System development and validation for the CMS Inner Tracker at High Luminosity LHC

Natalia Emriskova

To cite this version:

Natalia Emriskova. System development and validation for the CMS Inner Tracker at High Luminosity LHC. Physics [physics]. Université de Strasbourg, 2021. English. NNT : 2021STRAE020 . tel-03982095

HAL Id: tel-03982095
https://theses.hal.science/tel-03982095v1
Submitted on 10 Feb 2023

UNIVERSITÉ DE STRASBOURG

École Doctorale de physique et chimie-physique (ED 182)

Institut pluridisciplinaire Hubert Curien (IPHC) UMR 7178

THÈSE présentée par :

Natalia EMRISKOVA

Soutenue le : **22 juillet 2021**

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**
Discipline/ Spécialité : physique des particules élémentaires

Développement et validation du système de trajectométrie interne de CMS pour LHC à haute luminosité

THÈSE dirigée par :

Dr. ANDREA Jérémy	Institut Pluridisciplinaire Hubert Curien (IPHC)
Dr. MERSI Stefano	Organisation européenne pour la recherche nucléaire (CERN)

RAPPORTEURS :

Prof. DELAERE Christophe	Université Catholique de Louvain
Dr. SNOEYS Walter	Organisation européenne pour la recherche nucléaire (CERN)

EXAMINATEURS :

Prof. BAUDOT Jérôme	Institut Pluridisciplinaire Hubert Curien (IPHC)
Dr. SEIF EL NASR Sarah	Université de Bristol

MEMBRES INVITÉS :

Dr. ABBANEO Duccio	Organisation européenne pour la recherche nucléaire (CERN)
Dr. ORFANELLI Stella (co-encadrante)	Organisation européenne pour la recherche nucléaire (CERN)

A mes deux personnes préférées au monde:
maminka et Alinko.

Mojim dvom najobľúbenejším ľudom na svete:
maminke a Alinkovi.

Acknowledgements

First and foremost, I want to thank my thesis supervisors: Stefano Mersi, who helped me a lot with the thesis work and from whom I learned great many things and Jeremy Andrea, who was my best support during the hard thesis writing period and who kept me on track to finish it. I also want to thank Stella Orfanelli for her ideas, directions and help with the measurements, and for taking over my supervision when Stefano was away. I also would like to express my gratitude to Duccio Abbaneo, who was always available to provide useful insight and advice. Thank you for being such a kind and inspiring person. I thank my colleague and friend Dominik Koukola, with whom I worked on a daily basis, who was a great help with all the technical work and who thought me most of what I learned in this project.

A big and warm thanks to Jorgen Christiansen for integrating me in the RD53 team and for the numerous very pleasant and instructive coffees. I also want to thank Flavio Loddo and the RD53 designers from whom I learned a lot about the chip, as well as other RD53 members, in particular Luis Miguel Jara Casas and Sara Marconi for their great help and nice moments spent together. Thank you to Rafael Gajanec for being a great office mate for a while bringing a lot of fun to the working day. In general, thanks to all young people, mainly from building 14, who made my time spent at CERN really amazing, especially Leyre Flores, Mark Kovac, Angelos Zografos, Nikola Rasevic, Viros Sriskaran, Mykyta Haranko and other friends (sorry for not listing all of you).

I also wish to acknowledge Gervasio Gómez, Javier Gonzáles Sanchez, Andrea Garcia Alonso and Jory Sonneveld for their great help with the test beam at DESY and also nice time spent there, as well as Daniel Pitzl for sharing his analysis code and Simone Gennai and Jordi Duarte Campderros for their help with the code. Other people who contributed to this work and shall be acknowledged are: Csaba Soos for his valuable suggestions about signal integrity measurements and for letting me use the VNA, Simon Kulis and Pedro Leitao for the useful information and insight concerning the lpGBT and the e-links and for letting me use their equipment, Andromachi Tsirou and the CMS technicians for their precious help with the cooling systems and other lab equipment.

And last but not least, thank you to my mum and Alex for their great support in hard times during this thesis and for giving me the strength to hold on till the end. I am grateful for both of you.

Déclaration sur l'honneur

Declaration of Honour

J'affirme être informée que le plagiat est une faute grave susceptible de mener à des sanctions administratives et disciplinaires pouvant aller jusqu'au renvoi de l'Université de Strasbourg et passible de poursuites devant les tribunaux de la République Française.

Je suis consciente que l'absence de citation claire et transparente d'une source empruntée à un tiers (texte, idée, raisonnement ou autre création) est constitutive de plagiat.

Au vu de ce qui précède, **j'atteste sur l'honneur que le travail décrit dans mon manuscrit de thèse est un travail original et que je n'ai pas eu recours au plagiat ou à toute autre forme de fraude.**

I affirm that I am aware that plagiarism is a serious misconduct that may lead to administrative and disciplinary sanctions up to dismissal from the University of Strasbourg and liable to prosecution in the courts of the French Republic.

I am aware that the absence of a clear and transparent citation of a source borrowed from a third party (text, idea, reasoning or other creation) is constitutive of plagiarism.

In view of the foregoing, ***I hereby certify that the work described in my thesis manuscript is original work and that I have not resorted to plagiarism or any other form of fraud.***

Nom: **EMRISKOVA**
Prénom: **Natalia**
Ecole doctorale: **ED182**
Laboratoire: **IPHC**
Date: **20/12/2021**
Signature:

Table of contents

List of Figures

List of Tables

Résumé en français

Contexte physique

Toute la connaissance actuelle du monde subatomique est résumée dans le modèle standard de la physique des particules. Ce modèle décrit toutes les particules élémentaires connues jusqu'ici ainsi que leurs interactions qui sont elles-mêmes représentées comme un échange de particules élémentaires. Les particules constituant la matière sont appelées des fermions et celles permettant les interactions sont appelées des bosons. Le tableau du modèle standard a été complété en 2012 par la découverte expérimentale du boson de Higgs dont l'existence a été prédite par la théorie 50 ans auparavant. Bien que le modèle standard soit une théorie permettant d'expliquer la quasi-totalité des données expérimentales disponibles à ce jour et représente la meilleure théorie de physique subatomique élaborée jusqu'ici, il est encore incomplet. En effet, il ne permet pas d'expliquer un certain nombre de phénomènes observés dans l'Univers tels que la matière ou l'énergie noire ou encore l'asymétrie entre la matière et l'antimatière. C'est pourquoi les physiciens recherchent des signes de nouvelle physique au-delà du modèle standard à l'aide des expériences à hautes énergies auprès des accélérateurs/collisionneurs de particules.

Le grand collisionneur de hadrons, *Large Hadron Collider* (LHC) en anglais, est le plus grand et plus puissant accélérateur et collisionneur de particules du monde. Cet anneau de 27 km de circonférence près de Genève, accélère les protons (et les ions lourds), avant de les mettre en collision à la vitesse proche de celle de la lumière, pour mener des recherches en physique fondamentale. L'amélioration du LHC vers sa phase à haute luminosité va permettre aux expériences d'accumuler de très larges quantités de données, et par conséquent d'augmenter les chances d'observer de nouveaux phénomènes rares ainsi qu'étudier les phénomènes connus avec une grande précision. Mais la haute luminosité implique aussi un taux de particules produites très important, et donc une irradiation sans précèdent, présentant un défi technologique non seulement pour le collisionneur lui-même mais également pour les détecteurs de particules qui y sont installés. En particulier, le détecteur polyvalent appelé le solénoïde compact à muons, *Compact Muon Solenoid* (CMS) en anglais, va subir une importante amélioration avant de commencer à collecter les données du LHC à haute luminosité.

L’amélioration du trajectomètre interne de CMS

Le détecteur CMS est composé de plusieurs sous-détecteurs. En son cœur, au plus près du point de collisions, se trouve le trajectomètre interne: un détecteur à pixels au silicium permettant une reconstruction précise des trajectoires des particules chargées ainsi que l’identification des vertex d’interactions primaires et secondaires. Pour le LHC à haute luminosité, le détecteur à pixels va être entièrement remplacé par un nouveau détecteur plus performant, plus rapide et plus résistant à la radiation. Le composant principal du nouveau système est le module à pixels hybride, constitué d’un capteur au silicium, de deux ou quatre puces de lecture contenant l’électronique frontale et d’un circuit imprimé flexible contenant des composants passifs. Les modules à pixels vont être organisés en séries, comme c’est illustré sur la Figure 1 à gauche, afin de minimiser au maximum le nombre de câbles d’alimentation. Les données enregistrées par les modules seront transmises à haute vitesse par des câbles électriques très légers vers une puce d’agrégation, appelée *low power gigabit transceiver* (lpGBT) en anglais, puis les signaux électriques seront convertis en signaux optiques par le VTRx+ pour finalement être transmis par des fibres optiques hors du détecteur vers le système d’acquisition et le centre de stockage des données. Le schéma de cette chaîne d’acquisition est présenté sur la Figure 1. L’évaluation des différentes options technologiques est une phase importante pour faire les choix les plus adaptés des différents composants du système. Le sujet de cette thèse porte donc sur le développement du nouveau système de trajectométrie interne de CMS, et en particulier, sur les tests de performance qui ont permis de faire des choix technologiques pour différentes parties du système, à savoir le choix l’électronique frontale, de la géométrie des cellules des capteurs à pixels, et des câbles électriques.

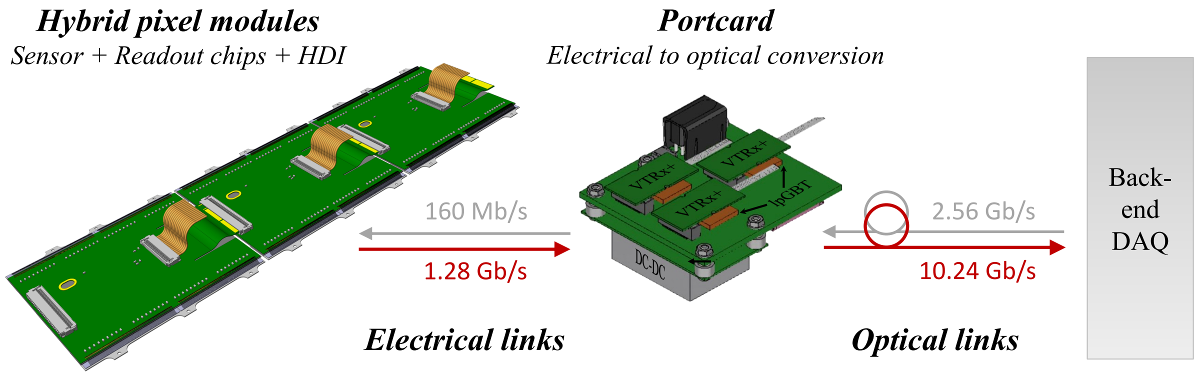


Figure 1: L’architecture du système d’acquisition des données du trajectomètre interne de CMS après son amélioration.

Sélection de l'électronique analogique frontale

L'un des composant les plus importants du nouveau trajectomètre interne de CMS est la puce de lecture qui implémente l'électronique frontale. Les puces de lecture, dédiées à l'amélioration des trajectomètres internes des expériences ATLAS et CMS, sont en train d'être développées par la collaboration RD53. Une puce de démonstration, appelée RD53A a été conçue et produite, afin de démontrer l'adéquation de la technologie CMOS de 65 nm choisie avec l'opération à bas seuil, avec un taux de faux signaux limité et avec une résistance suffisante à la radiation. La taille de la puce de lecture RD53A est de $20.0 \times 11.8\,\mathrm{mm}^2$ et elle contient 400×192 pixels carrés de 50 µm de côté. La majeure partie de la puce est occupée par les pixels et les fonctionnalités globales sont rassemblées à la périphérie de la puce.

RD53A est une puce mixte qui contient des circuits analogiques et aussi numériques. Chaque pixel est en fait à 50% analogique et à 50% numérique. La partie analogique du pixel, dénommée l'électronique analogique frontale, est une partie cruciale de la puce de lecture, puisque c'est le circuit qui reçoit le signal analogique du capteur, le traite et le convertit en signal numérique. Elle est composée principalement de deux grandes parties: le pré-amplificateur et le comparateur. Le signal analogique provenant du capteur est réceptionné par le pré-amplificateur où il est amplifié et mis en forme. Puis, il est comparé au seuil de détection dans le comparateur. Tant que le signal est en dessous du seuil le comparateur est au niveau logique bas (ou zéro) et lorsque le signal dépasse le seuil le comparateur affiche le niveau logique haut (ou un). Ainsi, le signal est converti en un signal numérique carré qui est ensuite traité par la partie numérique du pixel. En particulier, la partie numérique compare la durée du niveau logique haut avec l'horloge du LHC oscillant à 40 MHz. La durée du signal, exprimée en cycles d'horloge du LHC, est proportionnelle à la quantité de charge déposée par la particule dans le capteur. Cette technique, appelée *time-over-threshold* (TOT) en anglais, est utilisée pour avoir l'information sur la charge déposée en plus de la position d'impact de la particule sur le capteur sans devoir stocker les signaux analogiques complets. Le TOT permet d'une part, d'identifier la particule grâce à la quantité de charge déposée et d'autre part, il améliore la précision spatiale de la reconstruction de la trace.

Dans ce travail, la puce de démonstration RD53A a été testée pour vérifier l'opération correcte à bas seuil de détection, les niveaux de bruit électronique, le taux de faux signaux et d'autres fonctionnalités. Les résultats de ces tests ont été d'une grande utilité à l'équipe de concepteurs microélectroniques, puisqu'ils ont révélé quelques problèmes de conception dans les circuits électroniques, et ont ainsi permis d'améliorer le design de la puce. De plus, une des spécificités de la puce RD53A est qu'elle comprend trois électroniques analogiques frontales, offrant ainsi le choix aux deux expériences d'utiliser la solution la plus adaptée à leurs besoins. La puce RD53A est en effet divisée horizontalement en trois sections et chaque section comporte une électronique analogique différente.

Les trois électroniques analogiques frontales proposées sont dénommées: la Synchrone, la Linéaire et la Différentielle, basé sur leurs conceptions. L'électronique analogique synchrone implémente un amplificateur sensible à la charge avec un circuit de retour de Krummenacher et un comparateur synchrone avec l'horloge du LHC, ce qui veut dire que le signal à la sortie de l'amplificateur est comparé au seuil uniquement à chaque front montant de l'horloge. De plus, le comparateur contient un oscillateur local qui permet un comptage TOT plus fin que les 40 MHz de l'horloge standard du LHC. Une autre spécificité de l'électronique synchrone est que l'ajustement local du seuil de chaque pixel est automatique et fait de manière périodique par un circuit spécifique, ce qui élimine la nécessité de calibration régulière du détecteur qui prend du temps. L'électronique linéaire implémente elle aussi l'amplificateur sensible à la charge avec un circuit de retour de Krummenacher qui est suivi d'un comparateur classique asynchrone. L'ajustement local des seuils se fait de manière traditionnelle, grâce à un convertisseur numérique-analogique (*digital-to-analogue converter* (DAC) en anglais) en courant, avec une résolution de quatre bits, présent dans chaque pixel. L'électronique différentielle, quant à elle, implémente un amplificateur sensible à la charge avec un circuit de retour à la réinitialisation continue avec un circuit de compensation de courant de fuite qui peut être activé lorsque les courants de fuite du capteur deviennent trop importants, i.e. après l'irradiation. L'amplificateur est suivi d'un pré-comparateur ajoutant un gain supplémentaire avant que le signal soit comparé au seuil dans un comparateur différentiel. Comme son nom indique, ce comparateur implémente le seuil en tant que différence entre deux valeurs. L'ajustement local des seuils se fait aussi par un DAC en courant présent dans chaque pixel, mais dans cette électronique frontale le DAC d'ajustement a une résolution de cinq bits.

Une campagne de test a été organisée par CMS afin de vérifier les fonctionnalités des trois circuits, les évaluer vis-à-vis des besoins spécifiques de l'expérience et de comparer leurs performances en vue de choisir l'électronique analogique la plus appropriée pour le nouveau trajectomètre interne. Tout d'abord, une procédure précise d'alimentation et de calibration a été établie. Cette procédure a été appliquée lors de chaque test systématiquement sur les trois électroniques afin d'assurer qu'elles soient comparées dans les mêmes conditions. Les tests ont été menés en étroite collaboration avec les concepteurs et l'équipe de simulation et ont permis de mettre en évidence les points forts ainsi que les points faibles de chaque conception. Les critères de sélection, en termes de performances du détecteur, que l'électronique analogique frontale devait absolument satisfaire pour être choisie ont été:

- L'opération correcte à un seuil de détection de $1000\,\mathrm{e}^-$ et l'ajustement local des seuils avec une dispersion aussi petite que possible pour un seuil homogène dans le détecteur.
- Une résistance à la radiation suffisante pour assurer au maximum un remplace-

ment de la couche interne du détecteur pendant sa durée de vie d'environ 10 ans. D'après les simulations, la couche interne est censée accumuler une fluence de $3.4 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$ et une dose ionisante totale de 1.9 Grad, si le scénario d'ultime luminosité du LHC est atteint.

- Le bruit électronique causant une occupation de faux signaux maximale de 10^{-6} après avoir désactivé les pixels bruyants. Un pixel était considéré comme bruyant et désactivé si son occupation de signaux induit par le bruit a été supérieure à cent dans un million de déclenchements. Le nombre de pixels bruyants désactivés a été pris en compte aussi dans cette étude, puisqu'un nombre de pixels désactivés trop important affecterait l'efficacité du détecteur.

- Finalement, le temps mort, i.e. le temps pendant lequel l'électronique traite un signal et est donc insensible à un autre signal arrivant peu de temps après, devait être inférieur à 1% dans la couche interne du détecteur où l'occupation est la plus importante. Le temps mort causé par la partie numérique est dû aux mémoires tampon et est donc donné par la conception et ne peut pas être ajusté. Le temps mort causé par la partie analogie du pixel est ajusté par la calibration du TOT. Le critère du temps mort maximum de 1% correspondait à la calibration de TOT de $3000\,\mathrm{e^-}$ par cycle d'horloge.

Un des résultats marquants de cette campagne de test est montré sur la Figure 2. Cette figure montre la fraction des pixels désactivés pour chaque électronique à gauche et l'occupation de faux signaux des pixels restants à droite. Le test a été fait pour une calibration TOT standard représentée en vert et pour une calibration TOT permettant de réduire le temps mort à 1% représentée en bleu. L'électronique synchrone a montré un bruit électronique le plus élevé des trois conceptions et ce, même après avoir désactivé une fraction des pixels plus importante que dans les deux autres électroniques. De plus, ce bruit était encore exacerbé lorsque la performance de la puce a été mise à l'épreuve avec la calibration de TOT visant à limiter le temps mort, en particulier à bas seuil. Dans cette configuration (représentée en bleu foncée), nécessaire pour l'opération de la couche la plus interne du détecteur, l'électronique synchrone n'a pas satisfait le critère du bruit. De plus, la fraction des pixels bruyants devant être désactivés approchait 4%. L'origine de ce bruit n'a pas été entièrement comprise et reproduite en simulation, c'est pourquoi cette électronique a été considérée comme une option à risque pour le nouveau détecteur.

Un autre test décisif concernait l'ajustement local du seuil. Toutes les trois électroniques fonctionnaient correctement au seuil de $1000\,\mathrm{e^-}$ et le seuil a pu être ajusté de manière homogène à plus ou moins $100\,\mathrm{e^-}$. Le test a ensuite été fait avec une puce de lecture après l'irradiation. La puce connectée à un capteur avec des pixels rectangulaires a été irradiée avec des protons de 23 MeV jusqu'à une dose ionisante totale de 350 Mrad et une fluence de $3 \times 10^{15}\,\mathrm{n_{eq}/cm^2}$. La tension de polarisation du capteur a été augmentée jusqu'à ce que le courant de fuite atteigne 10 nA par pixel, ce qui est la valeur maximale spécifiée pour la

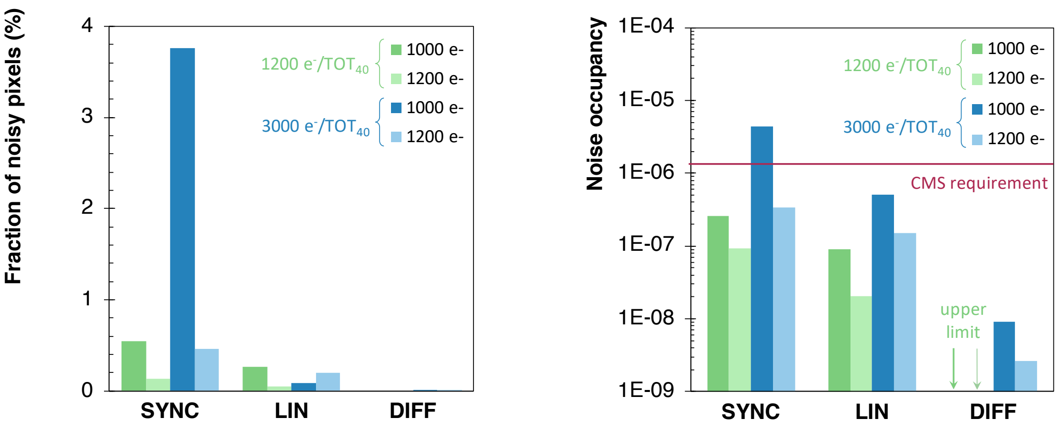


Figure 2: Fraction des pixels bruyants désactivés (gauche) et l'occupation de signaux induits par le bruit (droite) des trois électroniques analogiques frontales implémentées dans la puce de lecture RD53A, mesurées avec deux seuils de détection (1000 e^- et 1200 e^-) et deux calibrations de TOT (1200 e^-/TOT_{40} et 3000 e^-/TOT_{40}).

conception de l'électronique frontale. Dans cette configuration, l'électronique synchrone et linéaire fonctionnaient toujours correctement, tandis que l'ajustement local du seuil dans l'électronique différentielle est devenu problématique. En fait, il a été découvert que l'irradiation de cette électronique à basse température a causé une dispersion très importante des seuils d'un pixel à l'autre. La distribution des seuils avant l'ajustement est montrée sur la Figure 3 à gauche. Le seuil global avait été configuré à 1000 e^- et certains pixels avaient un seuil effectif jusqu'à quinze fois plus haut. Cette large dispersion était à l'origine des problèmes d'ajustement local, puisque la taille du pas d'ajustement est un paramètre global et il n'était pas possible de trouver un pas qui convenait à tous les pixels. L'activation du circuit de compensation de courant de fuite en combinaison avec un algorithme d'ajustement amélioré a permis d'ajuster les seuils, cependant la distribution des seuils après l'ajustement comportait de longues traînes, comme c'est montré sur la Figure 3 à droite. Ce comportement a été compris et reproduit en simulation de circuit. Une amélioration de cette électronique a été proposée par l'équipe de conception, néanmoins la version améliorée a été démontrée de ne fonctionner que jusqu'à 500 Mrad. Or, étant donné les niveaux de radiation attendus dans le trajectomètre interne de CMS, une résistance à la radiation jusqu'à 500 Mrad seulement impliquerait le remplacement de la couche interne du détecteur tous les 2 ans. C'est pourquoi cette électronique n'a pas été retenue pour l'amélioration du détecteur à pixel de CMS.

Finalement, l'électronique linéaire a satisfait tous les critères de sélection, mais un défaut a quand même été trouvé. En effet, cette électronique a eu une réponse plus lente que les deux autres, causant des signaux détectés en retard, apparaissant comme le bruit de fond. Une méthode dédiée a été développée afin de pouvoir prédire la fraction des signaux détectés en retard et le résultat est montré sur la Figure 4. La fraction des signaux détectés en retard par l'électronique synchrone (bleu) et différentielle (violet) est

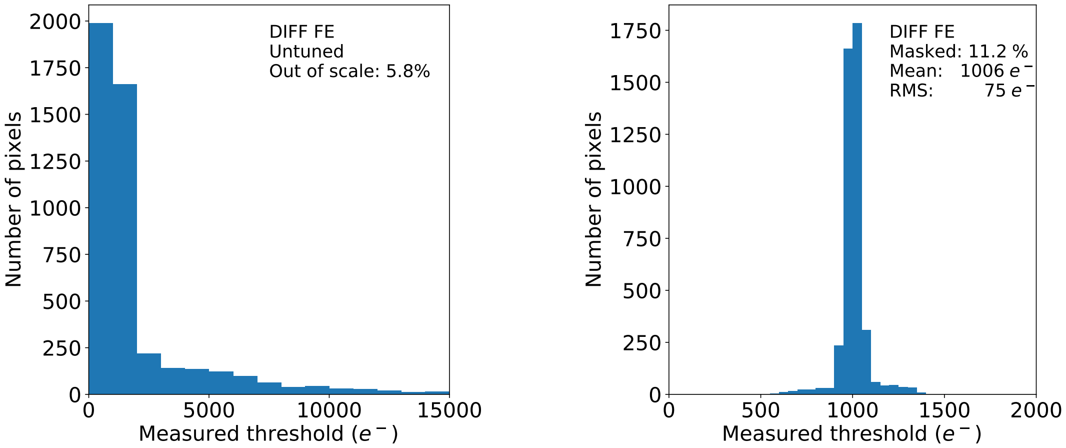


Figure 3: Distribution des seuils avant (gauche) et après l'ajustement local (droite) obtenue avec l'électronique analogique différentielle dans la puce de lecture RD53A après l'irradiation.

très proche du bruit de fond représenté en gris alors qu'elle est significativement plus élevée pour l'électronique linéaire (vert). Une amélioration du circuit a été proposée par les concepteurs. Les simulations du nouveau circuit ont prouvé une amélioration de temps de réponse et le circuit amélioré montre une fraction de signaux détectés en retard comparable aux deux autres électroniques. La version améliorée de l'électronique linéaire a aussi été produite dans une mini puce de test et testée avant et aussi après irradiation jusqu'à 1 Grad. Le temps de réponse a été amélioré dans les deux cas.

A l'issue cette l'évaluation, l'expérience CMS a choisi la version de l'électronique analogique linéaire avec un circuit amélioré pour l'intégration dans la puce de lecture finale prévue d'être disponible en 2021.

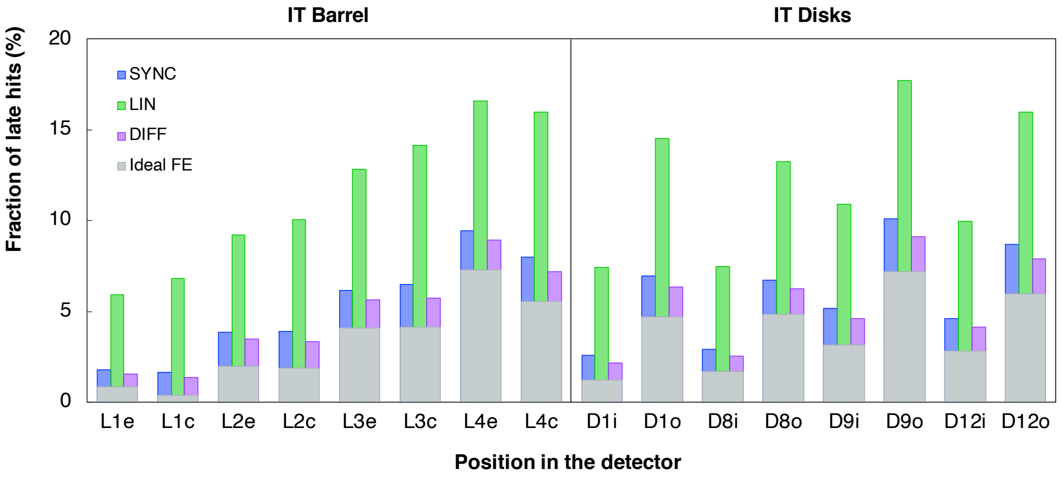


Figure 4: Fraction des signaux détectés en retard par chacune des trois électroniques analogiques frontales implémentées dans la puce de lecture RD53A évaluée pour différentes position des modules dans le détecteur.

Diaphonie dans les capteurs à pixels

Après le choix de l'électronique frontale pour la puce de lecture, un autre choix important pour le système concernait le capteur à pixels au silicium. En effet, CMS considère deux géométries possibles pour la forme des cellules du capteur: des pixels carrés de dimension de $50 \times 50\,\mu m^2$ ou des pixels rectangulaires de $100 \times 25\,\mu m^2$. Dans les deux cas, l'aire d'un pixel est de $2500\,\mu m^2$, ce qui correspond à la taille des pixels de la puce de lecture. Les deux types de pixels du capteur peuvent être connectés aux pixels carrés de la puce grâce aux électrodes de lecture appropriées. D'après les simulations, les pixels rectangulaires montrent une meilleure résolution des traces et ils sont surtout avantageux pour réduire la bande passante à l'avant du détecteur. En effet, dans les parties du détecteur où les particules arrivent sur les capteurs avec un angle d'incidence très faible, le signal créé s'étend à travers plusieurs pixels le long de l'axe des faisceaux. Les agrégats longs remplissent plus les mémoires tampon de la puce de lecture. L'utilisation des pixels rectangulaires avec un côté deux fois plus long dans cette direction réduit de moitié le nombre de pixels déclenchés en direction de l'axe des faisceaux et par conséquent réduit la bande passante. En revanche, les pixels rectangulaires sont aussi plus sensibles au phénomène de diaphonie (cross-talk en anglais). La diaphonie dans les capteurs est un effet dû au couplage capacitif entre pixels adjacents, qui peut induire des faux signaux sur les pixels avoisinant le pixel traversé par la particule. Avant de pouvoir décider quelle géométrie de pixel adopter pour le nouveau trajectomètre interne, il est donc important de quantifier la diaphonie.

Deux méthodes de mesure de diaphonie ont été élaborées dans le cadre de cette thèse. Les deux méthodes ont été appliquées à cinq capteurs différents, sélectionnés pour cette étude afin de comparer les résultats. Tous les cinq capteurs choisis ont été développés par la collaboration CMS pour l'amélioration du trajectomètre interne; quatre d'entre eux étaient des capteurs planaires et un était un capteur 3D. Parmi les quatre capteurs planaires, deux ont été fabriqués par *Hamamatsu Photonics K.K.* (HPK) et les deux autres par *Fondazione Bruno Kessler* (FBK) à Trente. L'un des capteurs HPK avait des pixels carrés et a été utilisé comme référence, puisque les pixels carrés ne sont pas censés être sujet à la diaphonie. L'autre capteur HPK avait des pixels rectangulaires. Les deux capteurs FBK avaient tous les deux des pixels rectangulaires, l'un était de conception standard tandis que l'autre a été amélioré dans le but de réduire la diaphonie. Ses implants ont été découpés pour éviter une superposition avec les électrodes de lecture entraînant un couplage capacitif. Ce nouveau type de capteur est de ce fait désigné comme "implant tronqué". Et finalement, un capteur 3D avec des pixels rectangulaires, fabriqué par *Centro Nacional de Microelectrónica* (CNM) à Barcelone, a aussi été étudié pour comparer la diaphonie dans les deux types de capteurs.

La première méthode de mesure de diaphonie utilisait le circuit de calibration intégré dans la puce de lecture. Ce circuit permet d'injecter dans chaque pixel différentes quantités de charge électrique, afin de simuler les signaux induits par les particules. La quantité de charge qu'il est nécessaire d'injecter dans un pixel pour induire un signal dans le pixel voisin, connaissant le seuil des pixels, permet de calculer la quantité de diaphonie. La diaphonie peut donc être visualisée et mesurée à partir des S-curves, c'est-à-dire l'occupation des pixels mesurée en fonction de la charge injectée. Dans un capteur sans aucune diaphonie entre pixels, la charge injectée en dessous du seuil ne permet pas de déclencher les pixels et l'occupation est nulle, tandis que l'injection de charge suffisamment au-dessus du seuil déclenche les pixels et cause une occupation de 100%, comme c'est représenté sur la Figure 5 à gauche, qui a été obtenue avec des pixels carrés. L'occupation en fonction de la charge pour un capteur avec des pixels rectangulaires est montrée sur la Figure 5 au milieu. Cette figure montre qu'une fois que la charge injectée est suffisamment grande, l'occupation atteint 200%, ce qui veut dire que chaque injection de charge déclenche deux pixels: le pixel injecté plus un voisin. Ceci met en évidence la diaphonie asymétrique dans les capteurs planaires avec des pixels rectangulaires. Lorsque ce test est fait avec un capteur 3D avec des pixels rectangulaires, comme c'est montré sur la Figure 5 à droite, l'occupation atteint 300%, ce qui indique que chaque injection de charge bien au-dessus du seuil déclenche trois pixels: le pixel injecté plus deux voisins et par conséquent la diaphonie dans les capteurs 3D est symétrique. De plus, si la charge moyenne à 50% est noté t (ce qui représente une mesure du seuil) et la charge moyenne à 150% est notée s, la diaphonie peut être calculée comme t/(t+s).

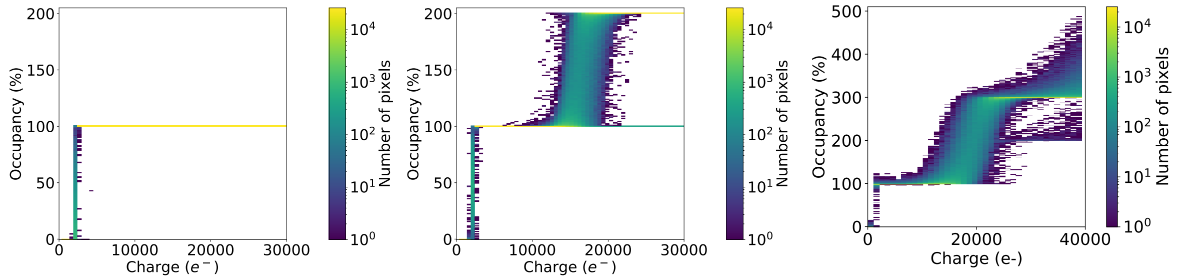


Figure 5: Occupation en fonction de la charge injectée dans un capteur planaire avec pixels carrés (gauche), avec pixels rectangulaires (milieu) et un capteur 3D avec pixels rectangulaires (droite).

Cette méthode de mesure de diaphonie avec injections de charge a été appliquée sur les cinq capteurs sélectionnés pour l'étude et le résultat est montré sur la Figure 6. Aucun signe de diaphonie n'est observé pour le capteur HPK avec des pixels carrés, en tout cas pas sur la gamme de charges qui peuvent être injectées par la puce RD53A. Ceci permet donc de mesurer la limite supérieure de la diaphonie dans ce capteur. La diaphonie la plus importante a été mesurée avec le capteur HPK avec des pixels rectangulaires. Les capteurs fabriqués à FBK montrent moins de diaphonie que celui de HPK et la diaphonie est réduite dans le capteur FBK aux implants tronqués. Finalement, la diaphonie dans le capteur 3D avec des pixels rectangulaires est symétrique et est donc représentée avec une autre couleur. De plus, le capteur 3D a la moindre diaphonie de tous les capteurs aux pixels rectangulaires. D'autre part, l'influence de différents paramètres sur la diaphonie a été vérifiée et les résultats ont montré que la diaphonie ne dépend ni du seuil, ni de la calibration TOT, mais elle dépend de la tension de polarisation du capteur atteignant sa valeur maximale lorsque le capteur et entièrement déplété. Il a été aussi observé que la diaphonie peut être réduite en augmentant le courant de polarisation à l'entrée du pré-amplificateur de l'électronique frontale.

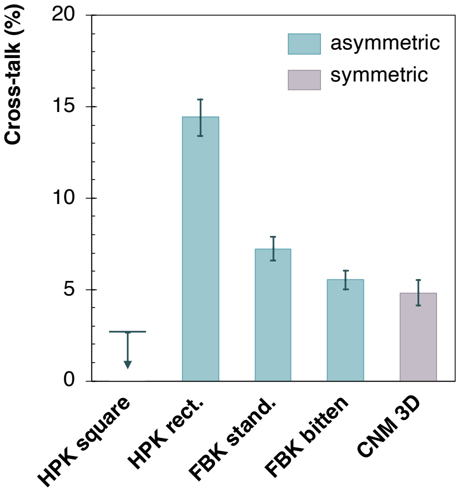


Figure 6: Diaphonie mesurée avec injections de charge dans les cinq capteurs sélectionnés.

La seconde méthode de mesure de diaphonie utilisait des particules issues d'un faisceau de test. Le faisceau d'électrons d'environ 5 GeV à DESY Hambourg a été utilisé pour implémenter et vérifier cette méthode et le télescope à pixels présent sur la ligne de faisceau à DESY a été utilisé pour avoir les traces de référence. Le télescope était composé de six capteurs MIMOSA26 placés perpendiculairement au faisceau. La méthode consistait à prendre des données avec le faisceau de particules en calibrant la puce aux différents seuils de détection. L'effet de diaphonie est en effet censé être visible dans le graphique de l'efficacité en fonction du seuil mesuré séparément pour le pixel traversé par la particule et pour les deux pixels adjacents, partageant le côté long du pixel. Les cinq capteurs choisis ont chacun été testés aux différents seuils de détection. A bas seuil d'environ 1000 e^-, requis comme le seuil pour l'opération du nouveau trajectomètre interne de CMS, tous les cinq capteurs ont démontré une très bonne efficacité. Les quatre capteurs planaires ont été efficaces à plus de 99%, tandis que le capteur 3D a atteint une efficacité légèrement au-dessus de 97%. En fait, l'efficacité est plus basse dans les capteurs 3D lorsque les particules arrivent sous incidence normale à cause de la présence des colonnes de polarisation où les particules ne peuvent pas être détectées. Cette légère inefficacité est mitigée lorsque le capteur est incliné de quelques degrés. La résolution spatiale des cinq capteurs a aussi pu être évaluée à partir des données prises à bas seuil. Pour le capteur avec des pixels carrés, la résolution a été d'environ 13 à 14 µm dans les deux directions, tandis que pour les pixels rectangulaires la résolution a été d'environ 30 µm sur du coté long et au tour de 7 µm sur le côté court du pixel.

