= \int_{-Y}^{Y}\int_{-X}^{X} P_{noNoise}\cdot dP_{hit} \\
&= \int_{-Y}^{Y}\int_{-X}^{X} \frac{1}{2\pi\sigma_x\sigma_y} e^{-a\cdot\rho}\cdot e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} dxdy \\
&= \frac{4}{2\pi\sigma_x\sigma_y}\int_{0}^{Y}\int_{0}^{X} e^{-\pi\cdot\left(\frac{\sigma_x^2\cdot y^2+\sigma_y^2\cdot x^2}{\sigma_x\cdot\sigma_y}\right)\cdot\rho}\cdot e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} dxdy \\
&= \frac{4}{2\pi\sigma_x\sigma_y}\int_{0}^{Y}\int_{0}^{X} e^{-\left(\frac{2\pi\sigma_x\sigma_y\rho+1}{2\sigma_x^2}\right)x^2}\cdot e^{-\left(\frac{2\pi\sigma_x\sigma_y\rho+1}{2\sigma_y^2}\right)y^2} dxdy \qquad (4.3.17)
\end{aligned}$$

where X and Y stand for the half width and half length of the hit searching window respectively. To simplify this expression, we introduce the error function $\mathrm{erf}(z)$ which is shown in equation 4.3.18 and Figure 4.7.

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_{0}^{z} e^{-x^2}dx \qquad (4.3.18)$$

Using equation 4.3.18, equation 4.3.17 leads to

$$P_{CorrectHit} = \frac{1}{2\pi\sigma_x\sigma_y\rho+1}\cdot\mathrm{erf}\left(\frac{X}{\sigma_x}\sqrt{\frac{2\pi\sigma_x\sigma_y\rho+1}{2}}\right)\cdot\mathrm{erf}\left(\frac{Y}{\sigma_y}\sqrt{\frac{2\pi\sigma_x\sigma_y\rho+1}{2}}\right) \qquad (4.3.19)$$

If the searching window is much larger than the resolution, the value of error function in equation 4.3.19 becomes 1, such that we find:

$$P_{CorrectHit} = \frac{1}{2\pi\sigma_x\sigma_y\rho+1} \qquad (4.3.20)$$

Figure 4.8 shows the right hit-track association in a searching ellipse which is defined by the convolution of track uncertainty and intrinsic spatial resolution of the sensor.

4.3.3 Estimation on tracking efficiency

The tracking efficiency of Monte-Carlo simulation can be defined as the number of tracks well reconstructed divided by the number of all the true tracks. The question is how to

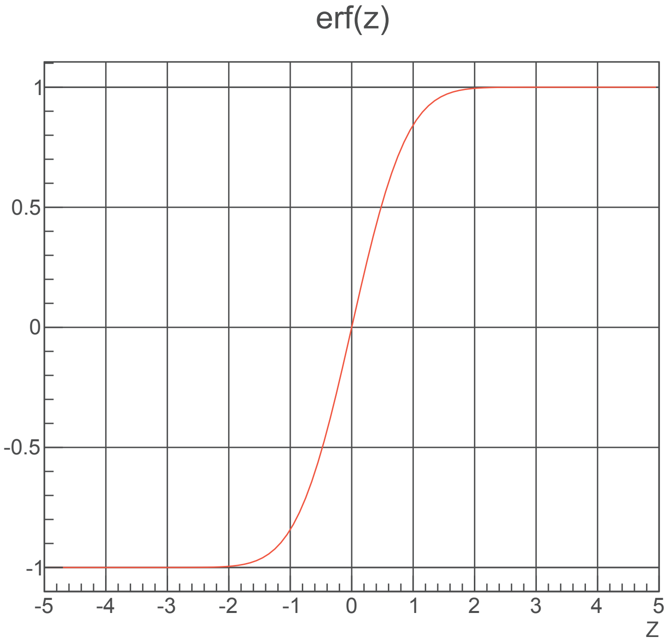


Figure 4.7: Error Function

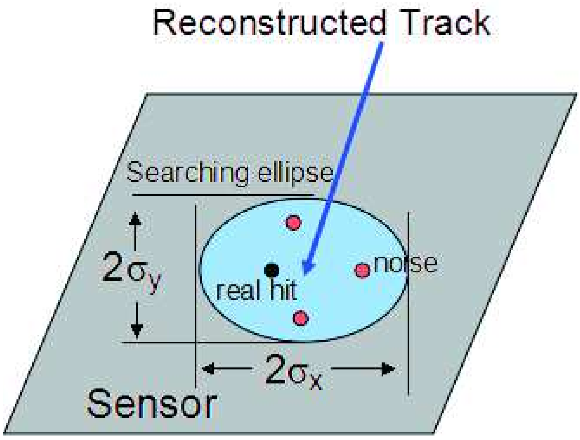


Figure 4.8: The right hit-track association in a searching ellipse.

define a good track with a proper criteria. For a given track found by the track finder, one can have requirements on the quality of each hit, on the total number of good hits, on the number of hits at silicon detector or on the uncertainty of the track. For a long trajectory that could penetrate silicon pixel tracker and reach a dozen of layers in a multilayer drift chamber (MDC), the tracking efficiency in MDC could be very close to 100% and at least three hits in silicon tracker could be a reasonable requirement. In this case, the tracking efficiency could be estimated by multiplying $P_{CorrectHit}$ of each silicon detecting layer. Although the results of this calculation may be quite different from the efficiency given by a specified track finding algorithm, this estimation could be used to compare different designs and then help to choose the tracker layout with a better tracking efficiency performance.

4.4 Summary and conclusion

In this Chapter, the generalized least squares fitting and the Kalman filtering and smoothing algorithm have been introduced with derivations. When multiple scattering could be neglected, the least squares could give the covariance matrix of optimized track parameters directly. Based on a parabola model and an uniform layer configuration with intrinsic spatial resolution σ along the lever arm of a length L, the curvature resolution of a tracker is proportional to $\sigma/(\sqrt{N}L^2)$, the track direction resolution is proportional to $\sigma/(\sqrt{N}L)$ and the intercept (δy at $x = 0$) resolution increases with $\sigma/\sqrt{N}$. In case of tracking with multiple scattering effects, an explicit solution is hard to derive. Usually, the resolutions are calculated numerically for a given tracker layout and trajectory properties. This is the same with a Kalman filter based estimation. To assess the performance of a tracker, one can use an inward tracking, in which the outermost hit is chosen as the seed and the track state is updated hit by hit towards the innermost layer. Because of this procedure, Kalman filter could also be used for track finding. The two tracking algorithms are identical for track fitting results when the same track model is used, and Kalman filter is popular now for its fast speed and flexible treatment between two hits. The main disadvantage of this recursive algorithm is at its initialization. A poor initialization leads to over/under-estimation of parameter variances and even a failure of track finding. Tracking efficiency is estimated by using the sum of the probabilities of right associations between a real hit (not a noise) and the reconstructed track for all the hits. It is named as pseudo-efficiency as its value is related to both tracking efficiency and track purity. The equations and results in this Chapter are used to develop the general strategy for tracker geometry optimizations in the next chapter.

Chapter 5

BESIII CPS Inner tracker design and optimization

To develop a tracker from scratch, the constraints from ladder design and tracker size have to be determined first. BESIII SPT, as an upgrade of a gas chamber, has already defined the tracker volume. As a nominal proposal, only the first four layers of the inner chamber will be replaced by silicon ladders. Therefore, we applied these constraints to the software we developed for this study. The helix trajectory is coped with in two projections: a circle segment in $r-\phi$ plane and a straight line in $r-z$[*] plane. The former is used to develop the general strategy for the optimization of a barrel silicon pixelated tracker, which mainly focuses on momentum measurements. The later is then used to discuss the designs for improved impact parameter resolutions.

5.1 Ladder design

The preliminary design of ladders have materials listed in Table 5.1. As for a double-sided ladder, additional materials with a budget $\sim 0.151\%$ are added for CPS and related flex cables glued on the opposite surface of the ladder. The geometries studied in this thesis mainly include the normal ladders as the plume-like ladders may need more studies for their stability and cooling considerations.

---

[*] For particles with a small P_T, the track length s in $r-\phi$ should be used. It is a straight track in $s-z$ plane.

Material	Thickness [µm]	Radiation Length $[X/X_0]$
Silicon(CPS)	50	0.053%
Epoxy	10	0.003%
kapton	100	0.036%
Aluminium	50	0.056%
Epoxy	10	0.003%
Carbon Fibre	500	0.211%
Total	720	0.362%

Table 5.1: Materials in one ladder.

5.2 Geometry optimization for tracking

The geometry or spacing of a bubble tracker has been studied by Gluckstern using weighted least squares and parabola approximation to estimate the covariances of the trajectory direction and curvature in early 1960s[5]. The optimal spacing for track curvature measurement was given with half of all the N measurements taken at the center of the tracker and the other half measured at the both ends of the detecting region equally when multiple scattering is not dominating, that is the layer configuration $N/4 - N/2 - N/4$ at positions $x = 0, L/2, L$ where L is the length of the tracker lever arm in a telescope system or the radial distance from the innermost layer to the outermost in cylindrical trackers. In 1990s, Karimäki developed the explicit formulae to calculate parameters covariances of circle tracks by using a non-linear extension of the linear least squares method[67, 68]. The spacing to achieve the best curvature accuracy in Karimäki's strategy is the same with Gluckstern's when N is a multiple of 4. In addition, Karimäki pointed out that the measurements should be taken symmetrically with respect to the center of the tracker and maximum the product $M(N - M)$ in which M is the number of the measurements at $x = L/2$ (the center). In this thesis, a new formula was proposed to get the optimal spacing for any $N \geq 3$ without the constraint to have the same number of measurements at the two ends of a tracker on condition of neglected multiple scattering. Compared with Karimäki's non-linear least squares calculation, the parabola approximation leads to overestimated curvature variance by a factor about $(L/R)^2/21$ [*] where R is the radius of the trajectory[68]. Nevertheless this overestimation is close to 0 in the case we tried to explore the optimal geometry of the tracker since the trajectory will have sufficient projected momentum that is high enough in the tracking plane to make sure that the multiple scattering effect is too small to be sensed, which leads to very large track radius compared with the tracker lever arm L. Therefore, the optimal spacing without considering multiple scattering could be studied by the convenient parabola approximation on the condition of extremely small (L/R).

5.2.1 Curvature variance and the optimal spacing without multiple scattering

We start from the parabola equation below:

$$y = a + bx + cx^2$$

in which, parameter a is the intercept at $x = 0$, b is the track direction and c is equal to $1/(2R)$ half of the track curvature at $x = 0$. In the case of no multiple scattering, the generalized least squares will be simplified to the weighted least squares and the variance matrix of the parameters are given below:

$$\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \sum_{j=1}^{N} 1/\sigma_j^2 & \sum_{j=1}^{N} x_j/\sigma_j^2 & \sum_{j=1}^{N} x_j^2/\sigma_j^2 \\ \sum_{j=1}^{N} x_j/\sigma_j^2 & \sum_{j=1}^{N} x_j^2/\sigma_j^2 & \sum_{j=1}^{N} x_j^3/\sigma_j^2 \\ \sum_{j=1}^{N} x_j^2/\sigma_j^2 & \sum_{j=1}^{N} x_j^3/\sigma_j^2 & \sum_{j=1}^{N} x_j^4/\sigma_j^2 \end{bmatrix}^{-1} \tag{5.2.1}$$

where $\hat{\boldsymbol{\theta}}$ is the vector of the three estimated parameters defined by $\hat{\boldsymbol{\theta}} = (a, b, c)^T$ and σ_j is the measurement error of the $j'th$ layer located at position x_j. To deal with the inversion in this equation and simplify the derivation, all the layers are supposed to have the same

---

[*] L is the length of the lever arm in Gluckstern's method, but it stands for track segment length in Karimäki's calculation. Therefore, the overestimated factor $(L/R)^2/21$ itself is only valid for small $|\frac{L}{R}|$

spatial resolution σ and the notation $X^{(k)}$ is introduced as follows:

$$X^{(k)} = \sum_{j=1}^{N} x_j^k \tag{5.2.2}$$

in which, $k = 0, 1, 2, 3, 4$ and $X^{(0)} = N$. And the auxiliary matrix $\boldsymbol{A_V}$ including all the geometry information is defined below:

$$\boldsymbol{A_V} \equiv \begin{bmatrix} N & X^{(1)} & X^{(2)} \\ X^{(1)} & X^{(2)} & X^{(3)} \\ X^{(2)} & X^{(3)} & X^{(4)} \end{bmatrix} \tag{5.2.3}$$

Therefore the covariance matrix could be written as

$$\boldsymbol{V_{\hat{\theta}}} = \sigma^2 \boldsymbol{A_V^{-1}} \tag{5.2.4}$$

Finally, the explicit expression of the upper-right elements in the symmetric matrix $\boldsymbol{V_{\hat{\theta}}}$ reads

$$\boldsymbol{V_{\hat{\theta}}} = \frac{\sigma^2}{|\boldsymbol{A_V}|} \begin{bmatrix} X^{(2)}X^{(4)} - X^{(3)}X^{(3)} & X^{(2)}X^{(3)} - X^{(1)}X^{(4)} & X^{(1)}X^{(3)} - X^{(2)}X^{(2)} \\ & NX^{(4)} - X^{(2)}X^{(2)} & X^{(1)}X^{(2)} - NX^{(3)} \\ & & NX^{(2)} - X^{(1)}X^{(1)} \end{bmatrix} \tag{5.2.5}$$

where $|\boldsymbol{A_V}|$ is the determinant of matrix $\boldsymbol{A_V}$.

$$|\boldsymbol{A_V}| = N\left[X^{(2)}X^{(4)} - (X^{(3)})^2\right] + X^{(1)}(X^{(2)}X^{(3)} - X^{(1)}X^{(4)}) + X^{(2)}\left[X^{(1)}X^{(3)} - (X^{(2)})^2\right]$$

And the track fitting performance is thus found to be determined by the layer spatial resolution and the spacing, the geometry of the tracker. Naturally, the problem to optimize the system for the most accurate measurement ability on track curvature is changed into the minimization of the curvature related variance V_{cc} with respect to layer positions, where V_{cc} is defined by the bottom-right element of the matrix $\boldsymbol{V_{\hat{\theta}}}$ in Equation 5.2.5. The related variance V_{cc} and its partial derivative with respect to the postion of k'th layer are shown as follows:

$$V_{cc} = \sigma^2 \frac{NX^{(2)} - X^{(1)}X^{(1)}}{|\boldsymbol{A_V}|} \tag{5.2.6}$$

$$\frac{\partial V_{cc}}{\partial x_k} = \sigma^2 \frac{num1 \cdot num2}{|\boldsymbol{A_V}|^2} \tag{5.2.7}$$

in which,

$$num1 = 2\sum_{i=1}^{N}\sum_{j=1}^{N}[x_i(x_k - x_j)(x_k - x_i)(x_j - x_i)] \tag{5.2.8}$$

$$num2 = \sum_{i=1}^{N}\sum_{j=1}^{N}[x_j(2x_k - x_j)(x_j - x_i)] \tag{5.2.9}$$

If the indices i and j are exchanged in Equation 5.2.8 and 5.2.9, the values of $num1$ and $num2$ will not be changed, such that we have:

$$\begin{aligned} num1 &= \sum_{i=1}^{N}\sum_{j=1}^{N}[x_i(x_k - x_j)(x_k - x_i)(x_j - x_i) + x_j(x_k - x_j)(x_k - x_i)(x_i - x_j)] \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\left[-(x_k - x_i)(x_k - x_j)(x_j - x_i)^2\right] \end{aligned} \tag{5.2.10}$$

$$\begin{aligned} num2 &= \sum_{i=1}^{N}\sum_{j=1}^{N}[x_j(2x_k - x_j)(x_j - x_i) + x_i(2x_k - x_i)(x_i - x_j)] \cdot \frac{1}{2} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\left[(x_j - x_i)^2(x_k - \frac{x_i + x_j}{2})\right] \end{aligned} \tag{5.2.11}$$

Usually to minimize V_{cc} requires $\partial V_{cc}/\partial x_k = 0$ for all the layers, however it is not possible in our situation. The minimization can only be solved under certain constrains:

- When $x_k = x_1$, x_k is the smallest value such that $num1 < 0$ and $num2 < 0$. The partial derivative

  $$\frac{\partial V_{cc}}{\partial x_1} > 0$$

  therefore, $V_{cc}(x1)$ is an increasing function of $x1$, which means $x1$ is the smaller the better to have the smallest variance on track curvature. In practice, the position of the very first layer is limited by beam pipe and background density in collider experiments.