Finalement, l’efficacité en fonction du seuil a été extraite et est présentée sur la Figure 7 pour tous les cinq capteurs testés. Le pixel pointé par la trace de référence fournie par le télescope est considéré comme étant le pixel effectivement traversé par la particule et est dénommé le pixel principal. L’efficacité en fonction du seuil de ce pixel est représentée en vert. Elle est grande à bas seuil, comme mentionné précédemment, et décroît lorsque le seuil est de plus en plus élevé. L’efficacité du pixel adjacent susceptible d’être déclenché par la diaphonie est représentée en rouge et celle de l’autre voisin n’étant pas sujet à la diaphonie à priori est représentée en bleu. Pour le capteur avec des pixels carrés (en haut à gauche), les deux voisins ont exactement la même efficacité, donc aucun signe de diaphonie n’est observé. Il en est de même pour le capteur 3D. En effet, cette méthode de mesure de diaphonie a été construite sur l’hypothèse que la diaphonie est asymétrique et induit un signal seulement dans un des voisins. De ce fait, cette mesure n’est pas sensible à la diaphonie dans le capteur 3D qui a été démontrée d’être symétrique par la première méthode. Pour les trois capteurs planaires avec des pixels rectangulaires (en haut au milieu et à droite et en bas à gauche), l’efficacité du pixel soumis à la diaphonie est plus élevée que l’efficacité du pixel sans diaphonie. En effet, la diaphonie asymétrique déclenche un des deux pixels, ce qui le fait apparaître plus efficace, tandis que l’efficacité de l’autre voisin est induite uniquement par le partage de la charge déposée, ce qui se

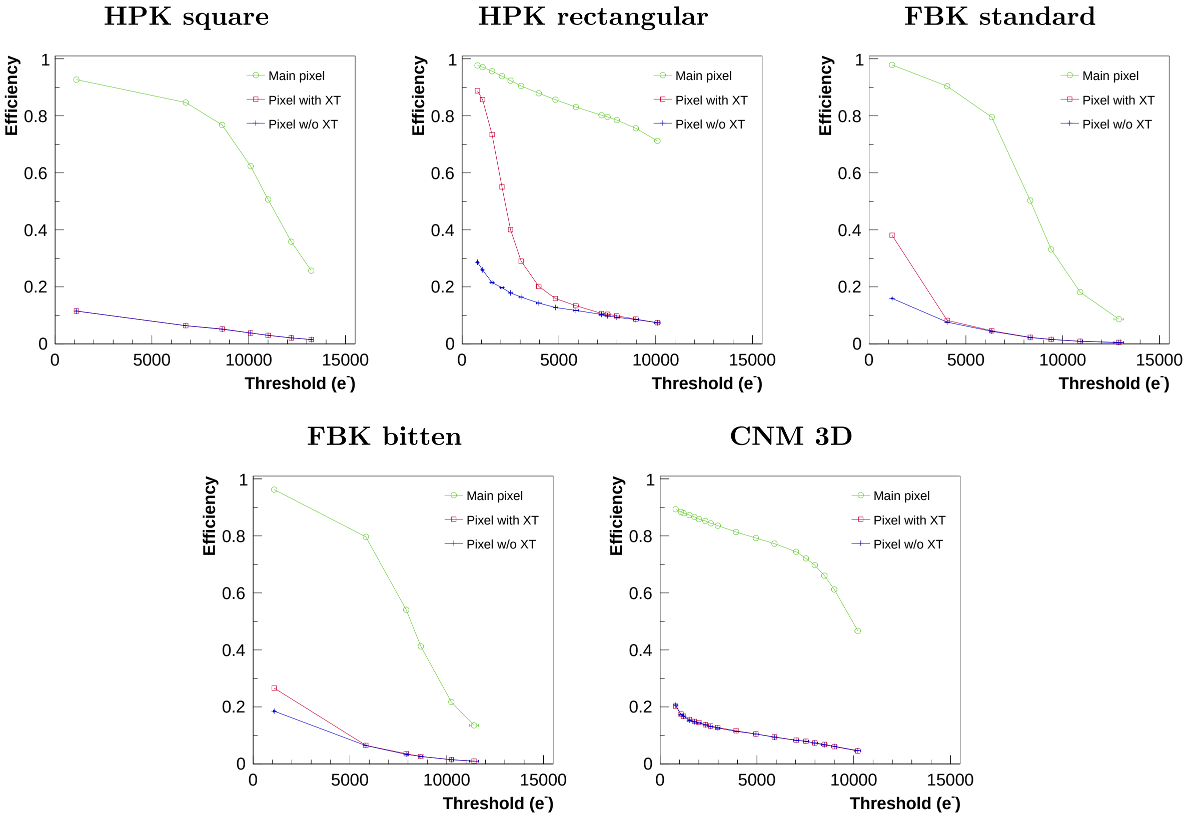


Figure 7: L’efficacité en fonction du seuil de détection mesurée avec le faisceau d’électrons de 5 GeV et les cinq capteurs sélectionnés mettant en évidence la diaphonie asymétrique.

produit notamment lorsque la particule traverse le pixel près du bord. La différence d'efficacité entre les deux pixels voisins est proportionnelle à la diaphonie. On remarque qu'elle est la plus grande dans le capteur HPK et la plus petite dans le capteur FBK aux implants tronqués, ce qui est en accord avec les résultats de mesure de diaphonie avec les injections de charge présentés sur la Figure 6.

La différence d'efficacité entre les deux voisins peut en principe être utilisée pour mesurer la quantité de diaphonie. Pour cela, il suffit de choisir une valeur d'efficacité et de relever le seuil pour lequel cette efficacité est atteinte par chacun des trois pixels. Une formule permet ainsi de calculer la diaphonie. Or cette méthode n'a pas pu être vérifiée dans ce travail puisque la mesure obtenue avec le capteur HPK avec des pixels rectangulaires ne comporte aucune efficacité simultanément atteinte par les trois pixels. Les deux capteurs FBK ont une petite gamme d'efficacités atteintes par les trois pixels, cependant comme les points de mesures ne sont pas nombreux les interpolations linéaires entre les points faussent le calcul. Une autre façon d'extraire la quantité de diaphonie à partir de ces données a donc été élaborée. Un modèle physique simple basée sur la distribution de charge attendue dans le capteur et tenant compte des processus de partage de charge déposée par les particules et ajoutant la diaphonie asymétrique, a été mis en place et implémenté dans une simulation Monter Carlo. Grâce à cela, les fonctions d'ajustement ont pu être élaborées. Dans un premier temps, un modèle simplifié a été utilisé pour faire un ajustement aux courbes d'efficacité en fonction du seuil. Cet ajustement ne donnait pas de très bons résultats, alors le modèle a été amélioré et au lieu d'ajuster l'efficacité en fonction du seuil, le profil d'efficacité a été extrait pour chaque seuil. Pour cela, l'efficacité de chacun des trois pixels étudiés a été représentée en fonction de la position de la trace de référence au sein du pixel principal, le long du côté court du pixel, comme c'est montré

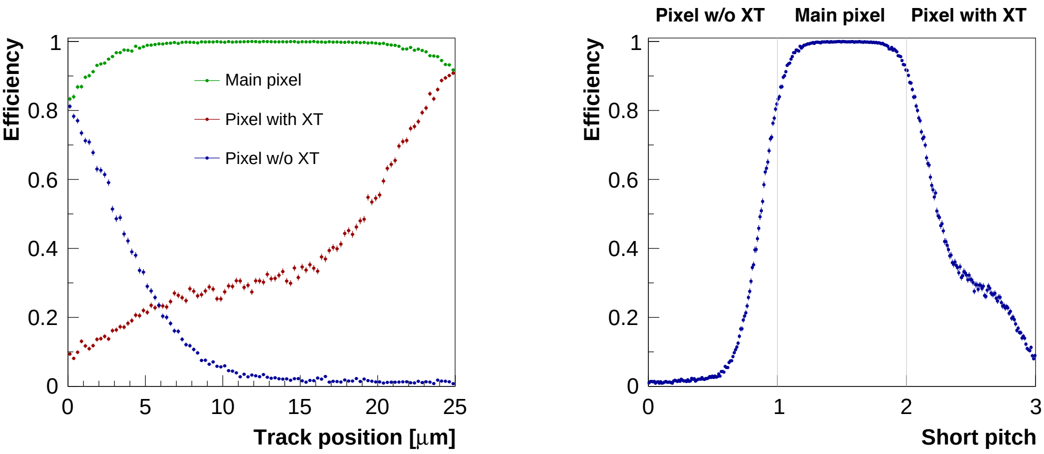


Figure 8: L'efficacité des trois pixels en fonction de la position de la trace au sein du pixel principal le long du côté court du pixel mesurée avec le capteur FBK standard avec un seuil d'environ 1100 e$^-$ (gauche) et l'efficacité en fonction de la position de la trace transformé en profile d'efficacité (droite).

sur la Figure 8 à gauche. Puis, les deux courbes pour les pixels voisins ont été inversées et placées de part et d'autre de la courbe du pixel principal afin d'obtenir un profil d'efficacité continu, comme c'est montré sur la Figure 8 à droite.

Le profil d'efficacité a été construit pour chaque capteur à bas seuil d'environs 1100 e^- et le résultat est montré sur la Figure 9 à gauche. Le côté gauche du profile représente l'efficacité induite par le partage de la charge déposée entre le pixel principal et le voisin et est très similaire d'un capteur a l'autre. Le côté droit du profil contient le même partage de la charge combiné avec la diaphonie asymétrique qui diffère d'un capteur a l'autre, comme attendu, confirmant une fois de plus la diaphonie la plus importante dans le capteur HPK avec des pixels rectangulaires et la moins importante dans celui avec des pixels carrés et dans le capteur 3D. Le profil d'efficacité a été extrait pour chaque capteur, pour chaque seuil séparément, comme c'est montré pour le capteur standard FBK sur la Figure 9 au milieu. L'ensemble des profils d'efficacité à différents seuils ont été ajustés simultanément par un ensemble de courbes d'ajustement où le seuil a été utilisé comme paramètre connu et fixe et cinq autres paramètres, parmi lesquels était aussi la quantité de diaphonie, ont été obtenus pour chaque capteur après une minimisation du χ^2 global. La quantité de diaphonie asymétrique ainsi extraite est représentée sur la Figure 9 à droite en jaune et comparaison avec la diaphonie obtenue par la méthode avec injections de charge représentée en vert. Les deux méthodes donnent des résultats similaires et montrent que les capteurs planaires aux pixels rectangulaires sont soumis a une diaphonie de l'ordre de 10% qui peut être réduite par une conception appropriée, en utilisant les implants tronqués. De plus, les simulations ont montré que la diaphonie de l'ordre de 10% a un impact négligeable sur la résolution des traces dans le trajectomètre interne et ne représente donc pas un argument suffisamment négatif pour contrecarrer les avantages des pixels rectangulaires. Le choix de la géométrie des pixels pour les nouveaux capteurs

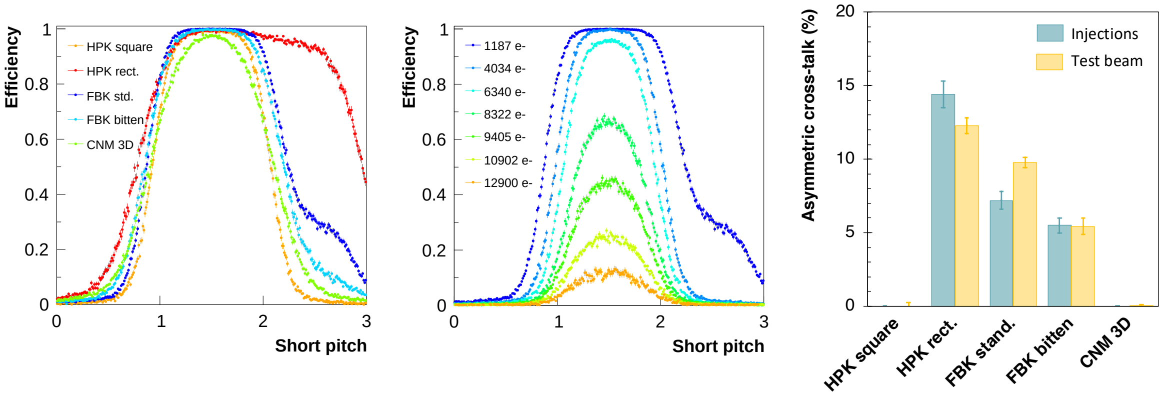


Figure 9: Profils d'efficacité pour chacun des capteurs sélectionnés obtenus à bas seuil (gauche), pour le capteur FBK standard obtenus aux seuils différents (milieu) et la diaphonie asymétrique extraite de l'ajustement des profils d'efficacité en comparaison avec celle mesurée avec des injections de charge (droite).

n'a pas encore été fait, mais à l'issue de cette étude de diaphonie, les pixels rectangulaires restent l'option favorite pour l'amélioration du détecteur.

Évaluation de la liaison électrique de données

Les capteurs à pixels au silicium et l'électronique frontale forment ensemble les modules à pixels où les particules sont détectées. Le composant suivant dans la chaîne d'acquisition est le câble électrique assurant la communication avec les modules à pixels et l'extraction des donnés, i.e. assurant la liaison de données. En apparence une partie simple du développement du détecteur, les câbles électriques présentent cependant quelques défis technologiques. En fait, plus de 4000 câbles électriques vont être nécessaires pour transmettre les signaux de contrôle et de calibration aux modules à pixels, et plus de 7000 câbles pour extraire les données. Étant donné l'espace limité dans le détecteur et afin d'éviter de perturber le passage des particules dont on veut mesurer la trajectoire, les câbles doivent être aussi fins que possible. Or plus un câble est fin, plus il est difficile de conserver une bonne intégrité des signaux électriques, en particulier lorsque ceux-ci sont transmis à haute vitesse comme ce sera le cas dans le nouveau trajectomètre interne de CMS.

Des prototypes de câbles électriques ont été conçus afin de minimiser leur masse tout en préservant l'intégrité des signaux transmis suffisante pour être correctement réceptionnés et interprétés par la puce d'agrégation lpGBT. Deux types de câbles sont considérés pour l'intégration dans le nouveau système d'acquisition: les paires torsadées non blindées et les câbles plats flexibles. Des prototypes de différentes longueurs, épaisseurs et topologies (dans le cas des câbles plats flexibles) ont été développés et devaient être évalués et comparés pour choisir la meilleure liaison de données pour le futur détecteur. En particulier, la transmission correcte des données des modules à pixels jusqu'aux puces d'agrégation lpGBT doit être assurée. Dans ce travail, plusieurs prototypes de câbles électriques ultra légers ont été évalués. D'une part, des paires torsadées non blindées étaient disponibles avec deux épaisseurs de fil différentes 34 AWG et 36 AWG et quatre longueurs différentes pour chaque épaisseur 35 cm, 1 m, 1.4 m et 2 m. D'autre part, trois différentes conceptions de câbles plats flexibles ont été évaluées, une topologie droite, une bifurquée et une en anneau. De plus, un câble plat flexible droit acheté dans le commerce a aussi été évalué et comparé aux prototypes. Pour chaque test, une paire de câbles standard commerciaux SMA de 20 cm a été mesurée et prise comme référence.

Dans un premier temps, ces prototypes ont été évalués en tant que composants électriques indépendants du reste du système. Pour cela, leurs caractéristiques ont été mesurées à l'aide d'un analyseur de réseau. Cet instrument de pointe a permis d'obtenir les paramètres S de chaque câble indiquant l'atténuation et les réflexions du signal. A partir des paramètres S de chaque câble l'impédance a pu être extraite. L'impédance est la caractéristique la plus importante pour un système de transmission de signaux électriques et elle doit être aussi homogène que possible entre différents composants du système. L'impédance différentielle cible pour la chaîne d'acquisition du nouveau trajectomètre interne de CMS et de 100 Ω et les câbles électriques doivent avoir une impédance aussi proche que possible de cette valeur pour minimiser des distorsions et réflexions des signaux. L'impédance instantanée mesurée avec l'analyseur du réseau en fonction du temps de propagation d'une sinusoïde le long de chaque câble est présenté sur la Figure 10. Le résultat montre que les câbles qui correspondent le mieux à l'impédance requise (vert) sont les paires torsadées de 36 AWG (bleu claire) et le câble plat flexible en anneau (rose et violet).

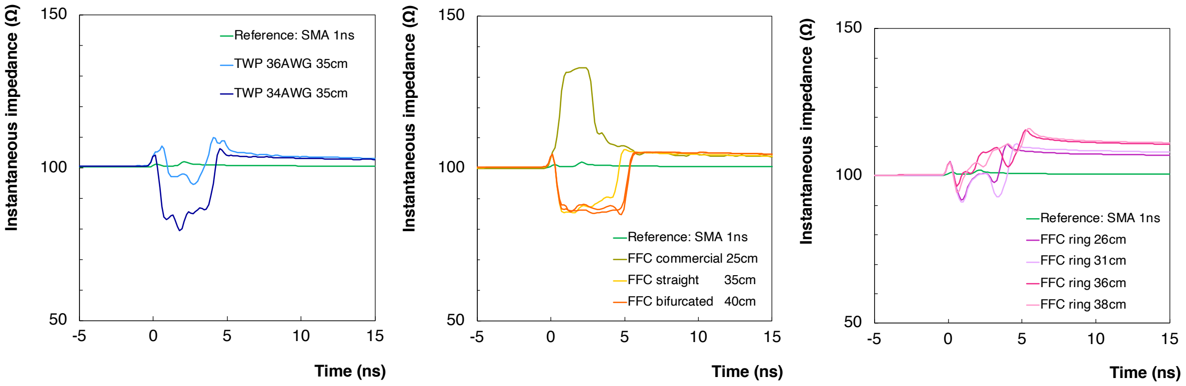


Figure 10: Impédance instantanée en fonction du temps de propagation du signal le long du câble mesurée avec l'analyseur de réseau pour différent prototypes des câbles électrique conçus pour l'amélioration du trajectomètre interne de CMS: paires torsadées de 35 cm (gauche), câbles plats flexibles droits et bifurqué (milieu) et câble plat flexible en anneau (droite).

Ensuite, les prototypes ont été évalués intégrés dans un système d'acquisition. La puce de lecture RD53A a été utilisée pour générer un signal de test bien défini qui a été transmis par différents câbles à 1.28 Gb/s pour étudier la qualité du signal après sa transmission à travers le câble à l'aide d'un oscilloscope. En particulier, l'amplitude du signal devait être vérifiée compte tenu de son atténuation dans les câbles, mais surtout parce que la puce d'agrégation lpGBT a une amplitude minimale requise pour recevoir les données correctement. Le jitter, c'est-à-dire la fluctuation du signal, est un autre facteur important qui ne doit pas dépasser un maximum requis par le lpGBT. Ces deux paramètres ont été mesurés à partir de diagrammes de l'œil pour différent câbles. De plus, la fonctionnalité appelée pré-accentuation implémentée dans la puce RD53A et son influence sur les diagrammes de l'œil a été étudiée. La pré-accentuation permet d'accentuer l'amplitude et les hautes fréquences du signal avant sa transmission par la liaison de données. Ainsi lors de la transmission à travers le câble, cette amplitude et hautes fréquences ajoutées sont atténuées et le signal reçu est égal au signal d'origine. La pré-accentuation dans la puce RD53A peut être configurée à l'aide de trois paramètres, dénommées TAP0, TAP1 et TAP2. Le premier ajuste l'amplitude du signal et les deux autres accentuent les hautes fréquences. Les résultats d'étude de la pré-accentuation ont montré que l'amplitude du signal augmente de manière linéaire avec le paramètre TAP0 tandis que ce paramètre n'a pas d'influence sur le jitter. Les paramètres TAP1 et TAP2 peuvent être optimisés selon l'atténuation induite par chaque câble. Une configuration optimale permet d'ouvrir le diagramme de l'œil, i.e. augmente l'amplitude et diminue le jitter, comme c'est montré sur un exemple sur la Figure 11 à gauche et au milieu. Par contre, une pré-accentuation trop importante dégrade la qualité du signal comme c'est montré sur la Figure 11 à droite.

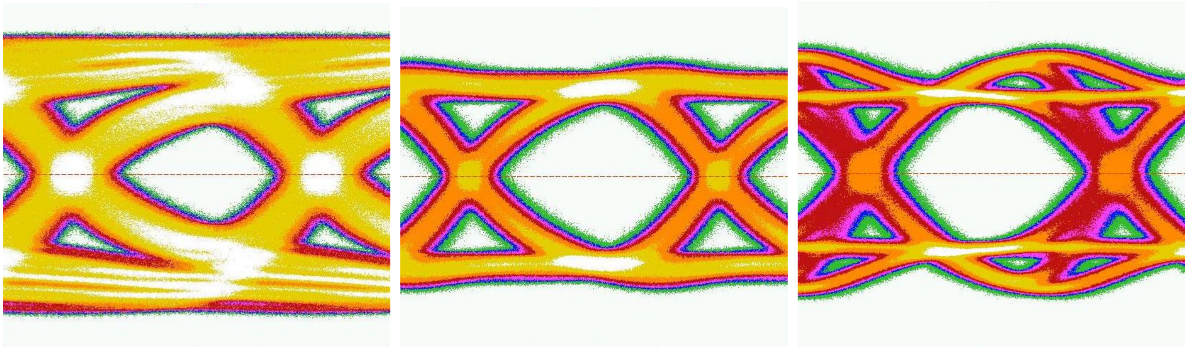


Figure 11: Diagrammes de l'œil obtenus avec une paire torsadée de 36 AWG et 2 m de longueur illustrant l'influence du paramètre TAP1 de la pré-accentuation de RD53A sur la forme du signal transmis: sans pré-accentuation (gauche), une pré-accentuation optimale (milieu) et trop de pré-accentuation (droite).

Les résultats de cette étude ont mené aux conclusions suivantes. La longueur du câble a une influence significative sur la qualité du signal; un câble plus long atténue plus l'amplitude et induit plus de jitter. A une longueur comparable et sans aucune pré-accentuation, les paires torsadées fournissent une meilleure qualité du signal. De plus, la dégradation du signal avec la longueur est plus importante dans les câbles plats flexibles. Pour cette raison, les paires torsadées sont de meilleurs candidats pour les parties du détecteur ou des câbles plus longs sont nécessaires. Entre les deux épaisseurs de paires torsadées, celles avec les fils plus fins de 36 AWG montrent une meilleure performance de transmission, ce qui est en accord avec une meilleure correspondance de l'impédance avec les 100 Ω requis pour le système. Les trois paramètres de pré-accentuation de RD53A ont été optimisés pour chaque câble étudié et l'amplitude obtenu est montrée sur la Figure 12. Tous les câbles jusqu'à 1 m satisfont l'amplitude minimale requise par le lpGBT (ligne rouge) sans la pré-accentuation lorsque l'amplitude du signal est poussée au maximum (vert). Les câbles plus longs ont besoin de la pré-accentuation (bleu) pour atteindre l'amplitude requise. Avec la pré-accentuation optimisée tous les prototypes jusqu'à une longueur de 2 m satisfont la spécification en amplitude du signal ce qui est un résultat très prometteur notamment parce que la longueur maximale des câbles dans le détecteur va être de 1.6 m. Ainsi différentes conceptions de câbles électriques ultra légers ont été validées dans ce travail jusqu'à une longueur de 2 m.

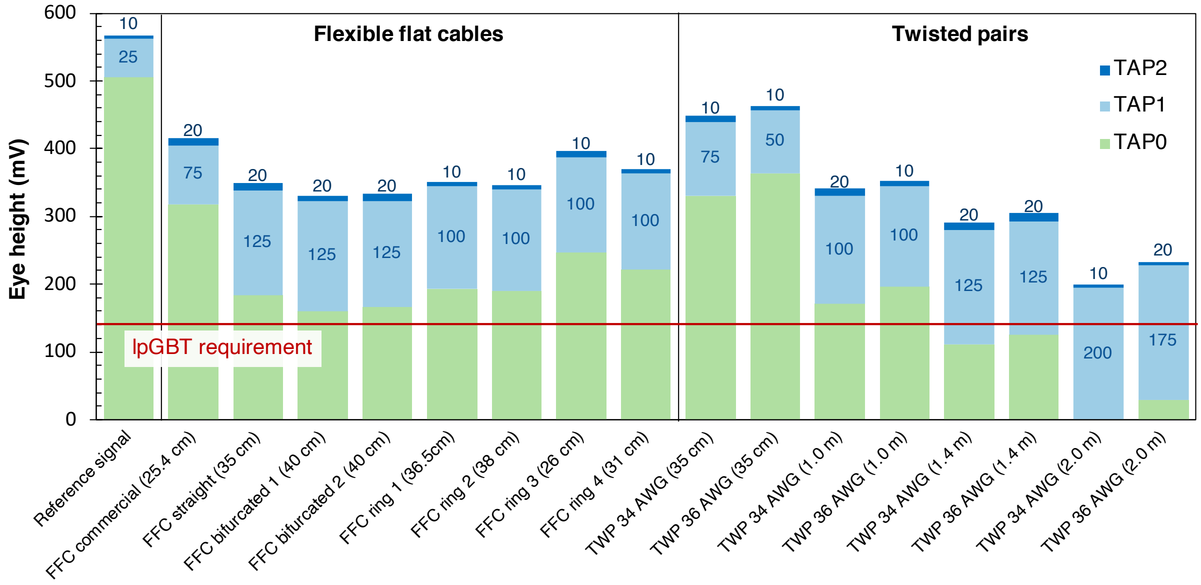


Figure 12: Amplitude du diagramme de l'œil obtenue avec chaque câble électrique avec l'amplitude du signal poussé au maximum et la pré-accentuation optimisée.

Conclusion

Cette thèse portait sur le développement du nouveau système d'acquisition pour l'amélioration du trajectomètre interne de CMS. Elle s'articulait en trois parties. Dans un premier temps, la puce de lecture RD53A a été testée et un particulier ses trois électroniques analogiques frontales. Les résultats des tests ont permis de choisir l'électronique linéaire avec un circuit amélioré pour l'intégration dans la nouvelle puce de lecture pour CMS. Ensuite, deux méthodes pour mesurer la diaphonie dans les capteurs à pixels ont été proposées, expliquées et vérifiées sur cinq capteurs différents dans le but de contribuer au choix entre les pixels carrés et les pixels rectangulaires pour les nouveaux capteurs. La diaphonie asymétrique de l'ordre de 10% a été mise en évidence dans les capteurs planaires avec des pixels rectangulaires et deux façons de la réduire ont été identifiées: les implants troqués et le courant de polarisation du pré-amplificateur. La conclusion de cette étude est que la diaphonie dans les pixels rectangulaires n'empêche pas leur utilisation dans le détecteur et ils restent l'option favorite. Et finalement, différents prototypes de câbles électriques ont été évalués et comparés pour implémenter la liaison de données du détecteur assurant une intégrité des signaux électriques suffisante pour être correctement réceptionnés par le lpGBT. Tous les prototypes pouvaient satisfaire le critère d'amplitude grâce à la pré-accentuation. Les paires torsadées non blindées ont été identifiés comme meilleurs candidats pour les parties du détecteur où de long câbles seront nécessaires alors que les câbles plat flexibles restent une option pour les parties où une topologie particulière pourrait faciliter l'intégration et une conception très dense réduirait la masse des câbles.

Introduction

Particle physics studies and aims at explaining the fundamental laws of our Universe governing at subatomic scale. Our current understanding is embodied in the standard model of particle physics, which provides a coherent picture of the fundamental constituents of matter, the elementary particles and their interactions, themselves described as an exchange of elementary particles. Remarkably, the Standard Model provides a successful description of the current experimental data and its development would not have been possible without a close interplay between theory and experiment.

Several breakthroughs in particle physics have come from experiments at high-energy particle accelerators. The largest and most powerful particle accelerator built to date is the Large Hadron Collider (LHC). This 27 km-long ring of superconducting magnets accelerates particles close to the speed of light and creates head-on collisions at unprecedented energy. The particles produced in these collisions are detected and identified by particle physics experiments relying on various concentric cylindrical detectors. The LHC was built to test the predictions of the standard model and besides its other important achievements, it was crowned with the discovery of the Higgs boson in 2012, adding the last missing ingredient to the standard model and fulfilling its primary goal.

Whilst the standard model is undoubtedly one of the great triumphs of modern physics, it is not an ultimate theory and many questions still remain unanswered. Therefore, the LHC will undergo a major upgrade in the upcoming years, which will boost the potential for discoveries of the LHC particle physics experiments. This so-called High Luminosity LHC upgrade will allow to collect ten times more data within a decade of data taking and hence to significantly increase the probability to observe new very rare phenomena.

The particle physics experiments at the LHC, such as the CMS experiment, need to be upgraded as well to cope with the challenging conditions of the High Luminosity LHC and to maintain their excellent performance. In particular, the CMS Inner Tracker pixel detector, which is the innermost detector of the CMS experiment, will be exposed to the most extreme radiation environment and the highest particle density. Hence, the current CMS Inner Tracker will be replaced with a totally new pixel detector developed to withstand the extreme conditions, while maintaining its excellent tracking and vertex reconstruction performance also at the High Luminosity LHC.

This thesis starts with a brief overview of the standard model, exposing the current status of the particle physics in Chapter 1, followed by a discussion of some of the major open questions not answered by the standard model and representing the main motivation for the High Luminosity upgrade of the LHC. The CMS experiment with its various subsystems is presented in Chapter 2 as well as its upgrade for the High Luminosity LHC with a focus on the upgrade of the Inner Tracker pixel detector representing the framework of this thesis.

This work was devoted to the readout system development for the new CMS Inner Tracker, and in particular to the evaluation of prototypes of the first three components in the detector readout chain: the pixel readout chips and pixel sensors, composing together the hybrid pixel modules, where the particles are detected and the electrical links carrying the data from the pixel modules to the rest of the data acquisition system. The study of these three components correspond to the three main topics of this thesis and each of them contributed to one important system choice. The analogue front-end evaluation presented in Chapter 3 represented an important contribution for the selection of the most suitable option for the new CMS pixel readout chip. The establishment of methods to evaluate cross-talk in pixel sensors explained in Chapter 4 will help the choice of the sensor pixel geometry for the new detector. And finally, the evaluation of different types of electrical links and validation of the signal integrity in these prototypes is addressed in Chapter 5 and will contribute to the final implementation of the data links for the detector.

Chapter 1

Physics motivations

Four fundamental interactions govern the Universe. They are called the *gravitational*, *electromagnetic*, *strong* and *weak* interactions. The gravitational and electromagnetic forces have significant long-range effects that we experience in our everyday life, while the strong and weak force act at subatomic distances and govern nuclear interactions. Each of the four fundamental interactions can be mathematically described as a field. The gravitational force is attributed to the curvature of space-time and described by Einstein's general theory of relativity. The other three forces are described as quantum fields with their interactions mediated by elementary particles described by the *standard model* (SM) of particle physics.

Decades of experiments have contributed to the development of the SM, which agrees very well with experimental results, but many important questions remain unanswered. What is the origin of matter in the Universe? Can the fundamental interactions be unified? What is the nature of dark matter and dark energy? These, and many other fundamental questions, are addressed by particle physics experiments studying the products of particle collisions at high energies. Nowadays more than ever, the collider-based *high energy physics* (HEP) experiments became essential to make a progress in our understanding of the Universe.

An overview of the SM particles is presented in Section 1.1 followed by the discussion of the main shortcomings of this theory. The world largest and most powerful particle accelerator that greatly contributed to the testing of the SM and other theories at the energy frontier is presented in Section 1.2. And the chapter closes with the discussion of the physics motivations, the design implementation and the challenges of the upcoming major upgrade of this accelerator.

1.1 The standard model of particle physics

All the current knowledge about the *elementary particles* and their *interactions* is embodied in the SM [1–3] of particle physics. This theoretical model was built in the early 1970's upon the mathematical foundations of the *quantum field theory* (QFT) [4], the gauge theories and experimental observations. The SM has proven not only the ability to describe (almost) all current experimental results, but also the capability to precisely predict a variety of new phenomena. Over time and through many experiments, the SM has become established as a well-tested physics theory representing one of the triumphs of modern physics.

The SM classification of all known elementary particles is presented in Figure 1.1. Each particle has a set of quantum numbers, which are discrete and conserved in particle interactions. The SM divides particles into two categories based on their intrinsic angular momentum called *spin*. The particles with half-integer spin are called *fermions* and compose the matter, while the particles with an integer spin are called *bosons* and describe particle interactions, except for the Higgs boson which has a special role in the SM.

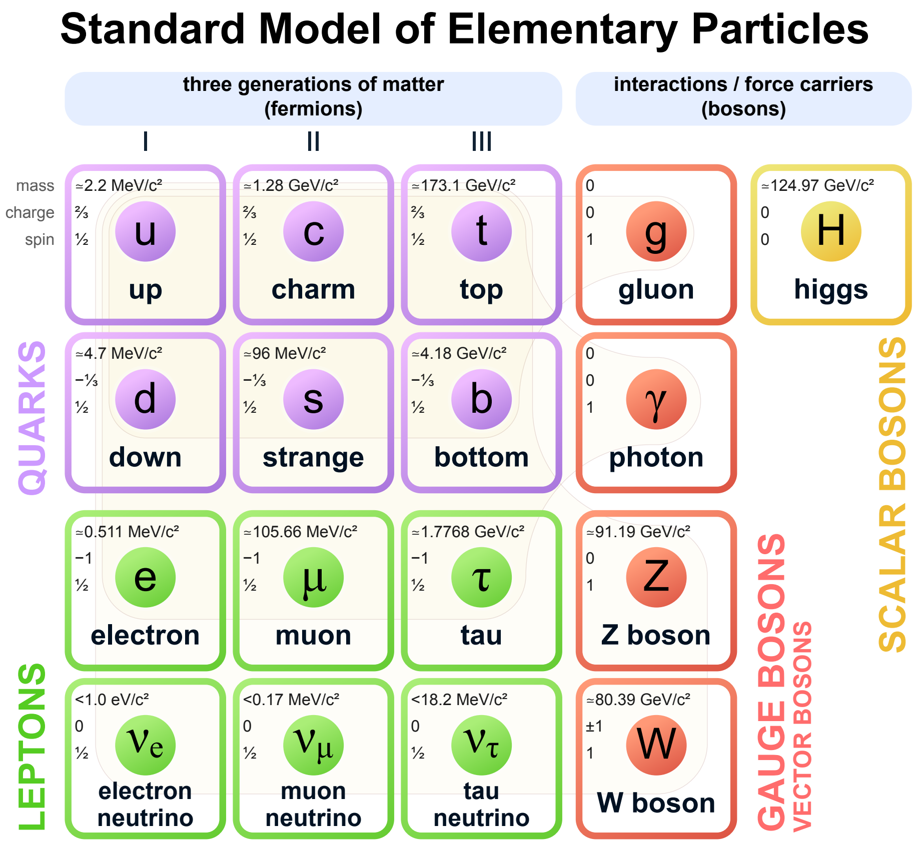


Figure 1.1: Standard model classification of known elementary particles [5].

1.1.1 Fermions

The twelve known fermions, which are the building blocks of matter, satisfy the Fermi-Dirac statistics and the Pauli exclusion principle. Their dynamics are described by the Dirac equation [3, Chapter 4] and the main consequence is that each fermion has a corresponding anti-particle with exactly the same mass, but opposite charge. The fermions are classified in two types: *leptons* and *quarks*.

The six leptons, represented in green in Figure 1.1, are arranged in three generations. The electron, muon and tau are charged leptons with an electric charge of -1 and a sizeable mass. They are sensitive to the electromagnetic and weak force. The neutrinos are electrically neutral and therefore, they only interact via the weak force, which makes them very difficult to detect. They are considered as massless particles within the SM.

The six quarks, depicted in violet in Figure 1.1, are also organised by pairs in three generations. The quark mass increases with increasing generation; the top quark being the heaviest known particle. Quarks carry electric charge equal to $2/3$ for the up, charm and top and $-1/3$ for the down, strange and beauty. Hence, they interact electromagnetically. They are also sensitive to the weak interaction and their defining property is their colour charge, which makes them interacting via the strong force as well. Unlike leptons, which are observable in a free state, quarks only exist in bound states, confined together in colour-neutral composite particles called *hadrons* because of the phenomenon of colour confinement. The hadrons composed of a quark and an anti-quark are called *mesons* and the hadrons made of three quarks are called *baryons*.

The lightest particles from the first generation are the most stable and they compose all the ordinary matter in the Universe: the up and down quarks bound into protons and neutrons form the atomic nuclei, which together with the electrons form the atoms. The heavier particles from the second and third generation are unstable and quickly decay into the stable lower mass particles. The study of higher-generation particles requires high-energy experiments, using either particle accelerators and detectors or cosmic rays.

1.1.2 Bosons

The SM describes fundamental interactions between fermions as the exchange of force-mediating particles called *gauge bosons* and these interactions are visually represented by *Feynman diagrams* [6]. The gauge bosons, represented in red in Figure 1.1, are vector bosons with a spin equal to one that obey the Bose-Einstein statistics.

Within the SM, the electromagnetic interaction, which acts between electrically charged particles, is carried by the *photons* (γ). It creates electric and magnetic fields, responsible for the attraction between orbital electrons and atomic nuclei which holds atoms together, as well as for the chemical bonds and the electromagnetic waves. The weak interaction responsible for the radioactive decay is acting on all fermions and is carried by the $W^{\pm}$ and Z bosons. The electromagnetism and the weak force are considered as two aspects of the electroweak interaction and have been unified with the electroweak theory.

The strong interaction acting between particles with colour charge is carried by *gluons* (g). It is responsible for quarks binding together to form hadrons and it also creates the nuclear force that binds the protons and neutrons together to form atomic nuclei. The gravitational force is not described by the SM, but its intensity is so weak at the mass scale of elementary particles compared to the other three interactions that it can be neglected.

And finally, the most recently discovered particle, the *Higgs boson* (H), depicted in yellow in Figure 1.1, is a scalar boson with a spin equal to zero. It is associated with the Higgs field and the Brout-Englert-Higgs mechanism [7, 8] through which the particles acquire their masses.

1.1.3 Beyond standard model

Although the SM provides a coherent picture of all the elementary particles known to date and describes their interactions via three fundamental forces to an amazing precision, it is still incomplete and some fundamental questions are remaining unsolved [9]. The QFT used to describe the electroweak and strong interactions at small scale is for the moment incompatible with the general relativity describing the gravity at large scale. The SM also fails to explain why the gravity is about 40 orders of magnitude weaker than the electromagnetic or nuclear forces. Another unsolved question of the SM is the mass of neutrinos, considered massless within the classical SM, while the neutrino oscillations observed in various experiments prove that they have a non-zero mass [10]. The SM also does not explain why there are so many different matter particles with such big differences in mass, up to five orders of magnitude between the mass of electrons and the mass of the top quark. Other open questions concern the large asymmetry between the amount of matter and antimatter in the Universe or the origin and nature of the so-called *dark matter* and *dark energy*, supposed to compose about 95% of the Universe.

The SM is certainly a very successful and effective theory to which many mechanisms were added ad-hoc. Nevertheless, it is not the ultimate theory capable of explaining everything and there are strong indications that the SM only corresponds to a low-energy approximation of a more fundamental global theory denoted as physics *beyond the standard model* (BSM) [11]. Many different BSM theories postulating the existence of new particles or mechanisms were proposed to address the previously discussed shortcomings of the SM. The collider-based particle experiments, in particular those installed on the CERN Large Hadron Collider, are searching for new physics to verify the proposed BSM theoretical predictions.

1.2 High Luminosity LHC

The European Organisation for Nuclear Research or *Conseil Européen pour la Recherche Nucléaire* (CERN) [12] is a research organisation founded in 1954 and based near Geneva, at the Franco-Swiss border. The laboratory was originally devoted to the study of atomic nuclei, while nowadays it is also dedicated to high-energy physics to study the fundamental particles and their interactions. For this purpose, CERN operates the largest particle physics laboratory in the world, equipped with particle accelerators and detectors, used for various high-energy physics experiments.

Since its foundation, several important scientific breakthroughs took place at CERN, among which: the discovery of the neutral currents [13, 14] in 1973, the discovery of the W and Z bosons [14–18] in 1983, the first production of anti-hydrogen atoms [19] in 1996, the anti-hydrogen confinement for more then 15 min [20] in 2011 and ultimately the discovery of the Higgs boson [21, 22] in 2012, finally observed almost 50 years after its existence was first postulated [7, 8]. Besides numerous scientific achievements, CERN has also a leading role in new technology development and the most famous example is the creation of the World Wide Web [23, 24].

CERN nowadays operates a complex of eight particle accelerators and two decelerators. They are either used as injectors for larger accelerators, or to directly supply experiments with particles, or both. The CERN accelerator complex is presented in Figure 1.2. Each machine in the acceleration chain boosts the energy of a particle beam before injecting it into the next machine in the sequence. The details about the various accelerators composing the CERN's main acceleration chain are provided in Appendix A.

1.2.1 Large Hadron Collider

The *Large Hadron Collider* (LHC) [26] is the world's largest and most powerful particle collider, the largest and the most complex experimental facility ever built, and the largest single machine in the world. It was built between 1998 and 2008 in collaboration with over 10 000 scientists and engineers from over 100 countries and hundreds of universities and laboratories. This circular accelerator and collider is the latest addition to the CERN accelerator complex. It has a circumference of 26.7 km and it is installed in an underground tunnel with a diameter of 3.7 m, situated between 45 m and 170 m below ground level. The LHC tunnel, previously occupied by the *Large Electron-Positron collider* (LEP) [27], is located in the region between the Geneva Airport and the nearby Jura Mountains.

Inside the accelerator, two high-energy particle beams travel at a velocity close to the speed of light, before they are made to collide. The beams travel in opposite directions in two separate beam pipes, both kept within an ultrahigh vacuum to avoid interactions with the molecules of air. The particle beams are guided around the accelerator ring by

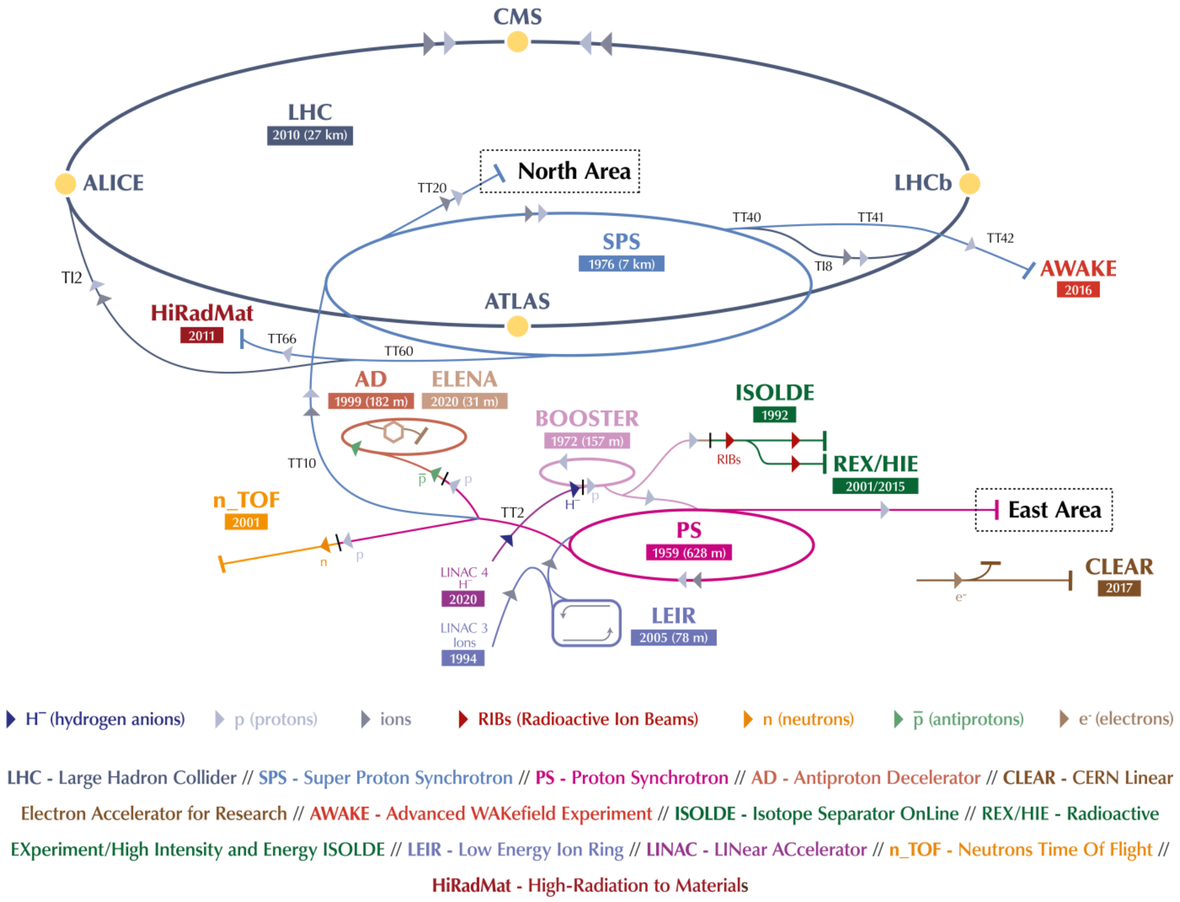


Figure 1.2: CERN accelerator complex [25].

a strong magnetic field maintained by superconducting magnets, cooled down to 1.9 K by superfluid helium. The LHC is made of 9593 niobium-titanium (Nb-Ti) magnets, among which 1232 dipole and 392 quadrupole magnets. The 15 m-long dipole magnets bend the beams to maintain them on a circular trajectory thanks to a strong magnetic field of 8.33 T. The quadrupole magnets, each 5 to 7 m long, focus the beams to keep the particles of same charge tightly together in bunches to maximise the number of collisions when the bunches collide. Besides these two key components, the accelerator is also equipped with many higher order magnets and radiofrequecy cavities. Figure 1.3 provides a view inside the LHC tunnel, showing a dipole magnet revealing its internal structure and the two beam pipes.

The LHC accelerates either protons or lead ions fully stripped of their electrons ($^{208}\text{Pb}^{82+}$) and produces proton-proton (p-p), proton-lead (p-Pb) and lead-lead (Pb-Pb) collisions*. The proton beams are made of 2808 proton bunches with 1.15×10^{11} protons per bunch. The proton bunches are spaced by 25 ns and are colliding at 40 MHz at four *interaction points* (IPs). The maximum design energy per proton beam is 7 TeV, resulting in col-

---

*Collisions of fully stripped xenon ions ($^{129}\text{Xe}^{54+}$) were also produced in the LHC during a special physics run in 2017 [29].

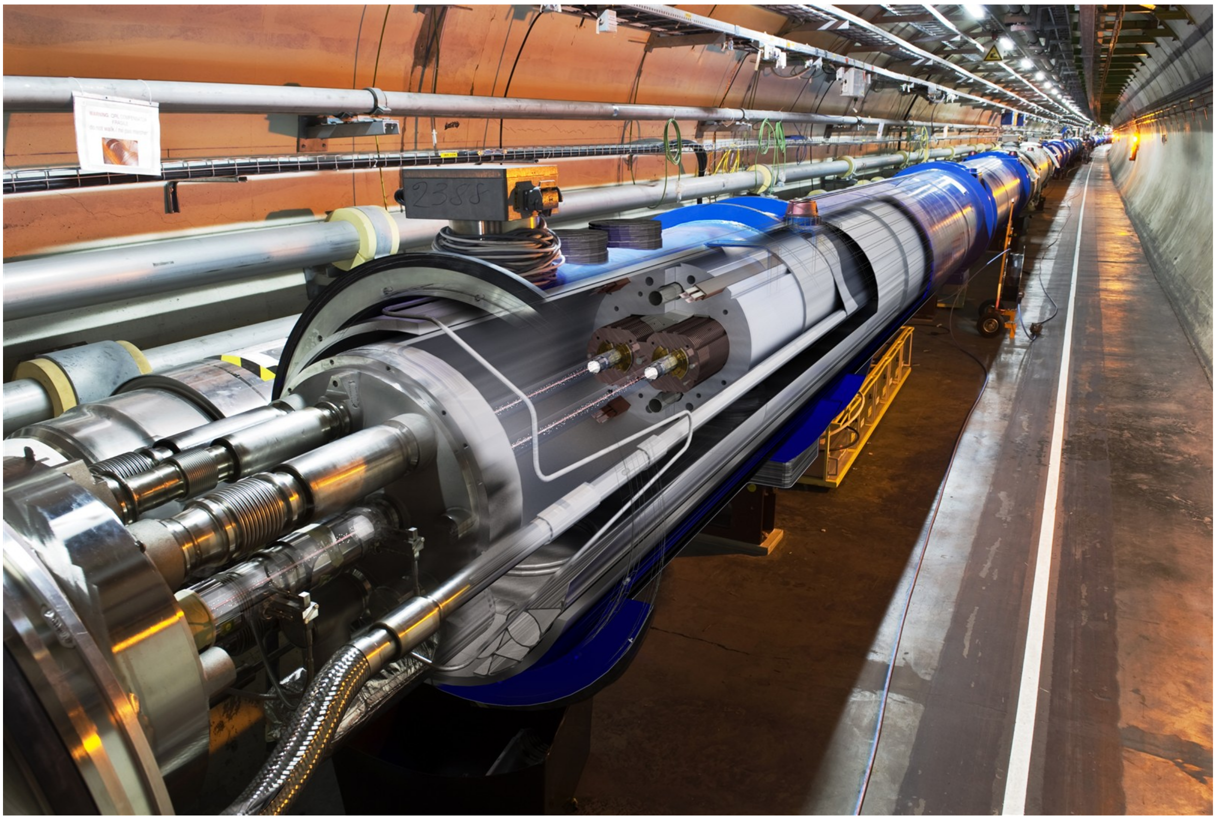


Figure 1.3: Photograph of the LHC in the underground tunnel with an opening in the dipole magnet showing the two beam pipes and the internal structure of the magnet [28].

lisions at the maximum design centre of mass energy of $\sqrt{s}$ =14 TeV. The products of these collisions are detected and recorded by different experiments. In the experimental caverns around the four IPs are installed the four major LHC experiments: ALICE [30], ATLAS [31], CMS [32] and LHCb [33]. *A Toroidal LHC ApparatuS* (ATLAS) and *Compact Muon Solenoid* (CMS) are general purpose experiments, covering a broad physics program and studying both proton and lead collisions. *A Large Ion Collider Experiment* (ALICE) is mainly dedicated to the heavy ion collisions to study the *quark-gluon plasma* (QGP). The *LHC beauty* (LHCb) experiment is focused on the heavy-flavour physics, studying the quarks of the 2^{nd} and 3^{rd} generation and in particular the beauty quark and the corresponding B mesons. An overview of the four experiments is presented in Figure 1.4.

The LHC first started its operation in September 2008. After only nine days of operation, an incident caused by an electrical failure mechanically damaged 29 magnets and released helium from the magnet cooling into the tunnel [34, 35]. This lead to a fourteen-month delay [36] in data taking and a lowering of the beam energy. During the first data-taking period, called *Run 1*, from 2009 to 2013, the LHC produced p-p collisions at the centre-of-mass energies of $\sqrt{s} = 7$ and 8 TeV. The Run 1 was followed by the *first long shutdown* (LS1) that lasted for two years. The LHC was restarted for the *Run 2* in 2015, and

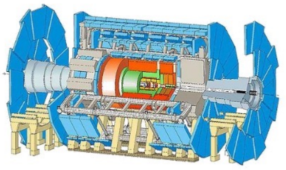
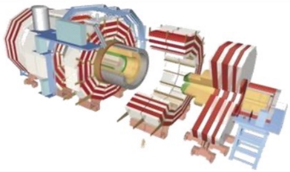
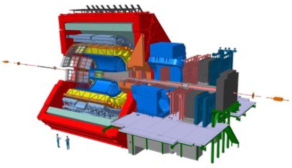
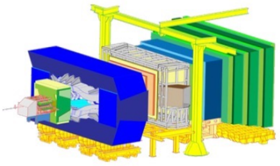
NAME	LOGO	DETECTOR	PURPOSE	LOCATION
ATLAS A Toroidal LHC ApparatuS	ATLAS EXPERIMENT		The biggest, general-purpose detector is probing for fundamental particles.	Point 1 Meyrin (CH)
CMS Compact Muon Solenoid	CMS		The heaviest, general-purpose detector, investigates the largest range of physics possible.	Point 5 Cessy (FR)
ALICE A Large Ion Collider Experiment	ALICE		ALICE studies quark-gluon plasma, a state of deconfined matter that existed just after the big bang.	Point 2 Sergy (FR)
LHCb LHC Beauty	LHCb		LHCb specializes in investigating the matter/antimatter asymmetry by studying physics of beauty quarks.	Point 8 Ferney-Voltaire (FR)

Figure 1.4: Overview of the four major LHC experiments.

produced p-p collisions at $\sqrt{s} = 13\,\text{TeV}$ until the late 2018. The currently ongoing *second long shutdown* (LS2) is devoted to preparations for the *Run 3* of the LHC data taking, during which the accelerator could finally reach its maximum design centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$. The LS2 period is also used for important upgrades of the four LHC experiments as well as for the upgrade of the accelerators of the LHC injection chain. According to the latest schedule, the first beams after this shutdown will circulate in the LHC in September 2021 and the Run 3 that will last for three years is foreseen to start in March 2022.

1.2.2 Luminosity and pileup

The two most important parameters characterising the performance of a particle collider are the beam energy and the rate of interesting collisions. The latter is directly proportional to the collider *luminosity* [37]. For a given process, the number of events per unit time is given by:

$$\frac{\partial N}{\partial t} = \mathcal{L} \times \sigma \tag{1.1}$$

where σ is the cross section, i.e. the occurrence probability of the process and $\mathcal{L}$ is the *instantaneous luminosity* expressed in $\mathrm{cm^{-2}\,s^{-1}}$ and defined as:

$$\mathcal{L} = \frac{f N_b N_p^2}{4\pi\sigma_x\sigma_y}\, F \tag{1.2}$$

where f is the revolution frequency of the particle bunches, N_b the number of bunches per beam, N_p the number of particles per bunch, $\sigma_{x,y}$ the transverse beam size along x and y axes and $F \leq 1$ is a geometrical correction factor for the crossing angle at the IP. In the LHC, the revolution frequency of proton bunches is $f = 11.245\,\mathrm{kHz}$, the number of proton bunches per beam is $N_b = 2808$, the number of protons per bunch is $N_p = 1.15 \times 10^{11}$ and the transverse beam size is $\sigma_x = \sigma_y = 16.6\,\mu\mathrm{m}$, which results in a design instantaneous luminosity of $\mathcal{L} = 1.2 \times 10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$.

The instantaneous luminosity integrated over time is called the *integrated luminosity* and can be written as:

$$L = \int_{t_1}^{t_2} \mathcal{L}(t)\; dt \tag{1.3}$$

It is expressed in inverse femtobarns ($\mathrm{fb^{-1}}$) with $1\,\mathrm{fb^{-1}} = 10^{-39}\,\mathrm{cm^{-2}}$ and it characterises the amount of collected data over a period of time. The total integrated luminosity delivered to date by the LHC amounts to almost $193\,\mathrm{fb^{-1}}$ [38] and it is expected to reach $350\,\mathrm{fb^{-1}}$ by the end of the Run 3.

Another important parameter for a particle collider such as the LHC, directly linked to the luminosity, is the pileup. The *pileup* (PU) arises from the fact that the beams are bunched and therefore, multiple interactions take place in each *bunch crossing* (BX). The PU corresponds to the number of inelastic proton–proton interactions per BX. Figure 1.5

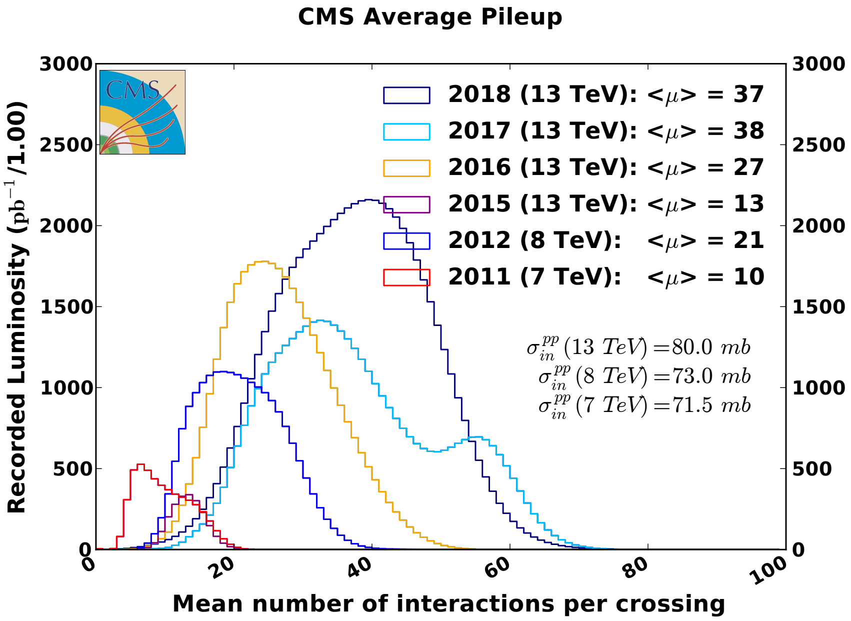


Figure 1.5: Pileup distributions for each year of the LHC proton-proton data taking [38].

shows the PU distributions recorded by the CMS experiment in different years of the p-p data taking. It demonstrates the increase of the average pileup throughout the years as a direct consequence of the increase in the luminosity. In the last year of data taking, before the LS2, the mean pileup was equal to 37 collisions.

1.2.3 High Luminosity Upgrade

The search for new physics at the LHC to answer some of the SM open questions require to collect very high statistics, either to increase the probability to observe new very rare phenomena with small cross-sections, or to perform high precision measurements of known processes searching for small deviations from the model. Increasing the statistics means increasing the amount of collected data and requires an increase in the integrated luminosity. To increase the current integrated luminosity by an order of magnitude in a timescale of about ten years requires to significantly increase the instantaneous luminosity of the collider. This will be achieved with the *High Luminosity LHC* (HL-LHC) [39, 40] upgrade, planned for installation during the *third long shutdown* (LS3) between 2025 and mid-2027, as it is indicated in the project schedule presented in Figure 1.6.

Figure 1.6: Official LHC and HL-LHC schedule [41].

The HL-LHC project [41] relies on several technological innovations. More powerful superconducting magnets made of niobium-tin (Nb_3Sn) will be used for the first time in a particle accelerator. Twenty-four Nb_3Sn quadrupoles will be installed on either side of the ATLAS and CMS experiments to strongly focus the particle bunches before they collide to increase the proton density within bunches and hence the number of collisions.

This will be achieved thanks to a higher magnetic field of 12 T, compared to the current Nb-Ti quadrupoles providing an 8 T field. In addition, two of the current dipoles will be replaced with four pairs of shorter and more powerful Nb_3Sn dipoles. They will bend particle trajectories over a shorter distance thanks to a higher magnetic field of 11 T, compared to the current Nb-Ti dipoles providing an 8.33 T field. This replacement will allow to free some space for new additional collimators and new injection magnets, which will improve the performance of the machine.

The particle bunches accelerated to very high velocities in the LHC become elliptical due to the relativistic effects. They cross each other at the IPs with a given crossing angle, which reduces the instantaneous luminosity and gives rise to the correction factor F in Equation (1.2). To mitigate this effect, a new type of superconducting radio-frequency cavities called *crab cavities* will be used for the first time in a particle collider. Sixteen crab cavities will be installed on both sides of the ATLAS and CMS experiments. They will tilt the bunches before the collision to produce head-on collisions and maximise the bunch overlap and thus increase the luminosity. After each collision, the crab cavities will also tilt the out going bunches back to the initial position. New beam optics for the HL-LHC will take care of maintaining a constant luminosity throughout the beam lifetime, which is referred to as the luminosity *levelling*. The performance of the HL-LHC will also rely upon the performance of the other accelerators in the accelerating chain. Hence, the HL-LHC upgrade will go in hand with the upgrade of the whole injection chain referred to as the *LHC Injectors Upgrade* (LIU) [42, 43].

After LS3, when the High Luminosity upgrade will be completed, the HL-LHC will start its operation with the *Run 4* and should provide p-p collisions at $\sqrt{s} = 14\,\mathrm{TeV}$, as it is indicated in the schedule in Figure 1.6. The HL-LHC proton beams will be twice brighter, with 2.2×10^{11} protons per bunch focused into more dense bunches prior to the collision. The instantaneous luminosity will be progressively increased during the Run 4 to reach $\mathcal{L} = 5.0 \times 10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$. If this nominal luminosity is maintained throughout the Run 4, 5 and 6, the HL-LHC will deliver about $250\,\mathrm{fb^{-1}}$ of p-p collisions per year, and up to $350\,\mathrm{fb^{-1}}$ per year after LS4 when the ion runs will be over. In this nominal scenario a total integrated luminosity of $3000\,\mathrm{fb^{-1}}$ is foreseen by the end of a ten-year operation period. An ultimate performance scenario is also envisaged, in which case the instantaneous luminosity will be further increased to $\mathcal{L} = 7.5 \times 10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$ during the Run 5 and 6 to deliver up to $470\,\mathrm{fb^{-1}}$ per year, eventually reaching a total of $4000\,\mathrm{fb^{-1}}$.

The HL-LHC will deliver at least ten times higher integrated luminosity, i.e. ten times more collisions then the LHC, which will considerably increase the discovery potential of its experiments. For instance, the HL-LHC will produce at least fifteen million Higgs bosons per year, compared to around three million produced by the LHC in 2017. This will allow to study the Brout-Englert-Higgs mechanism in greater details and to decrease the statistical uncertainty on some of the rare Higgs decays [44], e.g. the $t\bar{t}H$ production in the $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ channel with a di-lepton in the final state [45]. Having said that, an important limitation arising from the high luminosity operation of the LHC is a higher pileup. In the nominal luminosity scenario the pileup is expected to be on average 140 p-p collisions per BX, while in the ultimate scenario it will be pushed to up to 200. A higher pileup causes a higher detector occupancy, which makes it much more difficult for the detectors to disentangle all the simultaneous processes. Hence, a high pileup has an impact on the detection efficiency and the resolution. Figure 1.7 shows a simulated event display of the vertex region for a top-pair event with an additional 140 interactions and is illustrating how challenging the HL-LHC environment will be for the particle detectors.

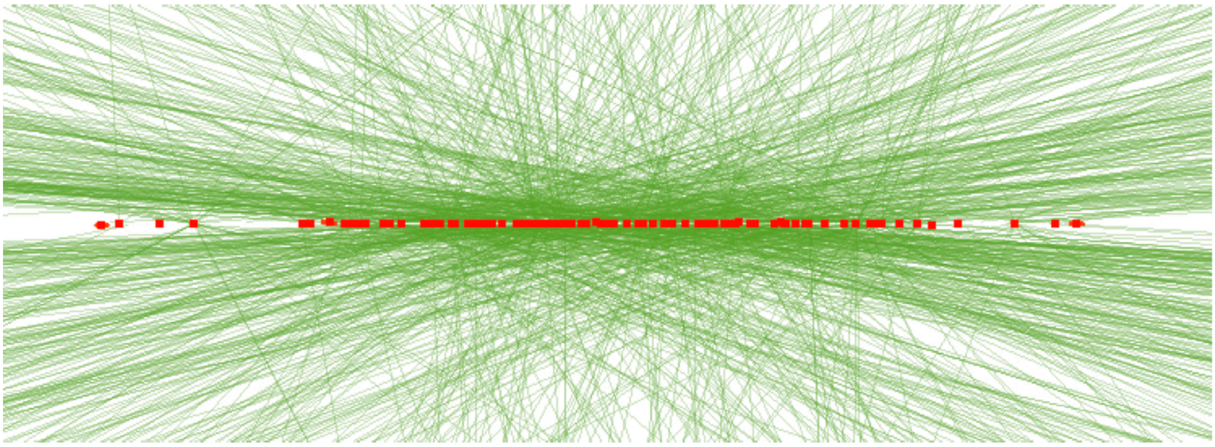


Figure 1.7: Event display showing reconstructed tracks and vertices of a simulated top-pair event overlaid with additional 140 interactions [46].

Chapter 2

The CMS Inner Tracker upgrade

The High Luminosity upgrade of the LHC will boost its potential for physics discoveries, but it will also impose extreme conditions for the experiments. Hence, the detectors will be upgraded along with the accelerator to achieve a sufficient radiation tolerance and to sustain the large particle rate, while maintaining the excellent detection performance. As one of the four major LHC experiments, the Compact Muon Solenoid (CMS) [32] will see many of its components replaced or refurbished.

The name of the experiment is related to the fact that it is really compact for all the material it is made of. In fact, even though it is not the biggest LHC experiment, it is the heaviest. It was designed to detect muons with great accuracy and its subdetectors are located inside the biggest superconducting magnet ever made providing a strong and uniform magnetic field in the tracker volume.

The CMS Collaboration is one of the largest international scientific collaborations in history. It counts over 5500 members, including particle physicists, engineers, computer scientists, technicians and students, coming from 229 institutes and universities, spread across 51 countries [47]. In July 2012, the CMS Collaboration along with its sister experiment ATLAS, announced the discovery of the Higgs boson [21, 22] and added another piece of puzzle to the standard model.

This chapter starts with an overview of the CMS detector and its subdetectors in Section 2.1 followed by an overview of the detector upgrade for the HL-LHC. In Section 2.2, the focus is put on the upgrade of the Inner Tracker, which is the main topic of this thesis.

2.1 The CMS detector

The CMS is a general-purpose particle detector covering a broad physics program. It is installed in the experimental cavern about 100 m underground, on one of the four

LHC collision points, denoted as *Point 5**, and located near to Cessy, a French village situated between the Geneva lake and the Jura mountains. With its total weight of about 12 000 t, CMS is the heaviest LHC detector. Its layout is shown in Figure 2.1. CMS is a 21.6 m-long cylinder with a diameter of 14.6 m. It is composed of several cylindrical layers coaxial to the beam direction, referred to as the *barrel*, closed at both ends by a set of discs orthogonal to the beam pipe, called the *endcaps*. This typical collider experiment structure offers a very good detector solid angle coverage.

The IP, where the particles are brought to collisions, is situated at the centre of the detector and defines the origin of the coordinate system. CMS adopts a right-handed cylindrical coordinate system. The x axis points towards the centre of the LHC, the y axis points vertically upwards and the z axis points along the beam direction. The azimuthal angle ϕ is measured from the x axis in the x-y plane, the radial coordinate in this plane is denoted as r and the polar angle θ is measured from the z axis. The pseudorapidity η is defined as $\eta = -ln\tan(\theta/2)$ [32].

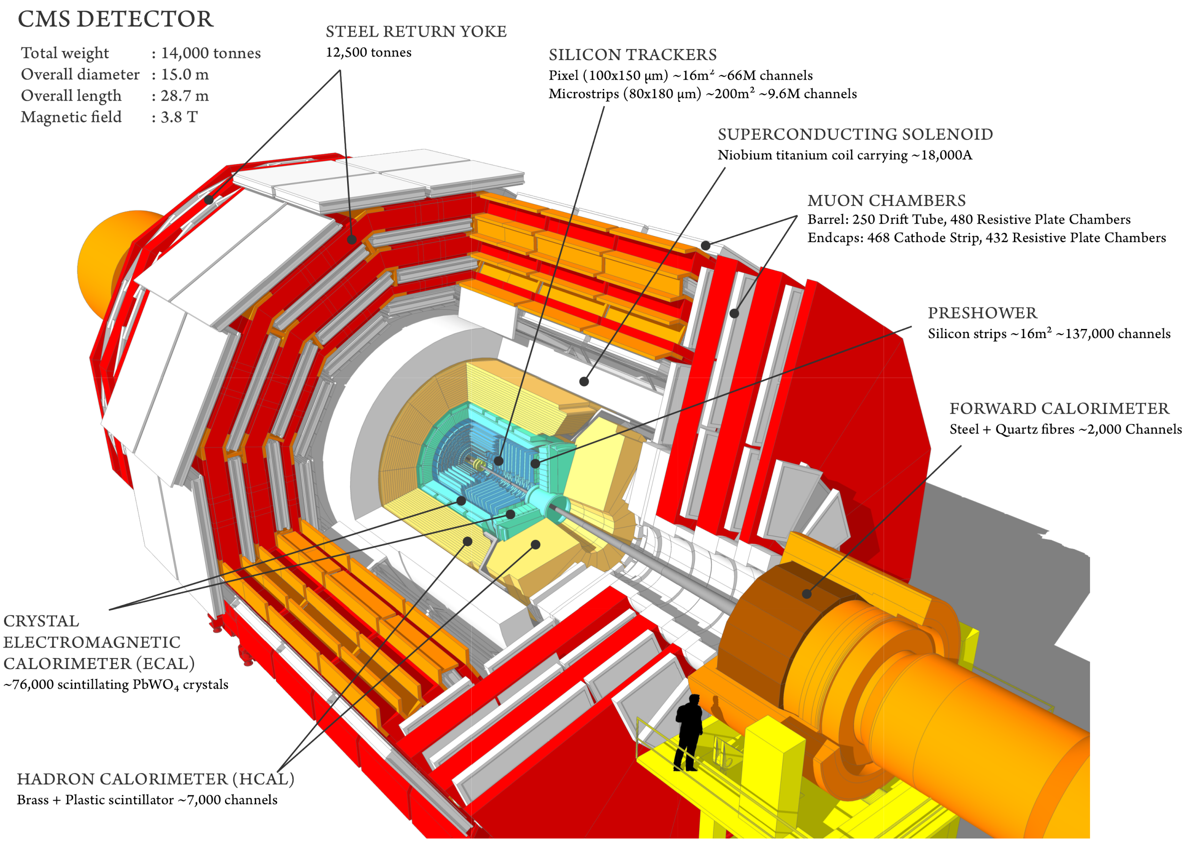


Figure 2.1: Layout of the CMS detector [48].

---

*The LHC is composed of eight arcs and eight straight sections. The straight sections comprise surface and underground installations and are referred to as the *LHC points*. The four main LHC experiments are installed at Point 1 (ATLAS), Point 2 (ALICE), Point 5 (CMS) and Point 8 (LHCb). The other LHC points host various technical infrastructures such as collimators, *radio frequency* (RF) systems or the beam dump.

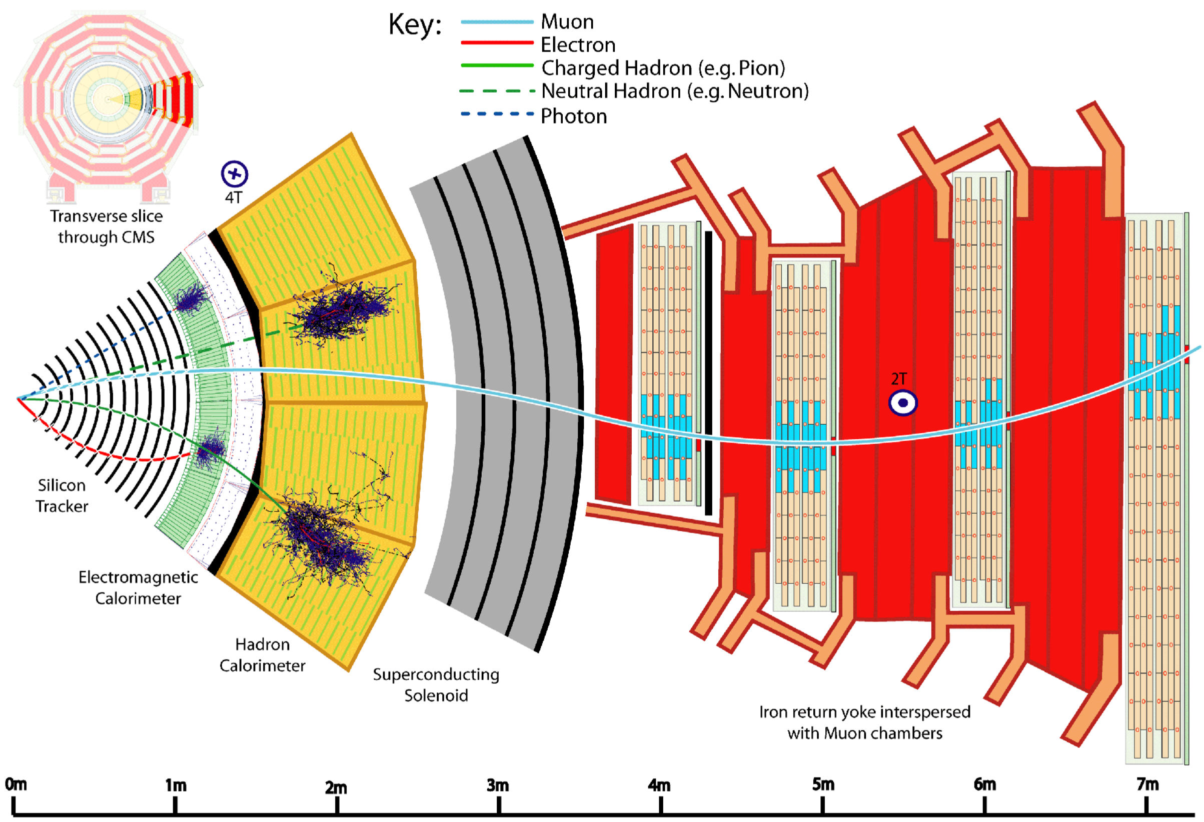


Figure 2.2: Slice of the CMS detector in the transverse plane showing its main subdetectors and the particle interaction with them [49].

The CMS detector is composed of several subsystems that are indicated in Figure 2.1. Similarly to the overall CMS structure, the subdetectors are composed of a barrel part enclosed on both sides by the endcaps. Ordered by increasing radial distance from the beam pipe, the main CMS subsystems are the silicon tracker, the electromagnetic and hadronic calorimeters, the solenoid magnet and the muon detectors. A slice of the CMS detector in the transverse plane is shown in Figure 2.2. Examples of detection of various particles are also represented.

2.1.1 Solenoid magnet

One of the main distinctive features of the CMS detector is a large superconducting solenoid magnet [50]. With its weight of 220 t, length of 12.5 m and a free bore diameter of 6.3 m, it is the largest superconducting magnet of its type ever built. It is formed by a cylindrical coil of superconducting wires made of niobium-titanium alloy. Since it is cooled down to a temperature of 4.7 K using liquid helium, the electric current of 18 500 A can circulate in the coil with almost no resistance, thus producing a strong magnetic field of 3.8 T. This axial magnetic field, mostly uniform within the magnet volume is used to bend the trajectories of charged particles as they fly outwards from the IP. This allows to measure the electric charge of the particles and their transverse momentum. Outside the solenoid, the magnetic field is confined to the detector volume thanks to a 21 m-long and 14 m-wide steel return yoke, inside which the magnet is enclosed. The magnet and the return yoke weight 12 500 t, representing by far the heaviest CMS component.

2.1.2 CMS subdetectors

Silicon tracker

The silicon tracking system [51, 52], located at the heart of CMS, is detecting the trajectories of charged particles. It is entirely equipped with silicon sensors, offering a high granularity, a fast time response and sufficient radiation hardness close to the IP. Charged particles deposit energy in silicon sensors by ionisation and create a signal on collection electrodes. The functioning of a silicon sensor is discussed in details in Chapter 4. The hits detected in several concentric tracking layers and discs allow a precise reconstruction of particle trajectories. Thanks to its high channel density, the silicon tracker also allows the reconstruction of the primary interaction vertices and the secondary decay vertices. The silicon tracker is 5.6 m long, 2.2 m wide and has an acceptance coverage up to $|\eta| < 2.5$. The tracker has a total silicon surface of more than 215 m^2, which makes it the largest silicon device ever built. It is a very light detector to avoid deflecting particles from their trajectories. The silicon tracker is composed of two main parts: the *Inner Tracker* (IT) and *Outer Tracker* (OT).

The Inner Tracker pixel detector is the inner most subdetector of CMS. The original pixel detector has been replaced with an upgraded one, called Phase-1 pixel detector [53, 54], during the LHC winter shutdown in 2016/17. The design of the upgraded CMS pixel detector allows to cope with the higher instantaneous luminosities that have been achieved by the LHC after the first long shutdown of the accelerator. It also has a higher tracking efficiency and lower mass than the original detector. The Phase-1 pixel detector is made of hybrid pixel modules offering a high granularity to cope with the extreme track density close to the IP. Consequently, the IT plays in important role in the vertex reconstruction. The Phase-1 IT is composed of four barrel layers and three endcap discs on each side. The total silicon area of the Phase-1 pixel detector is 1.9 m^2 and it comprises 124 million readout channels with a pixel size of $100 \times 150\,\mu\mathrm{m}^2$. The sensor thickness is 285 µm [54].

The Outer Tracker strip detector is made of silicon strip modules and is divided into the *Tracker Inner Barrel* (TIB) with four layers, surrounded by the *Tracker Outer Barrel* (TOB) with six layers. Three inner endcap discs on each side form the *Tracker Inner Disks* (TID), followed by nine additional discs on each side forming the *Tracker Endcap* (TEC). The silicon strip modules are designed differently for different parts of the OT and exist in 15 different geometries. Each module carries one or two sensors, with thickness of 320 µm or 500 µm. The strip length ranges from 8.5 cm to 20 cm and the distance between strips called the *pitch* varies from 80 µm to 205 µm, keeping a constant ratio between the strip width and the pitch. The total silicon area of the OT is 214 m^2 and it contains 9.3 million channels read out by about 73 000 microelectronic chips.

Calorimeters

The purpose of the calorimeters is to measure the energy of particles. The CMS detector features two different calorimeters, both highly granular and compact enough to fit inside the coil of the solenoid magnet. The *Electromagnetic Calorimeter* (ECAL) [55] is the inner layer of the two measuring the energy of electrons and photons by stopping them completely. The ECAL is a homogeneous calorimeter made lead tungstate ($PbWO_4$) crystals, which serve as both the absorber and the scintillator. High-energy electrons or photons undergo multiple electromagnetic interactions with the heavy nuclei of the scintillating crystals and deposit all their energy in the dense material of the ECAL. The excited atoms of the crystals relax by emitting blue scintillation light, proportional to the amount deposited energy, detected and converted into electrical signal by photodetectors.

The hadrons that fly through the ECAL are stopped and measured in the *Hadronic Calorimeter* (HCAL) [56]. The HCAL is a heterogeneous sampling calorimeter made of alternating layers of passive absorbers and active scintillators. It measures the energy, but also the position and the arrival time of hadrons, which interact in the passive layers and produce many secondary particles causing a hadronic shower. As this shower develops, the particles pass through the layers of scintillation tiles, causing them to emit blue-violet light collected by optical fibres and detected by photodetectors. The total amount of light is a measure of the particle's energy and/or can be an indicator of the particle type.

Muon system

The muons are charged leptons, 200 times heavier than electrons. They are produced in many interesting physics processes as for instance in the Higgs boson decay into four muons, considered as one of the "clearest" Higgs signatures. They are also well-suited for the experiment trigger decision addressed in Section 2.1.4. Hence, as it can be inferred from the experiment's name, detecting muons was one of the main design goals of CMS. Muons can penetrate several metres of iron without interacting and unlike most particles, they are not stopped by the CMS calorimeters. For this reason, the CMS muon system [57] is surrounding the calorimeters and represents the outermost CMS subsystem. Four muon stations sit outside the magnet coil and are interleaved with the iron plates of the magnet return yoke. The hits detected in the multiple layers of each station, combined with tracker information, allow to precisely reconstruct the muon tracks. Thanks to the high magnetic field created by the solenoid magnet, even the trajectories of high-energetic muons are bent and their momenta can be calculated. The muon chambers are composed of three types of gaseous detectors: the *drift tubes* (DTs), the *cathode strip chambers* (CSCs) and the *resistive plate chambers* (RPCs). The many detection layers and their different types make the muon system naturally robust and able to filter out background noise.