- When $x_k = x_N$, x_k is the position of the outermost layer in the tracker. $num1 < 0$ and $num2 > 0$ lead to
$$\frac{\partial V_{cc}}{\partial x_N} < 0$$
which is a decreasing function of x_N. To have smaller V_{cc}, x_N should be as large as possible. However it is not possible for particles with low transverse momentum to reach a layer too far away.

- When $x_1 < x_k < x_N$, the layer to be studied is an intermediaNNe one located between the innermost and the outermost. The formalization in Equation 5.2.10 and Equation 5.2.11 leads to the possible solution to this minimization problem. On condition of the same number of layers placed at x_1 and x_N, all the other layers should have the same position $x_k = (x_1 + x_N)/2$. Then we have:
$$\left.\frac{\partial V_{cc}}{\partial x_k}\right|_{x_k=\frac{x_1+x_N}{2}} = 0$$
$$\left.\frac{\partial V_{cc}}{\partial x_k}\right|_{x_k=\frac{x_1+x_N}{2}+\epsilon} > 0$$
$$\left.\frac{\partial V_{cc}}{\partial x_k}\right|_{x_k=\frac{x_1+x_N}{2}-\epsilon} < 0$$
in which $0 < \epsilon \ll (x_1 + x_N)/4$.

Nevertheless, the optimization up to now does not answer the question how many layers should be installed at x_1 and x_N. For instance, in the case $N = 8$, we have several optimal geometries for 6, 4 or 2 intermediate layers:

$$1 - 6 - 1$$

$$2 - 4 - 2$$

$$3 - 2 - 3$$

where the three numbers denotes the number of layers located at x_1, $(x_1+x_N)/2$ and x_N respectively. To find the optimal spacing overall, two positive fractions f_1 and f_2, which stand for the ratios of layer number at $(x_1 + x_N)/2$ over N and layer number at x_N over N respectively, are introduced. $f_1 + f_2 < 1$ should be guaranteed to have measurements at three different positions. Moreover, it is without loss of generality to set $x_1 = 0$ and $x_N = L$ where L is the length of the lever arm of the tracker. Therefore the curvature variance V_{KK} could be derived from Equation 5.2.6:

$$V_{KK} = 4V_{cc} = \frac{16\sigma^2}{NL^4}\left(\frac{4}{f_1} + \frac{1}{f_2} + \frac{1}{1-f_1-f_2}\right). \tag{5.2.12}$$

The minimization of V_{cc} with respect to f_1 and f_2 leads to $f_1 = 0.5$ and $f_2 = 0.25$, that is the same layer spacing configuration $N/4 - N/2 - N/4$ at begin-center-end of the tracker given by references[5, 68].

In the case $f_2 = 1 - f_1 - f_2$, numbers of measurements taken at the two ends of a tracker are equal. The curvature variance V_{KK} could be simplified as:

$$V_{KK} = \frac{64\sigma^2}{L^4}\frac{N}{Nf_1(N-Nf_1)}. \tag{5.2.13}$$

which is exactly the equation obtained by Karimäki [68]. Since Equation 5.2.12 is more general, it should be used as long as the total number of layers $N \geq 3$. When $f_2 \neq 1 - f_1 - f_2$, the number of measurements taken at $x = 0$ and that at $x = L$ are not equal, the variance calculated by Equation 5.2.12 is not optimal any more. The intuitive understanding is that the layers in the middle should be put closer to the tracker end with more measurements to balance the different weights from the two sides of the measured track segment. The optimal position for the middle layers could be found by solving x_k from the equation below:

$$2N(1-f_1-f_2)Nf_2L^2(x_k - \frac{L}{2}) + N(1-f_1-f_2)Nf_1x_k^3 + Nf_1Nf_2(x_k - L)^3 = 0. \tag{5.2.14}$$

5.2.1.1 Optimal spacing when N is not a multiple of 4

If the positions of the layers are always $x = 0, L/2, L$, the optimal geometry which is described by f_1 and f_2 could be determined by minimizing V_{cc} in Equation 5.2.12 in the definition domain* of the two fractions without the constraint of equivalent number of layers at $x = 0$ and $x = L$. The first optimal spacing which is unsymmetrical could be

---

*Nf_1 and Nf_2 have to be integers

found at $N = 6$. The comparison of the optimal[*] unsymmetrical spacing and symmetrical spacing is shown in Table 5.2, where the variances V_{aa}[†] and V_{bb}[‡] of the other two track parameters are also calculated. Although the unsymmetrical optimal spacing $2-3-1$ is only slightly better than the symmetrical one $2-2-2$ in terms of curvature variance, $2-3-1$ is the optimal when all track parameters variances are considered. By using this strategy, the geometries to have the most accurate curvature measurement with given number of layers $N \leq 12$ are listed in Table 5.3.

Configuration(begin-center-end)	Curvature variance($4V_{cc}$)	V_{aa}	V_{bb}
$2-3-1$	$272\sigma^2/NL^4$	$3\sigma^2/N$	$65\sigma^2/(NL^2)$
$1-3-2$	$272\sigma^2/NL^4$	$6\sigma^2/N$	$89\sigma^2/(NL^2)$
$1-4-1$	$288\sigma^2/NL^4$	$6\sigma^2/N$	$84\sigma^2/(NL^2)$
$2-2-2$	$288\sigma^2/NL^4$	$3\sigma^2/N$	$78\sigma^2/(NL^2)$

Table 5.2: Comparison of the symmetrical optimal spacing and unsymmetrical spacing when $N = 6$

.

Except this analytical method to optimize tracker geometry, numerical calculation with a scan of all possible spacing configurations is also possible to find the optimal layout. The advantage of analytical solution is its very fast speed, however the limitation is also obvious that it is still quite hard to consider material effects like multiple scattering which will bring correlations among the measurements to make the error matrix much more complex.

5.2.2 Numerical geometry scan with multiple scattering involved

The method of geometry or layer spacing scan is inspired by Equation 5.2.1 as the estimated track parameter variance is only a function of layer positions and their intrinsic spatial resolutions. Considering multiple scattering will only add contributions from material budget of each layer into the parameter covariance matrix. This material contribution

---

[*] Optimization is only taken for minimizing curvature variance.

[†] $V_{aa} = \frac{\sigma^2}{N}\frac{1}{1-f_1-f_2}$

[‡] $V_{bb} = \frac{\sigma^2}{NL^2}(\frac{16}{f_1} + \frac{1}{f_2} + \frac{9}{1-f_1-f_2})$

Number of layers	Configuration(begin-center-end)
3	$1-1-1$
4	$1-2-1$
5	$1-3-1$
6	$2-3-1$ or $1-3-2$
7	$2-3-2$
8	$2-4-2$
9	$2-5-2$
10	$3-5-2$ or $2-5-3$
11	$3-5-3$
12	$3-6-3$

Table 5.3: The optimal spacing for curvature measurement using N layers ($3 \leq N \leq 12$)

is a function of layer positions, layer thickness in unit of radiation length and momentum of the trajectory under a given magnetic field. To simplify the calculation and focus on the scan strategy itself, we require that the momentum direction is perpendicular to the magnetic field, such that the trajectory is simplified as a piece of circle arc.

5.2.2.1 Spacing scan of N layers and its implementation

The innermost layer and the outermost are fixed at their positions $x_1 = 0$ and $x_N = L$. And all the other layers could be moved by a user specified step ΔL which has to satisfy the condition that $L/\Delta L$ is a multiple of $2(N-1)$ for even N and a multiple of $(N-1)$ for odd N to make sure that the position at the center of the tracker and the uniform spacing configuration could occur. Once ΔL is chosen, the second layer should be placed at x_2 which ranges from x_1 to x_N and then the third layer will be placed at x_3 whose range is from x_2 up to x_N and the other unfixed layers are settled orderly. Therefore this process requires N loops. Since we could change the total number of layers in a tracker, N is a variable, such that the recursive algorithm has to be used in the software. And the possible positions $x = 0, \Delta L, 2\Delta L, \ldots, L$ could be related to a series of numbers, that is $0, 1, 2, \ldots, L/\Delta L$, thus all the geometries are changed into all the $(N-2)$-combinations

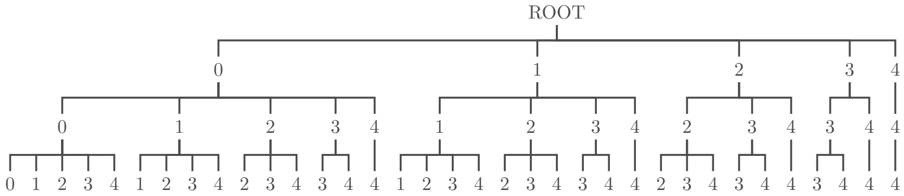


Figure 5.1: The recursion tree for $N = 5$ and $\Delta L = \frac{L}{4}$.

of the number set.

As the optimal spacing could contain layers placed at the same position, the repeated values in the number set are allowed. And the number of levels in the recursion tree is equal to the number of unfixed layers. The simplest tree for 5 detecting layers is shown in Fig. 5.1. One may notice that the only combination is $(1, 2, 3)$ if no overlap of layers is required in this example.

5.2.2.2 The choice of a stable track fitting algorithm

After a tracker layout is determined from one combination of the recursion tree, the curvature variance should be calculated by using a stable track fitting algorithm which satisfies the conditions below:

- The pull distribution of track parameters should be a standard normal distribution for the momentum range of interest.

- The tracking performance should be consistent for different track geometries.

The basic test to check the two conditions above is using a toy Monte-Carlo simulation which only contains measurement uncertainties. By removing multiple scattering influence, the chosen tracking algorithm should work very well in the interested curvature or momentum range. In this thesis, a generalized least squares (GLS) fitting and an extended Kalman filter (EKF) are studied and compared. The toy simulation is generated for pions which traverse 1 Tesla magnetic field and are detected by three silicon ladders with 10 µm spatial resolution located at 72.58 mm, 86.16 mm and 99.50 mm from the origin.

5.2.2.3 Generalized least squares (an extension of Gluckstern's method)

The matrix notation below is used to define the parameter vector $\boldsymbol{\theta}$, the measurement vector $\boldsymbol{y}$ and the geometry matrix $\boldsymbol{A}$ as follows:

$$\boldsymbol{\theta} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ & \vdots & \\ 1 & x_j & x_j^2 \\ & \vdots & \\ 1 & x_N & x_N^2 \end{bmatrix},$$

And the χ^2 reads

$$\chi^2 = (\boldsymbol{y} - \boldsymbol{A\theta})^T \boldsymbol{V}_{\boldsymbol{y}}^{-1} (\boldsymbol{y} - \boldsymbol{A\theta}), \tag{5.2.15}$$

where $\boldsymbol{V}_{\boldsymbol{y}}$ is the symmetric error matrix of $\boldsymbol{y}$ with the elements given below:

$$\boldsymbol{V}_{\boldsymbol{y}ij} = \sum_{l=1}^{i} (x_i - x_l)(x_j - x_l) \langle \delta\alpha_l \delta\alpha_l \rangle + \sigma_i^2 \delta_{ij}, \tag{5.2.16}$$

in which $i \leq j (x_i \leq x_j)$, σ_i is the intrinsic spatial resolution of the i'th layer and $\langle \delta\alpha_l \delta\alpha_l \rangle$ is the variance of multiple scattering angle at the l'th layer with its standard deviation given in Ref. [69]. In this simple implementation, the detecting layers are perpendicular to x-coordinate, which works well for ladder sector designs.

The results of toy simulation are shown in Fig. 5.2, where the bias of parabola model could be seen for the momentum below 0.3GeV. A similar plot could also be drawn with the position of the second layer changed to $96.16mm$ and no obvious difference can be found. The verdict of the generalized least squares and Gluckstern's method is summarized below:

1. The curvature estimation has bias for low momentum trajectories as $(L/R)^2$ is too large for parabola approximation.

2. The estimation of curvature variance is always good for all the tested momentums.

3. Tracking performance is stable for different tracker structures.

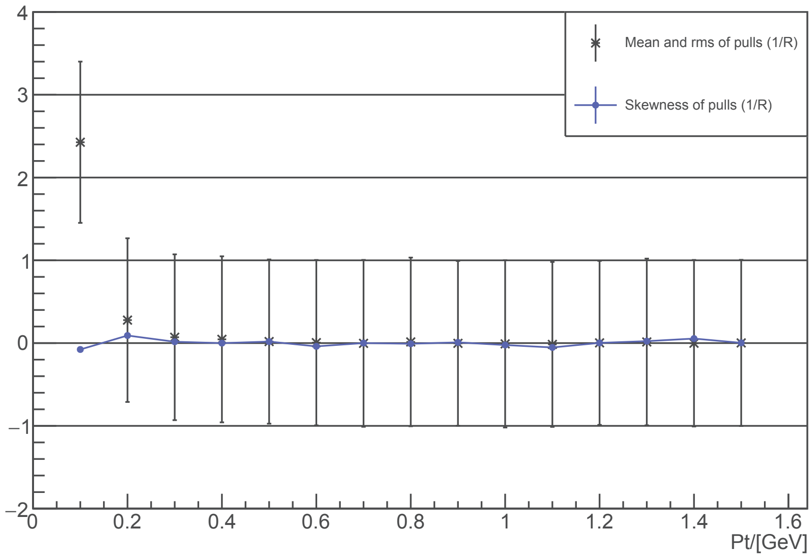


Figure 5.2: Curvature pulls vs momentum of GLS method. The pull of curvature is defined as $(k_{rec} - k_{true})/\sigma_k$, where k_{rec} is the reconstructed curvature, k_{true} is the true value and σ_k is the estimated curvature deviation given by the fitting algorithm. Mean values and standard deviations of the curvature pulls are shown with black asterisk and error bar, the skewness of pull distribution and its error are marked in blue.

5.2.2.4 The extended Kalman filter

Kalman filter based algorithms have been widely used in particle tracking for their fast speed, advantages in dealing with local process noise like material effects and the ability to do track finding. The extended Kalman filter (EKF) introduced in Ref. [70] is a non-line algorithm implemented in the *AliExternalTrackParam* class of *AliRoot* project, which is used as an inward tracking algorithm in this thesis. The tracker described in this implementation is cylindrical. The toy simulation of default geometry and the modified one with the second layer at $r = 96.16mm$ are compared in Fig. 5.3

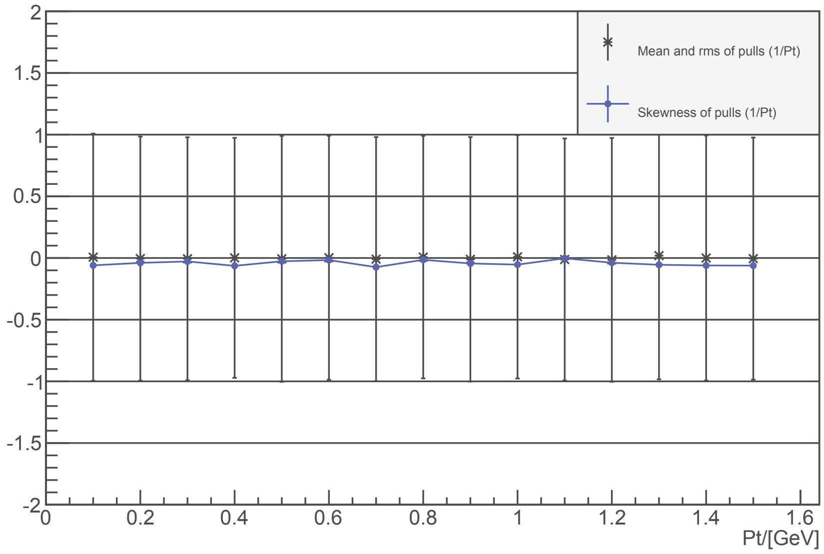
Using the same initialization, the variance of curvature estimated by the default geometry is well estimated while it is underestimated by the modified one. This result is reasonable, because one primary condition to derive Kalman gain is the proper initialization with the correct pre-knowledge about the state and its variance matrix in the first recursion. On the other hand, there is no bias of the curvature estimation, which verifies that this non-linear extension works well for circle reconstruction. Consequently, initialization values of EKF algorithm needs to be tuned for each geometry to avoid variance estimation problems, thus EKF is not convenient to be used in our geometry scan.