2.1.3 Particle identification at CMS

As it can be inferred from the previous sections, different types of particles leave distinctive signatures in the various subdetectors, based on what they can be identified [58]. The way different particles are detected in the CMS subdetectors is summarised in Figure 2.3. This sketch illustrates the CMS cross section in the transverse plane with the IP in the centre. The photons being electrically neutral, they do not leave any signal in the silicon tracker, but they produce an electromagnetic shower in the ECAL, where they are stopped. The electrically charged electrons leave a track in the tracker and produce an electromagnetic shower in the ECAL. The muons are also charged and their passage is detected in the tracker, the ECAL, the HCAL, as well as in the muon chambers, before they leave the detector unstopped. The hadrons generally deposit energy in both the ECAL and HCAL. The neutral hadrons, e.g. neutrons, are not visible for the tracker and are mainly detected in the HCAL, while the charged hadrons, e.g. protons, cause a track in the tracker, start showering in the ECAL and are stopped in the HCAL, where they deposit the most of their energy. And finally, some particles, such as neutrinos, do not interact at all with the detector material and leave CMS without depositing any signal. These particles can be identified from the missing transverse energy.

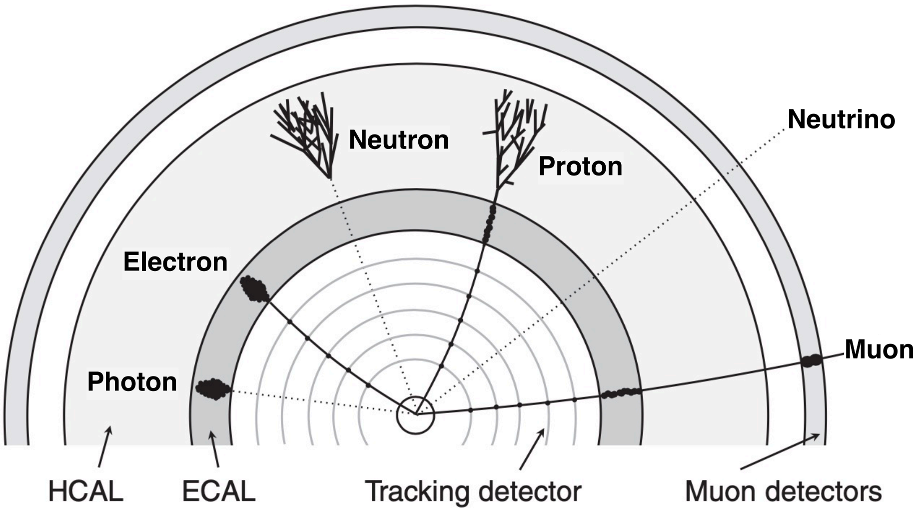


Figure 2.3: Particle signatures in different CMS subdetectors. Adapted from [3, p. 22].

2.1.4 Trigger system

The LHC collides proton bunches at 40 MHz and each BX leads to about 30 p-p collisions, resulting in a very high event rate. The short time between BXs of only 25 ns has major implications for the design of the detector read-out system. The large event rate is far beyond the limitation of the detector electronics readout bandwidth, which can support a readout rate of 100 kHz and beyond the storage capability of $\mathcal{O}$(MB/s). Moreover, most of the events are unlikely to reveal new phenomena and are not interesting for physics studies. Therefore, an online event selection process, called *trigger*, is necessary. The CMS trigger system [59] was designed to examine the detector information in order to select events with a potential physics interest, which reduces the readout rate and the amount of stored data. A large event rejection is complicated to achieve in a single step and therefore, the event selection is performed in two steps. The CMS trigger system consists of a hardware-based *Level-1 trigger* (L1) [60], followed by a software-based *High Level trigger* (HLT) [61]. The trigger thresholds are adjusted to the LHC instantaneous luminosity during data taking to restrict the L1 output rate to 100 kHz imposed by the detector readout electronics. The L1 has a fixed *latency* of 4 µs, i.e. the time needed for the L1 trigger logic to make a decision and to send a trigger to the detector electronics. During the trigger latency interval, all the event information is buffered, before it is either discarded or forwarded to the next stage. Upon a positive decision of the L1 trigger, the full information collected in the event, but not yet fully reconstructed, is forwarded to the HLT. It consists of a processor farm made of commercially available computers running high-level physics algorithms. The HLT further filters the events based on offline-quality reconstruction algorithms. The average event rate at the output of the HLT is 400 Hz. Events which successfully pass the trigger decision are then stored locally, before being transferred to an offline computing centre for permanent storage and distribution.

2.1.5 CMS Phase-2 Upgrade

The CMS collaboration needs to address the ageing of the detector due to the accumulated radiation damage and to improve its capability to meet the challenging requirements imposed by the unprecedented luminosity of the HL-LHC. For these reasons, the CMS detector will be substantially upgraded during LS3, starting in 2025 [41]. This upgrade is referred to as the CMS Phase-2 Upgrade [46]. The main goal of the Phase-2 Upgrade is to maintain or improve the excellent performance of the CMS detector in terms of efficiency, resolution, and background rejection in the harsh environment of the HL-LHC and to fully profit from the increased luminosity.

The increase in radiation levels requires improved radiation hardness, while the larger pileup and associated increase in particle density requires higher detector granularity to reduce occupancy. In addition, the bandwidth has to be increased to accommodate for the higher data rates, and the trigger capability needs to be improved to keep the trigger rate at an acceptable level without compromising physics potential. After the Phase-2 Upgrade, the trigger rates will be increased to 750 kHz at the L1 [62] and 7.5 kHz at the HLT and the L1 trigger latency will increase to 12.5 µs [46]. Given the radiation damage accumulated by LS3, a full replacement of the silicon tracker and the endcap calorimeters will be necessary, while the other subdetectors will be partially refurbished to cope with the expected higher particle and trigger rates, as well as with the increased L1 trigger latency. The details about the Phase-2 Upgrade of the different CMS subdetectors can be found in Appendix B. The focus in this work is put on the Phase-2 Upgrade of the Inner Tracker, which is discussed in details in the following section.

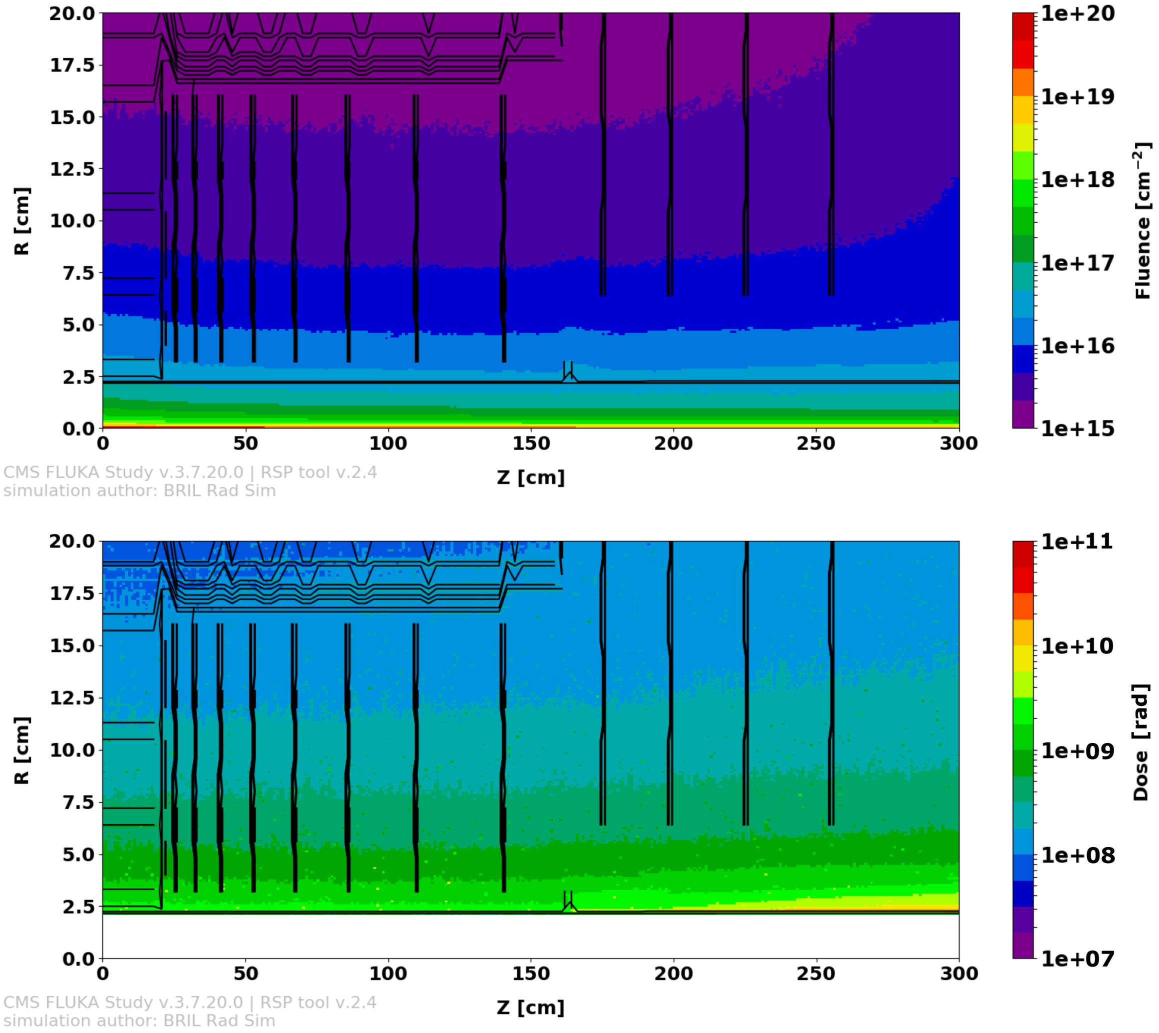
2.2 Inner Tracker Phase-2 Upgrade

The high luminosity upgrade of the LHC and the subsequent high pileup will bring extreme technological challenges to the CMS detector, in particular to the Inner Tracker, which is the subdetector closest to the IP exposed to the highest radiation levels and highest particle density. A detailed overview of the Phase-2 Upgrade of the IT, representing the framework of this thesis, is proposed in this section.

2.2.1 Design requirements

Ten times more radiation is expected from simulations of the HL-LHC environment, compared to the radiation requirements used for the design of the present pixel detector. Two scenarios are envisaged for the HL-LHC: in the “nominal” scenario, the accelerator would deliver a maximum of 140 p-p collisions per BX, to reach a total integrated luminosity of $3000\,\mathrm{fb}^{-1}$ by the end of the physics program. In the “ultimate” scenario, the number of p-p collisions per BX would be pushed up to 200, reaching an integrated luminosity of $4000\,\mathrm{fb}^{-1}$. A fluence of $2.6 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$ and a *total ionizing dose* (TID) up to 1.4 Grad are expected in the innermost layer in the nominal scenario, while the figures would scale up to $3.4 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$ and 1.9 Grad in the ultimate scenario. The FLUKA framework [63, 64] was used to simulate the radiation levels in different parts of the IT at the HL-LHC. The latest simulation maps for the ultimate luminosity scenario, showing the fluence and the TID in one quarter of the IT in the r-z view, are presented Figure 2.4.

Therefore, a sufficient radiation tolerance of the pixel modules is one of the main requirements for the new pixel detector. Almost two times thinner sensors will be used with a thickness reduced from the current 285 µm [54] to 150 µm [65]. The main advantage of thinner sensors is a better radiation tolerance, but the collected signal amplitude is smaller. Hence, the readout electronics, in addition to be radiation hard, must be capable of detecting small signals, in particular from highly irradiated sensors.



CMS FLUKA Study v.3.7.20.0 | RSP tool v.2.4
simulation author: BRIL Rad Sim

CMS FLUKA Study v.3.7.20.0 | RSP tool v.2.4
simulation author: BRIL Rad Sim

Figure 2.4: FLUKA simulation of the fluence (up) and the total ionizing dose (down) in one quarter of the CMS Phase-2 Inner Tracker.

The pileup will increase from 32 p-p collisions per BX reached in 2018 [66] up to 200 in the ultimate luminosity scenario. This factor six increase in pileup will translate to a six times higher hit rate. More precisely, the hit rate in the innermost layer of the IT barrel of the current detector is 580 MHz cm^{-2} [54], while it will reach up to 3.5 GHz cm^{-2} in the Phase-2 IT. The hit information from the pixel modules is sent to the back-end *data acquisition* (DAQ) system of the experiment only after receipt of a Level-1 trigger signal. The trigger rate will increase from 100 to 750 kHz and the trigger latency will increase from the current 4 to 12.5 µs [65]. Therefore, fast readout electronics with bigger buffers and more bandwidth are crucial for the Phase-2 IT.

The high granularity of the IT offers excellent spatial resolution, used for a precise three-dimensional reconstruction of particle trajectories, as well as the identification of primary interaction vertices and secondary decay vertices. At the high pileup of the HL-LHC, a high channel density is necessary to keep the occupancy below a per mille level and to pre-

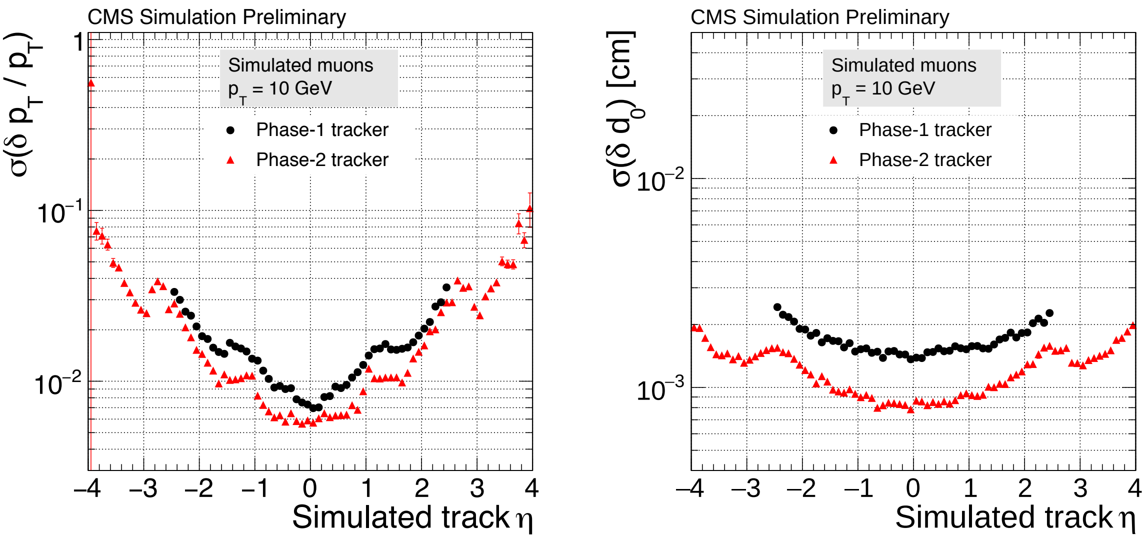


Figure 2.5: Relative transverse momentum resolution (left) and transverse impact parameter resolution (right) as a function of the pseudorapidity for the CMS Phase-1 (black dots) and Phase-2 (red triangles) Inner Tracker, simulated using single isolated muons with a transverse momentum of 10 GeV [65].

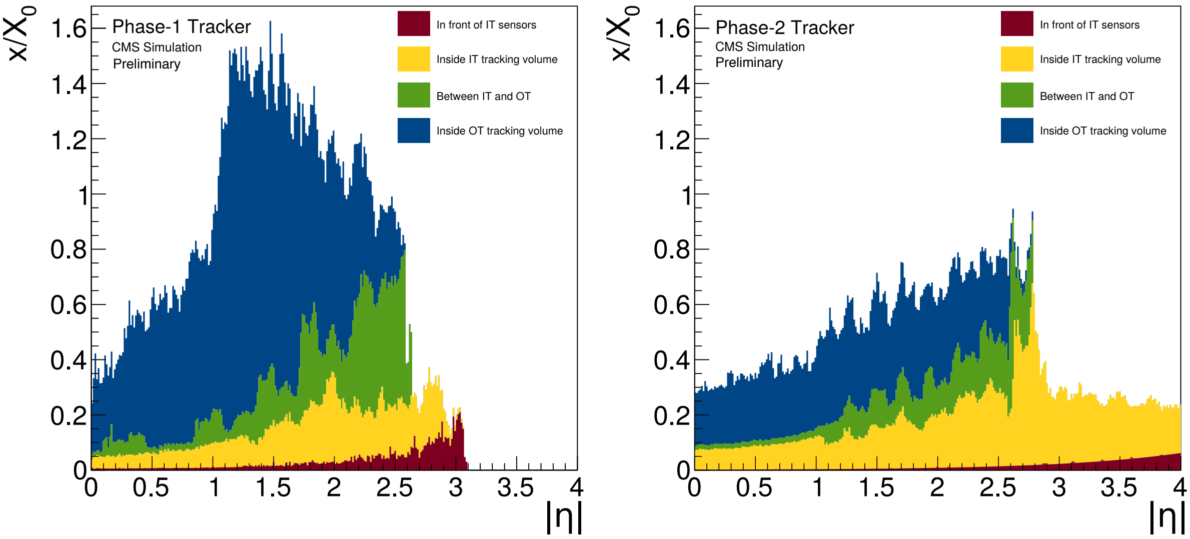


Figure 2.6: Material budget inside the tracking volume estimated in units of radiation lengths, comparing the CMS Phase-1 (left) and Phase-2 (right) Trackers [65].

serve the excellent tracking performance of the detector. Hence, the detector granularity will be increased using six times smaller pixels. The pixel size in the current detector is $100 \times 150\,\mu\text{m}^2$, while in the Phase-2 IT either $50 \times 50\,\mu\text{m}^2$ or $25 \times 100\,\mu\text{m}^2$ pixels will be used. The smaller pixel size will improve the track separation in jets, as well as the momentum and the impact parameter resolution. The improvement in the relative transverse momentum resolution and in the transverse impact parameter resolution of the Phase-2

IT compared to the current detector is shown in Figure 2.5. In addition, as it can be observed from the figure, the pseudorapidity coverage will be extended from $|\eta| < 2.5$ up to $|\eta| = 4$, which will improve the overall CMS physics performance by increasing the b-tagging efficiency of forward jets.

The detector design also strives for a minimal mass to avoid degradation of the tracking performance due to the interactions of particles with the detector material. Therefore, lightweight mechanical structures made of carbon fibre and two-phase CO_2 cooling will be used [65]. The high performance of the IT cooling is essential not only for the longevity of the sensors, but also for the radiation hardness of the readout chip. The cooling pipes will be shared among neighbouring modules targeting to keep the sensors below $-20\,^{\circ}\mathrm{C}$. An innovative powering scheme will be deployed to reduce the amount of power cables and the pixel data will be transmitted through low-mass electrical links and optical fibres [65] to further reduce the detector mass. The comparison of the material budget of the current CMS tracker with the upgraded one is presented in Figure 2.6. The material budget will be greatly improved in the OT and maintained similar in the IT. The additional material in the pseudorapidity region of $2.5 < |\eta| = 4$ confirms the detection coverage extension of the IT. Moreover, a simple installation and removal of the IT detector without removing the CMS beam pipe must be guaranteed to allow for potential replacement of degraded parts over an LHC extended year-end technical stop. To meet these stringent requirements the CMS IT will be replaced with a totally new pixel detector. A 3D picture of one quarter of the new detector is shown in Figure 2.7.

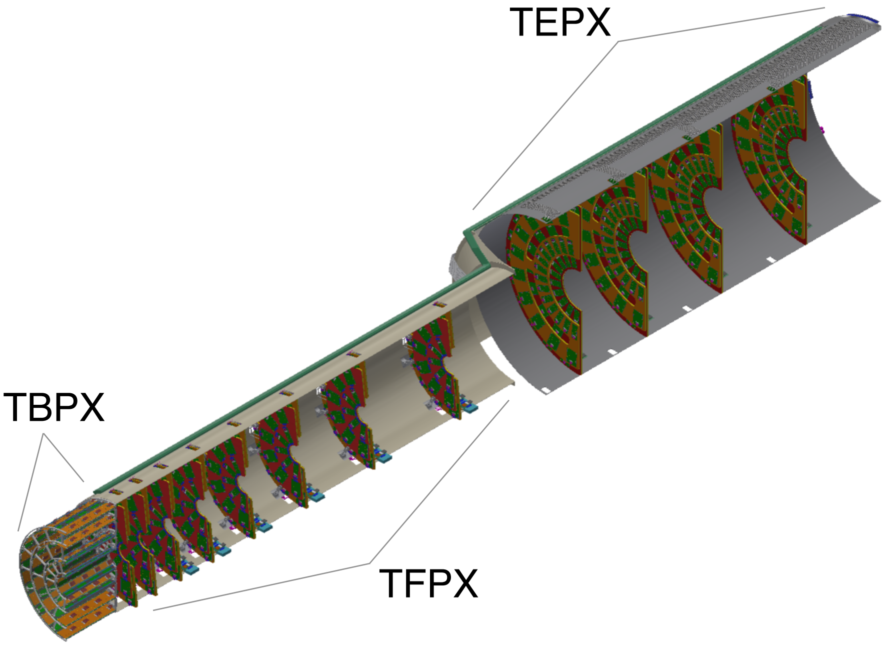


Figure 2.7: 3D picture of one quarter of the CMS Phase-2 Inner Tracker [65].

2.2.2 Hybrid pixel modules

The main building block of the IT system is a *hybrid pixel module*. It is composed of a pixel sensor, several pixel readout chips and a *high density interconnect* (HDI) [67]. A pixel sensor is a silicon diode in which one electrode is highly segmented in two directions. It is the active part of the system where particles are detected. Bias voltage is applied to enhance the electric field and to deplete the active volume of the sensor. When a charged particle passes through the depleted volume, it creates electron/hole pairs by ionization. The electric field allows to induce a signal pulse on one segment of the sensor electrode, i.e. on a *pixel*. Each pixel of the sensor is connected to a corresponding pixel of the readout chip with a metallic solder bump called a *bump bond* [68, 69], as it is shown in Figure 2.8. The sensor pixel area has to match with the readout chip pixel area to allow for a one-to-one connection. The pixel signal is amplified and digitized by the pixel *front-end* (FE) electronics and the data is stored in on-chip buffers during the L1 trigger latency. The hit information is sent out of the readout chip only after receipt of the L1 trigger signal, while the other hits are discarded.

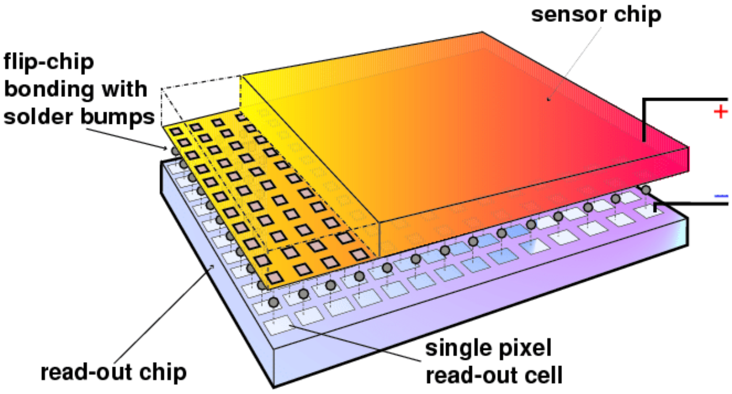


Figure 2.8: Schematic of a hybrid pixel detector with the sensor and the readout chip connected via bump bonds [70].

The assembly of the pixel sensor bump-bonded to the readout chip is wire-bonded to the HDI. This flexible printed circuit with passive components and connectors is used to distribute power (bias voltage for the sensor and supply voltage for the readout chip), clock and control signals. The HDI is also the first stage of the data transmission from the readout chip towards the DAQ system. An exploded view of a hybrid pixel module for the Phase-2 upgrade of the IT is presented in Figure 2.9. The module components visible from top to bottom of the figure are: the HDI with a power and a data readout connector, the pixel sensor, four pixel readout chips and the module rails allowing to mount the module on a support structure and ensuring the electrical isolation of the module.

The word *hybrid* qualifying the IT pixel modules is used to indicate that the sensor and the readout electronics are implemented in two separate silicon dies connected together. The electronics can also be implemented directly inside the silicon sensor to form a *monolithic* *active pixel sensor* (MAPS) [71, 72]. The MAPS offer a better noise performance than the hybrid detectors, avoiding the costly and complex hybridization process and consequently allowing to reduce the detector mass. They represent a more robust, lighter and less expensive technology promising for future particle physics experiments. However, given the extreme radiation levels and particle rates expected close to the IP at the HL-LHC, the more mature hybrid technology offering a better radiation tolerance and more complex logic will be maintained in the CMS IT also after LS3.

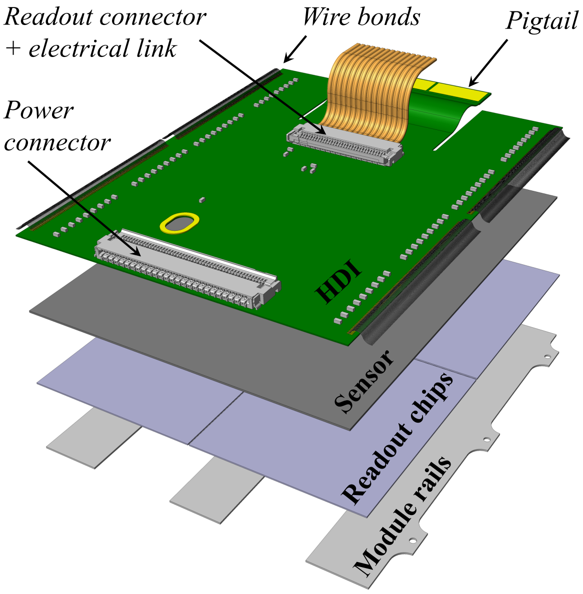


Figure 2.9: 3D exploded view of a CMS Phase-2 Inner Tracker hybrid pixel module.

2.2.3 Detector layout

The arrangement of the hybrid pixel modules to form the new pixel detector is shown in the latest layout of one quarter of the Phase-2 IT in the r-z view presented in Figure 2.10. The grey line represents the beam pipe envelope and the (0,0) coordinate in this figure corresponds to the IP. The green lines represent pixel modules with two readout chips called *double-chip modules* and the orange lines represent pixels modules with four readout chips called *quad-chip modules*. One end of the new IT will be about 2.6 m long, which makes the total length of the detector larger than 5 m. The maximum radius will be 25 cm, resulting in a total detector diameter of about 0.5 m. The Phase-2 IT will be composed of three subsystems: *Tracker Barrel Pixel detector* (TBPX), *Tracker Forward Pixel detector* (TFPX) and *Tracker Endcap Pixel detector* (TEPX).

TBPX. The IT barrel will be about 40 cm long and will be composed of four cylindrical layers. The layers are counted one to four from the interaction point outwards. The layer 1 will be placed very close to the beam pipe, only 3 cm from the interaction point, as it can

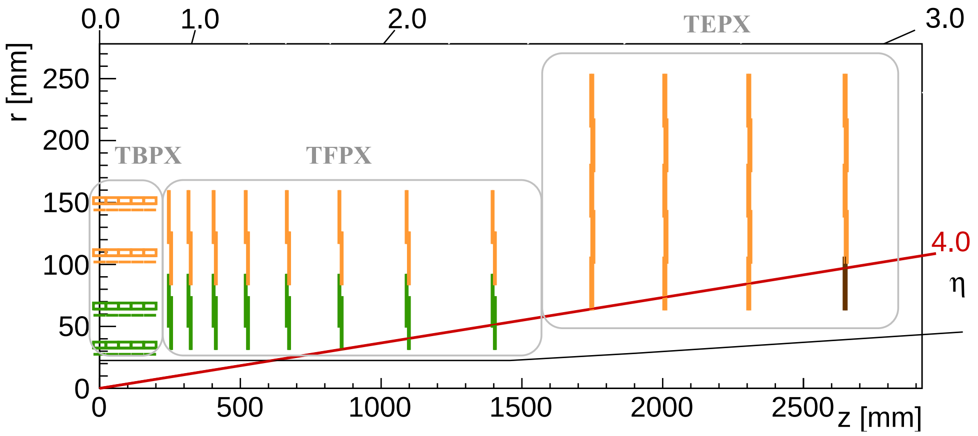
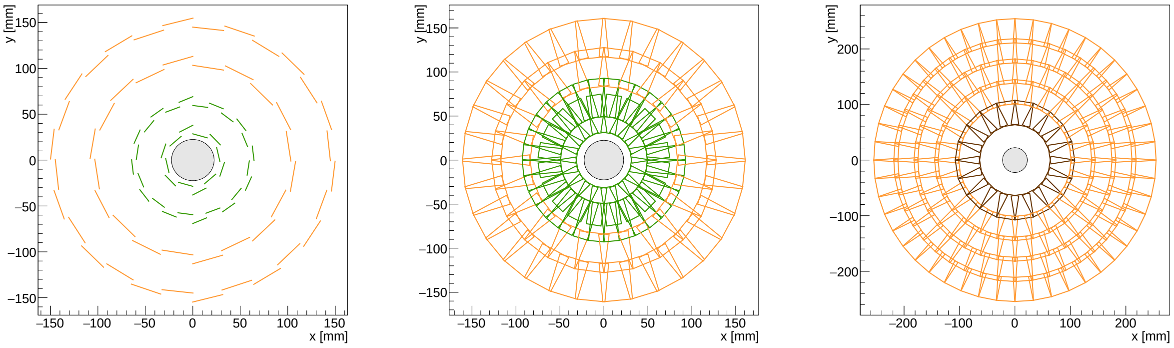


Figure 2.10: Layout of one quarter of the CMS Phase-2 Inner Tracker in the r-z view.

be seen in Figure 2.11(a). The outer radius of the TBPX will be about 15 cm. The two inner layers will be made of double-chip modules, shown in green and the two outer layers of quad-chip modules, shown in orange. The pixel modules in the TBPX will be arranged in so-called *ladders*. The neighbouring ladders in each layer will be mounted staggered in radius, as it is shown in Figure 2.11(a). The ladder overlap will ensure the full detection coverage in r-ϕ. The modules on a ladder will not overlap in z and a coverage gap at $\eta = 0$ will be avoided by using an odd number of modules per ladder. This makes the two halves of the TBPX slightly asymmetric. One half will be made of four and the other half of five pixel modules.

TFPX. Eight smaller double-discs forming the TFPX will be placed in the forward direction on each side of the barrel between $z = 25$ cm and $z = 140$ cm. The double-discs with an outer radius of 16 cm will be made of two discs further split into two halves referred to as “dees” because of their D-shape. The pixel modules of each TFPX double disc will be arranged in four concentric rings with overlaps in r and r-ϕ, as it is shown in Figure 2.11(b). The two inner rings will be made of double-chip modules depicted in green and the two outer rings will be made of quad-chip modules depicted in orange.

TEPX. Four larger double-discs placed between $z = 175$ cm and $z = 265$ cm on each side of the TFPX will form the TEPX. Each double-disc will have an inner radius of about 6 cm and an outer radius of about 25 cm and will be composed of five module rings, as it is shown in Figure 2.11(c). All five rings will be made of quad-chip modules shown in orange. The innermost ring of the last TEPX disc (D4R1) represented in brown in Figures 2.10 and 2.11(c) will not be a part of the tracking coverage of the detector because of an insufficient hit redundancy at this large pseudorapidity. Instead this ring will be used by the *Beam Radiation Instrumentation and Luminosity* (BRIL) [73] project for dedicated beam luminosity and background measurements. On top of the D4R1, the entire TEPX will receive extra triggers contributing to the BRIL luminosity monitoring.



(a) TBPX. (b) TFPX. (c) TEPX.

Figure 2.11: Layout of the Phase-2 Inner Tracker in the r-ϕ view showing the cross section of the Tracker Barrel Pixel detector (a), Tracker Forward Pixel detector (b) and Tracker Endcap Pixel detector (c).

The power, cooling, and data transmission services will be carried on a cylindrical shell enclosing the pixel detector called the *service cylinder*, visible in Figure 2.7, which will be independent for each quarter of the detector. This detector layout will allow to extend the forward acceptance up to a pseudorapidity of $|\eta| = 4$ [65], as indicated by the red line in Figure 2.10. The total active area of the Phase-2 IT will be $4.9\,\mathrm{m}^2$ and it will be composed of 3892 pixel modules. Two pixel sizes are considered for the Phase-2 IT: $100 \times 25\,\mu\mathrm{m}^2$ pixels and $50 \times 50\,\mu\mathrm{m}^2$ pixels. With a pixel area of $2500\,\mu\mathrm{m}^2$ the total number of readout channels will amount to about 2 billion.

2.2.4 Electronics system

The key feature of the new IT electronics system is the *serial powering* (SP) [74, 75] of the pixel modules. The 3892 IT modules will be made of 13256 pixel readout chips. With a supply voltage of 1.4–1.5 V per chip and a supply current of about 2 A, the total power consumption of the detector will be above 50 kW [76]. More than six tons of power cables would be necessary to deliver this amount of power to the pixel modules and still 2/3 of the delivered power would be lost in the cables [67]. The use of on-module DC–DC power conversion has been excluded due to insufficient radiation hardness and material budget constraints. Therefore, the only viable solution to power the Phase-2 IT, without significantly increasing the detector mass and disturbing particle trajectories with too many power cables, is to power the modules in series. The ATLAS and CMS pixel detectors for the HL-LHC represent the first large scale application of serial powering in a detector system.

In the new CMS IT, the pixel modules will be grouped into SP chains. The HDI pigtail of one module, visible in Figure 2.9, is plugged into the power connector of the following module and this way an SP chain is formed. An SP chain of three quad modules is shown in Figure 2.12 on the left. The entire system will be composed of 500 SP chains with a minimum of five and a maximum of twelve pixel modules in series. An SP chain in the TBPX corresponds to two neighbouring ladders, i.e. eight or ten modules per chain. The TFPX and TEPX modules will be grouped into SP chains per ring. Each side of the disc (front/back) will have independent chains. Therefore, the length of a chain equals to the number of the modules of a ring mounted on one side of the disc.

Within an SP chain the modules are powered in series, while the two or four readout chips of the same module will be powered in parallel, which means that the current of a module will be shared among them. A constant current sufficient to satisfy the highest load of the SP chain, plus 20% extra current to ensure a stable operation, will be provided by the power supplies placed at the back-end, as it is schematically shown in Figure 2.12. The supply current is propagated from one module to the other thanks to dedicated on-chip *Shunt low-dropout* (ShLDO) regulators [75, 77]. The bias voltage for the pixel sensors will be provided per SP chain by the power supplies located at the back-end, as indicated in Figure 2.12. No crossing of the bias voltage lines from one SP chain to another is allowed in the system. The bias of the sensors will be applied in parallel to the modules along the SP chain.

The power supplies will also provide power to another important component of the IT electronics system: the optoboards called *portcards*, taking care of the data merging and electrical-to-optical conversion of the signals. A portcard is a *printed circuit board* (PCB) hosting three *low power gigabit transceivers* (lpGBTs) [78] and three VTRx+ transceivers [79]. These portcard components will be powered independently by a two stage DCDC

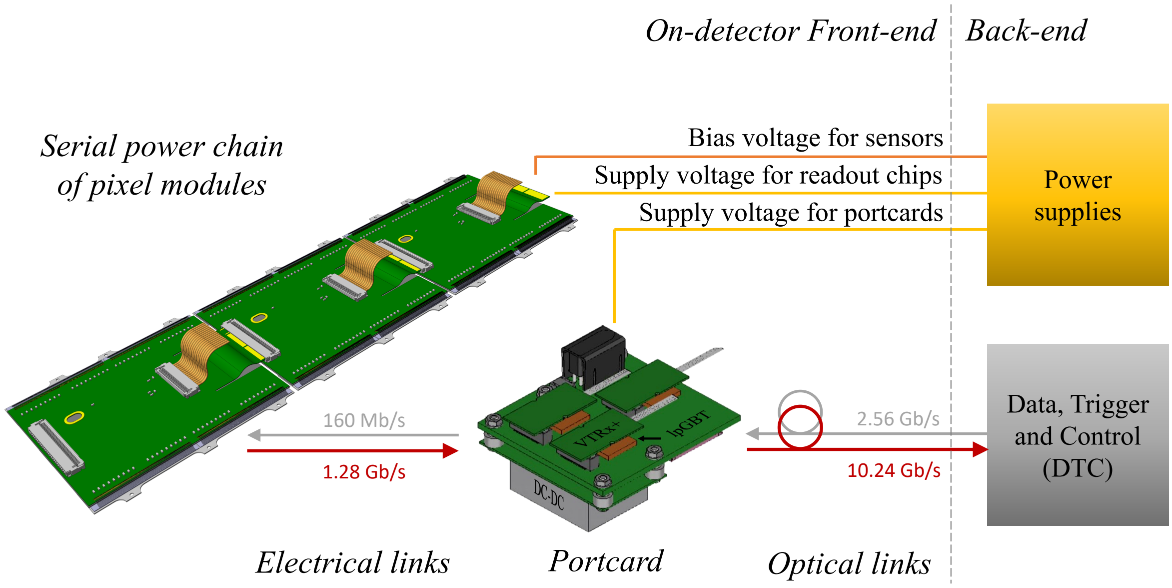


Figure 2.12: CMS Phase-2 Inner Tracker electronics system architecture.

converter scheme, based on the bPOL12V [80] and bPOL2V5 [81] DCDC converters placed on a DCDC mezzanine on each portcard. The portcards will be dedicated per SP chain to avoid mixing of the readout. The portcards reading out the TBPX modules will be placed on the service cylinder surrounding the TFPX, between discs number six and seven. The portcards reading out the TFPX and TEPX modules will be placed on a cartridge on the outer radius of each disc. This placement was chosen to protect the optoelectronics and the DCDC converters from the high-radiation levels.

The first part of the IT readout system will be based on electrical readout, followed by a second part based on the optical readout. The pixel modules will be connected to the portcards via low-mass electrical links and the portcards will be connected to the back-end DAQ via optical fibres. The optical fibres are flexible and can be bundled as cables, which makes them advantageous for long-distance communications, because light propagates through them with little attenuation compared to electrical copper cables. The command links will be sent from the back-end at 2.56 Gb/s and will be transmitted by the lpGBT to the pixel modules at 160 Mb/s encoded with a custom-made protocol. One control link per module will be used to transmit clock, trigger, commands and configuration data to the chips.

The hit data from events accepted by the L1 trigger as well as the monitoring information will be encoded with the Aurora 64/66b protocol [82] and transmitted at 1.28 Gb/s from the pixel modules through the electrical links to the portcards. An efficient data formatting called *binary tree encoding* will be applied by the readout chips to reduce the data rates by about a factor two and a 25% bandwidth headroom will be reserved on the links. The modularity of the data links depends on the hit rate, i.e. on the location of the modules in the detector. The maximum number of data links per module will be needed for the innermost layer of the barrel with six data links per module. The minimum number of data links will be used in the outer layers and rings, where one data link will be shared among the readout chips of one module, thanks to an on-chip data-merging functionality [76].