5.2.2.5 Geometry scan towards the optimal spacing

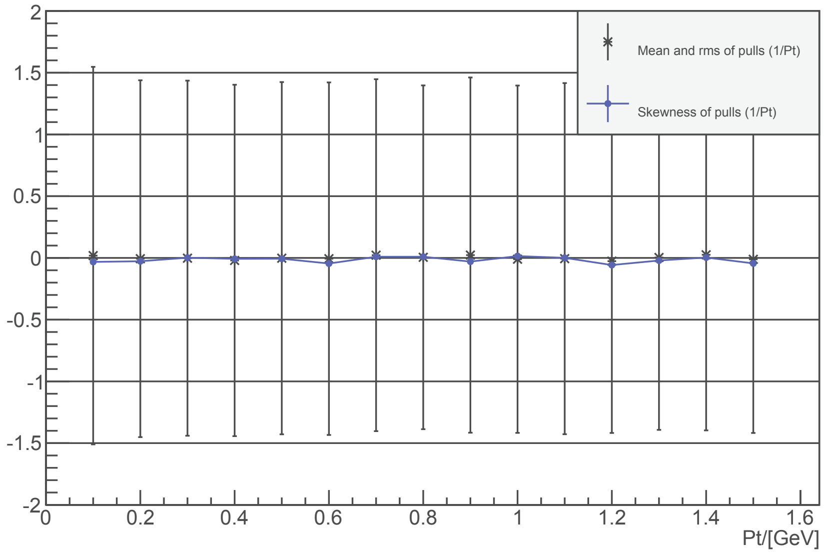
The GLS algorithm is chosen to calculate track parameter variances for geometry scan. The three-layer tracker in the previous toy simulation could be optimized very quickly. The optimal geometry is always three evenly placed layers regardless of multiple scattering. To verify the results in section 2, a six layer tracker in the same position range as that in the toy simulation is scanned in Fig. 5.4 with layer moving step $\sim 0.897\,\text{mm}$ and with a layer thickness of zero (without multiple scattering effect). The optimal configurations found from Fig. 5.4 are compared with the uniform spacing in Table 5.4. Both the layouts

Index	Configuration	Middle Position	$\sigma_K(p_T = 1GeV)$
27601	Uniform		$0.113\,m^{-1}$
4487	$2-3-1$	$85.143\,mm$	$0.092\,m^{-1}$
43330	$1-3-2$	$86.937\,mm$	$0.092\,m^{-1}$

Table 5.4: Results of the selected geometries



(a) Default geometry



(b) Modified geometry

Figure 5.3: Curvature pulls vs momentum of Kalman filter with the same initial values.

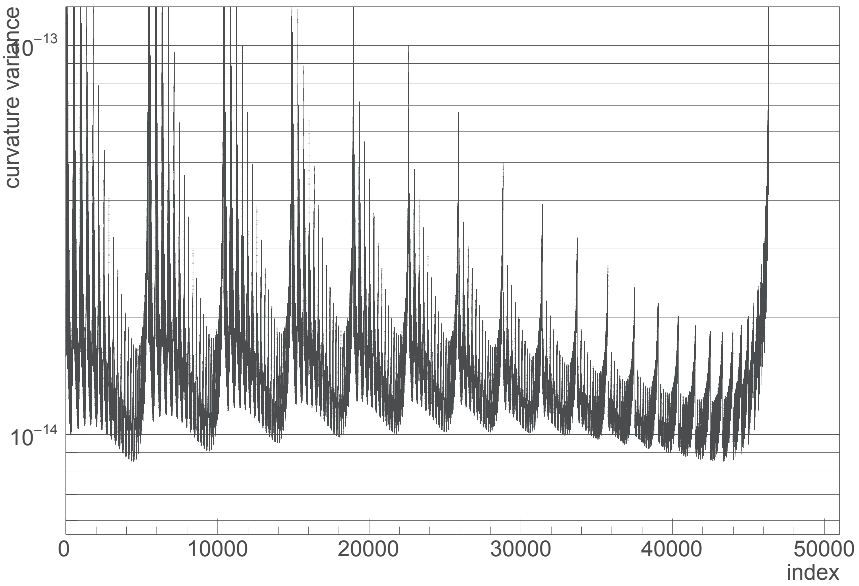


Figure 5.4: The curvature variance (μm^{-2}) as a function of geometry index given by the recursion tree for $N = 6$ and $\Delta L = \dfrac{L}{30}$

and the corrected positions[*] of the middle layers are coincident with the calculations in Section 2.

The same strategy involving multiple scattering could be processed by setting the correct layer thickness, and the application is introduced in section 4.

5.2.2.6 The choice of N, the total number of layers

Using the same spacing, Equation 5.2.12 indicates that the curvature variance is proportional to $1/N$ for very high momentum tracks. In this case, N is the larger the better if financial budget is not a problem. Nevertheless, this direct proportion is broken for relatively low momentum trajectories. Because of the rise of material effect, too many layers will smear the hits and make the tracking imprecise. To determine the number of layers, several values of N could be tried. By comparing the tracking performance of the optimal spacing for different N, the total number of layers and its related geometry could

---

[*] By using Equation 5.2.14, The corrected position of layout $2-3-1$ is $85.088\,mm$ while that of layout $1-3-2$ is $86.992\,mm$

be determined at the same time.

5.2.3 Application on BESIII Silicon pixel tracker

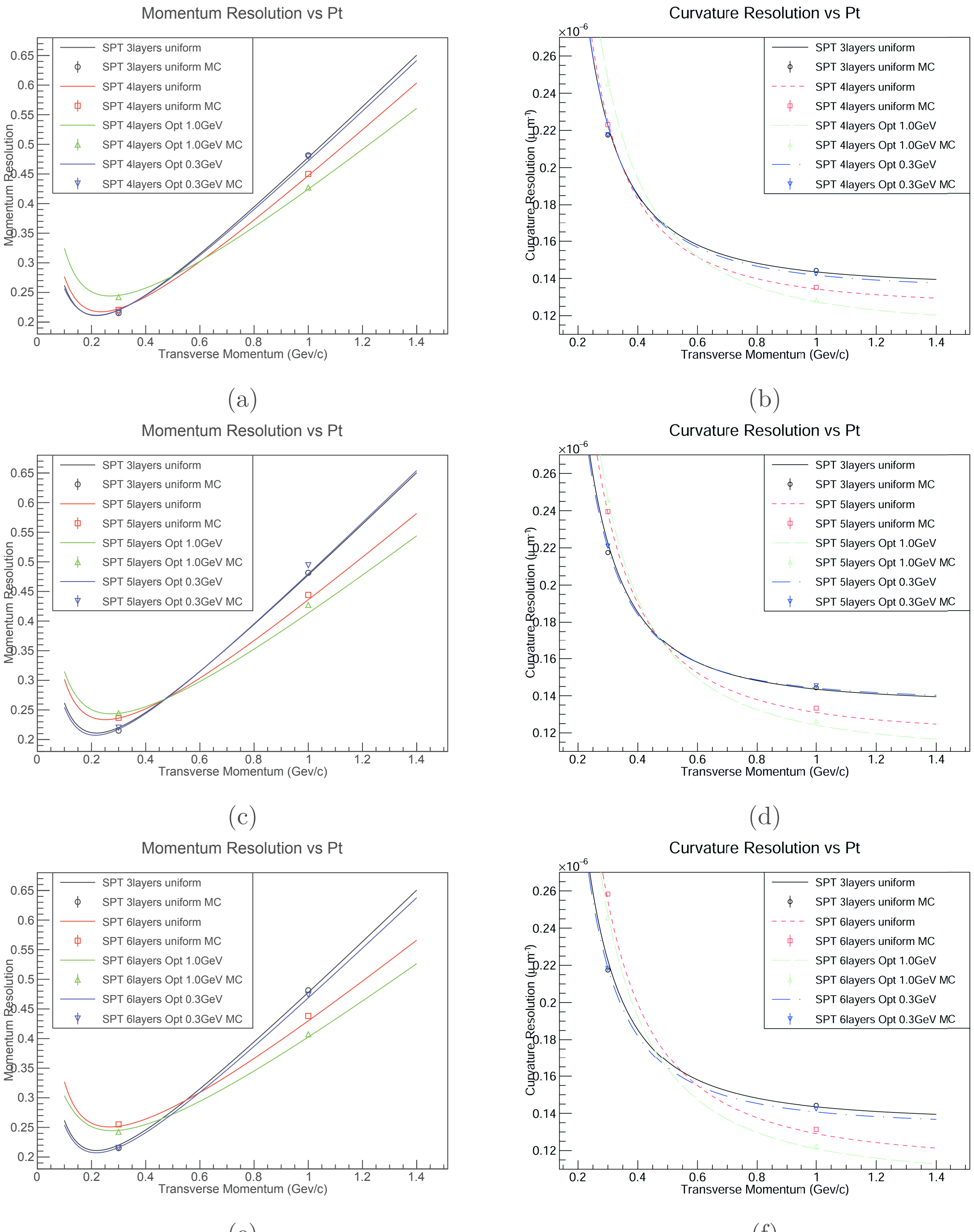
The silicon pixel tracker (SPT) is a proposal to replace the inner gas chamber of BESIII with CMOS pixel sensor ladders which have the spatial resolution around $10\,\mu m$ and the material budget $X/X_0 \sim 0.36\%$ [71]. The preliminary position range is from $72.58\,mm$ to $99.50\,mm$. In our study, the GLS tracking described in section 3 is used, therefore the ladders are simply placed perpendicular to x-axis. By using the geometry scan of N layers where N is 3, 4, 5 or 6, the layouts optimized for $0.3\,GeV$ and $1.0\,GeV$ trajectories with polar angle $\theta = 0$ are listed in Table 5.5.

N	P_T (GeV)	Step (mm)	Layer positions (mm)					
3	all	any	72.58,		86.04,			99.50
4	0.3	0.45	72.58,	72.58,		85.59,		99.50
4	1.0	0.45	72.58,	84.25,		87.83,		99.50
5	0.3	0.67	72.58,	72.58,	78.64,	90.75,		99.50
5	1.0	0.67	72.58,	72.58,	83.35,	87.39,		99.50
6	0.3	0.90	72.58,	72.58,	79.76,	92.32,	99.50,	99.50
6	1.0	0.90	72.58,	72.58,	84.25,	87.83,	99.50,	99.50

Table 5.5: Optimized geometries for SPT

The momentum resolutions σ_{p_T}/p_T calculated by GLS and the results from full Monte-Carlo simulations are shown in Fig. 5.5. The predictions given by our strategy are in accord with the simulated data, which indicate that the momentum resolutions of all the geometries optimized for 0.3 GeV are similar and the 3-layer layout or the 4-layer one with uniform spacing are more favoured by the tracks with the most probable momentum around 0.3 GeV. If the physics at 1.0 GeV is more interesting, the last geometry with 6 layers in Table 5.5 could be considered as the baseline design.

In addition, the usage of the aggressive double-sided ladders with a spatial resolution $\sim$ 10 μm was also studied. The comparison of momentum resolution among several layouts



(a) (b)

(c) (d)

(e) (f)

Figure 5.5: Comparison of momentum and curvature resolutions as a function of momentum of pions flying in the bending plane. The lines denote the predictions from GLS fitting of different geometries in which layer overlap is permitted while the dotes with error bars are used for MC results where the minimum layer interval is 1 mm.

is shown in Figure 5.6. When double-sided ladder used as the innermost layer [*], the performance on tracking is close to that of the 3-layer uniform configuration with a small improvement on high momentum range. When the PLUME-like ladder placed at the centre between the two normal ladders, the tracking performance is better than that of the 4-layer layout optimized for 1.0 GeV. Nevertheless, the 3-layer uniform geometry is still a better choice in the low momentum range (below 0.3 GeV for the default material budget and 0.5 GeV for ladders with doubled material budget).

5.3 Geometry optimization for vertexing

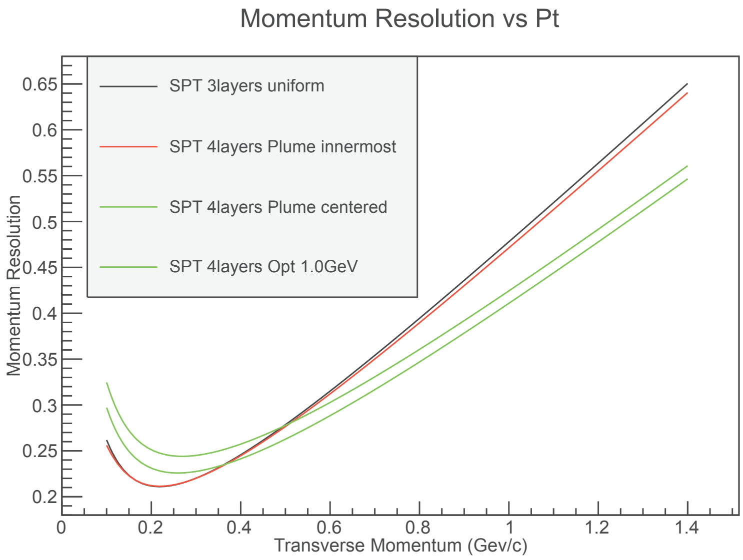
The momentum resolution and impact parameter resolution of a detector can not be optimized at the same time. In this section, the optimization of the inner tracker based on the impact parameter resolution is discussed. The related telescope systems were studied first with straight tracks, and the parabola model was used latter for the final decision. As the impact parameter resolution or the resolution of the distance to the closest approach (DCA) is the figure of merit in this study, we introduced the advanced ladder technique of the PLUME project, which has MIMOSA-26 glued on both surfaces of the ladder with a total material budget $\sim 0.35\% X_0$ (PLUME-2). As for our studies, $0.362\% X_0$ is the estimated budget for a normal ladder and $0.513\% X_0$ is used for a double-sided one. Two detector geometries were compared: The 3-layer uniform geometry used in the previous section and the 4-layer layout including one PLUME-like double-sided ladder placed just after the beam pipe. The two layouts including the beam pipe of BESIII, are shown in Figure 5.7

5.3.1 Radii optimization in a telescope system

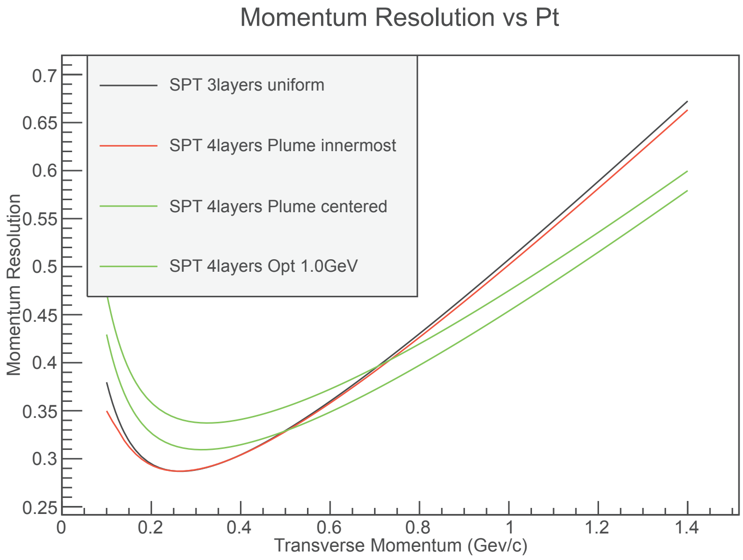
When multiple scattering is not dominating, the simplified telescope resolution equation could give some hints in detector design. In beam test, the tracks are straight lines, such

---

[*] In this case, the normal ladder in the middle is placed at $r \approx 85.4$ mm to achieve the best momentum measuring ability.

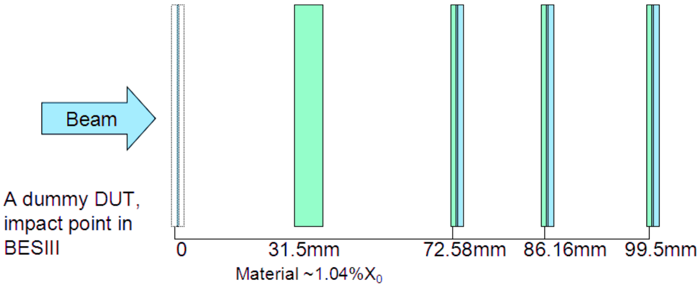


(a) Ladders with default material budget.

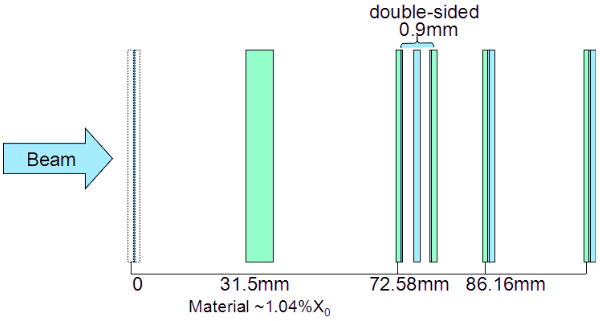


(b) Ladders with a doubled material budget.

Figure 5.6: Comparison of momentum resolution as a function of momentum of pions in the bending plane. The black solid line and the green dashed one denote the 3-layer uniform layout and 3-layer configuration optimized for 1.0 GeV respectively. For geometries including one double-sided ladder, the red solid and green solid curves are used for the PLUME-like ladder placed at the innermost position and at the centre respectively.



(a) Layout of three normal ladders.



(b) Layout of four layers.