After data merging and electrical-to-optical conversion, the data will be transmitted at 10.24 Gb/s from the portcards through optical fibres to the back-end DAQ. At the back-end, 28 *data, trigger and control* (DTC) boards will be needed to readout the entire IT system. These boards will be based on the custom Apollo platform [83] in *advanced telecom computer architecture* (ATCA) format based on commercial *field programmable gate arrays* (FPGAs) and multi-channel optoelectronic transceivers. Each DTC can receive up to 36 optical fibres [84] and will read out data from 48–704 readout chips.

The scope of this work is the development and validation of the first components of the Phase-2 IT readout system, i.e. the pixel modules and electrical links, contributing to some of the detector design choices.

Chapter 3

Analogue front-end selection

One of the crucial components of the Inner Tracker system is the readout chip and in particular its analogue front-end that receives the particle signal from the sensor and processes it. Three different analogue front-ends, called Synchronous, Linear and Differential, were designed by the RD53 Collaboration and implemented in a large scale demonstrator chip, called RD53A, offering a choice to the experiments. The selection of one of these three analogue front-ends for integration in the final CMS pixel chip represents an important system choice.

A dedicated evaluation campaign was carried out to select the most suitable design to build a radiation tolerant pixel detector able to sustain high particle rates with high efficiency and a small fraction of spurious hits. A systematic test procedure and a complete measurement program were devised to qualify the three front-ends, identify potential issues and give feedback to designers to allow for design improvements. The key performance parameters of the future pixel detector were identified and used to define the requirements against which the three designs were evaluated.

This chapter starts with the presentation of the RD53A chip in Section 3.1. Then, the analogue front-end is defined and explained in Section 3.2, followed by an overview of the functionalities of the three proposed designs. The CMS requirements, against which the three options were evaluated, are presented in Section 3.3 and the systematic test and calibration procedure, applied to all measurements, is explained in Section 3.4. Finally, the results of the measurement program, in terms of threshold, radiation tolerance, noise, dead time and time response, including the key measurements that allowed CMS to make the choice, are presented in Sections 3.5 to 3.8. The chapter closes with the discussion of the results and the conclusion on the analogue front-end choice.

3.1 RD53A readout chip

A high-performance radiation tolerant pixel readout chip is essential for good tracking performance of the IT operating at HL-LHC. Such a readout chip is being designed in TSMC [85] 65 nm CMOS technology by the RD53 Collaboration [86], a joint effort between the ATLAS and CMS experiments. A large-scale demonstrator chip called *RD53A* [87] was produced to prove the suitability of the chosen technology for low threshold, low noise, and low power operation at high hit rates and to verify sufficient radiation tolerance [88].

The RD53A is a mixed signal chip, having both analogue and digital circuits. The charge deposited in the sensor creates an analogue pulse in a pixel that indicates the hit position as well as the amount of collected charge using the *time-over-threshold* (TOT) technique. The signal amplitude is proportional to the time, expressed in clock cycles, during which the signal is above the threshold, as illustrated in Figure 3.1. In this example, the TOT of the signal corresponds to 3 clock cycles. In the RD53A chip, the TOT_{40} digitisation with 4-bit resolution is done with respect to rising edges of the 40 MHz LHC clock. Therefore, one TOT_{40} unit corresponds to 25 ns [87]. In the final pixel chip, the counting will be performed on both the rising and the falling edge of the clock, resulting in a finer TOT_{80} counting at 80 MHz with one TOT_{80} unit equal to 12.5 ns [89].

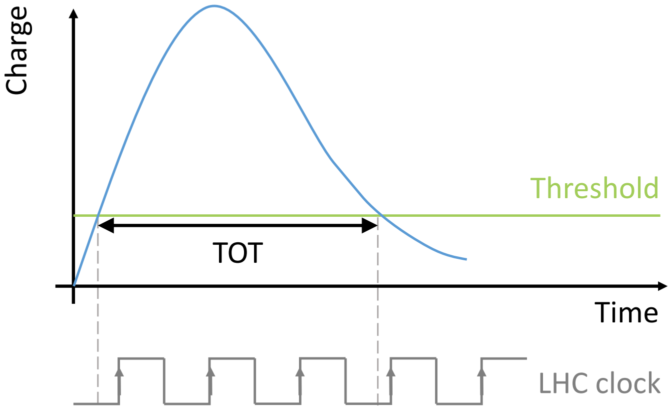


Figure 3.1: Illustration of the time-over-threshold digitisation technique.

The RD53A chip size is $20.0 \times 11.8\,\mathrm{mm}^2$, which is about half the size of the final chip, as it was sharing the chip reticle with CMS OT chips. The pixel matrix is composed of 400×192 square pixels with 50 µm pitch. A functional floor plan of the RD53A is shown in Figure 3.2. The chip is composed of three main parts: the top pad, the pixel matrix and the bottom chip periphery. A row of test pads is placed at the top edge of the chip for debugging purposes. This feature will be removed in the final chip. The bottom chip periphery gathers all the common analogue and digital circuitry needed to bias, configure, monitor, and read out the chip. It contains custom-designed *intellectual property* (IP) blocks important for the chip operation.

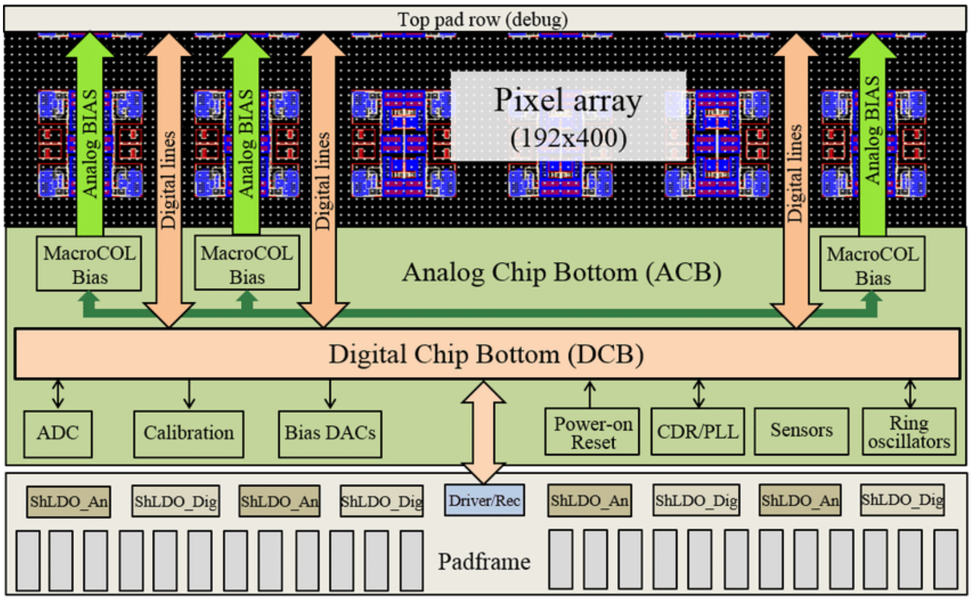


Figure 3.2: Functional floor plan of the RD53A chip (not to scale) [87].

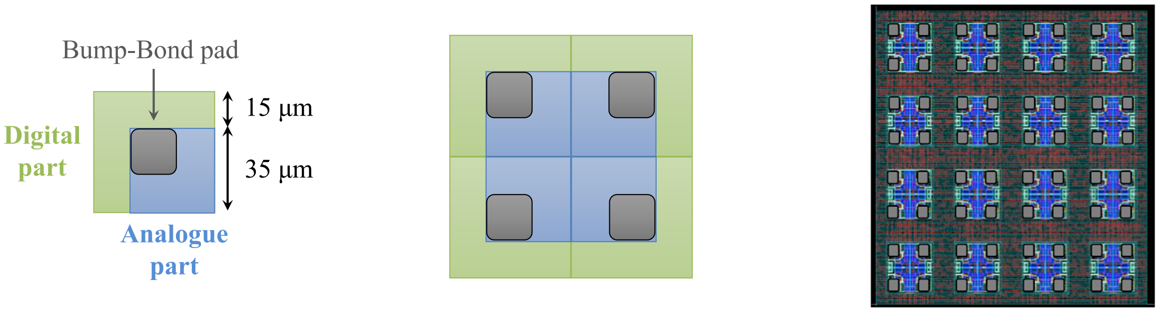
The analogue building blocks in the periphery are grouped in a macro-block called *analogue chip bottom* (ACB). The ACB provides the current reference to the pixel bias *digital-to-analogue converters* (DACs) and voltage reference for the calibration circuit addressed in Section 3.4.2. The analogue bias is distributed column-wise to the pixels, as indicated by the arrows in Figure 3.2. The temperature and radiation sensors enable the chip monitoring and an *analogue-to-digital converter* (ADC) takes care of the digitisation of the monitoring signals. The *power on reset* (POR) block takes care of the chip initialisation. It ensures that the chip has a correct configuration immediately after the start up and that the stored logic states are well defined. The *clock data recovery* (CDR) block [90] containing an internal *voltage controlled oscillator* (VCO) and a *phase locked loop* (PLL) allows to lock to the incoming command stream running at 160 Mb/s and to recover the clock from it. The VCO produces a primary 1.28 GHz clock locked to the phase of the command input stream transitions and this clock is divided to generate all clocks needed by the chip, including the 40 MHz clock distributed to the pixels.

A special voltage regulator, called ShLDO [77] is implemented for the serial powering of the pixel modules. This combination of a shunt and *low-dropout* (LDO) regulator, allows to generate a regulated voltage for the chip from a constant input current supply. The chip features two ShLDOs, one for the analogue and one for the digital power, each distributed over four active pads. The use of multiple pads in parallel, spread along the chip bottom avoids a power dissipation in a single spot in the chip.

The ACB is surrounded by a synthesised macro-block, called *digital chip bottom* (DCB), which implements the input, output and configuration digital logic. The main feature of the DCB is a high-speed serial output transmitter with a *current mode logic* (CML) cable driver [91]. All data, status and configuration read-backs are encoded with the Aurora

64 b/66 b protocol [82, 92] and send out of the chip at 1.28 Gb/s on up to four output data lanes. A row of wire-bond pads for the chip connection to the HDI and the rest of the system is located at the bottom chip edge.

The biggest part of the chip is occupied by the matrix of 76 800 pixels. A pixel is composed of 50% of analogue and 50% of digital circuits, arranged as shown in Figure 3.3 (left). In the centre of each pixel is a pad for bump-bond connection to the sensor. The analogue parts of the pixels are grouped by four into so-called *analogue islands*, which are embedded in a fully synthesised *digital sea*, as indicated in Figure 3.3 (middle). The analogue islands are further organised in *pixel cores*. A pixel core, shown in Figure 3.3 (right), is composed of 8 × 8 pixels or 16 analogue islands. It is a basic building block of the pixel matrix that can be simulated as a standalone circuit at transistor level.



Pixel *Analog island* *Pixel core*

Figure 3.3: Organisation of the pixel matrix in the RD53A chip.

The RD53A is a demonstrator chip, not intended for use in the experiments, since it contains design variations for testing purposes (e.g. the top pad) and to offer to the experiments a choice between two digital and three analogue architectures. The two proposed digital flavors are called the *central buffer architecture* (CBA) [93], implemented in the first 128 columns, on the left side of the chip, and the *distributed buffer architecture* (DBA) [87] in the remaining 272 columns. The basic read-out unit in both cases is a *pixel region*. The CBA implements pixel regions of 16 pixels (4 × 4) and the TOT information is not stored per pixel but per pixel region in a shared memory. Only TOT values of the hit pixels are stored to save the memory, but at the cost of having to store a hit map to assign the TOTs to the corresponding hit positions within the region. On the other hand, in the DBA, the pixel regions are made of four pixels only (4 × 1) and the TOTs are stored per pixel, assigning a TOT = 15 to pixels with no hit. Consequently, the CBA is efficient as long as the hit occupancy is low, while the DBA is more efficient at high hit occupancies. After testing and verification of both options, not in the scope of this work, the DBA architecture was selected for the final pixel chip.

3.2 Analogue front-ends in the RD53A chip

In this work, the focus is put on the analogue part of the pixel called the *analogue front-end* (AFE). The RD53A chip implements three different AFEs that will be presented in this section after the introduction of a generic AFE. The comparison of the three AFE flavours and the selection of one of them for implementation in the final CMS readout chip is the object of this chapter.

3.2.1 Generic analogue front-end

The AFE is a crucial part of the pixel, which receives the signal from the sensor and takes care of amplifying, shaping and digitizing it. The steps of the signal processing in different parts of a generic AFE are depicted in Figure 3.4, from left to right.

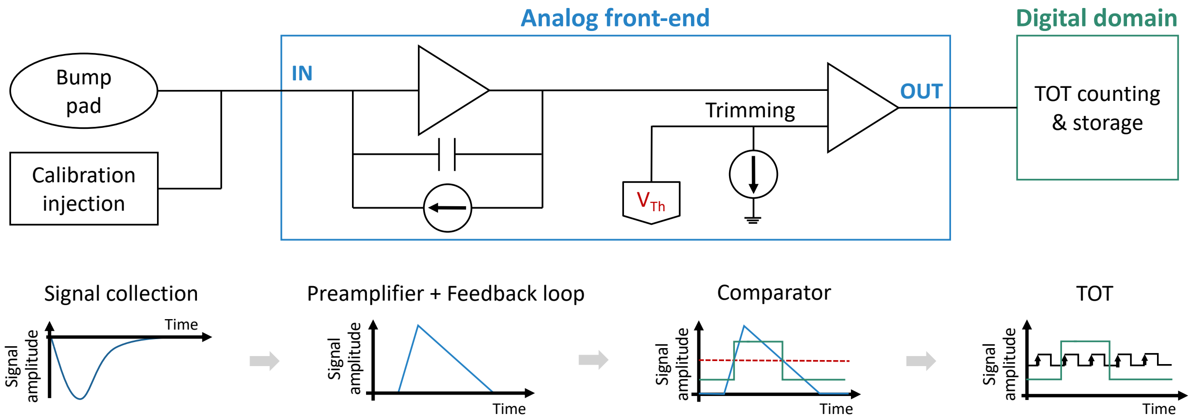


Figure 3.4: Signal processing steps in different parts of a generic analogue front-end, from signal collection to digitisation.

The signal originating from the sensor (negative in our case since collecting electrons) or from a charge injection is received by the *preamplifier* (PA), where it is amplified to obtain a better *signal-to-noise ratio* (SNR). The signal reception can be realised in three main ways: in current mode to preserve the pulse shape, in voltage mode to obtain a high signal amplitude or with a *charge sensitive amplifier* (CSA), which is the most widely used in the FE electronics [94, p. 29]. The CSA [94, Sec. 3.2] is an electronic current integrator that produces a voltage output with an amplitude proportional to the integrated input charge. The output voltage is inversely proportional to the value of the feedback capacitor that defines the gain of the PA. The value of the feedback capacitor C_f is chosen as a compromise, between a high gain (small C_f) and a large effective input capacitance (larger C_f). The typical values of feedback capacitors are of $\mathcal{O}(1\text{–}10\,\text{fF})$ [68, p. 137]. The feedback capacitor must be discharged after a hit by an appropriate feedback loop, to get ready for the next signal, which results in a triangle-shaped signal at the output of the PA. The feedback loop also takes care of the compensation of the leakage current from the sensor (explained in Appendix C). Some AFEs implement a bandwidth-limiting filter after the PA, to shape the signal, add more gain and to improve the SNR, as well as to shorten the pulse duration to reduce signal pileup [94, Sec. 3.3].

A discriminator [94, Sec. 3.6] is used to convert the analogue signal to a digital hit. The triangular signal from the PA is compared to a selected detection threshold. This results in a square signal at the output of the discriminator, which is fed to the digital part of the pixel, where the TOT counting is performed and can be used as a measure of the deposited charge. The discriminator threshold is set globally for all pixels. Yet, pixel-to-pixel variations of the threshold exist because of the transistor mismatch, voltage drops or PA gain variations. This causes an increased noise hit rate or a reduced sensitivity in different pixels. Therefore, it is common to include in every pixel a trimming circuit, which allows to tune the local threshold as close as possible to the target global threshold. In general, a simple threshold tuning DAC is used and the threshold is adjusted with trim bits stored in dedicated pixel registers.

3.2.2 Analogue front-end parameters

Several parameters in an AFE can be set by the user to achieve an optimal performance. While some parameters can be very specific to a particular design, four main parameters are usually present in a generic AFE and their role is illustrated in Figure 3.5. Two of them drive currents in the PA and define the shape of the signal at the PA output. The PA bias current determines the signal rise time. A higher PA bias results in a faster rise of the signal and also in a lower noise. In general, increasing the PA bias is beneficial for the performance, however, the PA bias represents the main contribution to the power consumption of the AFE. Hence, a good compromise between performance and power has to be found. The PA discharge current takes care of discharging the feedback capacitor after a hit, i.e. it drives the falling edge of the signal. Increasing the discharge current causes a faster PA output return to the baseline. Consequently, a different TOT is obtained for the same signal depending on the discharge current and this parameter is used

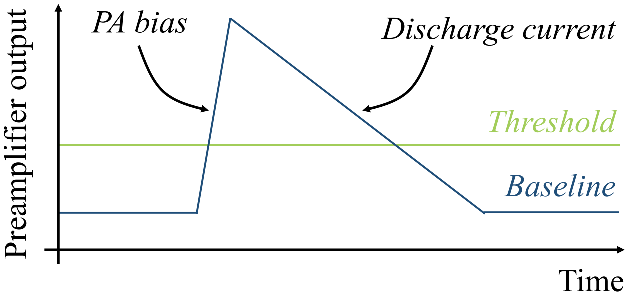


Figure 3.5: Role of the four main parameters in a generic analogue front-end.

for the TOT calibration. The other two parameters drive voltages and together define the detection threshold. The effective threshold is the difference between the threshold voltage, set in the discriminator, and the baseline voltage, setting the PA output DC baseline, i.e. the baseline of the signal. In practice, the signal baseline is kept fixed and only the threshold voltage is adjusted.

3.2.3 RD53A analogue front-ends

The RD53A chip is divided horizontally into three sections, each one having one AFE design, as indicated in Figure 3.6. The *synchronous* (SYNC) AFE is implemented between columns 0 and 127, the *linear* (LIN) AFE between columns 128 and 263, and the *differential* (DIFF) AFE between columns 264 and 399 [87]. It was not possible to have an equal area for all three designs, because the 400-pixels-wide matrix is built of 8 × 8 pixel cores, as it was explained in Section 3.1. The three AFEs share the digital logic and the chip periphery in the RD53A chip [87]. All three AFEs are based on a CSA with a feedback loop ensuring the return to baseline of the PA output pulse after each hit. The gain of the preamplifier can be chosen globally thanks to different feedback capacitors present in each AFE. The specific features of each AFE are discussed in the following paragraphs.

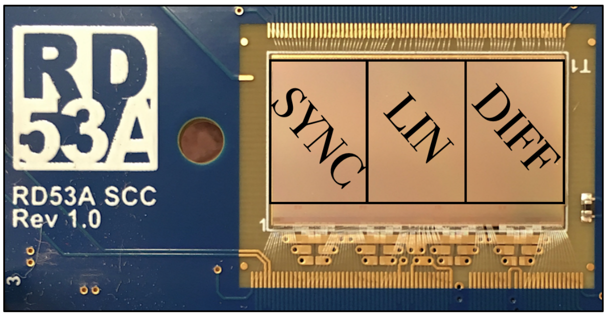


Figure 3.6: Photograph of the RD53A chip, wire-bonded to a test card, indicating the placement of the three analogue front-ends.

Synchronous front-end. The schematic of the Synchronous front-end is shown in Figure 3.7. It features a single-stage CSA with a Krummenacher feedback (I_{Krum}, $V_{\mathtt{REF\_Krum}}$) [95], which ensures both the sensor leakage current compensation and the constant current discharge of the feedback capacitor. The Krummenacher current (I_{Krum}) drives the speed of the PA signal return to the baseline. The PA is AC-coupled (C_{AC}) to a synchronous discriminator composed of a differential amplifier, providing a further small gain, and a positive feedback latch, which performs the signal comparison with threshold (V_{th}) and generates the discriminator output. The latter can also be switched to a local oscillator with selectable frequency higher than the standard LHC clock, in order to perform a fast TOT counting. The distinctive feature of this AFE is a so-called "auto-zero" functionality. In traditional designs, the transistor mismatch causing pixel-to-pixel variations of the threshold is compensated with a trimming DAC. In the SYNC AFE instead, internal capacitors (C_{az}) are used to compensate voltage offsets automatically. A periodic acquisition of a baseline (V_{BL}) is required, which can be done during LHC abort gaps [87, 96].

Linear front-end. The Linear front-end implements a linear pulse amplification in front of the discriminator. The schematic of this AFE is shown in Figure 3.8. As for the SYNC AFE, the PA of the LIN AFE is based on a CSA featuring a Krummenacher feedback (I_{Krum}, $V_{\mathtt{REF\_Krum}}$). The signal from the CSA is fed to a low power threshold discriminator based on current comparison, which compares the signal with the threshold (V_{th}). It is composed of a transconductance stage followed by a *transimpedance amplifier* (TIA) providing a low impedance path for fast switching. A 4-bit binary weighted trimming DAC with adjustable range (I_{DAC}) allows to reduce threshold dispersion across the pixel matrix [87, 97].

Differential front-end. The PA of the Differential front-end, shown in Figure 3.9, has a continuous reset (I_{ff}), unlike the other two designs, which use the Krummenacher feedback with constant current reset. This continuous feedback is able to prevent the input from saturation for a moderate leakage current. For higher currents, a dedicated *leakage current compensation* (LCC) circuit can be enabled. The LCC is disconnected from the input when disabled, which improves phase margin and noise performance. The DC-coupled precomparator provides additional gain in front of the comparator and acts as a differential threshold circuit, i.e. the global threshold is adjustable through two distributed threshold voltages (V_{th1} and V_{th2}) instead of one. The precomparator stage is followed by a time-continuous comparator. The threshold is trimmed in each pixel using a local 5-bit trimming DAC (TDAC) [87].

Summary. The common features and particularities of the three AFEs implemented in the RD53A chip are summarised in Table 3.1.

Table 3.1: Design features of the three RD53A analogue front-ends.

	SYNC AFE	**LIN AFE**	**DIFF AFE**
Preamplifier	CSA	CSA	CSA
Feedback loop	Krummenacher	Krummenacher	Fast feedback + LCC
Threshold tuning	Autozeroing	4-bit DAC	5-bit DAC
Discriminator	Synchronous	Asynchronous	Asynchronous differential
TOT counting	20 to 300 MHz	40 MHz	40 MHz

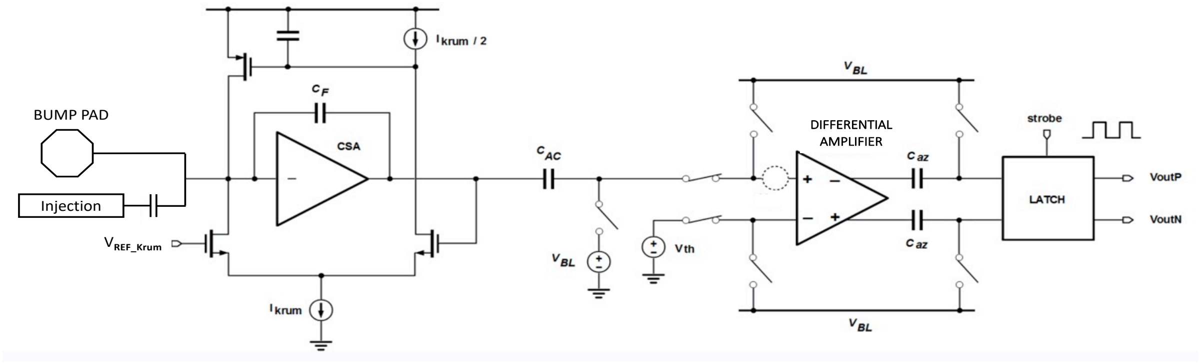


Figure 3.7: Schematic of the synchronous front-end implemented in the RD53A chip [87].

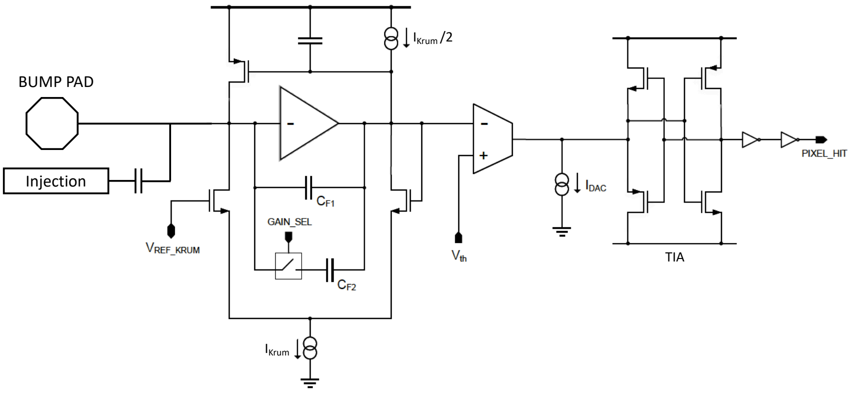


Figure 3.8: Schematic of the linear front-end implemented in the RD53A chip [87].

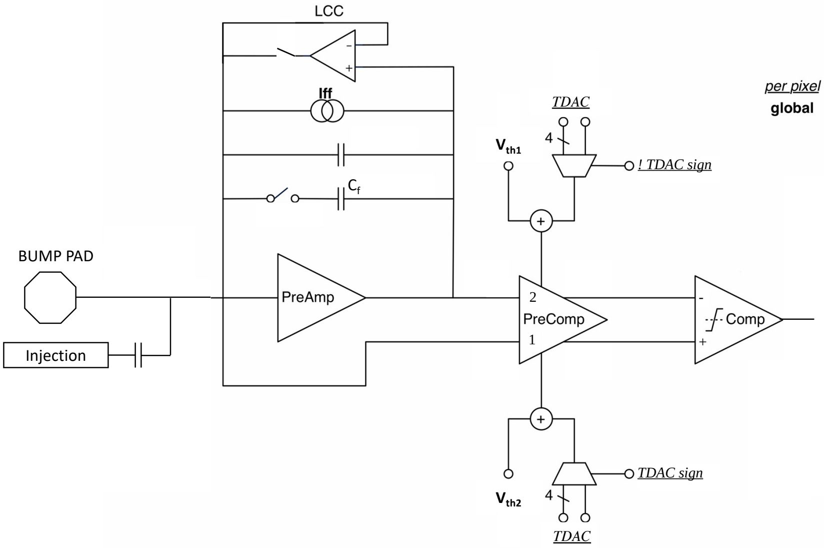


Figure 3.9: Schematic of the differential front-end implemented in the RD53A chip [87].

3.3 CMS requirements

The first step towards the choice of the AFE for the CMS final chip was the establishment of the evaluation criteria. In this section, the most important detector parameters are discussed and used to derive the CMS requirements, against which the three AFE designs were evaluated.

Detection threshold. The detection threshold is the most important AFE parameter, since it determines the smallest signal that can be detected. The threshold is usually set as low as possible to maximise the detection efficiency, but is should not be too low, to keep the rate of noise hits at an acceptable level. The typical signal expected from the sensor has to be estimated to set the requirement for the detection threshold. Silicon sensors with a thickness of 150 µm will be used in the new IT. The charge distribution obtained with 120 GeV protons from a test beam collected in a 130 µm-thick sensor with $100 \times 150\,\mu\text{m}^2$ pixels is shown in Figure 3.10. The *most probable value* (MPV) is about $7900\,\text{e}^-$ before irradiation and it decreases by about $2000\,\text{e}^-$ after irradiation to $1.2 \times 10^{15}\,\text{n}_\text{eq}/\text{cm}^2$ [65]. Based on the expected signal, a detection threshold of $1000\,\text{e}^-$ is required by CMS for the innermost layer of the IT to ensure sufficient detection efficiency, especially with irradiated sensors. A threshold of $1200\,\text{e}^-$ is sufficient for the outer layers of the detector, where the fluence is lower.

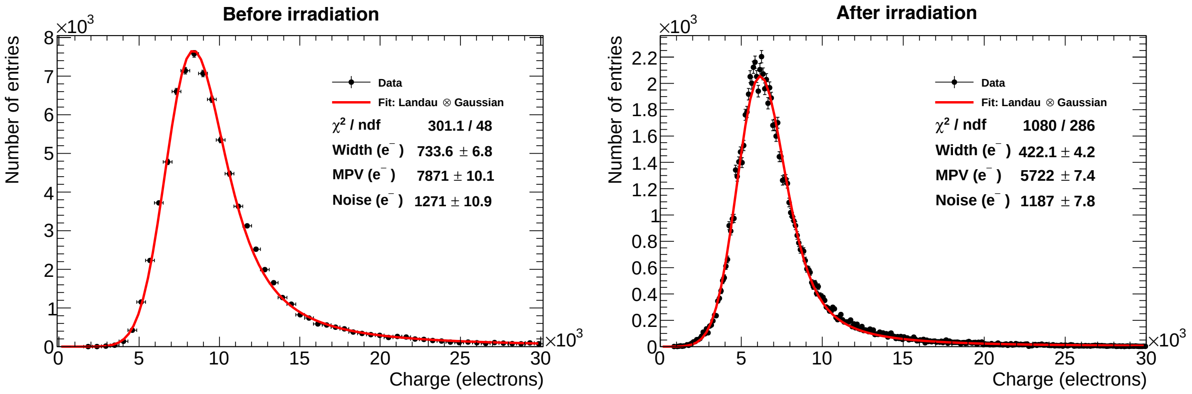


Figure 3.10: Test beam measurement of the collected charge before (left) and after (right) irradiation to $1.2 \times 10^{15}\,\text{n}_\text{eq}/\text{cm}^2$, using single pixel clusters, in a 130 µm-thick pixel sensor with $100 \times 150\,\mu\text{m}^2$ pixels [65].

Radiation tolerance. The IT is the CMS subdetector closest to the IP and therefore it is exposed to the highest radiation levels. As it was explained in Chapter 2, Section 2.2.1, two scenarios are envisaged for the HL-LHC: in the "nominal" scenario, the accelerator would deliver a maximum of 140 p-p collisions per BX, to reach a total integrated luminosity of $3000\,\text{fb}^{-1}$ by the end of the physics program. In the "ultimate" scenario, the number of p-p collisions per BX would be pushed up to 200, reaching an integrated luminosity of 4000 fb^{-1}. A fluence reaching $2.6 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$ and a TID up to 1.4 Grad are expected in the innermost layer in the nominal scenario, while the figures would scale up to $3.4 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$ and 1.9 Grad in the ultimate scenario. The RD53A chip was designed to withstand a TID of at least 500 Mrad and an average leakage current up to 10 nA/pixel [88]. However, with this specification the radiation levels expected in CMS, reaching 1.9 Grad in the ultimate luminosity scenario, would imply a replacement of the innermost layer of the IT barrel after every two years of operation. The CMS Collaboration aims for one replacement of the innermost layer during the ten-year lifetime of the detector, hence a higher radiation tolerance is privileged.

Noise occupancy. For a stable operation at low threshold it is important to minimise the front-end noise to have a limited fraction of spurious hits in the data. Single pixels that are too noisy can be disabled, to keep the overall noise occupancy low, however their fraction must be limited in order not to significantly affect the detector efficiency. The noise hit rate becomes particularly relevant in the parts of the detector with the lowest hit occupancy. Therefore, the noise requirement for the AFE evaluation was established based on the expected occupancy. The occupancy simulated as a function of pseudorapidity for different parts of the new IT is presented in Figure 3.11. The lowest expected occupancy is 10^{-4} and this value was taken as the limit to declare a pixel noisy. Hence, in this work

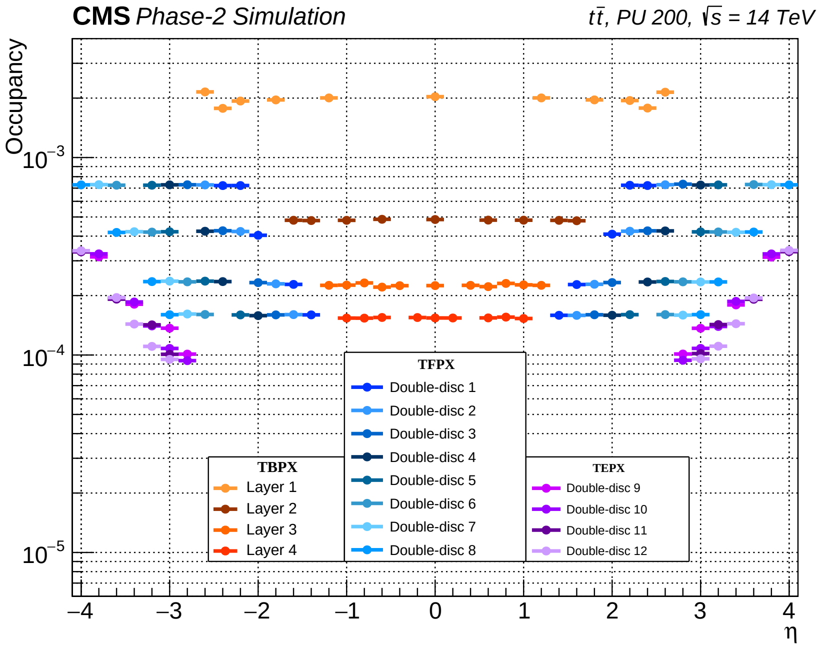


Figure 3.11: Hit occupancy as a function of pseudorapidity for all layers and double-discs of the Phase-2 Inner Tracker, simulated using top quark pair production events with a pileup of 200 events [65].

the pixels with more than 100 noise hits in 10^6 triggers were disabled. The average noise occupancy of the remaining pixels was required to be well below the hit occupancy. The maximum accepted average noise occupancy of the new AFE was set to 10^{-6}, which is two orders of magnitude below the lowest expected occupancy.

Dead time. After each event the hit pixel is busy processing the signal and is not able to record another event for a certain time called *dead time*. The dead time has to be kept as small as possible, especially at the high hit rates of the HL-LHC. The CMS requirement for the dead time in the innermost layer of the IT barrel was set to 1%, to ensure high detection efficiency even at the highest expected hit rate. This requirement translates to a maximum efficiency loss of 1% at maximum hit rate caused by the total dead time (digital + analogue). The dead time in the RD53A chip has a minor contribution from the digital buffering and a major contribution from the CSA of the AFE. While the digital contribution is due to the limited hit buffer size and cannot be reduced with the chip settings, the AFE dead time depends on the TOT calibration. The TOT response to a given input charge can be set in the chip to a certain number of TOT_{40} units (one TOT_{40} unit corresponds to one 40 MHz clock cycle, i.e. to 25 ns). The charge resolution is obtained by dividing the input charge by the corresponding number of clock cycles and can therefore be expressed in $\mathrm{e}^-/\mathrm{TOT}_{40}$ units.

A Monte Carlo simulation of hit efficiency losses due to the digital, analogue, and total dead time is shown in Figure 3.12 for two charge resolutions: 1500 $\mathrm{e}^-/\mathrm{TOT}_{40}$ and

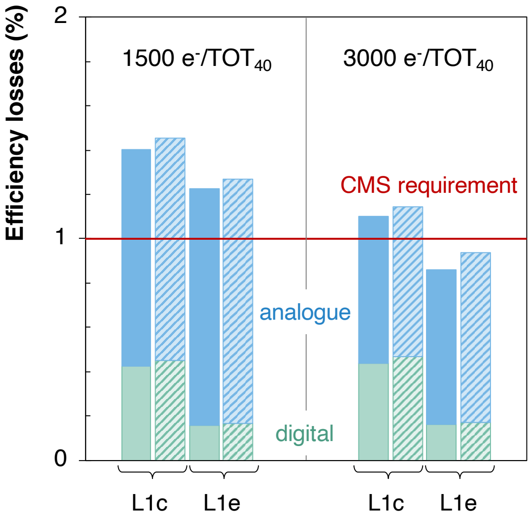


Figure 3.12: Hit efficiency losses due to the digital buffering (green) and analogue dead time (blue), simulated at 200 pileup for two charge calibrations: 1500 $\mathrm{e}^-/\mathrm{TOT}_{40}$ and 3000 $\mathrm{e}^-/\mathrm{TOT}_{40}$. The simulation was done for the centre (c) and edge (e) of the innermost layer (L1) of the IT barrel and for two pixel geometries: the solid bins represent the $100 \times 25\,\mu\mathrm{m}^2$ pixels and hashed bins represent the $50 \times 50\,\mu\mathrm{m}^2$ pixels.

$3000\,e^-/TOT_{40}$. The simulation was performed for two pixel module positions in the innermost layer of the IT barrel: the centre ($z = 0$), denoted L1c, and the edge, denoted L1e. For each position both pixel geometries were simulated. The rectangular pixels are represented with solid bins and the squared pixels with hashed bins. The squared pixels cause a slightly higher inefficiency. As expected, the TOT charge resolution has no influence on the digital dead time, and the hit losses caused by the AFE are smaller with the coarser charge resolution of $3000\,e^-/TOT_{40}$. The efficiency losses are higher in the centre making the dead time requirement difficult to meet. With the charge resolution of $1500\,e^-/TOT_{40}$ the requirement is not satisfied in any of the two module positions, while with $3000\,e^-/TOT_{40}$ the requirement is satisfied on average. The hit efficiency losses are slightly above the requirement in the centre and slightly below at the edge. The charge resolution of $3000\,e^-/TOT_{40}$ was therefore taken as the TOT calibration requirement for the AFE evaluation.

The impact of charge resolution on tracking performance was also evaluated. Simulation of the tracking performance for the reconstruction of single muons with a transverse momentum of 10 GeV was performed with planar 150 µm-thick sensors, with both sensor pixel geometries and at two different thresholds: $1200\,e^-$ and $2400\,e^-$. The resolution on the transverse (d_0) and longitudinal (z_0) impact parameters integrated over the full η range is shown in Figure 3.13 for three charge resolutions: $600\,e^-/TOT_{40}$, $3000\,e^-/TOT_{40}$ and $6000\,e^-/TOT_{40}$. The impact parameter resolution deteriorates for a higher threshold and appears to be insensitive to the charge resolution. Since a higher charge resolution does not affect the tracking performance, a charge resolution of $3000\,e^-/TOT_{40}$ was taken as the baseline calibration for the inner regions of the IT.

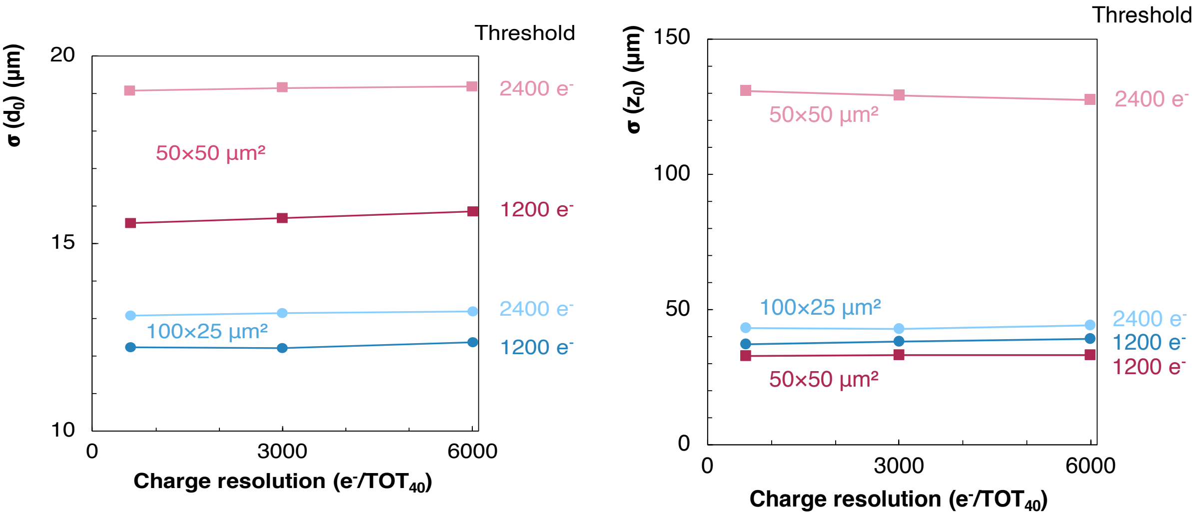


Figure 3.13: Influence of the charge resolution on the transverse (left) and longitudinal impact parameter resolution (right) obtained from simulation. The $100 \times 25\,\mu m^2$ pixels are represented in blue and the $50 \times 50\,\mu m^2$ pixels are represented in red. The dark colours indicate the threshold of $1200\,e^-$ and the light colours indicate the threshold of $2400\,e^-$.

3.4 Analogue front-end test procedure

Once the most important detector performance parameters were identified and the requirements set, a systematic test procedure was established. The same calibration procedure was applied to each AFE for each measurement to make sure that the three designs were tested in the same conditions.

3.4.1 Test set-up

Two readout systems exist within the RD53 Collaboration to interface, test and characterise the RD53A chip: YARR [98] developed by Lawrence Berkeley National Laboratory [99] and BDAQ53 developed by the University of Bonn [100]. The latter was used for all measurements presented in this chapter.

The BDAQ53 [101] is a versatile readout system and verification framework for the readout chips designed by the RD53 Collaboration. The Python-based software and Verilog-based firmware are open source and are available in a public repository [102]. The test system was composed of custom and commercial hardware, shown in Figure 3.14. The RD53A chip, with or without sensor, was glued and wire-bonded on a custom-made PCB, called *single chip card* (SCC). A power supply was used to power the chip and another one to bias the sensor if present. A set of external pins on the SCC allowed to measure or adjust important voltages and currents of the chip. The communication with the chip was made via a *display port* (DP) cable arriving to a custom-made *FPGA mezzanine card* (FMC) to make the connection with the FPGA. The commercially available Xilinx KC705 evaluation board [103], housing the Xilinx Kintex7 FPGA was used.

Different chip assemblies were used for the AFE evaluation campaign to compare different configurations: bare RD53A chips with no sensors and RD53A chips bump-bonded to sensors. The sensors selected for the evaluation campaign, based on the availability, were the HLL sensors [104], developed by the semiconductor laboratory of the Max Planck society [105]. HLL sensors with two pixel sizes were used: squared pixels of $50 \times 50\,\mu\text{m}^2$ and rectangular pixels of $100 \times 25\,\mu\text{m}^2$. If a measurement was performed only on one type of assembly, the rectangular pixels were used given that they have the highest capacitance.

The assemblies were tested at cold temperature to be close to the operation conditions of the future detector. The SCC card was placed in a cold box, where it was cooled down to about $-10\,^\circ\text{C}$, which is the lowest temperature that could be achieved with the cooling systems available for the lab test set-ups. Dry air was blowing into the cold box to reduce humidity and avoid condensation that could damage the wire-bonds. The temperature and humidity of the chip environment were monitored by external sensors. The temperature of the chip itself was monitored by a *negative temperature coefficient* (NTC) thermistor, mounted on the SCC very close to the chip.

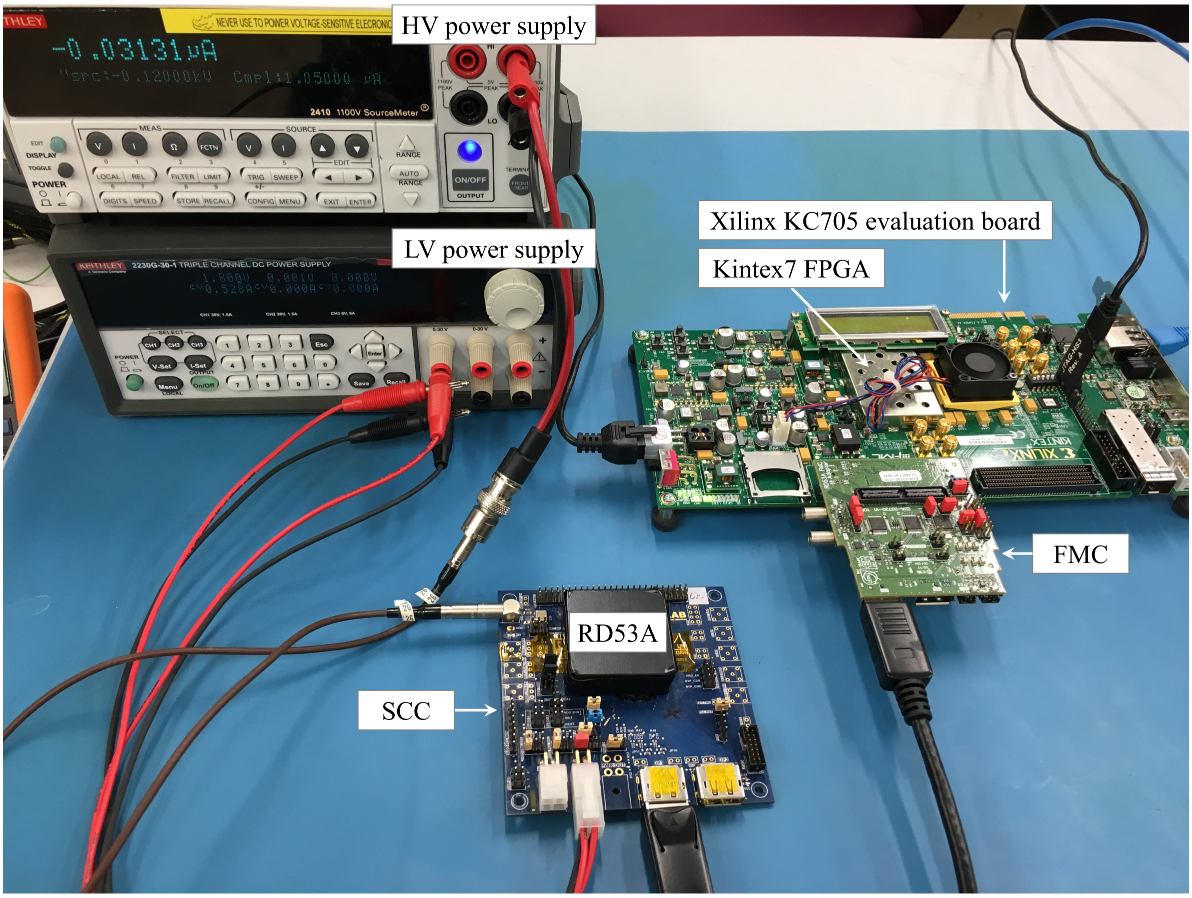


Figure 3.14: Photograph of the BDAQ53 test set-up used for the AFE evaluation.

3.4.2 Calibration injection circuit

Signals used for the AFE performance tests were generated internally in the chip by the *calibration injection circuit* [87], instead of being received from the sensor. This dedicated circuit, connected to the input of the PA, allows to inject a well-defined programmable charge to test the front-end functionalities and to calibrate it. Each pixel in the RD53A chip contains the same calibration injection circuit independently on the AFE flavor. The circuit is based on two switches that generate voltage steps fed to an injection capacitor, as shown in the block diagram in Figure 3.15. The switches are flipped by two control signals S0 and S1. The charge injected into enabled pixels is defined by two distributed DC voltages (V_{CAL_HI} and V_{CAL_MED}), which allows either to inject two successive pulses or a single differential pulse. The advantage of the second option used for all presented measurements is that the signal is created by a precision differential voltage (ΔV_{CAL}), which is not affected by local ground drops in the chip [87]. Thanks to the charge calibration, the injected charge in ΔV_{CAL} can be converted into electrons.

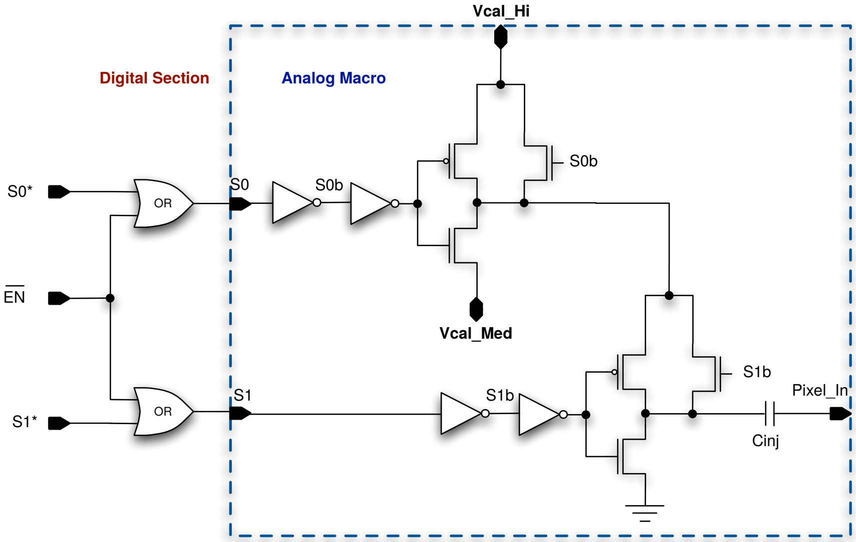


Figure 3.15: Calibration injection circuit present in every pixel of the RD53A chip [87].

3.4.3 Powering

The readout chip and the sensor had to be correctly powered to achieve good performance. In particular, the three AFEs were powered in the same way for a fair comparison.

Chip powering. The RD53A chip has two separate power domains: one for the digital and one for the analogue part. Therefore, two voltages have to be supplied to power the chip, one for each domain. They are denoted V_{DDD} and V_{DDA}. There are three ways of powering the RD53A chip: LDO mode, ShLDO mode and direct powering [87]. The LDO is an on-chip linear voltage regulator that supplies constant voltage to the chip from an input voltage. The value of the internal supply voltage provided by the regulator can be adjusted with trim bits, individually for V_{DDD} and V_{DDA}. In the ShLDO, mode a dedicated shunt circuitry is enabled in addition to the LDO to generate a regulated internal supply voltage from a constant input current. This powering mode is used for the serial powering of pixel modules. Finally, in direct powering mode the regulator is bypassed and the input voltage is directly supplied to the chip. This powering mode is risky since a too high supply voltage can damage the chip. The direct powering was introduced as a back-up option in case of issues with the regulator. It should be used for debugging purposes only. The recommended powering for single chip testing is the LDO mode and it was used for all measurements presented in this chapter. The values of the input voltage (V_{IN}) and the internal voltage supplied by the regulator (V_{DD}), used for the AFE evaluation were:

$$V_{IND} = V_{INA} = 1.8\,\mathrm{V}$$

$$V_{DDD} = V_{DDA} = 1.2\,\mathrm{V}$$

Reference current and voltage. For the AFE performance evaluation, two important references have to be adjusted in the RD53A. The reference current I_{REF} is important for all the AFE bias DACs, i.e. for the AFE parameters. The value of I_{REF} can be adjusted with four wire bonds, either connected to ground or to V_{DDA}. For the RD53A on the SCC, it can be set with jumpers via external pins. The reference voltage V_{REF} is important for the injection DACs. The calibration voltages V_{CAL_HI} and V_{CAL_MED}, which define the charge injected by the calibration injection circuit, are derived from the V_{REF}. Hence, it has an impact on the amount of injected charge and it has to be adjusted with trim bits. The nominal values of the two references, used for all measurements in this work were:

$$I_{REF} = 4.0\,\mu\text{A}$$

$$V_{REF} = 0.9\,\text{V}$$

The charge calibration used in this thesis corresponds to $V_{REF} = 0.9\,\text{V}$. It was determined based on test results obtained with a variable radioactive source consisting of Americium ($^{241}_{95}$Am) with different targets [106]. The conversion of the injected charge from ΔV_{CAL} to electrons is given by:

$$Q\ [e^-] = (10.4 \pm 0.1)\ \Delta V_{CAL} + (180 \pm 60) \tag{3.1}$$

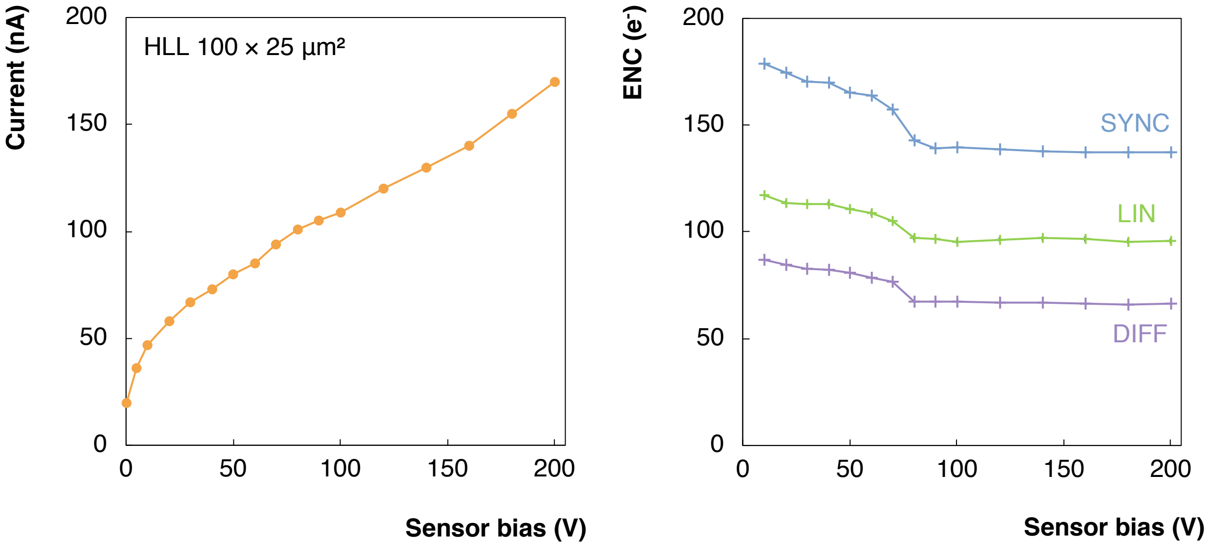
Sensor bias. The sensor was biased in tests performed on chips with sensors, even though it was not used for particle detection, the bias provided realistic operation conditions. More detailed explanation about the sensor bias is provided in Appendix C. For the AFE evaluation, the sensor bias voltage had to be high enough to reach full depletion and hence, to minimise the capacitive load on the AFE as well as the noise, and to maximise the leakage current. The HLL sensors are expected to be fully depleted at 30 V*. The full depletion voltage can usually be inferred from the sensor I/V-curve, as explained in Chapter 4, Section 4.1. The I/V-curve of the HLL sensor with rectangular pixels is presented in Figure 3.16(a). For this type of sensor, the full depletion voltage could not be verified from the I/V-curve because it does not show a plateau. Since the sensor bias voltage has an impact on noise, the noise of the same sensor was measured as a function of the sensor bias and the result is shown in Figure 3.16(b). The equivalent noise charge (ENC) is explained later in this chapter, in Section 3.6.1. For the moment, the important observation is that the noise reaches the minimum at around 90 V and then remains stable independently on the AFE flavor. Hence, for all AFE measurements performed with chips with sensors, the sensor bias was set to:

$$V_{sensor} = 120\,\text{V}$$

---

*The bias voltages for the n-in-p sensors are negative, as explained in Chapter 4, Section 4.1, but the absolute values are used for simplicity.

to make sure that the sensors were fully depleted. Moreover, the sensor has to be biased only after the chip is powered, otherwise the leakage current flowing into the unpowered AFE may cause damage.



(a) (b)

Figure 3.16: I/V-curve of an HLL sensor with rectangular pixels connected to an RD53A chip (a) and the equivalent noise charge of the three analogue front-ends measured with the RD53A chip with same sensor as a function of the sensor bias (b).

3.4.4 Calibration procedure

After the correct powering of the readout chip and biasing of the sensor (if applicable), the following calibration procedure was established and applied systematically prior to every measurement:

1. **Enable the AFE to be tested.** Only one AFE should be enabled at a time, otherwise some interference may occur at the borders between the AFEs. This can induce extra noise and affect the performance and thus bias the results.

2. **Set the AFE parameters.** The optimal parameters for each AFE to achieve the best performance within a precise power budget were obtained from circuit simulations and provided by designers [107–109]. Only two parameters are adjusted during the calibration procedure: the threshold and the PA discharge current for TOT calibration.

3. **Find and disable faulty pixels.** First, the digital functionality needs to be verified. The digital scan allows to send digital pulses to each pixel without involving the AFE. Pixels not responding to 100% of digital injections are disabled. Then, a

selected charge, well above threshold, is injected to verify the analogue part of each pixel and faulty pixels are disabled.

4. **Set and tune the threshold.** A first rough tuning at high threshold is done to homogenise the thresholds across the matrix. Then, the threshold is set to the required value of 1000 e^-/1200 e^- and a fine tuning is performed. For the SYNC AFE the threshold tuning is done automatically by the auto-zeroing circuit, therefore it is enough to set the threshold.

5. **TOT calibration.** The TOT response is calibrated by injecting a fixed charge to the pixels and adjusting the PA discharge current to obtain a target mean TOT value. The TOT calibration for the evaluation campaign was set to a mean TOT of 5.3 for an injection of 6000 e^-.

6. **Readjust the threshold if needed.** The PA discharge parameter has an influence on the threshold depending on the AFE design. Hence, in some cases the threshold needs to be readjusted after the TOT calibration.

7. **Mask noisy pixels.** Once the AFE is calibrated to optimal conditions, noisy pixels are masked. Empty triggers are sent to each pixel, with no injection and noise-induced hits are counted. Pixels with more than 100 hits in 10^6 triggers are disabled and their number is reported.

An extra precaution was taken for the evaluation of the DIFF AFE. The first tests of the RD53A chip revealed a periodic pattern in the this AFE. After investigation, a mistake in the layout has been identified. A pixel-to-pixel variation of the parasitic capacitance at the output of the comparator causes timing variations across the DIFF AFE matrix. This induces also variations in threshold, TOT and other performance parameters. The layout was corrected and verified in simulation for future chips, but for the RD53A AFE

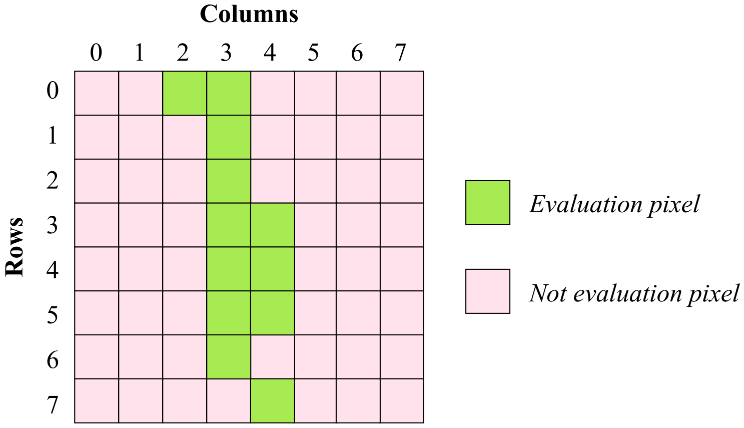
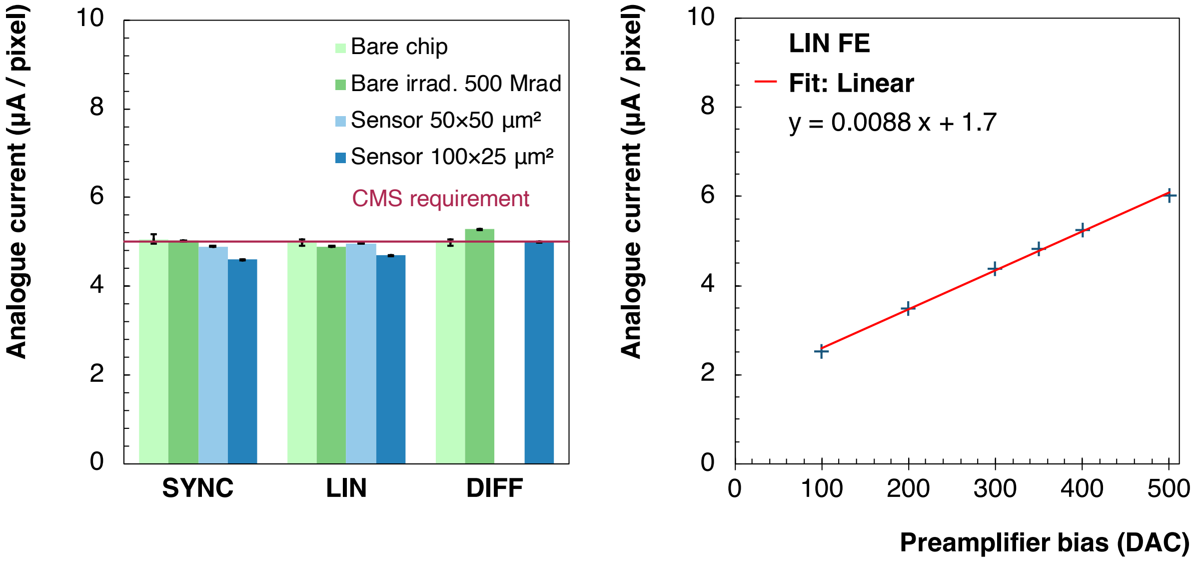


Figure 3.17: Pixel mask applied to each pixel core of the Differential AFE implemented in the RD53A chip for the AFE evaluation.

evaluation, only a subset of pixels, which supply a correct charge information, was selected. Therefore a so-called *good pixel mask* [109], shown in Figure 3.17, was applied for any performance measurement of the DIFF AFE.

3.4.5 Power consumption

The power consumption of the readout chip is an important factor for the detector to ensure that the cooling system can dissipate the heat. Hence, the AFE performance had to be verified within a certain power budget. The power budget for the AFE evaluation campaign was set to an analogue current consumption of 5 µA/pixel and the corresponding AFE parameters were used. The average analogue current consumption was measured after the full calibration of each AFE and before any further tests. The result for the three AFEs is shown in Figure 3.18(a) for four different chip assemblies: two bare chips shown in green (one not irradiated and one irradiated up to 500 Mrad) and two chips with sensors with different pixel geometries shown in blue. All three AFEs display an average analogue current consumption very close to the target 5 µA/pixel, indicated by the red line. This confirms that circuit simulations provide reliable values of the AFE parameters to reach a given power consumption. Moreover, the power consumption can be adjusted with the PA bias current, which is the AFE parameter with the biggest contribution to the total current consumption. The other AFE parameters have a negligible effect on the power consumption. The measured average analogue current consumption as a function



(a) (b)

Figure 3.18: Average analogue current consumption of four RD53A chips, two of them without sensor and two with sensors (a) and the average analogue current consumption as a function of the preamplifier bias measured with the Linear AFE (b).

of the PA bias is shown in Figure 3.18(b) for the LIN AFE. The measurement shows that the analogue current increases linearly with increasing PA bias. The fit with a linear function indicates that an increase of the PA bias by 100 increases the power consumption by almost 0.9 µA/pixel. This confirms that the PA bias represents a handle to adjust the current consumption of the AFE.

3.5 Threshold and threshold dispersion

3.5.1 Threshold measurement

The threshold of an AFE is set via a global bias DAC, i.e. an AFE parameter setting the discriminator threshold DC voltage. The value of the threshold DAC does not provide directly the information about the effective threshold, the latter has to be measured. The calibration injection circuit can be used to inject a range of charges to each pixel, which enables the measurement of the threshold. The occupancy as a function of the injected charge of each pixel is a sigmoid commonly called an *S-curve*. An example of an S-curve plot is shown in Figure 3.19. Each S-curve is fitted with an error function expressed as:

$$f(x) = \frac{A}{2}\left[erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)+1\right] \tag{3.2}$$

where $A/2$ is the maximum amplitude, i.e. 100%, μ is the mean, i.e. the charge for which the 50% occupancy is reached and σ represents the smearing of the S-curve. The calibration charge at which 50% occupancy is reached is taken as a measurement of

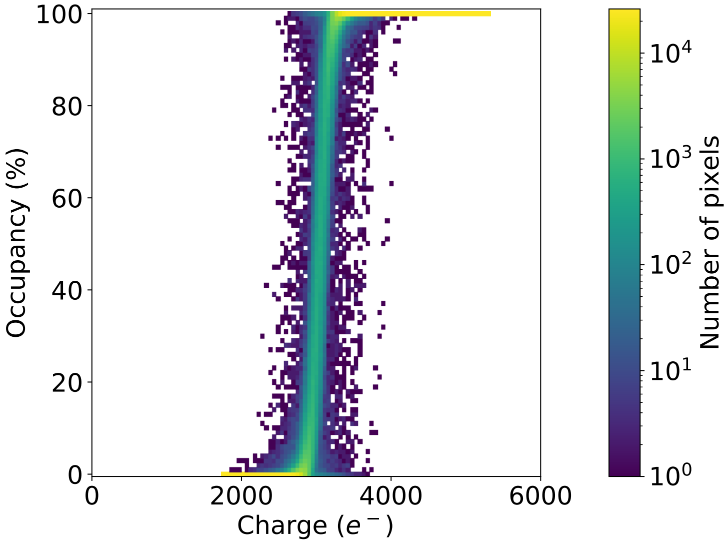
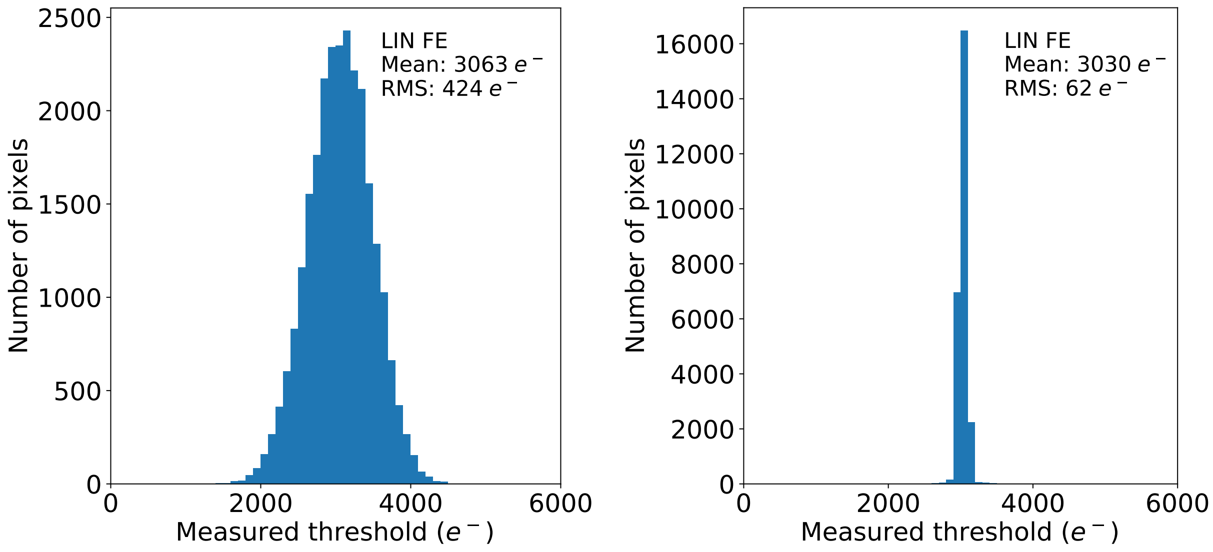


Figure 3.19: Example of the S-curves obtained with the Linear AFE.

the input charge equivalent of each pixel threshold. Examples of the pixel threshold distributions are shown in Figure 3.20. The mean of the distribution represents the global detection threshold.

3.5.2 Threshold dispersion and tuning procedure

The *root mean square* (RMS) of the threshold distribution indicates the threshold dispersion across the pixel matrix. The dispersion can be compensated with the threshold tuning, which consists in setting optimal trim bits for each pixel with a dedicated tuning algorithm. A binary search algorithm is used in BDAQ53 to iteratively adjust the trim bits of all pixels to an optimal value that brings the local threshold of a pixel as close as possible to the global threshold. Typically pixel-to-pixel variations result in a threshold dispersion of several hundred electrons, which can be reduced to less than hundred electrons after the threshold tuning. An example of a threshold distribution before and after tuning is shown in Figures 3.20(a) and 3.20(b), respectively.



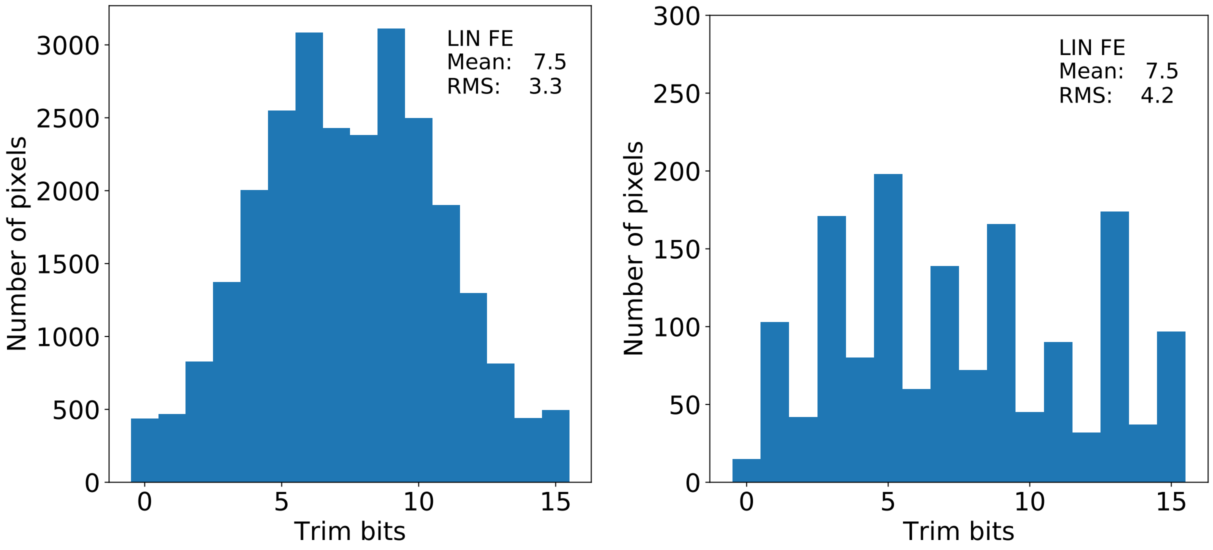
(a) Untuned threshold distribution.

(b) Tuned threshold distribution.

Figure 3.20: Example of threshold distribution before (a) and after tuning (b) obtained with the Linear AFE.

The trim bit range can be set with a global parameter for the whole AFE. If the threshold dispersion is moderate, a low dynamic range is sufficient and a precise fine tuning can be achieved. If the threshold dispersion is larger, for instance due to cold temperature or irradiation, a higher dynamic range allows to compensate for the threshold variations. A satisfactory threshold tuning results in as narrow threshold distribution as possible, with no tails, as it is shown in Figure 3.20(b). In addition, the distribution of trim bits among the pixels should be a centred Gaussian-like distribution, with no saturated bins, as illustrated in Figure 3.21(a).

The first RD53A tests, performed with an old tuning algorithm, revealed a periodic pattern in the trim bit distribution of the LIN AFE. The threshold tuning was repeated with different samples and the pattern persisted. During the tuning procedure the odd values of trim bits were privileged with respect to the even values as it is shown in Figure 3.21(b). This even/odd unbalance was suggesting an issue with the *least significant bit* (LSB) of the trimming DAC. After investigation it was found out that the even/odd unbalance was due to the lack of dummy transistors in the layout. The LSB was implemented close to the boundary of the circuit and some edge effects were causing a different LSB step size for even and odd trim bits [110]. The layout was corrected and verified in a test chip [110]. Moreover, the even/odd pattern was observed when the tuning was performed with an old algorithm. A new optimised tuning algorithm was implemented, allowing the tune the LIN AFE and to obtain a satisfactory trim bit distribution as the one presented in Figure 3.21(a). The new more efficient algorithm was used for all the following measurements.



(a)

(b)

Figure 3.21: Example of a good trim bit distribution obtained with the Linear AFE (a) and the even/odd pattern in the trim bit distribution observed when using the old tuning algorithm (b).

Another issue was observed when the tuning the LIN AFE was performed at cold temperature, as it was decided for the AFE evaluation. At cold temperature, the threshold dispersion of the LIN AFE increases and a bigger trim bit range is needed to compensate for the threshold variations. However, it was observed that the trim bit range could not be increased sufficiently to achieve a good tuning. A saturation of the current DAC setting the dynamic range of the trim bits was identified and confirmed by circuit simulations, as demonstrated in Figure 3.22. An improvement of the design, allowing to increase the dynamic range, was proposed, simulated and implemented in a test chip [110]. The test results of the new design, matching the simulation are also shown in Figure 3.22. The test results of the improved design confirm that a higher dynamic range can be reached. In addition, one more trim bit was added to the layout, which should improve the threshold dispersion after tuning. The test results also indicate a higher radiation tolerance of the new circuit [110]. For the AFE evaluation at cold temperature a work around was proposed by the designer, consisting in increasing the digital supply voltage from $V_{DDD} = 1.2\,\mathrm{V}$ to $V_{DDD} = 1.3\,\mathrm{V}$. This allows to extend the trimming range in the LIN AFE implemented in the RD53A and to achieve very good threshold tuning also at cold temperature. Increasing the digital voltage does not have an impact on the analogue power budget, discussed previously, and therefore, this strategy was adopted for the AFE evaluation campaign.

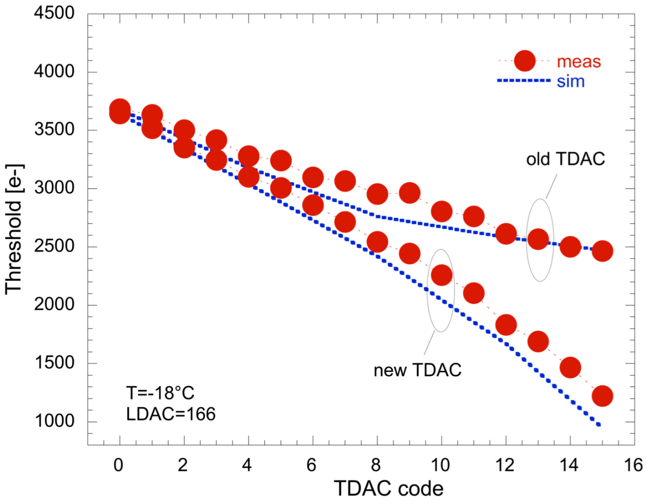
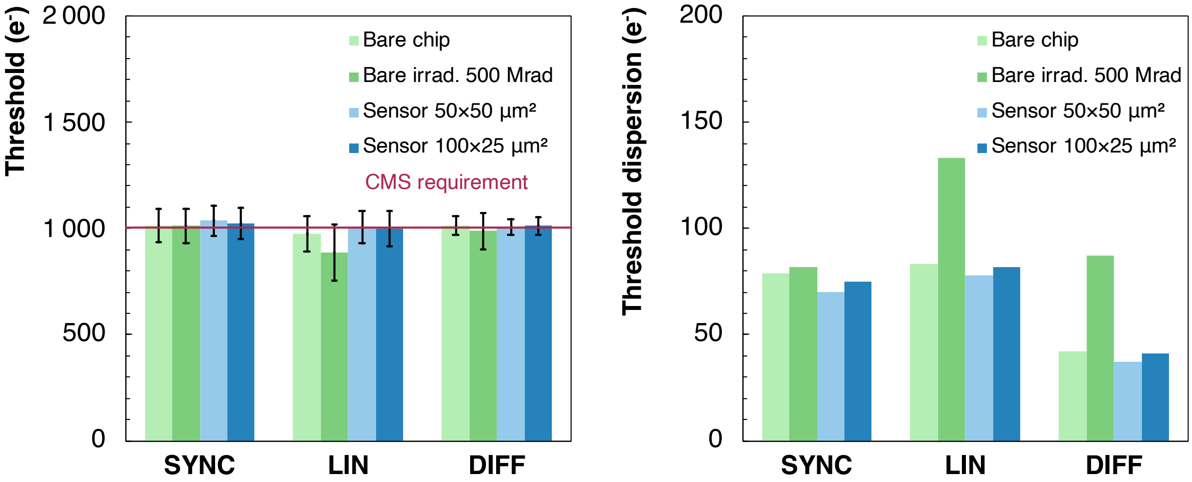


Figure 3.22: Effective threshold of the LIN AFE as a function of the trim bit value (denoted TDAC code), showing the saturation effect in the old design compared to the improved design [110].

3.5.3 Tuning results

The threshold was set and tuned to the required $1000\,\mathrm{e}^-$ for all three AFEs in four different RD53A assemblies: one bare chip, one bare chip irradiated to 500 Mrad and two chips with sensors; one with square and one with rectangular pixels. The threshold after tuning is shown in Figure 3.23(a) and the error bars represent the threshold dispersion RMS. The bare chips are shown in green and the chips with sensors in blue. All three AFEs could be tuned to the required low threshold indicated by the red line. The LIN AFE in the bare irradiated chip was tuned to even lower threshold and it was working well. For the sake of comparison, the threshold dispersion is also presented standalone in Figure 3.23(b). In the three not irradiated samples the dispersion in the SYNC and the LIN AFE was around $80\,\mathrm{e}^-$ and in the DIFF AFE a very small threshold dispersion of about $40\,\mathrm{e}^-$ was achieved.

In the irradiated bare chip the threshold dispersion increased, but is still reasonably low. The threshold tuning to the required threshold was performed on many different RD53A chips and the target threshold could always be reached.