Figure 5.7: Telescope configurations extracted from BESIII inner tracker.

that the covariance matrix in Equation 4.1.9 will only keep the upper left 2×2 sub-matrix.

$$\begin{aligned}\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} &= \begin{bmatrix} \sum\limits_{j=1}^{N} 1/\sigma_j^2 & \sum\limits_{j=1}^{N} x_j/\sigma_j^2 \\ \sum\limits_{j=1}^{N} x_j/\sigma_j^2 & \sum\limits_{j=1}^{N} x_j^2/\sigma_j^2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \sum\limits_{j=1}^{N} x_j^2/\sigma_j^2 & -\sum\limits_{j=1}^{N} x_j/\sigma_j^2 \\ -\sum\limits_{j=1}^{N} x_j/\sigma_j^2 & \sum\limits_{j=1}^{N} 1/\sigma_j^2 \end{bmatrix} \frac{1}{\sum\limits_{j=1}^{N} 1/\sigma_j^2 \sum\limits_{j=1}^{N} x_j^2/\sigma_j^2 - (\sum\limits_{j=1}^{N} x_j/\sigma_j^2)^2}\end{aligned} \tag{5.3.1}$$

At $x = x_{DUT}$, the extrapolated resolution is given below:

$$\begin{aligned}\sigma_{DUT}^2 &= V_{11} + x_{DUT}^2 V_{22} + 2x_{DUT}V_{12} \\ &= \frac{\sum\limits_{j=1}^{N} (x_{DUT} - x_j)^2/\sigma_j^2}{\sum\limits_{j=1}^{N} 1/\sigma_j^2 \sum\limits_{j=1}^{N} x_j^2/\sigma_j^2 - (\sum\limits_{j=1}^{N} x_j/\sigma_j^2)^2}\end{aligned} \tag{5.3.2}$$

The design optimization of an inner tracker in BESIII is similar to that of a single arm telescope used for a device located at $x_{DUT} = 0$. To simplify the problem, a proper assumption to choose detector layers with the same intrinsic spatial resolution σ and to set up them at $x_j > 0$ in ascending order ($x_1 < x_2 < \ldots < x_N$), has been applied. Therefore, the simplified telescope resolution at x_{DUT} reads

$$\sigma_{DUT}^2 = \sigma^2 \frac{\sum\limits_{j=1}^{N} x_j^2}{N \sum\limits_{j=1}^{N} x_j^2 - (\sum\limits_{j=1}^{N} x_j)^2}$$

Its partial derivative with respect to the position x_k of a target layer in the telescope could be written.

$$\frac{\partial \sigma_{DUT}^2}{\partial x_k} = \sigma^2 \frac{2 \sum\limits_{j=1}^{N} x_j \sum\limits_{j=1}^{N} [x_j(x_j - x_k)]}{\left[N \sum\limits_{j=1}^{N} x_j^2 - (\sum\limits_{j=1}^{N} x_j)^2\right]^2} \tag{5.3.3}$$

Equation 5.3.3 could be used to optimize the position of the target layer when the other layers are fixed.

- When the target layer is the first one ($k = 1$), we have

  $$\frac{\partial \sigma_{DUT}^2}{\partial x_1} > 0$$

  σ_{DUT}^2 is an increasing function of x_1, so the first layer should be as close to the impact point as possible.

- When the last layer is to be optimised ($k = N$), we have

  $$\frac{\partial \sigma_{DUT}^2}{\partial x_N} < 0$$

  σ_{DUT}^2 is a decreasing function of x_N, such that the radius of the last layer is the larger the better.

- When k index stands for an intermediate layer ($1 < k < N$), the partial derivative of σ_{DUT}^2 in Equation 5.3.3 changes from a positive value to a negative one as x_k increases in the range $x_1 < x_k < x_N$. Additionally this x_k is also constrained by the radii of its neighbour layers, so the optimised position of the k'th layer should be close to one of its neighbour.

In the case that multiple scattering cannot be ignored, numerical calculations could be processed for the comparison among different geometries. The corrections on track parameter variances have been introduced in the generalized least squares method described in Chapter 4.

5.3.2 Pointing resolutions of straight tracks

This study on the pointing resolution of straight tracks helps to optimize the design of a telescope system and a barrel tracker on its tracking performance in the $R - Z$ plane [*]. The pointing resolution of a DUT at $x = 0$ for the 3-plane layout shown in Figure 5.7a has been studied for different radii of the ladder in the middle. The calculations

---

[*] For low momentum tracks, the coordinates are Z and the track length in $R - \phi$

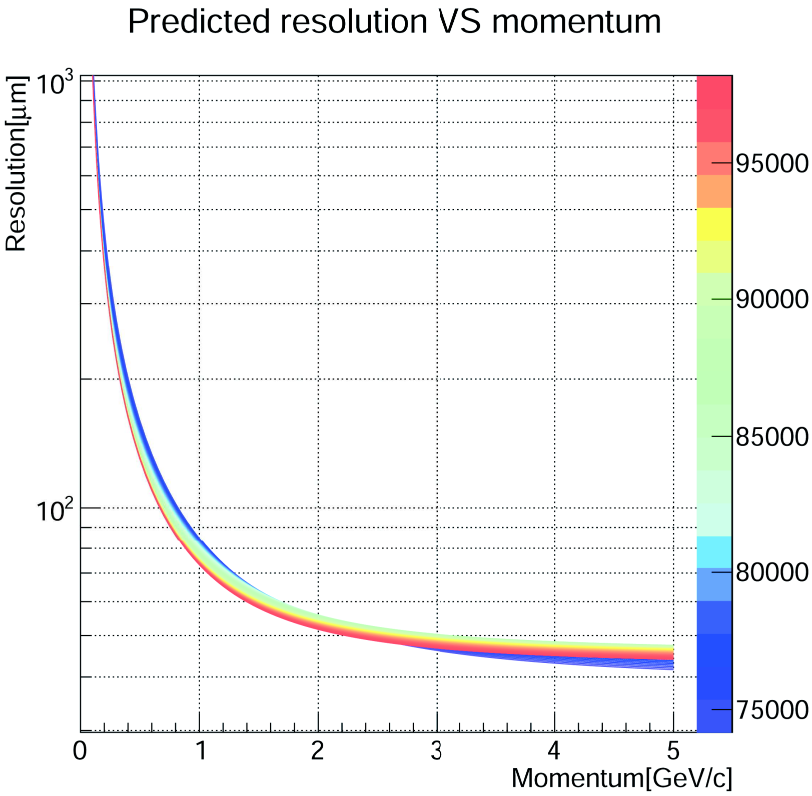


Figure 5.8: The telescope resolution as a function of momentum for the configuration of 3 normal ladders. Different colours are used for the relevant radii of the ladder in the middle. The numbers on colours of the palette are positions of the middle ladder in micros. The spatial resolutions of all layers are 10 µm.

without multiple scattering lead to the middle ladder with a position close to the first ladder or the outermost. When the material effects are considered, a simple geometry scan to change the position of the middle layer has been processed. Figure 5.8 shows the results with different colours indicating the relevant radii. The rainbow palette with colours changing from purple to red, is chosen for ladder positions (from 74.16 mm to 98.16 mm) in ascending order. The conclusions of the animation to draw this figure are listed as follows:

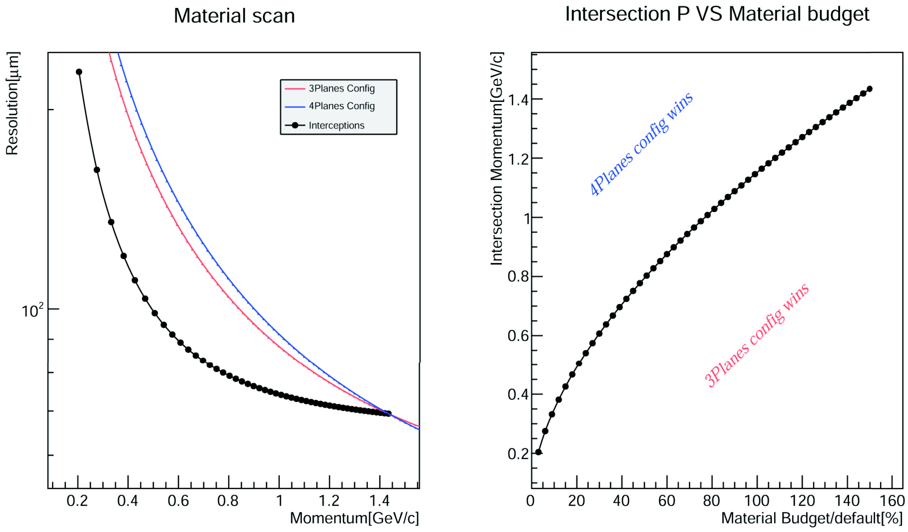
- In the low momentum range below 1 GeV, the resolution curve goes down as the radius increases. This means that the radius of the middle layer is the larger the better when multiple scattering dominates.

- In the high momentum range above 4 GeV, the resolution curve goes up first, then goes down a little, but cannot go lower than the first one. This behaviour happens for the case of negligible material effects. The best resolution is achieved for the middle ladder placed close to the innermost one, and a local minimum can be found in the case of the largest radius of the middle ladder.

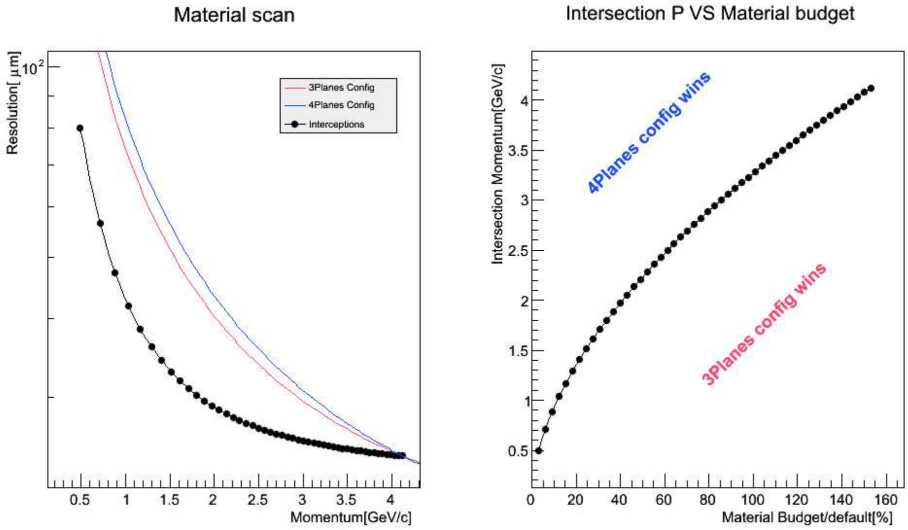
- In the moderate momentum range, the smallest radius is related to the local minimum of the resolution, and the optimal configuration is still the one with the largest radius for the middle ladder.

As for the momenta of interest for BESIII inner tracker, the layout with a middle ladder positioned close to the innermost is favoured.

The limit to place two ladders close to each other is just the double-sided ladder design. When the innermost ladder of the 3-layer geometry is replaced by a double-sided ladder, the new 4-layer configuration will have a better tracing performance when multiple scattering is not severe. However, the additional material may cause problems for low momentum tracks. To study the trade-off between the added hit information and the additional material, a material budget scan is processed to compare the pointing resolutions of the two telescope configurations shown in Figure 5.7. Moreover, the influence of different incident angles of the tracks could also be studied using this method. Figure 5.9a and Figure 5.9b show the results of pion beams in the material budget scan of the telescope systems with layer spatial resolution of 10 µm and 3.5 µm respectively. The plots on the left show the method to find an intersecting momentum at which the resolutions of the two configurations are the same for a given material budget. As the ladder budget goes down from 150% of the default design to 3%, the blue curve goes down faster than the red one, which leads to smaller intersecting momenta (in descending order). The plots on the right are shown to illustrate the momentum range in which one telescope system has a better pointing resolution. In the bottom right region of the black curve, the 3 normal ladder geometry overcomes the 4-layer layout, which means the effects from the additional materials eliminate the contributions of the added hit. As the material budget increases, the momentum range in which the 3-layer configuration wins grows larger. Comparing these results of layers with different intrinsic spatial resolutions, the system is more sensitive to multiple scattering for more accurate sensors. The momentum range favoured by the 3-layer geometry with 3.5 µm sensor resolution is therefore larger than



(a) Spatial resolution of 10 µm.



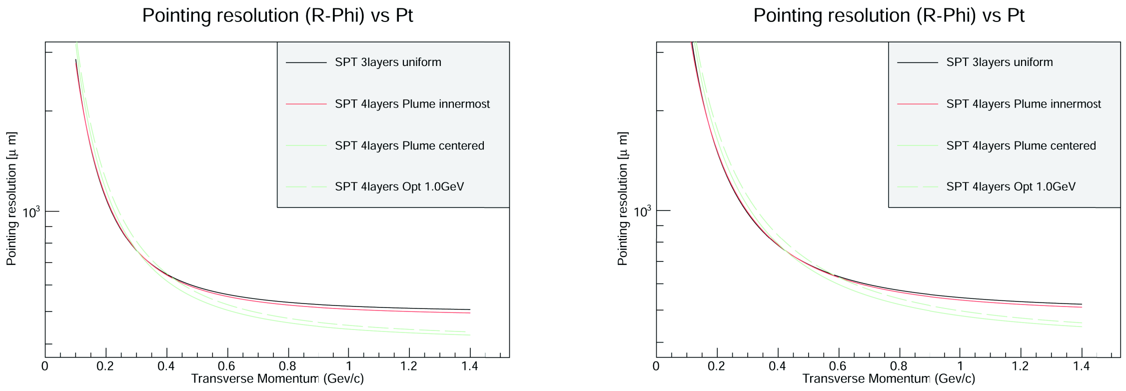
(b) Spatial resolution of 3.5 µm.

Figure 5.9: Results of material budget scan using pion beams. The blue curve is the resolution as a function of momentum for the 4-layer telescope while the red is for the 3-layer one. The black dots are the intersecting momenta at which the two telescope systems give the same pointing resolution at $x = 0$ for a certain material budget. The material budget of each ladder changes from 150% of the default design down to only 3%.

that with 10 µm resolution for the same material budget. In addition, the resolutions for a high material budget is also related to the tracking performance of tracks with an incident angle [*] larger than 0.

5.3.3 Pointing resolutions from parabola tracks

In the transverse momentum range above 0.3 GeV, the parabola track model is a good approximation for tracking in the bending plane. Therefore, the vertexing performance of BESIII inner silicon tracker can be estimated using the least squares fitting. Since the materials effects of the beam-pipe does not influence the choice of tracker layouts, it is removed from the study in this section. Figure 5.10a shows the pointing resolution in $R-\phi$ for pion trajectories with a fixed polar angle $\theta = 90^\circ$ in BESIII coordinate system, and Figure 5.10b shows that resolution of ladders with doubled material budget to estimate the performance of a thicker ladder design or the tracking performance of a smaller polar angle[†].



(a) Ladders with default material budget.

(b) Ladders with a doubled material budget.

Figure 5.10: Pointing resolution in $R-\phi$ for pion tracks in the bending plane.

Consequently, the pointing resolution of the 3-layer uniform layout is similar to that of the 4-layer configuration containing one double-sided ladder placed at the innermost position. The two configurations have a better performance for low momentum pions, and

---

[*] The incident angle is zero for the tracks perpendicular to the ladders.

[†] For a track with a polar angle θ, the multiple scattering contributions on the impact parameter resolution in $R-\phi$ is $\sim \theta_0/\sqrt{\sin\theta}$, which is related to a material budget $X/\sin\theta$.

their advantage is enlarged at small θ or with higher material budget. The two geometries marked in green have a better resolution when momentum is higher than 0.5 GeV.

5.4 The performance including MDC outer chamber

The standalone performance of the inner tracker has been studied in the previous sections. Combined with MDC outer chamber, the overall performance leads to the final decision of the design geometry. The outer chamber redefines the total radial length L, therefore changes the results dominated by L. The estimation of the overall tracking performance was studied using the semi-analytical tool first developed in STAR collaboration and then improved by the ALICE collaboration [44]. The P_T and pointing resolutions are evaluated based on EKF inward tracking, and the efficiency is estimated using the pseudo-efficiency introduced in Chapter 4.

Estimations of four BESIII tracking designs are compared: the results of MDC inner and outer chamber * are drawn with dashed blue lines, the performance of the 3-layer uniform SPT with MDC outer chamber is shown with dashed red, the curves of the 4-layer SPT (using a double-sided ladder as the innermost) with MDC outer chamber are identified with solid red, and the results of the two double-sided ladder SPT with MDC outer chamber are presented using solid green lines.