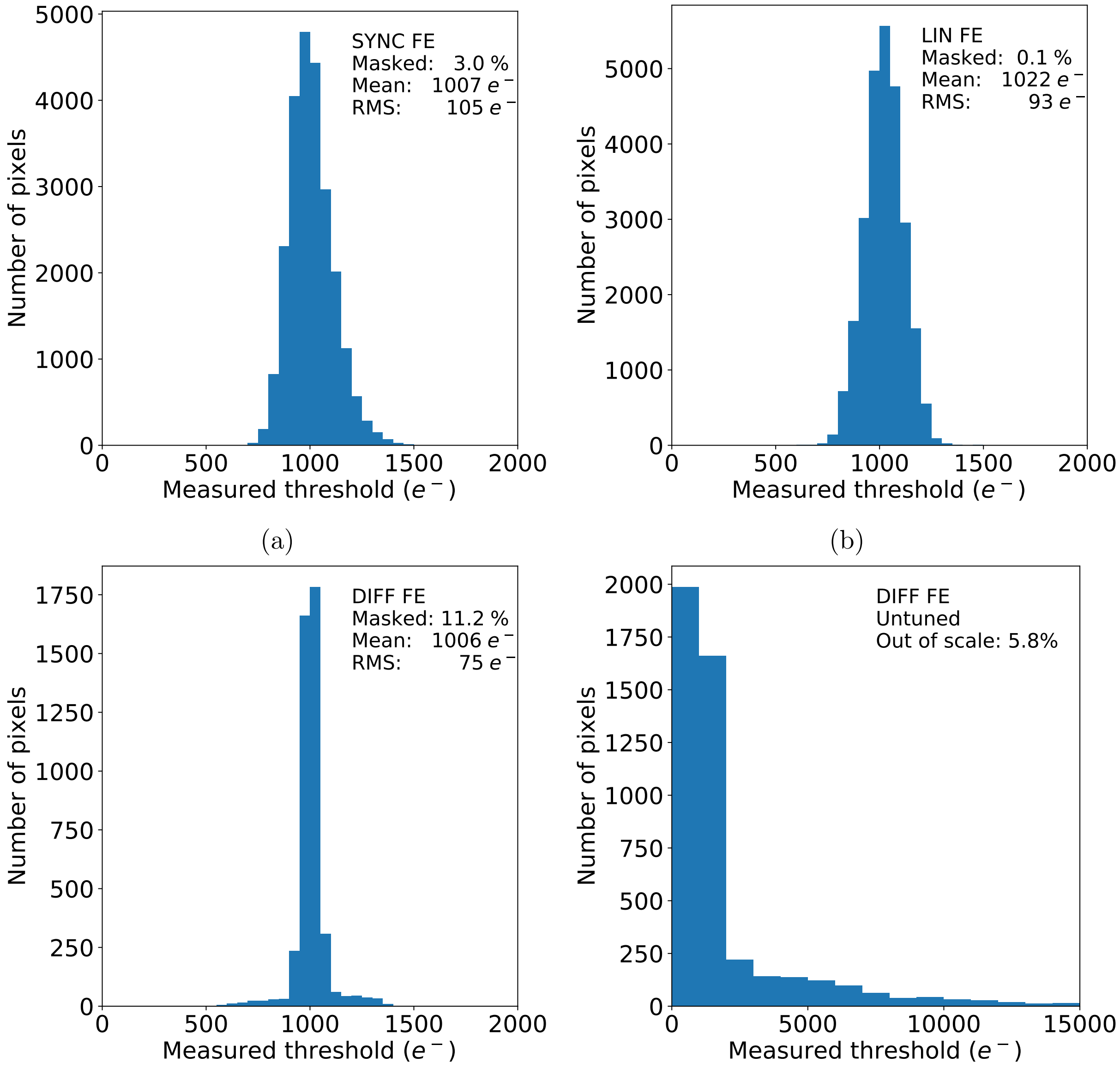


(a) (b)

Figure 3.23: Threshold (a) and threshold dispersion (b) after tuning for the three AFEs in four RD53A chip assemblies.

3.5.4 Influence of the irradiation

Furthermore, an assembly of an RD53A chip and a sensor with square pixels was irradiated at Karlsruhe Institute of Technology [111] with 23 MeV protons up to a fluence of 3×10^{15} $n_{eq}cm^{-2}$, corresponding to a TID reaching 350 Mrad. The RD53A chip was not powered during irradiation. The sample was irradiated at room temperature and kept constantly at cold temperature after irradiation to avoid annealing. It was tested at −10 °C in a dry environment. The sensor bias voltage was increased to reach an average leakage current of 10 nA/pixel, which is the maximum specification for the RD53A chip [88]. The pixels with a noise occupancy higher than 10^{-4} were masked and the remaining pixels were tuned to a threshold of 1000 e^-. Pixels with an anomalously high threshold that could not be adjusted with the range of the trim bits were masked in the tuning procedure. The threshold distributions of the three AFEs after tuning are shown in Figures 3.24(a), 3.24(b) and 3.24(c). All three AFEs were functional after irradiation and could reach the required threshold with a dispersion of about 100 e^- or lower. The threshold tuning of the LIN AFE worked well and only 0.1% of pixels were masked. The auto-zeroing in the SYNC AFE worked well too, however the leakage current caused a higher noise in this front-end and 3% of the pixels were masked. The threshold distribution of the DIFF AFE features a narrow core and long tails, and 11.2% of pixels were masked.



(a) (b) (c) (d)

Figure 3.24: Tuned threshold distribution of the Synchronous (a) the Linear (b) and the Differential AFE (c) and the untuned threshold distribution of the Differential AFE (d) obtained with an irradiated assembly of an RD53A and a sensor with square pixels.

The large fraction of masked pixels in the DIFF AFE is the consequence of a long tail in the untuned threshold distribution, shown in Figure 3.24(d). In fact, some pixels have a local threshold 15× larger than the threshold of 1000 e^- to which there were set. Pixels with a too high threshold cannot be tuned to the desired threshold value because the dynamic range of the trim bits is a global setting for the whole AFE. As the threshold dispersion depends on several AFE parameters, many parameter combinations were tried to mitigate the problem, and the 11.2% of masked pixels was the best result that could be achieved with this irradiated sample.

This study triggered an investigation and the design team found out that the combined effect of irradiation and cold temperature provoked a PMOS threshold increase in the DIFF precomparator, resulting in a small voltage margin. Simulations showed that the voltage margin is smaller in cold and decreases with irradiation, reaching a value close to zero for the DIFF AFE design implemented in RD53A after irradiation to 200 Mrad, which explains the problematic threshold tuning observed after irradiation to 350 Mrad. A design improvement of the DIFF precomparator was proposed and simulated, obtaining an extension of the expected operation range up to 500 Mrad, which remains marginal compared to the expected dose in the CMS detector. With such operation range, replacements of the innermost layer would be required every two years once the ultimate luminosity is reached, while CMS is aiming at one replacement during the whole high-luminosity program. For this reason the choice was oriented towards the other two AFEs, which seem promising candidates for a higher radiation tolerance.

3.6 Noise evaluation

Well-controlled noise level is crucial to build a highly efficient pixel detector, with a low fraction of spurious hits in the data. In the pixel readout chip, noisy pixels can be disabled, to reduce the overall amount of noise. Once noisy pixels are masked, the noise performance of the front-end can be evaluated by measuring either the *equivalent noise charge* (ENC) or the noise occupancy. Both methods were used to evaluate the noise levels in the RD53A AFEs and the results are presented in this section.

3.6.1 Equivalent noise charge

The equivalent noise charge is a common way of expressing noise in terms of input charge in electrons. The noise of a channel can be defined as the RMS of the voltage fluctuation at the end of the analogue processing chain divided by the gain. The ENC corresponds to the input charge, for which the signal voltage at the output of the PA is equal to the output voltage noise, i.e. the signal-to-noise ratio is equal to one. The analogue signal cannot always be measured directly, but the ENC can also be determined from the S-curves, defined in Section 3.5.1. The ENC distribution is obtained from the slope s of the transition region of the S-curves, using the following formula [68]:

$$ENC = \frac{1}{\sqrt{2\pi}s} \tag{3.3}$$

There are several sources of noise in the front-end. Different noise contributions are random and uncorrelated, hence they sum up in quadrature. As a result, the ENC can be expressed in a simplified way as:

$$ENC^2 = N_p\tau + \frac{N_s C_d^2}{\tau} + N_f C_d^2 \tag{3.4}$$

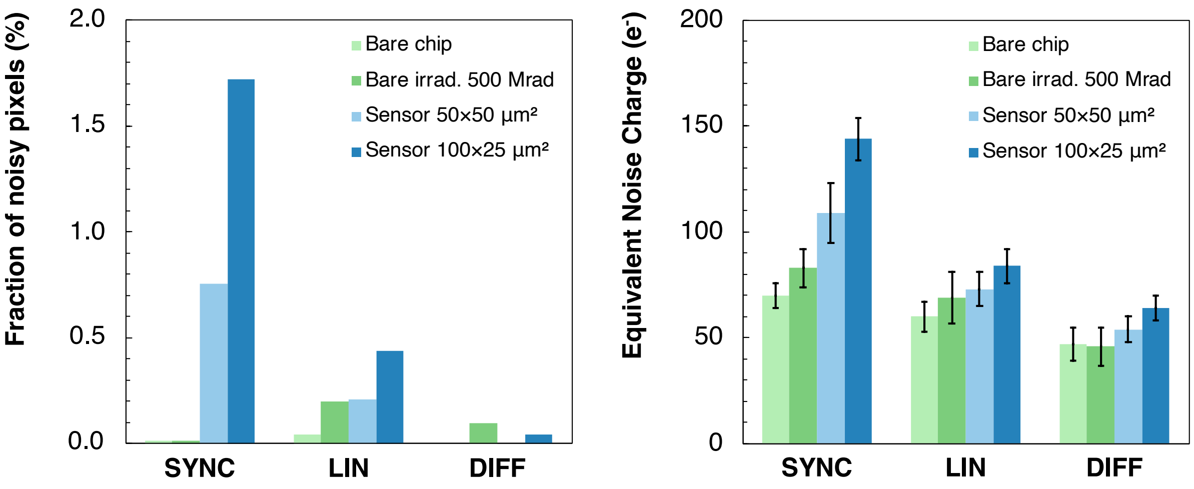
where N_p, N_s and N_f are respectively the current noise (parallel), voltage noise (series) and 1/f-noise [112, 113]. This formula shows that the noise is mainly determined by two factors: the total input capacitance, dominated by the load capacitance of the pixel sensor C_d and by the characteristic time of the PA τ. The latter is mainly influenced by the preamplifier bias current of the front-end, driving the signal rise time. Although the detection threshold has no influence on the ENC, it is an important noise factor as well. In fact, a lower threshold is closer to the noise level and the noise hit rate is higher. The influence of these three factors on the noise level was evaluated for the three AFEs and is presented in the following.

3.6.2 Influence of the detector load capacitance

The ENC of the three AFEs was measured for four RD53A assemblies with different detector loads: two bare chips with no sensor, one not irradiated and one irradiated to 500 Mrad[†], and two chips bump-bonded to sensors with different pixel geometries, one with squared and the other one with rectangular pixels. First, the noisy pixels were counted and masked. A pixel was declared noisy, if it fired more than 100 times in 10^6 triggers at the required threshold of 1000 e$^-$, as explained in Section 3.3. The fraction of masked pixels per front-end for the four samples is shown in Figure 3.25(a) and the average ENC of the remaining pixels in Figure 3.25(b). The error bars indicate the ENC RMS.

An exception was made for the irradiated bare chip (dark green), in which the pixels were masked if they fired 10 times in 10^7 triggers. Although the noisy pixels cut was more strict for this sample, the number of noisy pixels is below 0.2% for all three AFEs and the resulting ENC is close to the ENC of the bare chip before irradiation (light green). When a sensor is connected to the chip, independently on the AFE flavor, the ENC increases, due to the sensor load capacitance. Moreover, it can be noticed from Figure 3.25(b) that the sensor with rectangular pixels (dark blue) induces higher noise than the one with squared pixels (light blue), which confirms a higher capacitance of rectangular pixels. The total capacitance is dominated by inter-pixel capacitance, which is higher with the longer pixel edge. In addition, the ENC is systematically higher in SYNC and this front-end presents a significantly larger number of noisy pixels when connected to a sensor, increasing from 0.01% in a bare chip to 1.7% when a sensor with rectangular pixels is present. Circuit simulations show that it is due to a smaller phase margin in this front-end design.

[†] The initial chip specification for the radiation tolerance was 500 Mrad [88], however, based on the latest simulations CMS aims for a higher radiation tolerance.



(a) (b)

Figure 3.25: Fraction of masked noisy pixels (a) and equivalent noise charge (b) measured with four RD53A chips, two of them without sensor and two with sensors.

3.6.3 Noise occupancy

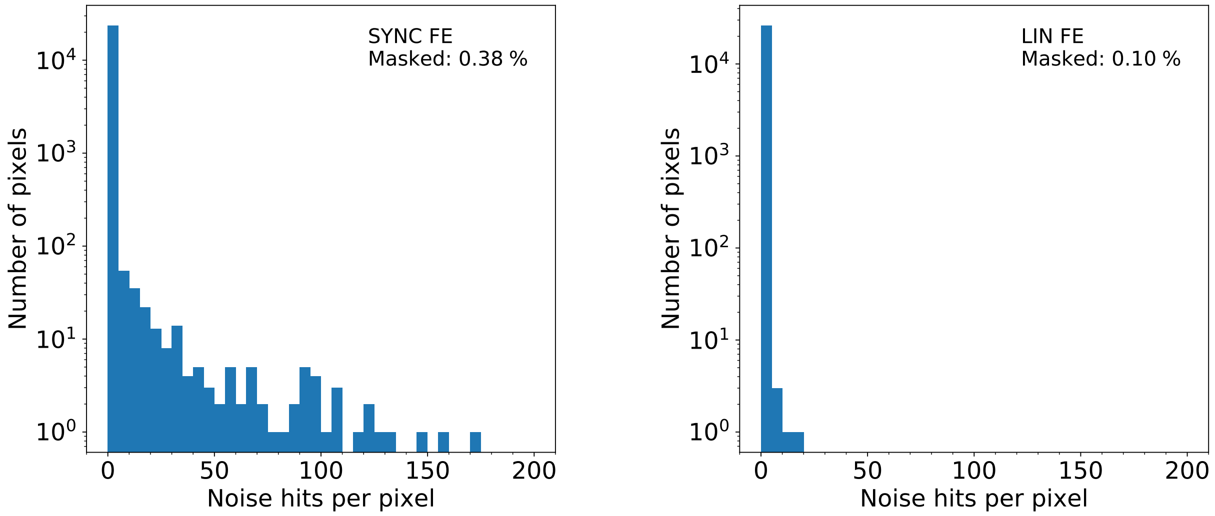
In addition to the ENC measurement, another method was implemented to better assess the noise in the proposed AFEs. First, the noisy pixels are masked and then a new set of 10^6 triggers is sent to each remaining pixel with no charge injection. This way each recorded hit is purely induced by the noise. The noise hits are counted for each pixel and the pixels are sorted by the number of noise hits. The distribution of noise hits per pixel, obtained with an RD53A connected to a sensor with rectangular pixels, is shown in Figure 3.26 for the SYNC and the LIN AFE tuned to a threshold of 1200 e^-. For the DIFF AFE the distribution is not interesting to show, since no noise hits were recorded in 10^6 triggers, and therefore, all pixels are placed in bin 0 of the distribution.

The bulk of the pixels are contained in bin 0, since they recorded no noise hits and pixels with some noise hits are distributed in a tail. The noise hit distribution of the SYNC AFE shows a longer tail than the LIN AFE. Some pixels in the SYNC AFE recorded more than 100 hits in 10^6 triggers and so their noise occupancy is above the limit of 10^{-4}, hence some noisy pixels are present also after masking. The tail of the LIN AFE is shorter with a maximum of 17 hits per pixel. This measurement confirms the higher noise in the SYNC AFE.

Moreover, the integral of the noise hit distribution gives the average noise occupancy, denoted n, and defined as:

$$n = \frac{N_{hits}}{N_{pixels}\ N_{triggers}} \tag{3.5}$$

where N_{hits} is the number of recorded hits, N_{pixels} the number of enabled pixels in the AFE and $N_{triggers}$ the number of sent triggers.



(a) Synchronous front-end.

(b) Linear front-end.

Figure 3.26: Distribution of noise hits per pixel obtained with the Synchronous AFE (a) and Linear AFE (b) in the RD53A chip at a threshold of 1200 e^-.

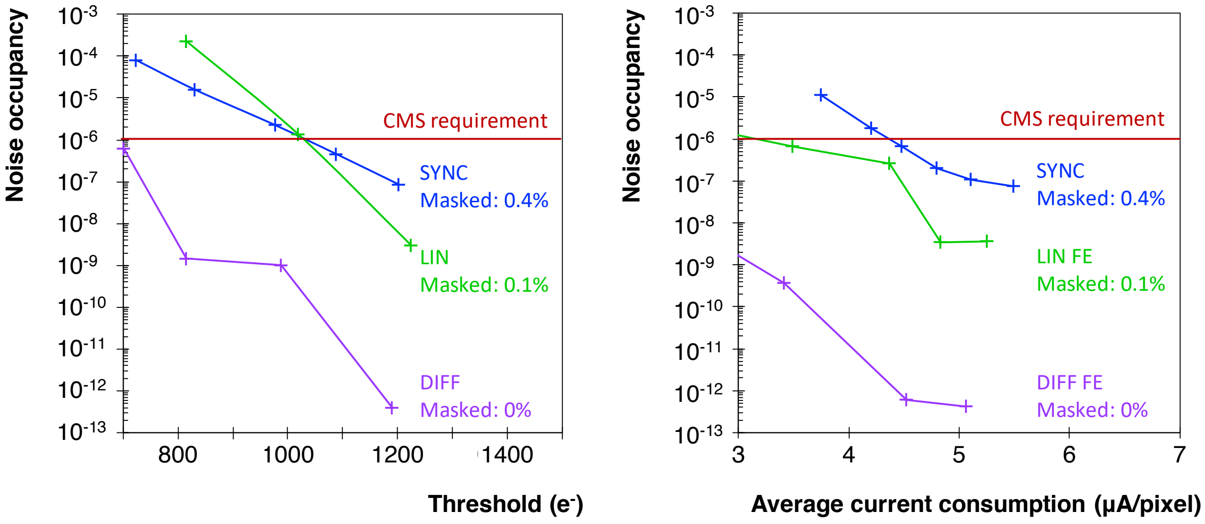
3.6.4 Influence of the threshold

Since the influence of the threshold on noise cannot be evaluated with the ENC, the noise occupancy measurement explained in the previous section was performed. The average noise occupancy was measured as a function of the threshold for the three AFEs, using an RD53A chip with a sensor with the highest capacitance, i.e. rectangular pixels. The masking of noisy pixels was done at a threshold of 1200 e^- and the number of masked pixels is indicated between parenthesis in the legend. Then the threshold was gradually decreased, keeping the same noisy pixel mask.

The result is shown in Figure 3.27(a), where the maximum noise occupancy of 10^{-6}, required by CMS, as explained in Section 3.3, is indicated by the red line. As expected, the average noise occupancy decreases with increasing threshold, regardless of the front-end design. The DIFF front-end shows excellent noise performance, with the noise occupancy several orders of magnitude below the requirement, even for low thresholds. In fact, in this front-end no hits were found in 10^6 triggers, hence a higher number of triggers had to be sent to evaluate the noise occupancy. The other two AFEs satisfy the noise requirements down to the required threshold of 1000 e^-. Nonetheless, in the SYNC front-end the number of masked pixels is still higher than for the other two, confirming previous observations.

3.6.5 Influence of the preamplifier bias

As it was inferred from Equation (3.4), some contributions to the noise, in an AFE, depend on the peaking time of the preamplifier. The latter can be modified using the PA



(a) (b)

Figure 3.27: Influence of the threshold (a) and the preamplifier bias current (b) on the noise occupancy of the three AFEs implemented in the RD53A chip.

bias current. When this current increases, the transconductance of the input transistor is increased, which results in lower noise with a penalty of an increase in the analogue current consumption. Indeed, the PA bias current is the main contribution to the total analogue power consumption of the chip. When the PA bias DAC is increased, the analogue current consumption increases proportionally.

The influence of the preamplifier bias current on the noise was studied for the three AFEs. The same measurement method and conditions, as described in the previous section, were used and the same noisy pixels were masked. The noise occupancy was measured for different PA bias currents. Since the PA bias DAC settings are different from one design to another, the measured noise occupancy is presented in Figure 3.27(b), as a function of the measured analogue current consumption per pixel. All the other front-end settings, that could contribute to the current consumption were kept constant during this measurement.

All three AFEs can meet the CMS requirement, indicated by the red line, if the bias current is adapted. The DIFF front-end shows very good noise performance. The noise occupancy is well below the requirement, even when operated with low PA bias current. The LIN and SYNC AFE need at least 3.5 µA and 4.5 µA per pixel, respectively, to reach the required noise level. This result demonstrates, that the front-end noise decreases when the PA bias increases. This parameter can therefore be used as a handle to reduce the front-end noise at a price of an increase in the power consumption.

3.7 Time-over-threshold calibration

An important consideration for a highly efficient particle detector is the event loss due to the dead time, especially at high luminosity and high pileup. As explained in Section 3.3, the dead time caused by the AFE depends on the TOT charge calibration, which is the topic of this section.

3.7.1 Time-over-threshold measurement and calibration

The RD53A AFEs provide the information about the signal amplitude, i.e. the input charge, as the TOT expressed in 40 MHz clock cycles. The calibration injection circuit can be used to measure the TOT response to a given signal. For that a fixed input charge is injected 100 times into each enabled pixel. The input charge selected for all TOT measurements and calibrations presented in this chapter is 6000 e^-, which represents the charge deposited by a *minimum ionizing particle* (MIP) in a 100 µm thick sensor.

The TOT response to the selected input charge can be calibrated by adjusting the discharge current of the PA, as mentioned in Section 3.2.2. When the PA discharge current increases, the PA output returns faster to the baseline and the corresponding TOT is smaller, as illustrated in Figure 3.28. The TOT calibration is a part of the AFE calibration procedure, explained in Section 3.4.4, and can be expressed by the measured mean TOT obtained for the selected input charge. The baseline calibration used in this thesis is a mean TOT of 5.3 clock cycles for an injection of 6000 e^-.

An example of a TOT distribution after such calibration is shown in Figure 3.29(a) for

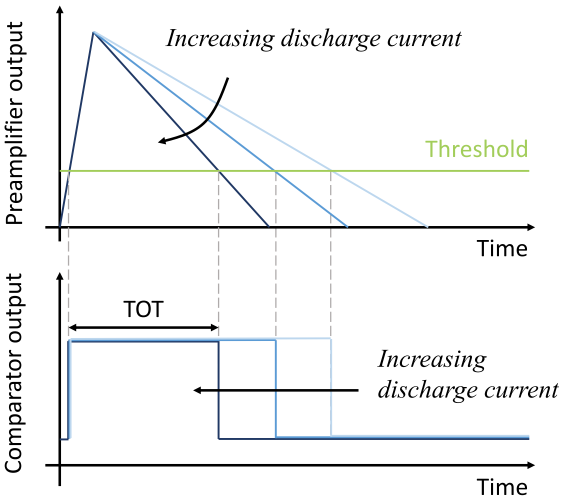
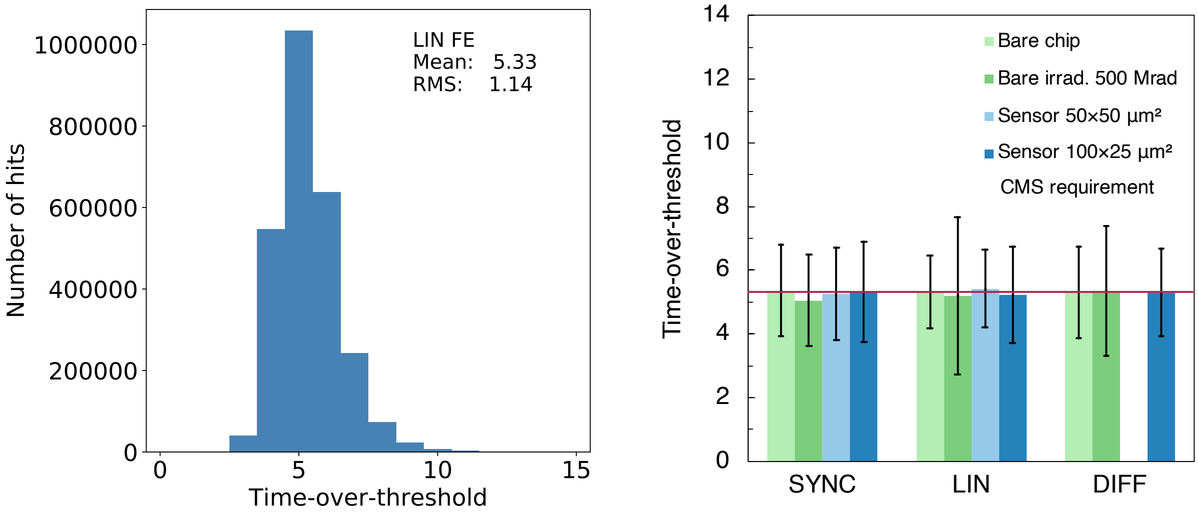


Figure 3.28: Influence of the preamplifier discharge current on the signal shape at the preamplifier output and on the corresponding TOT.

the LIN AFE. The process variation in combination with the threshold dispersion cause a spread of the TOT response to a fixed charge, which can be observed in the TOT distribution. The TOT cannot be adjusted locally as it is the case for the threshold, however, the TOT accuracy has a negligible impact on the tracking performance according to the simulation presented in Figure 3.13.

The TOT calibration was performed and verified for the four RD53A assemblies: two bare chips, one not irradiated and one irradiated to 500 Mrad and two chips with sensors, one with square and one with rectangular pixels. The mean value of the TOT distribution after calibration is reported in Figure 3.29(b) and the error bars represent the RMS. All three AFEs can reach the baseline TOT calibration indicated by the red line.



(a)

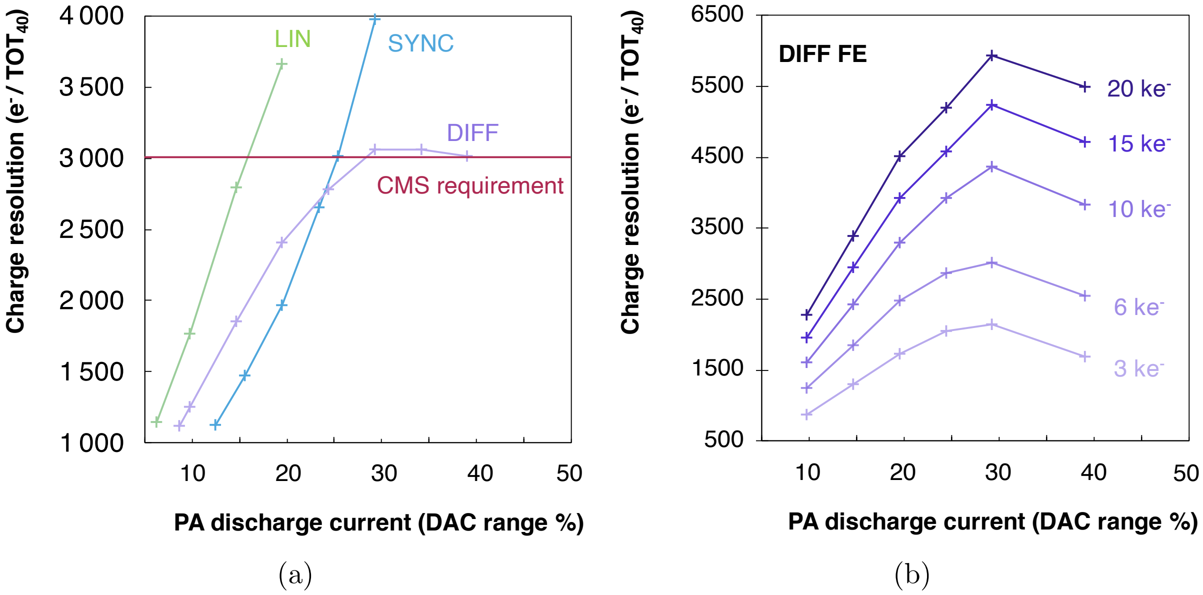
(b)

Figure 3.29: Example of TOT distribution after calibration obtained with the Linear AFE (a) and mean TOT of the three AFEs measured in four RD53A assemblies after calibration (b).

3.7.2 The dead time requirement

The TOT calibration can also be expressed in terms of the charge resolution, obtained as the ratio between the input charge and the measured mean TOT. The baseline calibration of μ TOT$_{40}$ = 5.3 for 6000 e^- corresponds the a charge resolution of about 1100 e^-/TOT$_{40}$, referred to as the *slow discharge* in the following. A charge resolution of at least 3000 e$^-$/TOT$_{40}$ is necessary to achieve the 1% dead time required for the innermost layer of the IT barrel, as explained in Section 3.3. The required charge resolution of 3000 e$^-$/TOT$_{40}$ is also referred to as the *fast discharge* in the following.

The TOT charge resolution of the three AFEs was measured as a function of the PA discharge current. First, all three AFEs were calibrated to the slow discharge, as can be observed in the lower left corner in Figure 3.30(a). This calibration is not reached for the same current in different AFEs. The SYNC AFE needs more current than the other two, which contributes to a higher power consumption. The PA discharge current was increased to verify the front-end compliance with the dead time requirement. As expected, when the discharge current increases the preamplifier discharges faster and the TOT resolution is coarser. All three AFEs can reach the required charge calibration indicated by the red line. The SYNC and the LIN AFE can also discharge faster, while the DIFF AFE shows a saturation of the preamplifier discharge current DAC and would be operated at its limit to reach the dead time required for the inner layers.



(a) (b)

Figure 3.30: Charge resolution as a function of the preamplifier discharge current measured with a constant charge injection of 6000 e^- for the three RD53A AFEs (a) and measured with different input charges for the Differential AFE only (b).

A dedicated measurement was carried out on the DIFF AFE, to better understand the observed saturation effect. The charge resolution of the DIFF AFE as a function of the discharge current was measured for different input charges, ranging from 3 to 20 ke$^-$. The result, presented in Figure 3.30(b), confirms the saturation of the discharge current DAC in this AFE, occurring at 30% of the DAC range, regardless of the input charge. This implies a marginal operation of this particular AFE to reach the dead time requirement.

The charge resolution as a function of the PA discharge current was also verified at two different temperatures: 35 °C and −8 °C. The result obtained with the SYNC AFE is shown in Figure 3.31. For the same discharge current, the charge resolution is less coarse at cold temperature, which means that with decreasing temperature the signal return to baseline is slower. This implies that to reach the same TOT calibration, more current is needed at cold temperature.

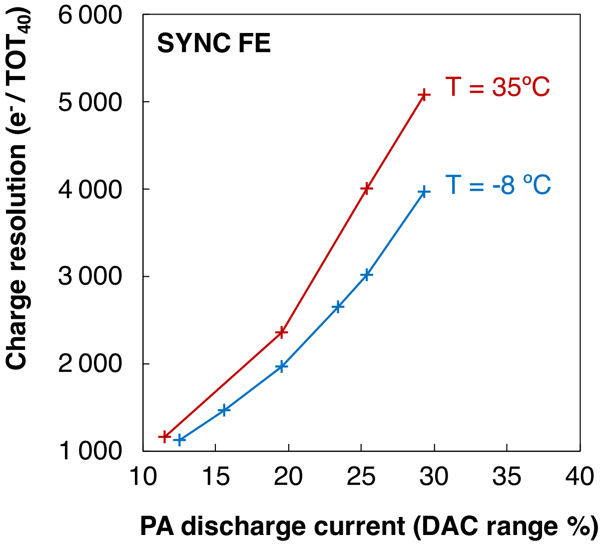
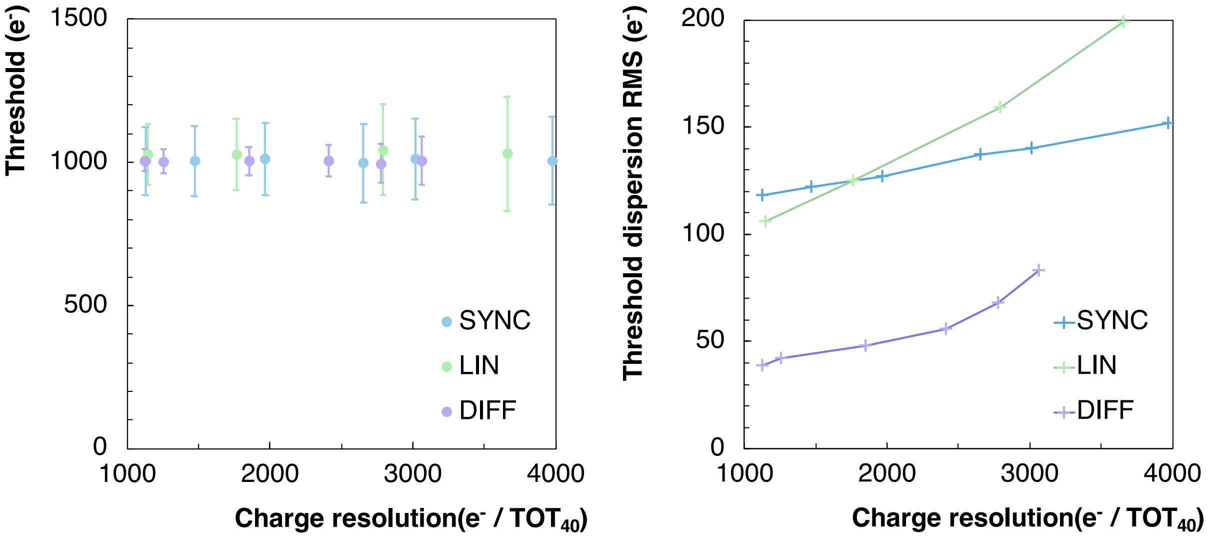


Figure 3.31: Charge resolution as a function of the preamplifier discharge current obtained with the Synchronous AFE at two different temperatures.

3.7.3 Impact on threshold

Increasing the discharge current reduces the dead time, but it also makes the AFE less stable because it reduces its phase margin. This may have an impact on the AFE performance. After the TOT calibration the threshold of each AFE was readjusted to the required threshold of 1000 e$^-$. The threshold was verified for different charge resolutions and the result is presented in Figure 3.32(a). The error bars indicate the threshold dispersion RMS. For any TOT calibration the required threshold of 1000 e$^-$ could be reached



(a)

(b)

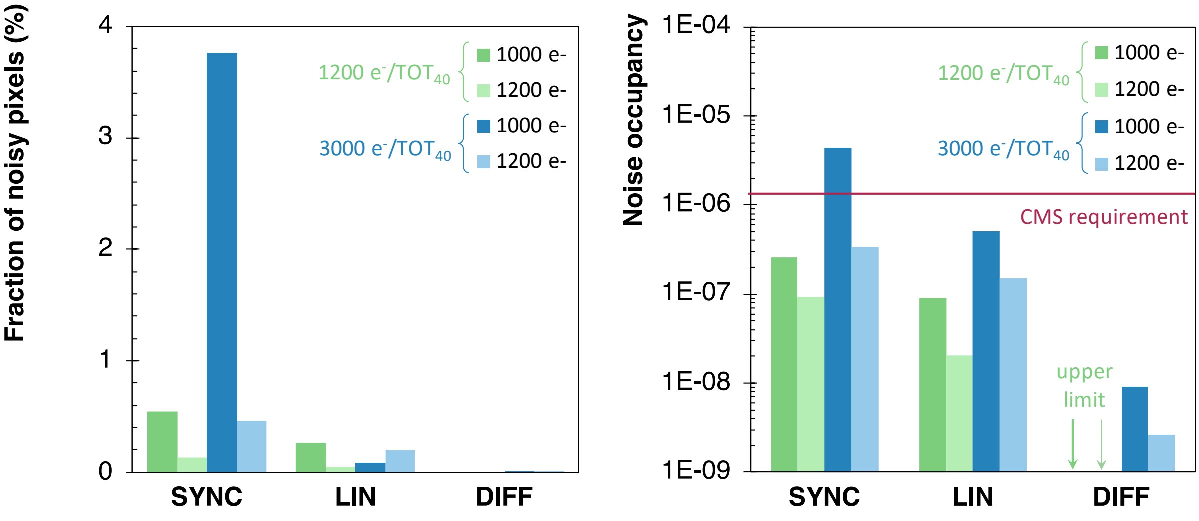
Figure 3.32: Threshold (a) and threshold dispersion (b) as a function of the charge resolution measured with a constant charge injection of 6000 e$^-$ for the three RD53A AFEs.

by all three AFEs. The threshold dispersion increases slightly at faster discharge. For better visibility the threshold dispersion RMS is presented standalone in Figure 3.32(b) as a function of the charge resolution. For all three AFEs the threshold dispersion increases at faster discharge, but remains below 200 e^-. Moreover, it could be compensated with a new threshold tuning.

3.7.4 Impact on noise

As mentioned previously, a faster PA discharge reduces the phase margin of the AFE, and therefore it is likely to induce more noise. Hence, the noise was re-evaluated at the fast discharge, while the noise measurements, exposed in Section 3.6, were obtained at slow discharge. The noise was measured for two detection thresholds, 1000 e^- and 1200 e^-, and two charge resolutions, 1100 e^-/TOT_{40} and the required 3000 e^-/TOT_{40}. The combination of these four parameters defined four measurement scenarios for which the average noise occupancy was measured. The measurement method was the same as in Section 3.6.3. In brief, pixels with more than 100 hits in 10^6 triggers were declared noisy and masked, then the average noise occupancy of non-masked pixels was defined as the number of noise hits per pixel and per trigger, measured over 10^6 events.

The fraction of masked pixels is shown in Figure 3.33(a) and the average noise occupancy in Figure 3.33(b) for the four considered scenarios. Green colours represent the slow discharge and blue colours the fast discharge. The darker colours are for the lower threshold of 1000 e^- and the lighter colours for 1200 e^-. The average noise occupancy



(a) (b)

Figure 3.33: Fraction of masked noisy pixels (a) and the average noise occupancy after masking (b) of the three AFEs in the RD53A chip, measured for two detection thresholds and two TOT calibrations.

of all three AFEs is higher at fast discharge and is the highest at fast discharge and low threshold. The DIFF AFE demonstrates again excellent noise performance, with almost no noisy pixels and the average noise occupancy well below the requirement, even at fast discharge. At slow discharge the noise in this AFE was so low that only an upper limit was estimated. The LIN AFE had few noisy pixels and the average noise occupancy satisfies the requirement for any scenario. The SYNC AFE appears to be the noisiest of the three, reaching almost 3.8% of noisy pixels when operated at fast discharge and low threshold. The higher noise in this AFE, significantly further degrading with more aggressive chip settings, was considered a critical aspect for the operation in the innermost layer of the CMS IT.

3.8 Late-detected hits

The time response of the pixel readout chip is of high importance to assign detected hits to their corresponding LHC BXs and to limit the out-of-time pileup causing spurious hits in the reconstruction.

The time response of the AFE, i.e. the combination of the PA rise time and the discriminator speed, is a function of the input charge. Pulses with the same peaking time but different amplitude pass the discriminator threshold at different times. High amplitude signals, depicted in blue in Figure 3.34, pass the threshold within one BX, i.e. within 25 ns. If the deposited charge is just above the threshold instead, the signal is rising slower and is detected later by the discriminator. Such a hit, shown in red, might be assigned to the following BX, and appears as a spurious hit in another event. The smallest charge (Q_{min}) which can be detected within the correct BX (dark blue signal) is equivalent to the so-called “in-time threshold” that is higher than the threshold of the discriminator.