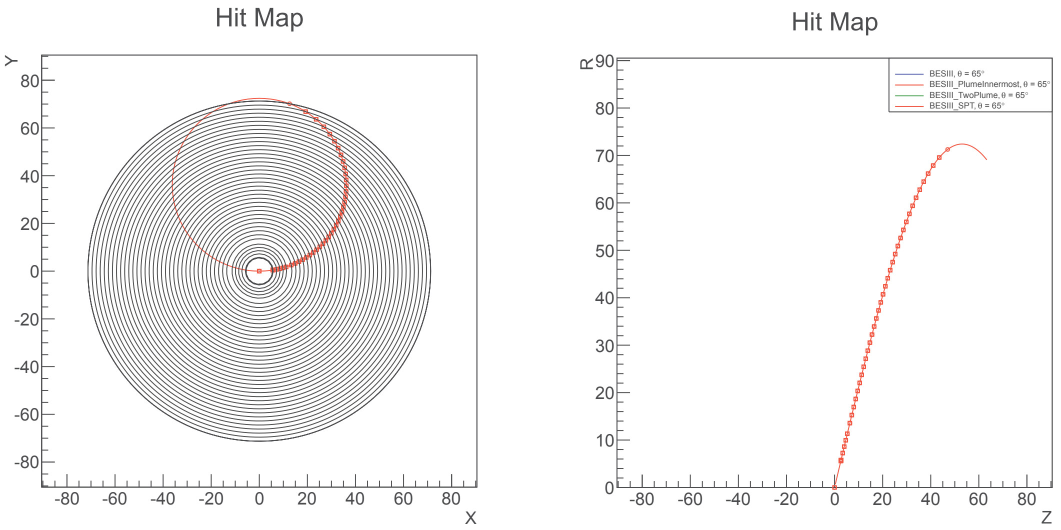
In this tool, hit positions are calculated for each tracker geometry. Since the original version may generate a wrong hit position due to numerical or geometry issues, some modifications including a hit map implementation are added into this fast tool. Figure 5.11 shows a track in the 3-layer uniform SPT geometry.

The pointing resolutions in $R-\phi$ and in Z of these designs are shown in Figure 5.12. The performance improvement from the double-sided ladders can be seen only in high momentum range ($P_T > 1$ GeV). The SPT with three normal CPS ladders have the best pointing resolutions in the momentum range of interest for BESIII.

The momentum resolutions of the four geometries are shown in Figure 5.13. The nominal SPT inner tracker has the best performance for $P_T < 1$ GeV. Since the parameters of the pure MDC chamber used in this study are from BESIII technical design report[1], the performance shown with the blue curves is better than the current MDC which suffers

---

*Single wire resolutions of 130 µm in $R-\phi$ and 6 mm in Z are applied for all the layers of MDC for the sake of simplicity.



(a) Hit map in $R-\phi$.

(b) Hit map in $R-Z$.

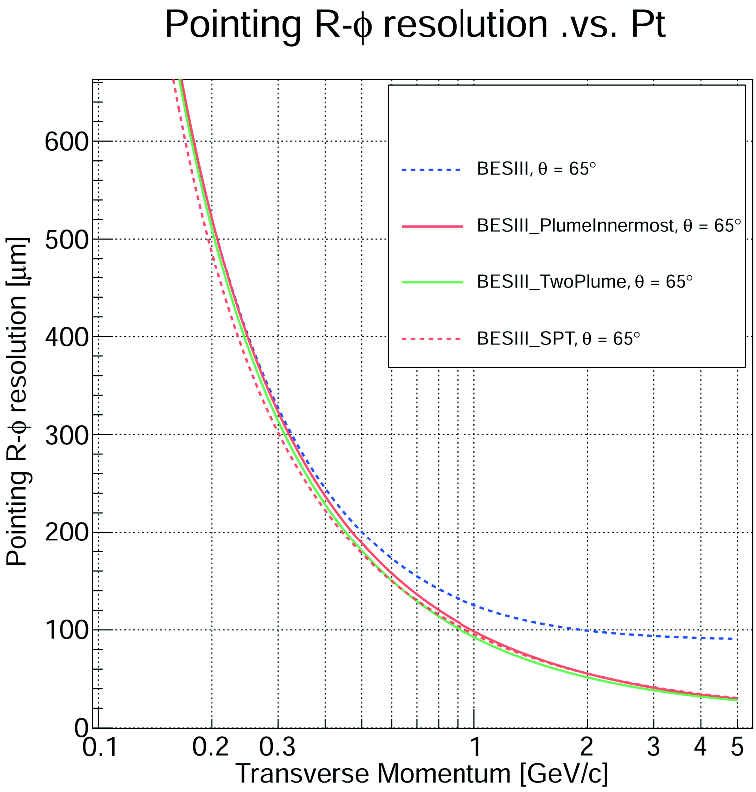
Figure 5.11: Hit maps of a track in BESIII SPT+MDC geometry. The results are calculated from pion tracks with a fixed $\theta = 65°$. R is the coordinate of the projected hit in $R-\phi$ plane. The unit is cm.

from ageing effects. Compared with the pure MDC design, a 10% improvement on the relative momentum resolution can be found at $P_T = 1\,\mathrm{GeV}$ for SPT.

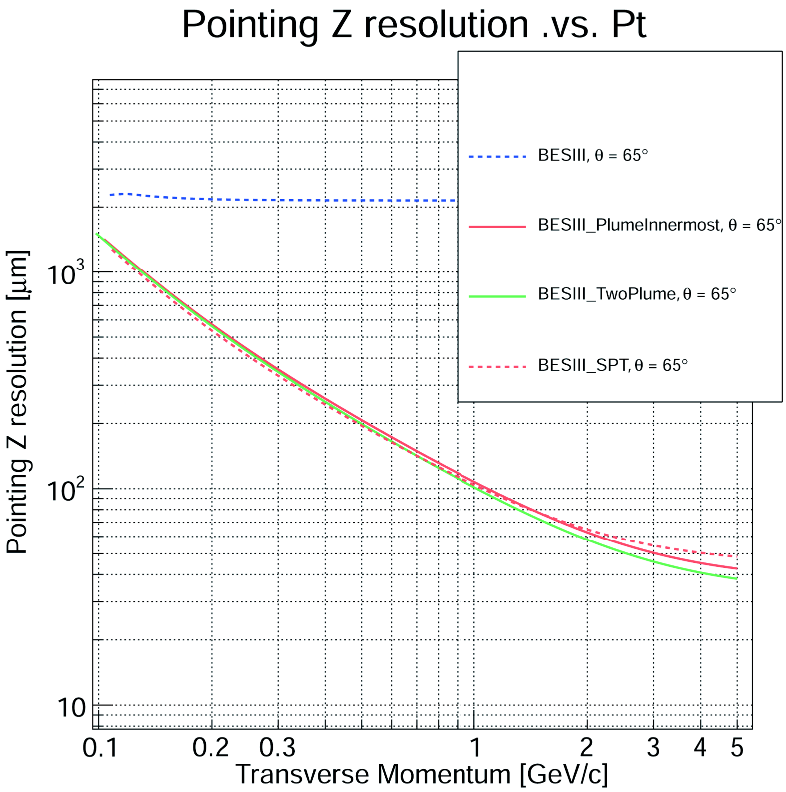
Finally, the pseudo-efficiencies of tracking for designs using silicon trackers have been compared in Figure 5.14. The results are estimated using a combination of inward and outward tracking, thus the pseudo-efficiency of double-sided ladder designs is better even if good hit-track associations are required for all the sensing layers (one more good association than the 3 normal ladder layout). This effect can also explain the advantage of the mini-vector reconstructed from the double-sided ladder for its ability to reduce noise hits. If only one good hit (not a noise) is required for a double-sided ladder, the efficiency improvement is even higher than that shown in Figure 5.14.

5.5 Summary and conclusion

A general strategy to optimize the design of a barrel pixelated tracker has been presented in this Chapter. The methods to search for the optimal spacing of N-layer silicon trackers with or without multiple scattering have been derived and verified by comparing the



(a) Pointing resolution on $R-\phi$.



(b) Pointing resolution on Z axis.

Figure 5.12: Pointing resolutions of different BESIII tracker designs. The results are estimated from pion tracks with a fixed $\theta = 65°$

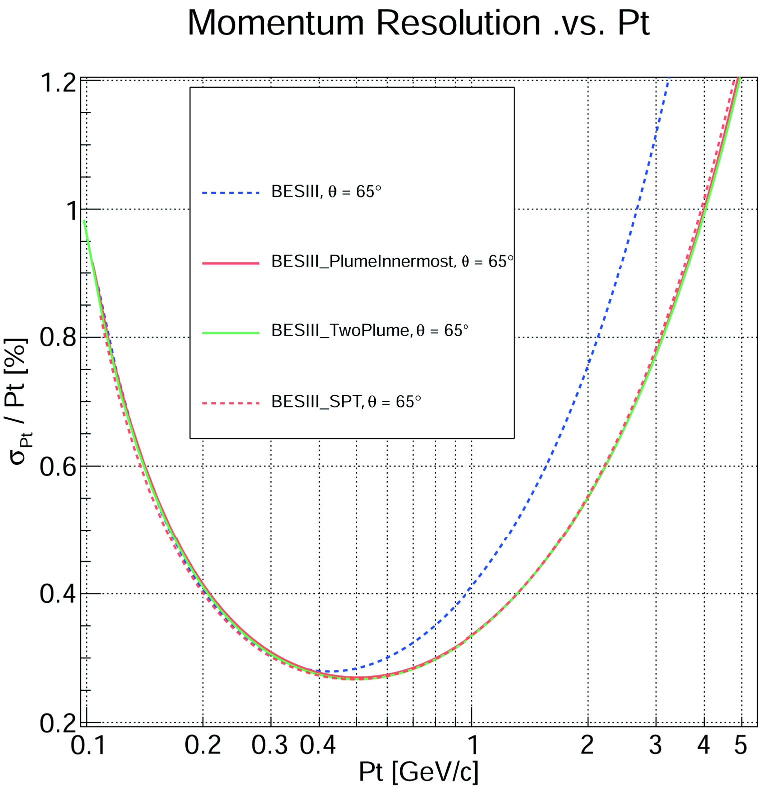


Figure 5.13: Relative P_T resolution as a function of P_T for BESIII tracker designs. The results are estimated from pion tracks with a fixed $\theta = 65°$.

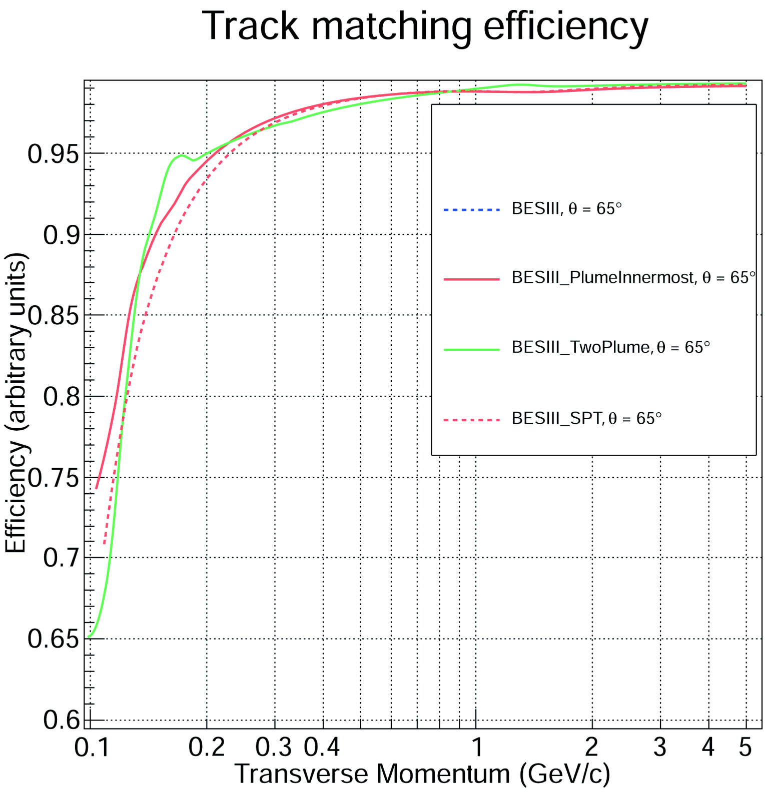


Figure 5.14: Pseudo-efficiency as a function of transverse momentum for different BESIII tracker designs. The results are estimated from pion tracks with a fixed $\theta = 65°$.

results of GLS tracking and the analysis of full Monte-Carlo simulations in the simplified situation where trajectories are circle segments. As for the helix tracks, the corrections of the track length in material could be calculated from the azimuth and the polar angle to estimate multiple scattering more accurately. The requirements of tracking algorithms to be used in geometry scan have also been proposed. Since the numerical scan is compatible with the other tracking methods, the momentum range of the tracks could be extended and the energy loss effect could also be involved by choosing a proper fitting algorithm. Finally, the total number of layers could also be determined by comparing the optimal spacing of different N in the geometry scan.

Our methods have been applied to optimize the design of BESIII inner silicon tracker. The standalone performance of the CPS tracker was studied. In the case that only normal ladders are used, the optimal geometry for momentum (below 0.5 GeV) measurements is the layout including three layers with uniform spacing. This geometry is used as the nominal design of SPT. The impact of the aggressive double-sided ladders has also been estimated. Although the improvements on resolutions are for the momentum range which is higher than the range of BESIII final states, the improvements on tracking efficiency could be foreseen.

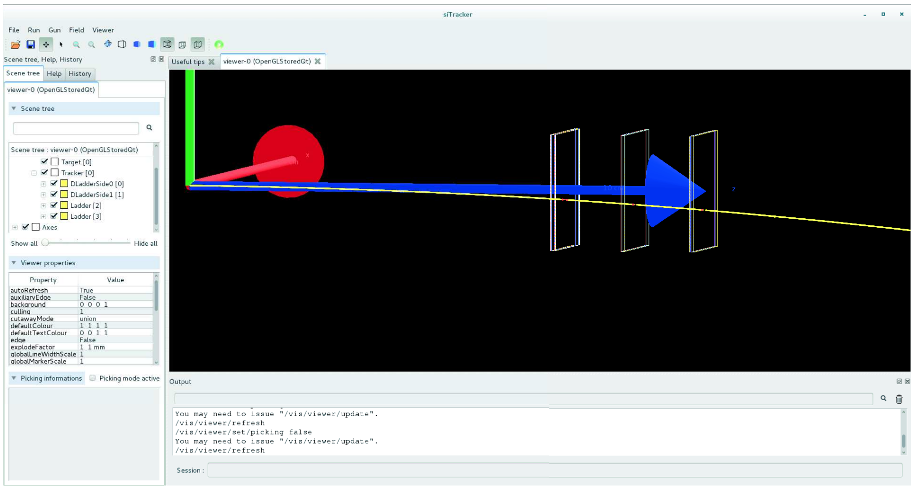
Chapter 6

Full Monte-Carlo Simulation studies on CPS inner tracker

The GEANT4 based simulations have been processed to validate the optimized BESIII silicon pixel tracker. The results for the standalone performance of SPT have already been shown in previous chapters, and the software (SiTracker) developed for this study will be presented in this chapter. The nominal design of SPT has been implemented and simulated in BESIII offline software system by authors in literature [71]. In this thesis, simulations using single pion events have been processed for the nominal SPT with the position of the middle ladder modified. This study is used to validate the conclusions of the analytical tools and to explore the possible improvements of the three-layer structure.

6.1 Software for the standalone simulation of the inner tracker

The software SiTracker is developed based on GEANT4[72, 73, 74] and ROOT [75]. Normal ladders are constructed inside of a tracker volume filled with air. The double-sided ladder is created using two normal ladders placed back-to-back with a complete overlap for the carbon fibre support. Since all the ladders have the same structure, they are constructed at the positions stored in a C++ vector. Therefore, new ladders could be inserted through GEANT4 command line interface in *SiDetectorMessenger* class (derived from *G4UImessenger* class). The simulation of one track in the 4-layer geometry with a double-sided ladder as the innermost, is shown in Figure 6.1.

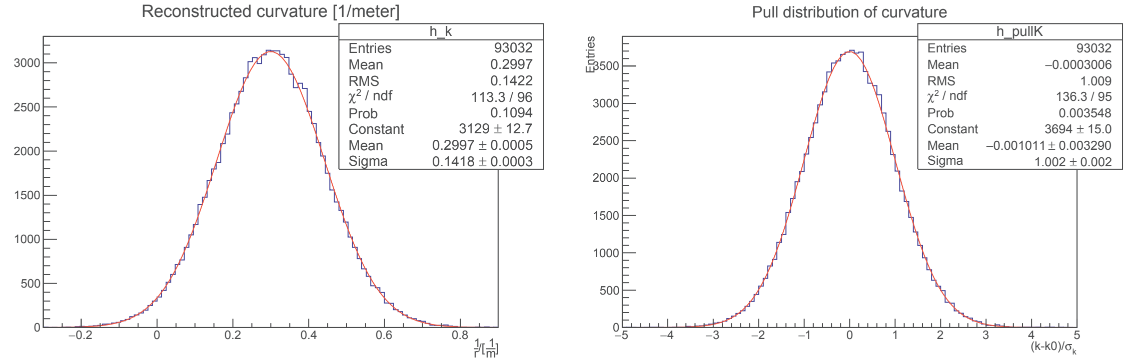


Property	Value
autoRefresh	True
auxiliaryEdge	False
background	0 0 0 1
culling	1
cutawayMode	union
defaultColour	1 1 1 1
defaultTextColour	0 0 1 1
edge	False
explodeFactor	1 1 mm
globalLineWidthScale	1
globalMarkerScale	1

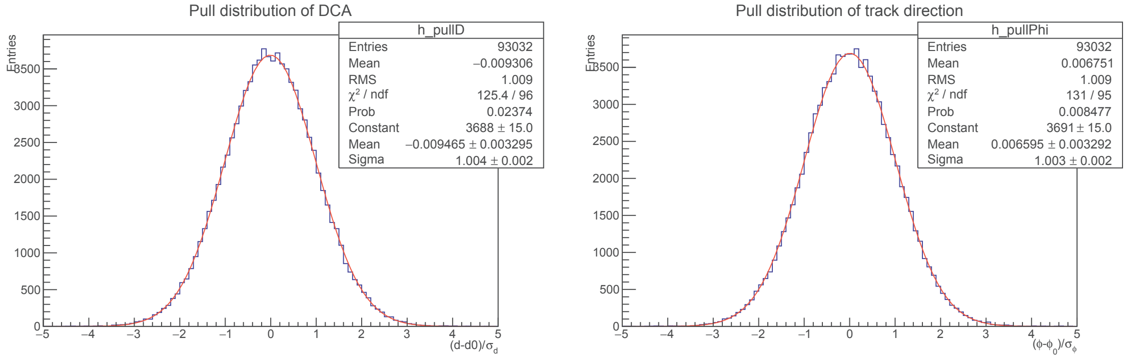
```
You may need to issue "/vis/viewer/update".
/vis/viewer/refresh
/vis/viewer/set/picking false
You may need to issue "/vis/viewer/update".
/vis/viewer/refresh
```

Figure 6.1: One event in the software: SiTracker.

Compiled with ROOT, SiTracker can generate a *ROOT* file which contains a *TTree* of run number, event number, number of hits, particle id, ladder number, the position and the momentum vector of each hit. The simulation for the intrinsic spatial resolution of CPS is applied later in the analysis using a Gaussian smearing on MC true hit. Finally, the data is analysed with the tracking software developed in the optimisation studies. The tracking performance of the geometry shown in Figure 6.1 for 1 GeV pions has been summarized in Figure 6.2. Curvature is used as a track parameter because it is a Gaussian variable, which can be found from Figure 6.2a. A pion with $P_T = 1\,\text{GeV}$ has a large radius around 3.3 m, therefore, it is possible to reconstruct the track with an opposite (a wrong) bending direction, a negative curvature, due to the limited lever arm length of SPT. To estimate curvature resolution in a good accuracy, these tracks (the tail at $k < 0$ in Figure 6.2a) are not excluded. The pull distributions of the three track parameters are also shown in the sub-figures. All of them have a mean value around 0 and a standard deviation ~ 1, which means the tracking algorithm works well.



(a) Curvature distribution (b) Pull distribution of curvature.



(c) Pull distribution of DCA (d) Pull distribution of track direction

Figure 6.2: Analysis results of MC data.

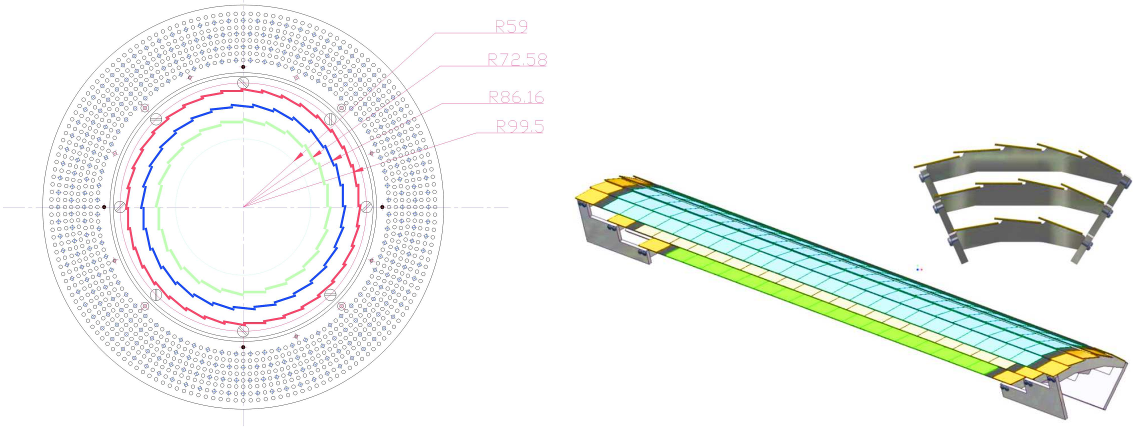
6.2 Simulations in BESIII offline software system

6.2.1 BESIII Offline Software System

BESIII Offline Software System (BOSS) [76] is the data processing and analysis software developed for BESIII experiment. The framework is built on the Gaudi [77] architecture to handle the GEANT4-based simulation, the calibration developed from the GLAST's scheme [78], the reconstruction system for all sub-detectors and the final physics analysis.

6.2.2 Implementation of CPS inner tracker

The silicon pixel tracker (SPT) based on simple ladder structures and the related reconstruction algorithms have been implemented into a branch of BOSS, named as SUPERBOSS [71]. The SPT cross-section view and a sector diagram are shown in Figure 6.3. In the preliminary design, the radii of the three layers are 72.58 mm, 86.16 mm and 99.5 mm respectively. The related ladder lengths are 38 cm, 45 cm and 52 cm. For track reconstruction, the combinatorial Kalman filter (CKF) [79, 80] is implemented for SPT.



(a) $R - \phi$ view (b) SPT sector

Figure 6.3: Ladders in SPT.

6.2.3 Simulation Samples and results

Single pion events are used in the simulation of SPT-MDC tracking system for geometry optimization studies. The pions are shot with a polar angle $\theta = 80°$ in the 1 Tesla magnetic field. To achieve the averaged performance of the tracker, the transverse momentum has a ϕ ranges from 0° to 360°. Three values of the radius of the middle layer are tried: 78 mm, 87 mm and 95 mm. A sensor resolution 10 µm is applied.

In the analysis, all the distributions are fitted using double Gaussian distributions to cope with the influence from non-Gaussian tails. The function used in this thesis is shown below:

$$f(x) = \frac{N}{\sqrt{2\pi}} \left[\frac{f_1}{\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \frac{1-f_1}{\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \right] \tag{6.2.1}$$

Figure 6.4 shows an example of the fit for DCA of 1 GeV pions in $R-\phi$ plane. The standard deviation is then given by:

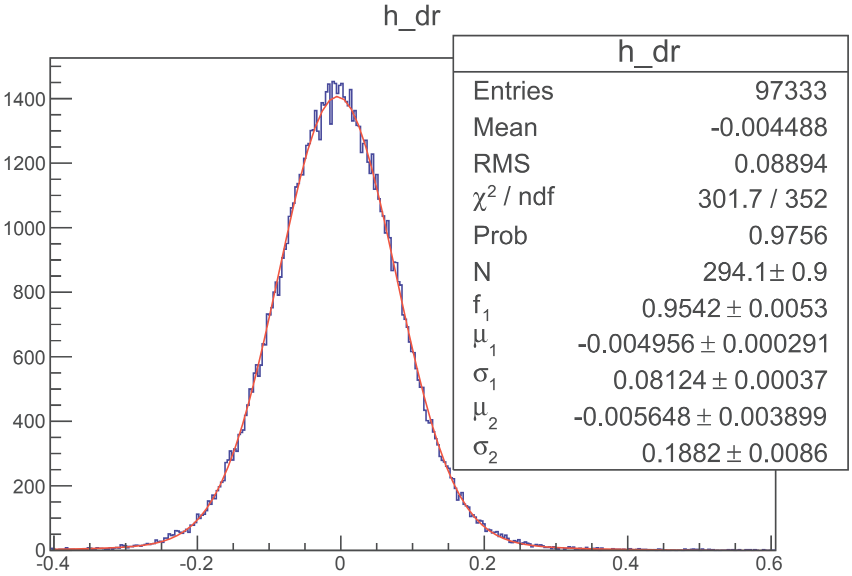


Figure 6.4: DCA [mm] of 1 GeV pions in $R-\phi$.

$$\sigma = \sqrt{f_1\sigma_1^2 + (1-f_1)\sigma_2^2} \tag{6.2.2}$$

The results of P_T resolution are shown in Figure 6.5, together with the BESIII design goal in a dashed line which is described by [81]:

$$\sigma_{P_T}/P_T = (0.32\% P_T) \oplus (0.37\%/\beta) \tag{6.2.3}$$

The performances of all the tree geometries have fulfilled BESIII requirements in the range

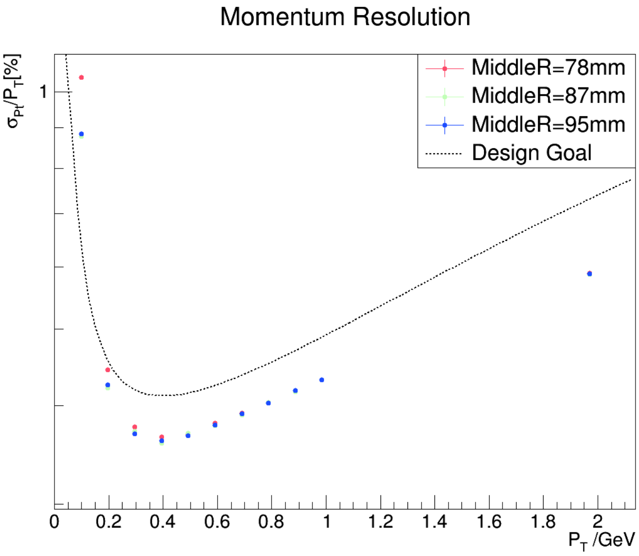


Figure 6.5: P_T resolution as a function of transverse momentum. The tracks are from pions of a fixed $\theta = 80°$.

$P_T > 0.2\,\mathrm{GeV}$. A smaller radius of the middle layer may lead to a worse P_T resolution when $P_T < 0.3\,\mathrm{GeV}$. A similar study using single muon events, has been published for the original MDC (without ageing) in article [81]. Its result is shown in Figure 6.6. The original MDC only fulfilled the design goal at $P_T \geq 0.3\,\mathrm{GeV}$.

DCA resolutions in $R - \phi$ are shown in Figure 6.7a. As the radius of the middle layer increases, the DCA resolution improves at $P_T < 2\,\mathrm{GeV}$. In Figure 6.7b, the factor of this improvement is only a few per cent at $P_T \sim 1\,\mathrm{GeV}$. The results are coincident with the conclusions of the rainbow plot drawn for a radius scan in the previous chapter.

Efficiencies based on EKF inward tracking are shown in Figure 6.8. The results of different geometries are similar. But a slightly worse performance can be found for a smaller radius of the middle layer.

6.3 Summary and conclusion

In this Chapter, the program for the standalone simulation and the related analysis of the silicon ladders have been introduced. We developed each piece of the codes almost from scratch. The analysis results have been compared to the calculations, and they are in very

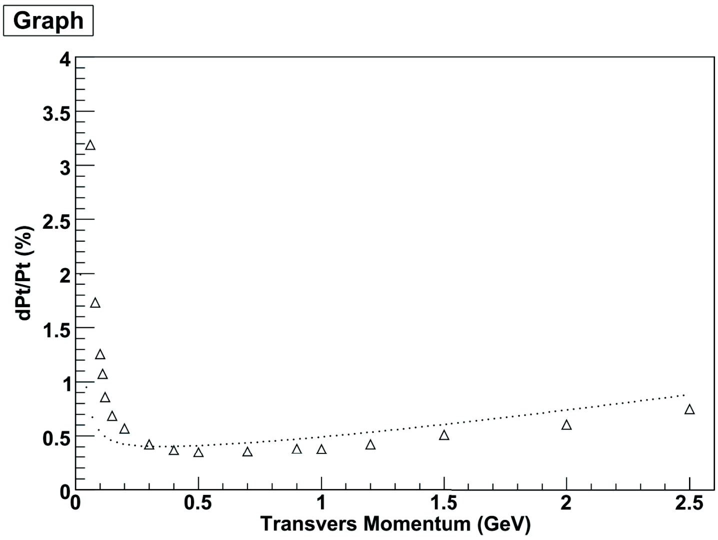
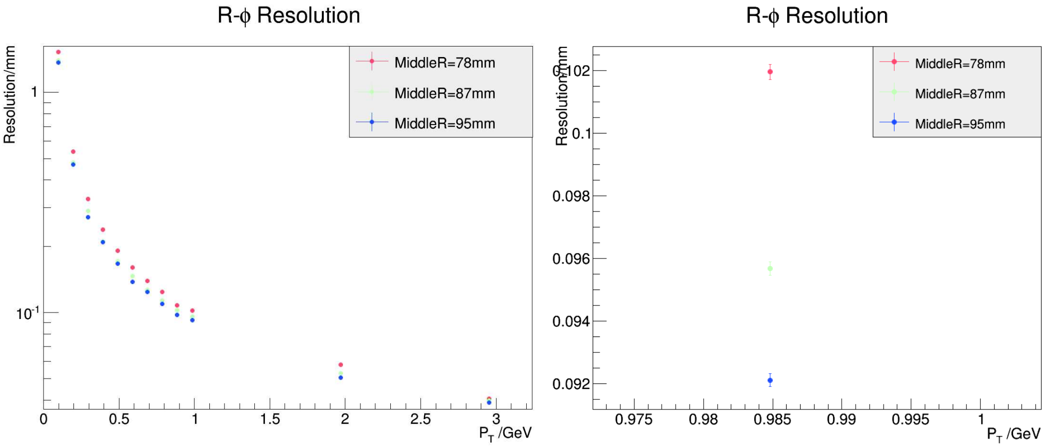


Figure 6.6: P_T resolution as a function of transverse momentum for single muons in MDC [81]. The results were published in 2007 for BESIII MDC calibration.



(a) DCA resolution on $R - \phi$

(b) Zoomed in

Figure 6.7: DCA resolution in $R - \phi$ as a function of momentum. The tracks are from pions of a fixed $\theta = 80°$.

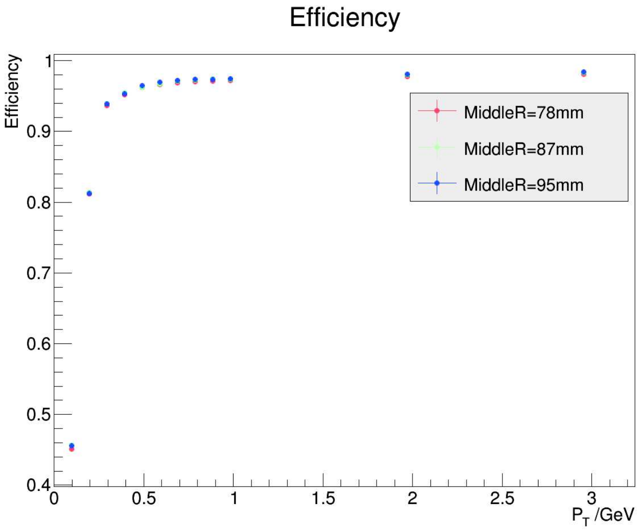


Figure 6.8: Efficiency as a function of transverse momentum. Only inward tracking is used.

good agreement. Full MC simulations using single pion events have also been processed and analysed in BESIII offline software system. The results show a better performance of the SPT design as predicted by our calculations and numerical scans. Therefore, the general methods developed in this thesis have been fully validated in Geant4 based simulations.

Conclusions and Perspectives

Conclusions

The thesis addresses the upcoming particle physics programme foreseen at the Beijing Electron Positron Collider II (BEPCII), for which the BESIII experimental set-up is being upgraded. The work presented focused on the upgrade of the central tracker of the experiment by exploring the possibilities offered by a new approach using CMOS Pixel Sensors (CPS) instead of the present gas drift chamber. To develop a silicon pixelated inner tracker for BESIII in such a low energy range for $\tau - c$ physics, both sensor performance and geometry designs have been studied in this thesis.

CPS design

Considering the test results of sensor MIMOSA22THRb with a spatial resolution $\sim$10 µm, the performances of CPS with larger pixel pitch and a lower power consumption have been guaranteed for the TowerJazz 0.18 µm process, which will help to develop long ladders used for trackers. Besides, the related beam test technique and analysis results have been shown for CPS calibrations.

Layout design

A general strategy and software to optimise the design of a barrel pixelated tracker has been developed, and its application on the BESIII inner silicon tracker indicates that the layout of three evenly placed CPS layers is favoured by BESIII physics for sensor spatial resolution $\sim$10 µm and ladder material budget $\sim 0.36\% X_0$. To achieve higher tracking efficiency, a double-sided ladder may be considered to replace the innermost ladder with acceptable tracking performance degradation for low momentum tracks.

The Geant4-based simulation tool and the related analysis programs have also been developed for this study, which offer the data analysis results to validate the calculations in our optimization strategy and to do a cross-check for different algorithms.

Perspectives

The studies in this thesis are also helpful and instructive to tracker and vertex detector designs for future experiments like circular electron positron collider (CEPC). As the large pixel (50 µm) technique in CPS has been verified, CPS ladders could be considered to cover the gas tracker to improve tracking and vertexing performance with lower power consumption and faster integration speed at the same time. For the momentum distributions of the interested final states from physics decays, it is better to optimise vertex detector and tracker simultaneously as one detector with proper trade-off between momentum resolutions and impact parameter resolutions.

Appendix A

Useful mathematical formulae

A.1 The sum of the k-th powers

A.1.0.1 The triangle numbers

$$\sum_{i=1}^{n} i = 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \tag{A.1.1}$$

A.1.0.2 The square pyramidal numbers

$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} \tag{A.1.2}$$

To prove equation A.1.2, one can start with expansions below:

$$\begin{aligned}
&\sum_{i=1}^{n} [(i+1)^3 - i^3] \\
&= \sum_{i=1}^{n+1} i^3 - 1 - \sum_{i=1}^{n} i^3 \\
&= (n+1)^3 - 1
\end{aligned}$$

At the same time, considering binomial theorem and equation A.1.1, we have:

$$\begin{aligned}\sum_{i=1}^{n}[(i+1)^3 - i^3] &= \sum_{i=1}^{n}(i^3 + 3i^2 + 3i + 1 - i^3)\\ &= 3\sum_{i=1}^{n} i^2 + \frac{3n(n+1)}{2} + n\end{aligned}$$

Since the expansions of the same expression are identical, one can find:

$$3\sum_{i=1}^{n} i^2 + \frac{3n(n+1)}{2} + n = (n+1)^3 - 1,$$

$$\begin{aligned}\sum_{i=1}^{n} i^2 &= \frac{1}{3}[(n+1)^3 - 1 - \frac{3n(n+1)}{2} - n]\\ &= \frac{1}{3}(n+1)n(n+\frac{1}{2})\\ &= \frac{n(n+1)(2n+1)}{6}\end{aligned}$$

A.1.0.3 Sum of k-th powers

With a similar technique of proving equation A.1.2 and all the equations for the sum of lower powers, it is easy to derive the formula of $\sum_{i=1}^{n} i^k$ for $k > 2$. The general formula for the sum of k-th powers is Faulhaber's formula:

$$\sum_{i=1}^{n} i^k = \frac{1}{k+1}\sum_{j=0}^{k}(-1)^j C_{k+1}^{j} B_j n^{k+1-j}, \tag{A.1.3}$$

where C_{k+1}^{j} is Binomial coefficient and B_j is the first Bernoulli numbers:

$$C_{k+1}^{j} = \frac{(k+1)!}{j!(k+1-j)!},$$

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_j = -\sum_{k=0}^{j-1} C_j^k \frac{B_k}{j-k+1}$$

For example, we could write the equations for $k = 3$ and $k = 4$ as follows:

1. Sum of the third powers:

$$\begin{aligned}\sum_{i=1}^{n} i^3 &= \frac{1}{4}\sum_{j=0}^{3}(-1)^j C_4^j B_j n^{4-j} \\ &= \frac{1}{4}(B_0 n^4 - 4B_1 n^3 + 6B_2 n^2) \\ &= \frac{n^2(n+1)^2}{4}\end{aligned} \tag{A.1.4}$$

2. Sum of the fourth powers:

$$\begin{aligned}\sum_{i=1}^{n} i^4 &= \frac{1}{5}\sum_{j=0}^{4}(-1)^j C_5^j B_j n^{5-j} \\ &= \frac{1}{5}(B_0 n^5 - 5B_1 n^4 + C_5^2 B_2 n^3 - C_5^3 B_3 n^2 + 5B_4 n) \\ &= \frac{1}{30}(6n^5 + 15n^4 + 10n^3 - n) \\ &= \frac{(n+1)}{30}[6(n+1)^4 - 15(n+1)^3 + 10(n+1)^2 - 1]\end{aligned} \tag{A.1.5}$$

A.2 Matrix calculus

A.2.0.1 The derivative of a scalar

The derivative of a scalar f with respect to a matrix $\boldsymbol{A}$ with dimension $(n \times m)$ is defined below

$$\frac{df}{d\boldsymbol{A}} = \begin{bmatrix} \frac{df}{da_{11}} & \cdots & \frac{df}{da_{1i}} & \cdots & \frac{df}{da_{1m}} \\ \vdots & & \vdots & & \vdots \\ \frac{df}{da_{j1}} & \cdots & \frac{df}{da_{ji}} & \cdots & \frac{df}{da_{jm}} \\ \vdots & & \vdots & & \vdots \\ \frac{df}{da_{n1}} & \cdots & \frac{df}{da_{ni}} & \cdots & \frac{df}{da_{nm}} \end{bmatrix}$$

The trace of a square matrix $\boldsymbol{C}$ with dimension $(m \times m)$ is defined by

$$tr(\boldsymbol{C}) = \sum_{i=1}^{m} C_{ii}$$

It is easy to find that for an arbitrary matrix $\boldsymbol{B}$ with dimension $(m \times n)$ we have

$$\frac{d\left[tr(\boldsymbol{AB})\right]}{d\boldsymbol{A}} = \boldsymbol{B^T} \tag{A.2.1}$$

Since the trace of a square matrix is equal to the trace of its transpose, it is straightforward to write

$$\begin{aligned}\frac{d\left[tr(\boldsymbol{ACA^T})\right]}{d\boldsymbol{A}} &= (\boldsymbol{CA^T})^T + (\boldsymbol{C^T A^T})^T \\ &= \boldsymbol{AC^T} + \boldsymbol{AC}\end{aligned} \tag{A.2.2}$$

A.3 Joint Gaussian probability density

A.3.0.1 Joint Gaussian density

The multivariate Gaussian distribution of a column vector $\boldsymbol{\theta}$ with dimension m:

$$\boldsymbol{\theta} \sim \mathcal{N}\left(\hat{\boldsymbol{\theta}}^-, \boldsymbol{C}^-\right),$$

$$\mathcal{N}\left(\hat{\boldsymbol{\theta}}^-, \boldsymbol{C}^-\right) = \frac{1}{\sqrt{(2\pi)^m \left|\boldsymbol{C}^-\right|}} \exp\left[-\frac{1}{2}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^-\right)^T \left(\boldsymbol{C}^-\right)^{-1} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^-\right)\right] \tag{A.3.1}$$

where $\hat{\boldsymbol{\theta}}^-$ is the mean values and $\boldsymbol{C}^-$ is the covariance matrix.

If a new vector $\boldsymbol{y}$ is connected to $\boldsymbol{\theta}$ by a linear function below

$$\boldsymbol{y} = \boldsymbol{H\theta} + \boldsymbol{v}$$

where $\boldsymbol{H}$ is a constant matrix and $\boldsymbol{v}$ is a random noise with variance $\boldsymbol{V}$, the probability density of $\boldsymbol{y}$ conditioned on $\boldsymbol{\theta}$ is shown as:

$$p(\boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{H\theta}, \boldsymbol{V}\right)$$

The joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{y}$ reads

$$\begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \hat{\boldsymbol{\theta}}^{-} \\ \boldsymbol{H}\hat{\boldsymbol{\theta}}^{-} \end{bmatrix}, \begin{bmatrix} \boldsymbol{C}^{-} & \boldsymbol{C}^{-}\boldsymbol{H}^{\boldsymbol{T}} \\ \boldsymbol{H}\boldsymbol{C}^{-} & \boldsymbol{H}\boldsymbol{C}^{-}\boldsymbol{H}^{\boldsymbol{T}} + \boldsymbol{V} \end{bmatrix} \right) \tag{A.3.2}$$

The marginal distribution of $\boldsymbol{y}$ is also given:

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{H}\hat{\boldsymbol{\theta}}^{-}, \boldsymbol{H}\boldsymbol{C}^{-}\boldsymbol{H}^{\boldsymbol{T}} + \boldsymbol{V} \right) \tag{A.3.3}$$

A.3.0.2 Property of joint Gaussian density

Assuming that the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfy the joint Gaussian distribution below:

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{x} \\ \boldsymbol{\mu}_{y} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{xx}} & \boldsymbol{\Sigma}_{\boldsymbol{xy}} \\ \boldsymbol{\Sigma}_{\boldsymbol{xy}}^{\boldsymbol{T}} & \boldsymbol{\Sigma}_{\boldsymbol{yy}} \end{bmatrix} \right)$$

The marginal and conditional densities of $\boldsymbol{x}$ and $\boldsymbol{y}$ read:

$$\begin{aligned} \boldsymbol{x} &\sim \mathcal{N}\left(\boldsymbol{\mu}_{x}, \boldsymbol{\Sigma}_{xx}\right) \\ \boldsymbol{y} &\sim \mathcal{N}\left(\boldsymbol{\mu}_{y}, \boldsymbol{\Sigma}_{yy}\right) \\ \boldsymbol{x}|\boldsymbol{y} &\sim \mathcal{N}\left(\boldsymbol{\mu}_{x} + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\left(\boldsymbol{y} - \boldsymbol{\mu}_{y}\right), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{xy}^{T}\right) && \text{(A.3.4)} \\ \boldsymbol{y}|\boldsymbol{x} &\sim \mathcal{N}\left(\boldsymbol{\mu}_{y} + \boldsymbol{\Sigma}_{xy}^{T}\boldsymbol{\Sigma}_{x}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}_{x}\right), \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{xy}^{T}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\right) && \text{(A.3.5)} \end{aligned}$$

A.4 Matrix inversion lemma

The matrix inversion lemma, also named as Woodbury matrix identity or Sherman Morrison Woodbury formula, is given below:

$$(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1} \tag{A.4.1}$$

in which, the dimensions of matrices $\boldsymbol{A}$, $\boldsymbol{U}$, $\boldsymbol{C}$ and $\boldsymbol{V}$ are $n \times n$, $n \times k$, $k \times k$ and $k \times n$ respectively.

By using the matrix inversion lemma, several useful formulae could be derived.

- With $\boldsymbol{A} = \boldsymbol{C_f}^{-1}$, $\boldsymbol{C} = \boldsymbol{C_b}^{-1}$ and $\boldsymbol{U} = \boldsymbol{V} = \boldsymbol{I}$, one can write the equations below:

$$\begin{aligned}
&(\boldsymbol{C_f}^{-1} + \boldsymbol{C_b}^{-1})^{-1} \\
&= \boldsymbol{C_f} - \boldsymbol{C_f}(\boldsymbol{C_b} + \boldsymbol{C_f})^{-1}\boldsymbol{C_f} && \text{(A.4.2)} \\
&= \boldsymbol{C_f}(\boldsymbol{C_b} + \boldsymbol{C_f})^{-1}(\boldsymbol{C_b} + \boldsymbol{C_f}) - \boldsymbol{C_f}(\boldsymbol{C_b} + \boldsymbol{C_f})^{-1}\boldsymbol{C_f} \\
&= \boldsymbol{C_f}(\boldsymbol{C_b} + \boldsymbol{C_f})^{-1}\boldsymbol{C_b} && \text{(A.4.3)}
\end{aligned}$$

All the equations above are symmetric, it is obvious to have

$$\begin{aligned}
(\boldsymbol{C_f}^{-1} + \boldsymbol{C_b}^{-1})^{-1} &= \boldsymbol{C_f} - \boldsymbol{C_f}(\boldsymbol{C_b} + \boldsymbol{C_f})^{-1}\boldsymbol{C_f} && \text{(A.4.4)} \\
&= \boldsymbol{C_b}(\boldsymbol{C_f} + \boldsymbol{C_b})^{-1}\boldsymbol{C_f} && \text{(A.4.5)}
\end{aligned}$$

- With $\boldsymbol{A} = \boldsymbol{V} = \boldsymbol{I}$, $\boldsymbol{U} = \boldsymbol{C_f}$ and $\boldsymbol{C} = \boldsymbol{C_b}^{-1}$, the matrix inversion lemma yields

$$(\boldsymbol{I} + \boldsymbol{C_f}\boldsymbol{C_b}^{-1})^{-1} = \boldsymbol{I} - \boldsymbol{C_f}(\boldsymbol{C_b} + \boldsymbol{C_f})^{-1} \qquad \text{(A.4.6)}$$

- With $\boldsymbol{A} = \boldsymbol{C}^{-}_{\boldsymbol{k+1}}$, $\boldsymbol{U} = -\boldsymbol{I}$, $\boldsymbol{C} = \boldsymbol{C}^{s}_{\boldsymbol{k+1}}$ and $\boldsymbol{V} = \boldsymbol{I}$, we find

$$\begin{aligned}
&(\boldsymbol{C}^{-}_{\boldsymbol{k+1}} - \boldsymbol{C}^{s}_{\boldsymbol{k+1}})^{-1} \\
&= (\boldsymbol{C}^{-}_{\boldsymbol{k+1}})^{-1} + (\boldsymbol{C}^{-}_{\boldsymbol{k+1}})^{-1}\left[\boldsymbol{C}^{s}_{\boldsymbol{k+1}}{}^{-1} - (\boldsymbol{C}^{-}_{\boldsymbol{k+1}})^{-1}\right]^{-1}(\boldsymbol{C}^{-}_{\boldsymbol{k+1}})^{-1} && \text{(A.4.7)}
\end{aligned}$$

- With $\boldsymbol{A} = \boldsymbol{C}^{-}_{\boldsymbol{k}}$, $\boldsymbol{U} = \boldsymbol{C}^{-}_{\boldsymbol{k}}\boldsymbol{H}^{T}_{\boldsymbol{k}}$, $\boldsymbol{C} = -\left(\boldsymbol{H_k}\boldsymbol{C}^{-}_{\boldsymbol{k}}\boldsymbol{H_k}^{T} + \boldsymbol{V_k}\right)^{-1}$ and $\boldsymbol{V} = \boldsymbol{H_k}\boldsymbol{C}^{-}_{\boldsymbol{k}}$, the matrix inversion lemma leads to

$$\begin{aligned}
&\left[\boldsymbol{C}_{\boldsymbol{k}}^{-} - \boldsymbol{C}_{\boldsymbol{k}}^{-}\boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}}\left(\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{k}}^{-}\boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}} + \boldsymbol{V}_{\boldsymbol{k}}\right)^{-1}\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{k}}^{-}\right]^{-1} \\
&= \left(\boldsymbol{C}_{\boldsymbol{k}}^{-}\right)^{-1} - \left(\boldsymbol{C}_{\boldsymbol{k}}^{-}\right)^{-1}\boldsymbol{C}_{\boldsymbol{k}}^{-}\boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}}\Big[-\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{k}}^{-}\boldsymbol{H}_k{}^{T} - \boldsymbol{V}_{\boldsymbol{k}} \\
&\quad + \boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{k}}^{-}\left(\boldsymbol{C}_{\boldsymbol{k}}^{-}\right)^{-1}\boldsymbol{C}_{\boldsymbol{k}}^{-}\boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}}\Big]^{-1}\boldsymbol{H}_k\boldsymbol{C}_{\boldsymbol{k}}^{-}\left(\boldsymbol{C}_{\boldsymbol{k}}^{-}\right)^{-1} \\
&= (\boldsymbol{C}_{\boldsymbol{k}}^{-})^{-1} + \boldsymbol{H}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{V}_{\boldsymbol{k}}^{\boldsymbol{-1}}\boldsymbol{H}_{\boldsymbol{k}}
\end{aligned} \tag{A.4.8}$$

Bibliography

[1] M. Ablikim *et al.*, “Design and Construction of the BESIII Detector,” *Nucl. Instrum. Meth.*, vol. A614, pp. 345–399, 2010.

[2] Q. Ouyang, “Research on maps detection technology for the upgrade of the inner chamber of the multilayer drift chamber of besiii,” 2012, application for National Natural Science Foundation of China, U12182099.

[3] G. Contin, “The maps-based vertex detector for the star experiment: Lessons learned and performance,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 831, pp. 7 – 11, 2016, proceedings of the 10th International “Hiroshima” Symposium on the Development and Application of Semiconductor Tracking Detectors. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900216303539

[4] C. Patrignani *et al.*, “Review of Particle Physics,” *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.

[5] R. L. Gluckstern, “Uncertainties in track momentum and direction, due to multiple scattering and measurement errors,” *Nuclear Instruments and Methods*, vol. 24, pp. 381–389, 1963.

[6] A. Nomerotski, O. Bachynska, J. Baudot, N. Chon-Sen, G. Claus, R. D. Masi, M. Deveaux, W. Dulinski, R. Gauld, M. Goffe, J. Goldstein, I.-M. Gregor, C. Hu-Guo, M. Imhoff, U. Koetz, W. Lau, C. Muntz, C. Santos, C. Schrader, M. Specht, J. Stroth, M. Winter, and S. Yang, “Plume collaboration: Ultra-light ladders for linear collider vertex detector,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated*

Equipment, vol. 650, no. 1, pp. 208 – 212, 2011, international Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016890021002824X

[7] A. Besson, A. P. Pérez, E. Spiriti, J. Baudot, G. Claus, M. Goffe, and M. Winter, "From vertex detectors to inner trackers with cmos pixel sensors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 845, pp. 33–37, 2017.

[8] N. Cabibbo, "Unitary symmetry and leptonic decays," *Physical Review Letters*, vol. 10, no. 12, p. 531, 1963.

[9] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics*, vol. 49, no. 2, p. 652, 1973. [Online]. Available: http://dx.doi.org/10.1143/PTP.49.652

[10] R. J. Morrison and M. S. Witherell, "D mesons," *Annual Review of Nuclear and Particle Science*, vol. 39, no. 1, pp. 183–230, 1989.

[11] M. Ablikim, C. Jiangchuan, R. Gang, and Z. Dahua, "Cabibbo suppressed semileptonic decay $\mathrm{D}^0 \to \pi^- \ e^+ \ \nu_e$," *High Energy Physics and Nuclear Physics*, vol. 29, no. 1, pp. 5–8, 2005.

[12] M. Ablikim, J. Bai, Y. Ban, J. Bian, X. Cai, J. Chang, H. Chen, H. Chen, H. Chen, J. Chen *et al.*, "Direct measurements of the branching fractions for $\mathrm{D}^0 \to \mathrm{K}^-\mathrm{e}^+\nu_e$ and $\mathrm{D}^0 \to \pi^-\mathrm{e}^+\nu_e$ and determinations of the form factors f_+^K (0) and f_+^π (0)," *Physics Letters B*, vol. 597, no. 1, pp. 39–46, 2004.

[13] BESIII collaboration, "The Preliminary Design Report of the BES-III Detector," Nov. 2003.

[14] S.-Z. Wang, R. Liu, and Y.-L. Chi, "Beam energy online measurement of BEPCII LINAC," 2015.

[15] D. M. Asner *et al.*, "Physics at BES-III," *Int. J. Mod. Phys.*, vol. A24, pp. S1–794, 2009.

[16] J. LIU, “Experimental study of the besiii drift chamber performance,” Ph.D. dissertation, Institute of High Energy Physics, Chinese Academy of Sciences, 2005.

[17] Z. Qin, Y. Chen, H. Sheng, L. Wu, J. Liu, B. Zhuang, X. Jiang, Y. Zhao, K. Zhu, Z. Yan *et al.*, “Beam test of a full-length prototype of the BESIII drift chamber with the readout electronics,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 571, no. 3, pp. 612–621, 2007.

[18] L. Malter, “Thin film field emission,” *Physical Review*, vol. 50, no. 1, p. 48, 1936.

[19] M.-Y. Dong, Q.-L. Xiu, L.-H. Wu, Z. Wu, Z.-H. Qin, P. Shen, F.-F. An, X.-D. Ju, Y. Liu, K. Zhu, Q. Ou-Yang, and Y.-B. Chen, “Aging effect in the besiii drift chamber,” *Chinese Physics C*, vol. 40, no. 1, p. 016001, 2016. [Online]. Available: http://stacks.iop.org/1674-1137/40/i=1/a=016001

[20] M. Szelezniak, “Upgrade of the STAR Silicon Detectors,” *PoS*, vol. Vertex2014, p. 015, 2015.

[21] G. Van Buren, L. Didenko, Y. Fisyak, L. Hajdu, J. Lauret, V. Perevotzchikov, D. Smirnov, F. Videbaek, J. Webb, X. Dong *et al.*, “Developments in tracking with star’s heavy flavor tracker,” 2015.

[22] J. Lindhard, A. H. So *et al.*, “Relativistic theory of stopping for heavy ions,” *Physical Review A*, vol. 53, no. 4, p. 2443, 1996.

[23] H. Wyckoff, “Stopping powers and ranges for protons and alpha particles,” *ICRU Report*, no. 49, 1993.

[24] M. Berger, M. Inokuti, H. Anderson, H. Bichsel, and J. Dennis, “Stopping powers for electrons and positrons,” *ICRU report*, no. 37, 1984.

[25] H. Bichsel, “Stopping power and ranges of fast ions in heavy elements,” *Physical Review A*, vol. 46, no. 9, p. 5761, 1992.

[26] M. J. Berger and W. H. Barkas, “Tables of energy losses and ranges of heavy charged particles,” NTIS, Tech. Rep. NASA-SP-3013, 1964.

[27] D. E. Groom, N. V. Mokhov, and S. I. Striganov, “Muon stopping power and range tables 10 mev–100 tev,” *Atomic Data and Nuclear Data Tables*, vol. 78, no. 2, pp. 183–356, 2001.

[28] R. Sternheimer and R. Peierls, “General expression for the density effect for the ionization loss of charged particles,” *Physical Review B*, vol. 3, no. 11, p. 3681, 1971.

[29] L. D. Landau, “On the energy loss of fast particles by ionization,” *J. Phys.*, vol. 8, pp. 201–205, 1944.

[30] P. Vavilov, “Ionization losses of high-energy heavy particles,” *Soviet Phys. JETP*, vol. 5, 1957.

[31] H. Bichsel, “Straggling in thin silicon detectors,” *Reviews of Modern Physics*, vol. 60, no. 3, p. 663, 1988.

[32] ——, “A method to improve tracking and particle identification in tpcs and silicon detectors,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 562, no. 1, pp. 154–197, 2006.

[33] Y.-S. Tsai, “Pair production and bremsstrahlung of charged leptons,” *Reviews of Modern Physics*, vol. 46, no. 4, p. 815, 1974.

[34] B. B. Rossi, “High-energy particles,” 1952.

[35] L. Gerhardt and S. R. Klein, “Electron and photon interactions in the regime of strong landau-pomeranchuk-migdal suppression,” *Physical Review D*, vol. 82, no. 7, p. 074017, 2010.

[36] P. H. Barrett, L. M. Bollinger, G. Cocconi, Y. Eisenberg, and K. Greisen, “Interpretation of cosmic-ray measurements far underground,” *Reviews of Modern Physics*, vol. 24, no. 3, p. 133, 1952.

[37] L. Gerward, “X-ray attenuation coefficients: current state of knowledge and availability,” *Radiation Physics and Chemistry*, vol. 41, no. 4-5, pp. 783–789, 1993.

[38] W. T. Scott, “The theory of small-angle multiple scattering of fast charged particles,” *Rev. Mod. Phys.*, vol. 35, pp. 231–313, 1963.

[39] J. Motz, H. Olsen, and H. Koch, “Electron scattering without atomic or nuclear excitation,” *Reviews of modern physics*, vol. 36, no. 4, p. 881, 1964.

[40] V. L. Highland, “Some Practical Remarks on Multiple Scattering,” *Nucl. Instrum. Meth.*, vol. 129, p. 497, 1975.

[41] G. R. Lynch and O. I. Dahl, “Approximations to multiple coulomb scattering,” *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 58, no. 1, pp. 6–10, 1991.

[42] G. Moliere, “Theory of the scattering of fast charged particles. 2. Repeated and multiple scattering,” *Z. Naturforsch.*, vol. A3, pp. 78–97, 1948.

[43] H. A. Bethe, “Moliere’s theory of multiple scattering,” *Phys. Rev.*, vol. 89, pp. 1256–1266, 1953.

[44] ALICE Collaboration, “Technical design report for the upgrade of the alice inner tracking system,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 41, no. 8, p. 087002, 2014. [Online]. Available: http://stacks.iop.org/0954-3899/41/i=8/a=087002

[45] M. Deveaux, “Development of fast and radiation hard monolithic active pixel sensors (maps) optimized for open charm meson detection with the cbm - vertex detector,” Ph.D. dissertation, Université Louis Pasteur Strasbourg 1, 2007.

[46] G. D. A. Himmi *et al.*, *Ultimate 2 User Manual*.

[47] A. Himmi, A. Doziere, O. Torheim, A. Winter, and C. Hu-Guo, “A zero suppression micro-circuit for binary readout cmos monolithic sensors,” 2009.

[48] M. Mager *et al.*, “Alpide, the monolithic active pixel sensor for the alice its upgrade,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 824, pp. 434–438, 2016.

[49] P. Yang, G. Aglieri, C. Cavicchioli, P. Chalmet, N. Chanlek, A. Collu, C. Gao, H. Hillemanns, A. Junique, M. Kofarago *et al.*, “Low-power priority address-encoder and reset-decoder data-driven readout for monolithic active pixel sensors for tracker

system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 785, pp. 61–69, 2015.

[50] T. Kameshima, S. Ono *et al.*, "Development of an x-ray pixel detector with multi-port charge-coupled device for x-ray free-electron laser experiments," *Review of Scientific Instruments*, vol. 85, no. 3, p. 033110, 2014.

[51] Marc Winter and PICSEL team of IPHC-Strasbourg, "Development of cmos pixel sensors for high-precision vertexing and tracking devices," 2015, terascale Workshop, Berlin.

[52] G. Lutz, "Optimum track fitting in the presence of multiple scattering," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 273, no. 1, pp. 349–361, 1988.

[53] R. G. Brown and P. Y. Hwang, *Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions.* John Wiley & Sons Inc, 1997.

[54] Y. C. Ho, "Explaining filtering (estimation) in one hour, ten minutes, one minute, and one sentence," 2014. [Online]. Available: http://blog.sciencenet.cn/blog-1565-851754.html

[55] A. Gelb, *Applied optimal estimation.* MIT press, 1974.

[56] J. S. Meditch and J. Meditch, *Stochastic optimal linear estimation and control.* McGraw-Hill New York, 1969, vol. 8.

[57] H. Sorenson, "Kalman filtering techniques," *Advances in Control Systems Theory and Applications*, vol. 3, pp. 219–292, 1966.

[58] 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*, vol. 262, no. 2, pp. 444–450, 1987.

[59] J. L. Junkins and J. Crassidis, *Optimal estimation of dynamic systems.* Taylor & Francis, 2004.

[60] H. Rauch, "Solutions to the linear smoothing problem," *Automatic Control, IEEE Transactions on*, vol. 8, no. 4, pp. 371–372, 1963.

[61] H. Rauch, C. Striebel, and F. Tung, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, pp. 1445–1450, 1965.

[62] E. Wolin and L. Ho, “Covariance matrices for track fitting with the kalman filter,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 329, no. 3, pp. 493–500, 1993.

[63] R. Mankel, “Ranger a pattern recognition algorithm for the hera-b main tracking system part iv: The object-oriented track fit,” *HERA-B note,(98-079)*, 1998.

[64] H. Wieman, “Pileup probability,” 2006. [Online]. Available: http://www-rnc.lbl.gov/~wieman/PileupProbabilityDerivation.htm

[65] ——, “Ghost track probability,” 2006. [Online]. Available: http://www-rnc.lbl.gov/~wieman/GhostTracks.htm

[66] H. Wieman and V. Perevoztchikov, “Probability of finding the correct hit using weighted chi-square hit selection.” Dec. 2006. [Online]. Available: http://www-rnc.lbl.gov/~wieman/HitFinding2DXsq.htm

[67] V. Karimäki, “Effective circle fitting for particle trajectories,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 305, no. 1, pp. 187–191, 1991.

[68] ——, “Explicit covariance matrix for particle measurement precision,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 410, no. 2, pp. 284–292, 1998.

[69] G. R. Lynch and O. I. Dahl, “Approximations to multiple coulomb scattering,” *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 58, no. 1, pp. 6–10, 1991.

[70] 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*, vol. 262, no. 2, pp. 444–450, 1987.

[71] Q. Xiu *et al.*, “Study of the Tracking Method and Expected Performance of the Silicon Pixel Inner Tracker Applied in BESIII,” 2015.

[72] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand *et al.*, “Geant4—a simulation toolkit,” *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[73] J. Allison, K. Amako, J. e. a. Apostolakis, H. Araujo, P. A. Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek *et al.*, “Geant4 developments and applications,” *IEEE Transactions on nuclear science*, vol. 53, no. 1, pp. 270–278, 2006.

[74] J. Allison, K. Amako, J. Apostolakis, P. Arce, M. Asai, T. Aso, E. Bagli, A. Bagulya, S. Banerjee, G. Barrand *et al.*, “Recent developments in geant4,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 835, pp. 186–225, 2016.

[75] “Root,” https://root.cern.ch/.

[76] W.-D. Li, H.-M. Liu, Z. Deng, K. He, M. He, X. Ji, L. Jiang, H. Li, C. Liu, Q. Ma *et al.*, “The offline software for the besiii experiment,” in *Proceeding of CHEP*, vol. 27, 2006.

[77] G. Barrand, I. Belyaev, P. Binko, M. Cattaneo, R. Chytracek, G. Corti, M. Frank, G. Gracia, J. Harvey, E. Van Herwijnen *et al.*, “Gaudi—a software architecture and framework for building hep data processing applications,” *Computer Physics Communications*, vol. 140, no. 1-2, pp. 45–55, 2001.

[78] “Glast ground software calibration index,” http://www-glast.slac.stanford.edu/software/calib/.

[79] R. Mankel, “A concurrent track evolution algorithm for pattern recognition in the hera-b main tracking system,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 395, no. 2, pp. 169–184, 1997.

[80] R. Frühwirth, P. Vanlaer, and W. Waltenberger, “Adaptive vertex fitting,” CERN-CMS-NOTE-2007-008, Tech. Rep., 2007.

[81] Z. Yao, L. Qiuguang, W. Jike, and W. Linghui, “The reconstruction and calibration of the besiii main drift chamber,” *CHEP2007, Victoria, BC, Canada*, 2007.

UNIVERSITÉ DE STRASBOURG

Qingyuan LIU

SHANDONG UNIVERSITY 1901

Development of a CMOS pixel based inner tracker for the BES-III experiment at BEPC-II

Résumé

La thèse se place dans le contexte du programme de physique prévu auprès du collisionneur Beijing Electron Positron Collider II, pour lequel l'expérience BESIII (Beijing Electron Spectrometer III) doit subir une jouvence. Mon travail concerne l'amélioration des performances de la trajectométrie centrale de l'expérience, en explorant une nouvelle approche utilisant les capteurs à pixels CMOS (CPS) en remplacement partiel de la chambre à dérive actuelle. Il est validé, qu'un pixel de 39.0X50.8 μm^2 permet d'atteindre une résolution spatiale ~10 μm. Cette confirmation ouvre la voie pour le développement d'échelle de grande longueur. Un logiciel générique proposant une stratégie d'optimisation pour un trajectomètre pixellisé a été mis au point et appliqué au cas particulier du BESIII. Il a permis de montrer qu'une configuration à trois couches optimise la résolution sur les trajectoire alors qu'une option incluant une couche double-face atteint une efficacité de reconstruction supérieure.

Mots-clés: détecteur de pixels, optimisation du tracker, CPS test, algorithme de suivi de trajectoires

Résumé en anglais

The thesis addresses the upcoming particle physics program foreseen at the Beijing Electron Positron Collider II, for which the Beijing Electron Spectrometer III (BESIII) is being upgraded. The work presented focused on the upgrade of the central tracker of the experiment by exploring the possibilities offered by a new approach using CMOS Pixel Sensors (CPS) instead of the present gas drift chamber. The performance of CPS with a larger pixel (39.0X50.8 μm^2) and a lower power consumption has been validated, which will help to develop long ladders. A general strategy and software to optimise the design of a barrel pixelated tracker has also been developed, and its application on the BESIII silicon tracker indicates that the layout of three evenly placed single-sided layers is favoured by BESIII physics for the best momentum resolution, with an option using one double-sided ladder to achieve a higher tracking efficiency.

Keywords: pixel detector, tracker optimization, CPS test, tracking algorithm