The time behaviour of an AFE is typically described by its time walk curve, i.e. the response delay of the discriminator as a function of the input charge. An example of a simulated time walk curve of the LIN AFE is shown in Figure 3.35. Given that the discriminator of one of the three AFEs is synchronous to the clock, the time walk of this AFE cannot be directly measured and another method had to be used to compare the timing of the three AFEs.

3.8.1 Time response measurement

The charge injection in the RD53A chip can be delayed with respect to the rising edge of the clock with a step size of 1.5625 ns [87]. The front-end time response was measured by injecting calibration pulses with different amplitudes and with different time delays. The detection threshold was set to 1000 e^- and the full range of available charges up to 35 ke^- was scanned, using a finer charge step for low charges where timing is critical.

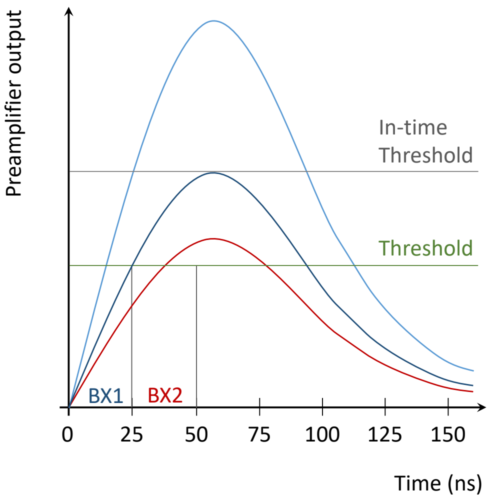


Figure 3.34: Illustration of the discriminator time response for different signal amplitudes.

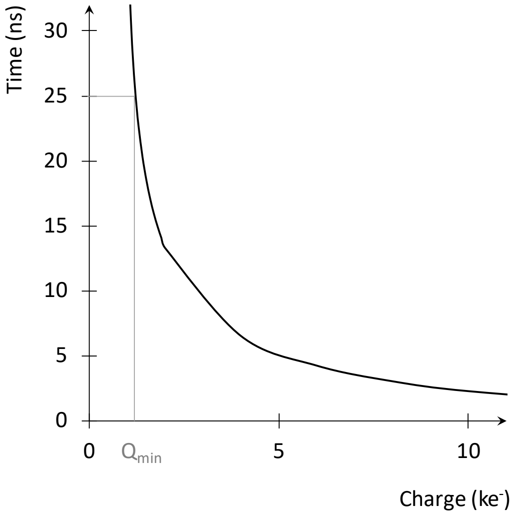


Figure 3.35: Simulated time walk curve of the Linear analogue front-end in the RD53A chip.

Each charge was injected 50 times to a pixel for each time delay. Figure 3.36 shows the two-dimensional plot of charge versus time for all three AFEs. The x axis represents time in nanoseconds and $t = 0$ indicates the time when the highest charge is detected. The y axis, showing the injected charge in electrons, is limited to 10 ke$^-$ in this figure. The colour code indicates the detection probability for a given BX, for each combination of charge and injection delay. The yellow zone corresponds to 100% detection efficiency, while in the white part no hit is detected. The left edge of the coloured region corresponds to the time walk curve. In the upper part of the plot, the coloured region is a straight rectangle with a time width of 25 ns, which confirms that high charges are always detected within one BX. Small charges instead are detected later, resulting in a tail in the detection region. This tail extends up to about 40 ns in the SYNC and DIFF AFE, indicating that these two AFEs have a comparable time response. The DIFF AFE is able to correctly assign slightly smaller charges than the SYNC AFE. The LIN AFE appears to be the slowest of the three, with the largest time walk of more than two BX.

3.8.2 Combination with time of arrival simulation

A Monte Carlo simulation was performed within the standard CMS simulation and reconstruction software framework called CMSSW [114] to evaluate the influence of the time response of each AFE on the detector performance and to estimate the resulting fraction of spurious hits. The time of arrival of particles was simulated for different locations in the IT detector, given that it depends on the position of the pixel module with respect to the interaction point. Sixteen different locations were studied: the centre ($z = 0$) and edge module of each barrel layer, the innermost and outermost module of the first and last

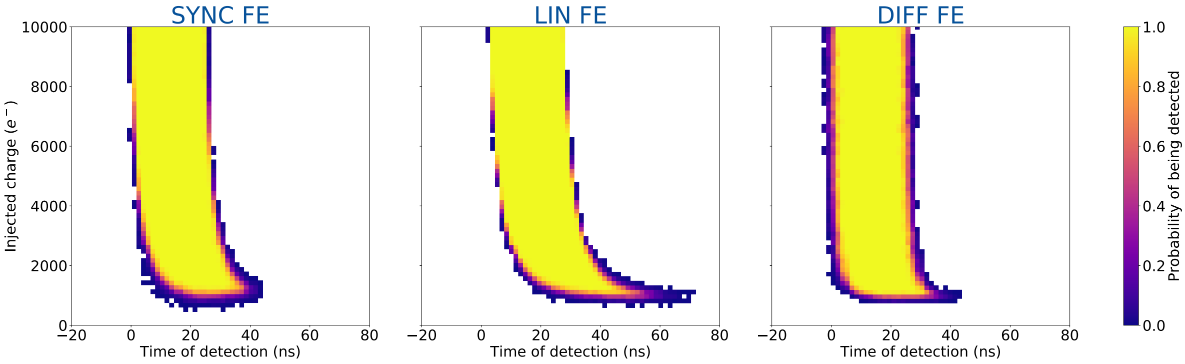


Figure 3.36: Time response measurement of the three analogue front-ends in the RD53A chip.

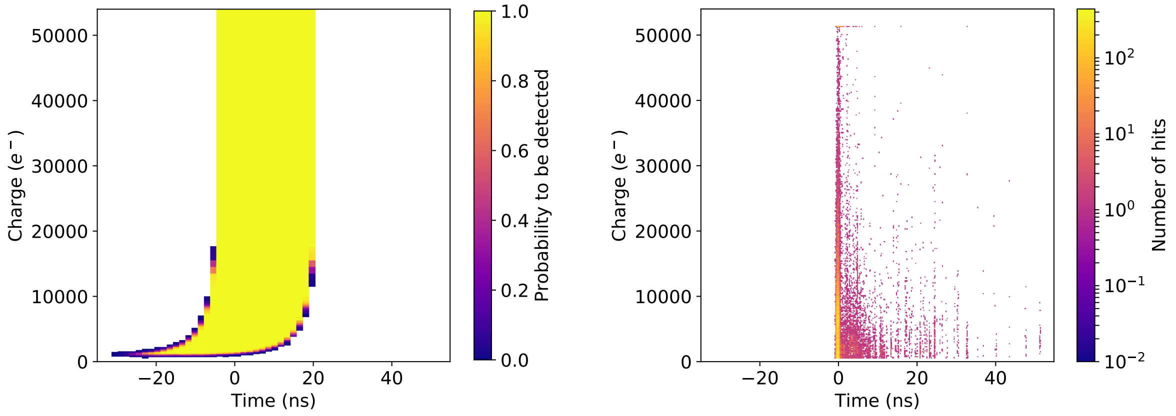
small disc, as well as of the first and last large disc (Figure 2.10). For each location about 2000 minimum bias QCD events, without pileup and without a transverse momentum cut, were simulated. The simulated pixel hits, corresponding to single pixels with deposited charge, were sorted by released charge, ranging from 600 e^{-}[‡] to 50 ke^{-}, with a granularity of 150 e^{-} and a time resolution of 0.25 ns. The simulated time of arrival versus charge distribution for the central module ($z = 0$) of the innermost layer of the IT barrel is presented in Figure 3.37(b).

Such a distribution was combined with the time response measurement introduced in the previous section. The x axis of the time response is reversed, obtaining the acceptance region, in time and charge, giving the probability of a charge to be detected in the correct BX. This way, instead of showing when a hit is detected by the electronics, the figure indicates when a hit has to occur to be detected in a given BX. The y axis has to be extended to match the charge range in the simulation. Assuming that the time response remains constant for very large signals, the yellow region with sharp edges is extended up to 50 ke^{-}. For illustration, the time response of the LIN AFE, after such modifications, is presented in Figure 3.37(a).

When the acceptance region of the front-end is superimposed with the hit distribution from simulation, as indicated in Figure 3.37(c), the hits that are inside the yellow part of the acceptance region have a 100% probability to be assigned to the correct BX. On the other hand, hits that are outside of the detection region have zero probability to be detected in time. Figure 3.37(d) shows the hits that will be assigned to a wrong BX, obtained from the exclusion of the two overlaid plots. The integral of the exclusion plot divided by the total number of hits gives the fraction of late-detected hits in a given location of the future detector.

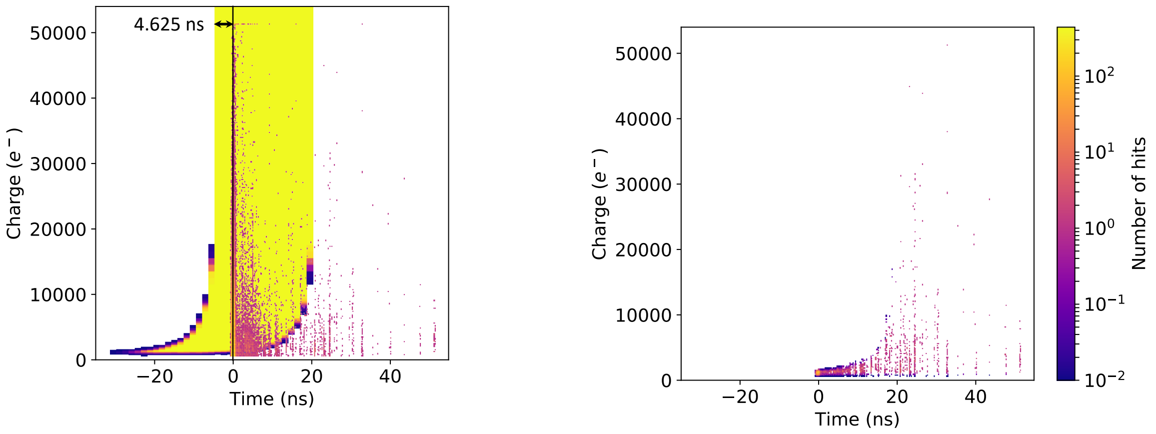
An important part of this method is the time alignment of the two overlapping plots.

[‡] Given that charges smaller than 600 e^{-} are not expected to be detected because of the threshold, the simulation started at this charge to avoid overloading the computing time.



(a) AFE acceptance region for a given BX.

(b) Time of arrival simulation.



(c) Overlay between measurement and simulation.

(d) Late-detected hits.

Figure 3.37: Different steps of the time response evaluation method.

The origin of the time axis of both the measurement and the simulation have to be correctly aligned. The $t = 0$ of the simulation corresponds to the time when the two proton bunches overlap in the interaction region, corrected with the expected time of flight from the interaction point to the given module. The zero of the chip acceptance can be shifted to maximise the overlap, as it would be done in the detector by calibration. For this measurement, the peak of the simulation is placed three fully efficient bins from the left edge of the acceptance region, i.e. 4.625 ns, as it is indicated in Figure 3.37(c). This estimate of about 5 ns was used to account for the imperfect time alignment in the detector due to the variations in the length of the electrical links, jitter and other contributions, adding also some margin.

3.8.3 Fraction of late hits

The method described above was used to evaluate the fraction of hits detected late by the three AFE designs. The result is shown in Figure 3.38 for the selected detector locations. The left half of the histogram corresponds to the IT barrel layers, numbered from the centre outwards L1 to L4. For each layer the study was done for two pixel modules, one at the edge (e) and the one in the centre (c) of the barrel. The fraction of late hits increases with the distance from the interaction point. The right half of the histogram is dedicated to the discs, numbered D1 to D12 with increasing distance from the interaction point. For each disc one module on the innermost (i) and one on the outermost (o) ring is presented. For any given disc the fraction of late detected hits is higher on the outer ring.

An ideal front-end with infinitely fast time response was also simulated and the fraction of hits detected late was estimated using the same method described above. Results are shown in grey in Figure 3.38 overlaid to the estimates of the actual analogue front-ends, because they represent the irreducible background. For the considered positions, this fraction is between 0.38% and 7.26%. These are hits generated by particles whose travel time up to the sensor is more than 25 ns longer than the minimum, for which the detector is tuned. The SYNC and DIFF AFE have similar performance, causing few percent of misassigned hits on top of the background. The DIFF is slightly faster. The LIN AFE instead is significantly slower, causing up to additional 11% of late hits in the detector, on top of the intrinsic worst case 7%.

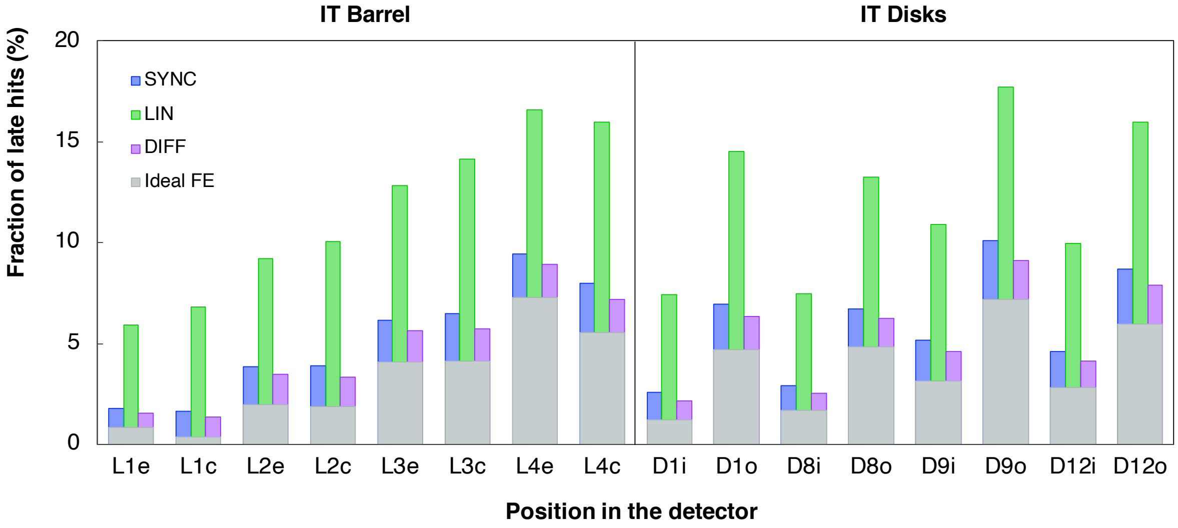


Figure 3.38: Fraction of hits detected late by the three RD53A AFEs for 16 pixel module positions. The IT barrel layers are numbered from the centre outwards L1 to L4 and for each layer the module at the edge is denoted “e” and the one in the centre is denoted “c”. The IT discs are numbered D1 to D12 with increasing distance from the interaction point and for each disc the innermost ring is denoted “i” and the outermost one is denoted “o”.

3.8.4 LIN FE slow time response mitigation

Following the outcome of the previous measurement, a modification of the discriminator circuit was proposed by the design team to improve the time response of the LIN AFE. The discriminator is composed of two stages: a transconductance stage and a transimpedance amplifier (TIA). In the TIA two diode-connected transistors, initially introduced to minimise the static current consumption at the output of the discriminator, were forcing other transistors to operate in the deep sub-threshold regime, consequently making them slower. A significant improvement in time walk at the cost of a marginal increase in static current consumption was achieved by removing those two transistors. This led to a simpler TIA stage in the new design of the LIN AFE [115], for the next version of the chip, called RD53B [89].

Circuit simulations were used to extract the time walk curves of both the original and the improved LIN AFE designs. They were transformed into time response plots and combined with the time of arrival simulations to estimate the fraction of late hits for the simulated designs. The simulated RD53A design was compared to the measurement and the difference in late hits is shown in Figure 3.39. The simulated AFE gives a slightly higher number of late hits. Nevertheless, the simulation demonstrates a very good agreement with the measurement, the difference in late hits being below 1%. This confirms the validity of the simulation, which can therefore be used to predict the fraction of late hits induced by the improved design. The difference in late hits between the original design (RD53A) and the new one (RD53B) is also shown in Figure 3.39. The new LIN

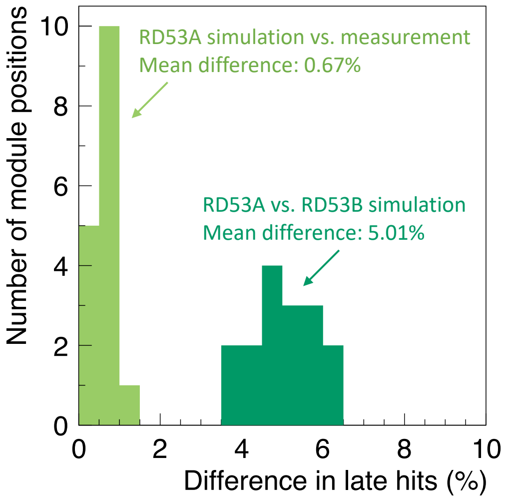


Figure 3.39: Difference between the fraction of hits detected late by the simulated and measured RD53A design of the LIN AFE (light green) and difference between the fraction of hits detected late by the simulated RD53A version of the LIN AFE and the improved RD53B version (dark green).

AFE demonstrates on average 5% less misassigned hits. The improved design of the LIN AFE was also implemented in a test chip and verified before and after irradiation. The simulation and measurement results after an irradiation up to 1 Grad show that the degradation of time walk after irradiation is minor [115].

3.8.5 Late-hit occupancy

The fraction of late-detected hits was converted to the occupancy due to late hits, using the simulated hit occupancies extracted from Figure 3.11. The result is shown in Figure 3.40 for all the positions in the detector. The irreducible background of misassigned hits is almost uniform in the tracker and amounts to between 10^{-5} and 10^{-4}, regardless of the AFE design. The late-hit occupancy levels are at least one order of magnitude above the required noise level, indicated by the red line in the figure. Hence the spurious hit rate in the detector is dominated by the time response of the AFE, not by the noise. Moreover, the performance of the improved design of the LIN AFE is comparable to the other two AFEs, although it remains slightly higher.

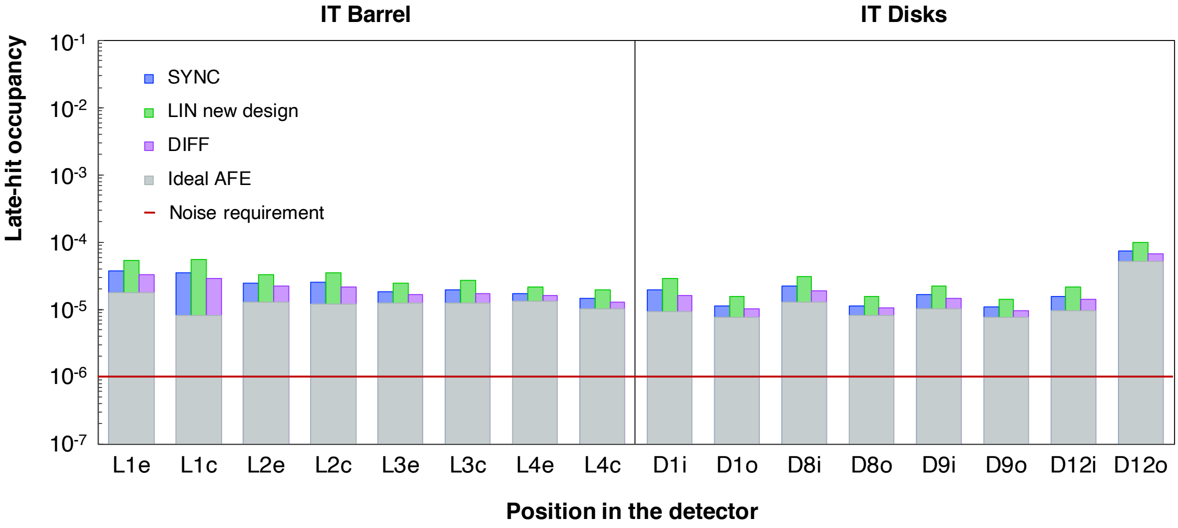


Figure 3.40: The occupancy due to hits detected late by the RD53A AFEs for different module positions in the detector. The IT barrel layers are numbered from the centre outwards L1 to L4 and for each layer the module at the edge is denoted “e” and the one in the centre is denoted “c”. The IT discs are numbered D1 to D12 with increasing distance from the interaction point and for each disc the innermost ring is denoted “i” and the outermost one is denoted “o”.

3.9 Conclusion

A new generation pixel readout chip is being designed for the upgrade of the CMS Inner Tracker to cope with stringent requirements imposed by unprecedented radiation levels and hit rates. Three different analogue front-ends were designed by the RD53 Collaboration and implemented in a large scale demonstrator chip, called RD53A. The three designs were characterised and the expected detector performance was evaluated against the requirements to choose the most suitable option for CMS.

The differential analogue front-end showed the best noise performance, with the noise occupancy several orders of magnitude below the requirement, as well as a very good time response. Nevertheless, this analogue front-end showed a problematic threshold tuning after irradiation, at cold temperature of $-10\,^{\circ}\mathrm{C}$ and with high leakage current. An improved design was proposed, expected to extend the operation with effective threshold tuning up to 500 Mrad, according to the simulation results. However, given the radiation levels expected in CMS, such an operation range would require several replacements of the innermost layer of the Inner Tracker barrel during the physics program. In addition, a saturation in the preamplifier feedback requires to operate at the limits of the time-over-threshold response in order to match the dead time requirement for the innermost layer of the Inner Tracker.

The synchronous analogue front-end features an automatic threshold tuning performed periodically by the auto-zeroing circuit and offers a very good timing performance. However, it appeared to be the noisiest of the three analogue front-ends. The noise increased for lower thresholds and fast preamplifier return to baseline, becoming critical for the operation settings of Layer 1.

The linear analogue front-end satisfied all the requirements, but featured a slower time response. However, an improved design was developed and is expected from simulation to reach a timing performance almost equivalent to the other two analogue front-ends. Since all the performance parameters of this analogue front-end were within CMS requirements and the main drawback was addressed and mitigated, the linear analogue front-end was identified as the lowest-risk option for the future pixel detector. CMS selected the linear analogue front-end with an improved design for the integration into the next version of the RD53 pixel chip for CMS, named C-ROC. A prototype is expected to become available in 2021.

Chapter 4

Sensor cross-talk

After the pixel readout chip and the associated choice of the AFE addressed in Chapter 3, another important system choice concerned the pixel sensor. The sensor is the part of the detector where the interaction with particles takes place, producing the signal by ionization, which is delivered to the readout electronics. The CMS Collaboration is considering two options for the geometry of the sensor pixels: either square pixels of $50 \times 50\,\mu\mathrm{m}^2$, exactly matching the pixels of the RD53 readout chip, or rectangular pixels of $100 \times 25\,\mu\mathrm{m}^2$, which can be connected to the square pixels of the readout chip thanks to electrode routing in the sensor.

The rectangular pixels show better resolution in simulations and their main advantage is to generate hits that require a lower bandwidth, mainly in the forward part of the IT barrel. On the other hand, the cross-talk effect observed in rectangular pixels may induce spurious hits and spoil the tracking resolution. Hence, methods to evaluate the amount of cross-talk is different pixel sensors were needed to help the choice of the pixel aspect ratio for the new detector.

This chapter starts with the introduction of a typical pixel sensor in Section 4.1, followed by the presentation of sensor properties and the discussion of different design options adopted by CMS. After the introduction of the cross-talk effect and the presentation of five CMS sensors selected for the cross-talk study, two methods to measure cross-talk are proposed and explained in Sections 4.2 and 4.3 together with the associated test results.

4.1 CMS pixel sensors

The basic building block of a pixel sensor is the reversely biased PN junction and the depletion zone constitutes the sensitive volume, where the particle ionization is detected. Some basic semiconductor physics, with emphasis on silicon, needed to understand the PN junction are reminded in Appendix C. This section describes how a simple PN junction is transformed into a pixel sensor. The sensor properties and different design options adopted by CMS are also explained.

4.1.1 Silicon pixel sensor

In a typical silicon sensor, the PN junction is usually realised by a shallow and highly-doped implant in a low-doped bulk material. In such configuration, the depletion region is reaching much deeper into the bulk of the sensor, as it is shown in Figure 4.1 for n^+-implant in p-bulk.

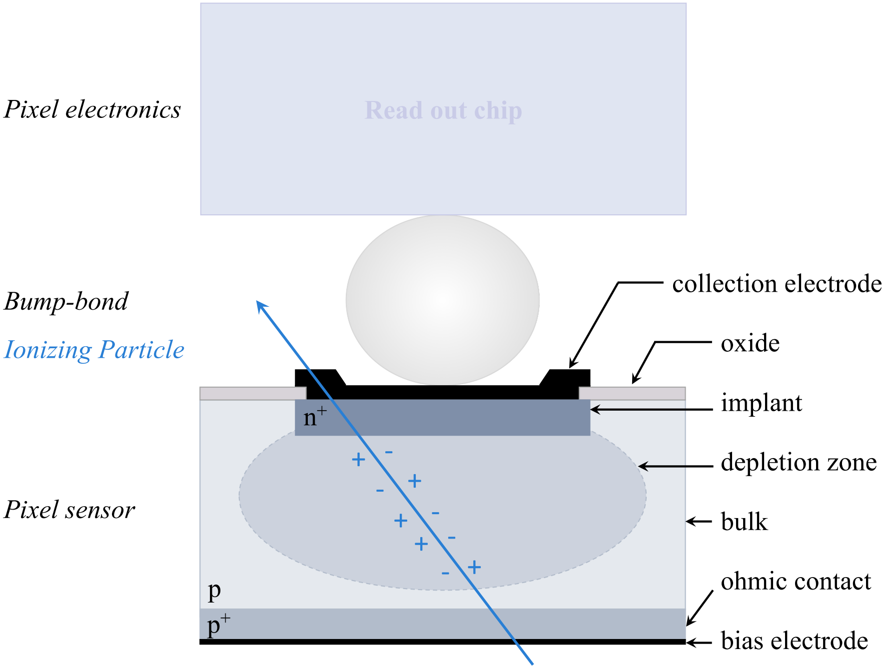


Figure 4.1: Cross-sectional view of a hybrid pixel detector showing different parts of a sensor pixel.

The highly-doped implant can be segmented in two directions to form pixels and to obtain a silicon pixel sensor able to measure the particle position in two dimensions. The individual pixel implant, separated by an insulating oxide, is connected to the readout electronics thanks to a metal *collection electrode* and a bump bond added during the flip-chip process [68, 69]. The implant connected to the readout chip is at the potential of the AFE preamplifier. An ohmic contact is created on the backside by a high doping and a metal layer for the *bias electrode*. When a negative bias voltage is applied to the bias electrode, the depletion zone starts to grow from the junction into the bulk. Since only the charge deposited in the depleted volume will be detected, the bias voltage is increased until the full sensor depletion is reached to maximise the signal. The width of the depletion zone can be approximated by:

$$W \approx \sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{e}\frac{V}{N}} \tag{4.1}$$

where $\varepsilon_0 = 8.89 \times 10^{12}\,\mathrm{F/m}$ is the permittivity of free space, ε_{Si} is the relative dielectric constant of silicon, N is the doping concentration of the bulk material and V the externally applied reverse bias voltage [68, p. 42]. Thus, the width of the depletion zone is proportional to the square root of the applied voltage. When the *depletion voltage* V_{depl} is reached the depletion region extends over the full thickness of the sensor. Beyond the full depletion the IV-curve displays a plateau region where the leakage current is increasing very little, as it is illustrated in Figure 4.2. If the applied bias is further increased the sensor is *over depleted* and at very high voltage values, the leakage current suddenly increases by several orders of magnitude and electrical *break down* occurs at the junction, which may destroy the sensor.

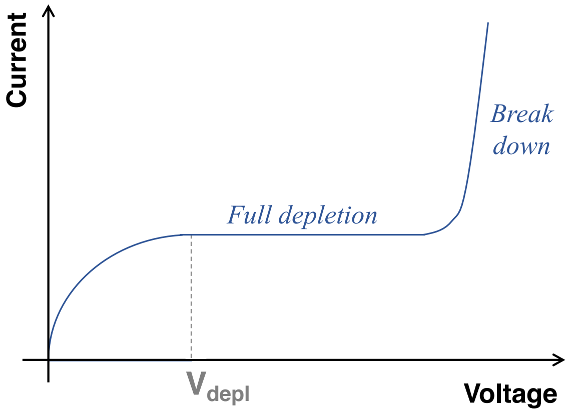


Figure 4.2: Typical IV-curve of a pixel sensor showing the depletion voltage, the full depletion plateau and the electrical break down.

4.1.2 Signal formation

An overview of the interactions of charged particles with matter, with fundamental processes at the origin of the signal used for particle detection, is proposed in Appendix D for completeness. In short, when a charged particle passes through the depleted silicon sensor, it creates electron-hole pairs by ionization. In the presence of the applied electric field, the electrons and holes are drifting towards the corresponding electrodes. The moving charge carriers induce electric current on the electrodes, according to the *Shockley-Ramo theorem* [116–118]. The amplitude of the current is proportional to the amount of charge carriers created by the passing particle. The signal starts when the charges start drifting and lasts until all the charges are collected at the electrodes. The collection time depends on the sensor type and is $\mathcal{O}(10\,\mathrm{ns})$ [68, 119]. In hybrid pixel detectors, only the segmented electrode is connected to the readout electronics, so only one type of carrier is collected. The p^+-type pixels collect holes and the resulting signal is a positive pulse. The n^+-type pixels, illustrated in Figure 4.1, that will be used for CMS Phase-2 Upgrade, collect electrons and the resulting signal is a negative pulse. The collection of electrons has an advantage, especially after irradiation, since the mobility of electrons is three times higher than the mobility of holes and so they are less prone to trapping (explained in Section 4.1.4).

4.1.3 Spatial resolution

The type of readout (analogue or digital) and the reconstruction algorithm play a role in the spatial resolution. In a fully binary readout, the spatial resolution of a pixel sensor is mainly determined by the size of the pixel pitch p and is given by [68, p. 62]:

$$\sigma_{position} = p/\sqrt{12} \tag{4.2}$$

If the information about the charge collected in each pixel is available, as it is the case with the TOT measurement in the RD53A chip, the spatial resolution can be further improved thanks to the *charge sharing* between neighbour pixels. For events in which the particle passes between pixels, the charge is distributed over several pixels and if the charge information is available the impinging position of the particle can be calculated with a higher precision using the *center-of-gravity* (COG) algorithm. The region of charge sharing shall be expanded as far as possible. The optimal spatial resolution is obtained when the number of one pixel clusters and two pixel clusters equals, i.e. when the charge sharing between two pixels occurs up to a quarter of the pixel pitch. The charge sharing is mostly dependent on the particle impinging position within the pixel and its incidence angle. In presence of a magnetic field, the drift path of the released charges is deflected by the so-called *Lorentz angle*, which also affects the charge sharing and hence the spatial resolution.

4.1.4 Radiation-induced effects

As silicon and other semiconductors are crystals, they are sensitive to radiation damage. The radiation damage can be divided into two categories: surface and bulk damage. The surface damage, occurring already at the doses of $\mathcal{O}(100\,\text{krad})$, causes the increase of the oxide charge and is mostly relevant for electronic devices. At higher fluences, the bulk damage becomes also important. High energetic particles do not only interact with atomic electrons forming the signal, but also with the nuclei, potentially displacing them out of the lattice position. In contrast to ionization, such interactions are usually not reversible and cause defects in the crystal lattice affecting the electric properties of the material. The main radiation-induced effects in silicon sensors are:

- **Increase of the leakage current:** Point defects in the crystal create deep energy levels in the band gap and act as generation–recombination centres. Therefore, more charge carriers are generated in the depletion zone after irradiation, which increases the leakage current, introduced in Appendix C.

- **Charge trapping:** Radiation-induced defects can also act as trapping centres, that are mostly unoccupied in the depletion region due to the lack of free charge carriers. They can hold or trap parts of the signal charge for a time longer than the charge collection time and therefore reduce the signal height. After a fluence of $\mathcal{O}(10^{15}\,\mathrm{n_{eq}/cm^2})$ only about 50% of the signal charge can be collected in a 300 µm silicon sensor [68, p. 78]. The charge trapping can be partially mitigated with a higher sensor bias voltage increasing the electric field.

- **Thermal runaway:** The combination of a higher leakage current and higher reverse bias voltage in an irradiated sensor causes an increased power dissipation and therefore a higher temperature. An increase of temperature causes an increase in the leakage current. This positive feedback loop quickly diverges, if an appropriate cooling is not provided.

- **Type inversion:** As radiation-induced defects can be charged, the bulk damage has an impact on the effective doping concentration. In fact, most bulk defects in irradiated silicon exhibit acceptor like behaviour and act as effective *p*-type doping. In a *p*-doped sensor bulk, this just increases the effective doping, while in an *n*-doped bulk the effective doping progressively converts from *n*-type to *p*-type. The type inversion occurs after a fluence of $\mathcal{O}(10^{12}\,\mathrm{n_{eq}/cm^2})$ and has an impact on the full depletion voltage and thus on the sensor operation [68, p. 72].

4.1.5 Sensor layout variations

n-in-p sensors

To create a silicon pixel sensor, *n*-type or *p*-type doping can be used for the substrate as well as for the implants, which results in four possible combinations. The CMS Phase-2 pixel detector will be equipped with n-in-p sensors, as the one detailed in Figure 4.1. The other three sensor types are described in Appendix E for completeness. The n-in-p sensors are the most recently developed type of sensor and are promising candidates to withstand the high fluence of the HL-LHC, given that they collect electrons with a higher mobility and hence less prone to trapping. In addition, these sensors do not undergo the type inversion and can be operated partially depleted during the whole lifetime. Moreover, the cheaper single-sided manufacturing process, makes them suitable for a large area silicon tracker. On the other hand, the main drawback specific to the n-in-p sensors is the need for spark protection. The guard rings are placed on the pixel side, which makes the sensor bias voltage propagate though the cut edge of the sensor to the vicinity of the readout chip. The resulting big voltage difference between the sensor and the readout chip, across the air space of about 20 µm, may cause destructive sparks [120]. The two standard options for spark protection are being evaluated by CMS: benzocyclobutene (BCB) polymer deposition and parylene coating [46].

Sensor thickness

Due to the high radiation environment of the HL-LHC thin sensors are preferred over thick ones. The initial advantage of a larger signal in thicker sensors disappears after irradiation, when severe charge trapping reduces the signal. In addition, thinner sensors need smaller operational voltages to reach full depletion and they produce less leakage current. For these reasons, the active thickness of pixel sensors for the Phase-2 Upgrade of the CMS IT was set to 150 µm [121, 122], which is about half the thickness used in the current CMS IT detector.

Planar vs. 3D sensors

The sensor described up to now, with shallow electrodes implanted on the top and bottom surfaces of the bulk, are denoted as *planar sensors*. The word "planar" refers to the electrode implantation depth compared to the thickness of the silicon substrate. Typically, electrodes are implanted with a depth of a few microns [123]. Thin planar n-in-p type silicon sensors, with an active thickness of 150 µm, will be used to build the new CMS IT. Different design options are being studied through several R&D submissions launched at *Hamamatsu Photonics K.K.* (HPK) [124] and *Fondazione Bruno Kessler* (FBK) Trento [125].

Planar sensors will be used throughout most of the IT detector, while *3D sensors* [126] are being considered for the innermost parts of the detector with the highest fluences. The electrodes of the 3D sensors are implemented as narrow columns penetrating the bulk perpendicularly to the sensor surface, as it is illustrated in Figure 4.3. In the 3D design, the distance between electrodes is decoupled from the sensor thickness. Thicker sensors can be used obtaining a larger signal, while maintaining small depletion voltages and efficient charge collection. In particular, the impact of charge trapping, which is a limiting factor for planar sensors at high fluences, is reduced in 3D sensors. Therefore, 3D sensors can in principle achieve higher radiation hardness than the planar sensors. Moreover, the lower operating voltage reduces the risk of edge sparking and lowers power consumption, making the 3D sensors less prone to thermal runaway issues. For these reasons, the 3D sensors are promising candidates for the very high radiation environment of the HL-LHC. However, the fabrication process is more complex and costly, hence they are only considered for the inner layers.

Different design options of 3D sensors are under investigation within CMS through R&D submissions at FBK Trento and *Centro Nacional de Microelectrónica* (CNM) Barcelona [127], using different processes. The FBK 3D sensors are fabricated using a single-sided process and the columns have less than 5 µm in diameter. The CNM 3D sensors use both single and double-sided process and the columns have a diameter of 8 µm. The diameter of the columns is important, because they induce an inefficiency when the particles arrive with normal incidence. This effect is mitigated at a tilt angle of few degrees.

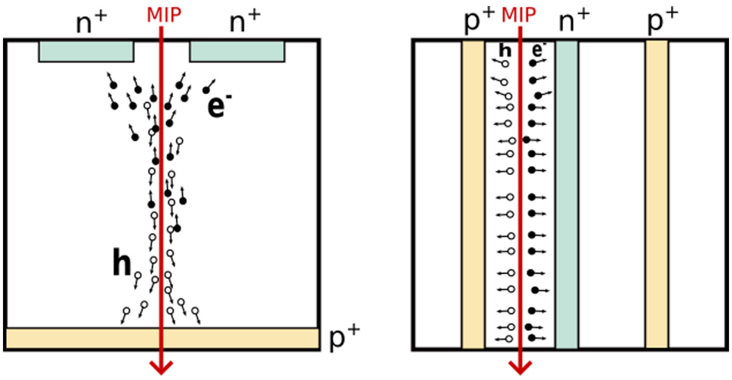


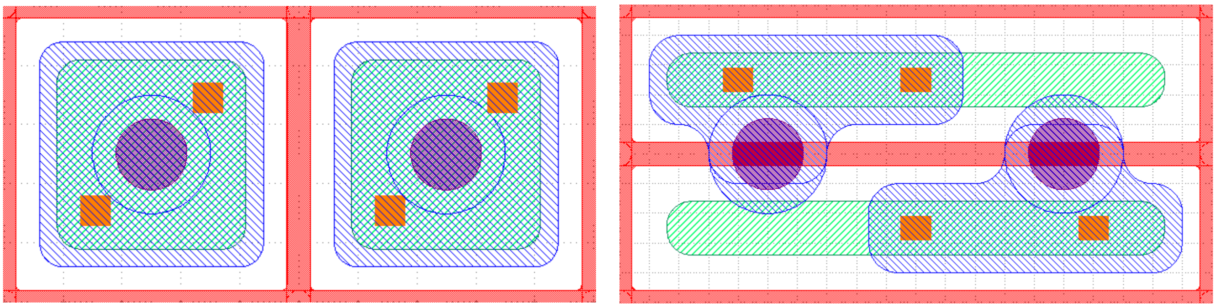
Figure 4.3: Comparison between a planar (left) and a 3D pixel sensor (right) [128].

The sensors from the different CMS submissions, bump bonded to RD53A chips, were evaluated in test beams before and after irradiation. The results obtained with both the planar [120, 121] and 3D sensors [128, 129] are very promising.

Pixel aspect ratio

The pixel size of the CMS IT will be reduced after the Phase-2 Upgrade by a factor six with respect to the current detector. Two different pixel geometries are considered for a total pixel area of 2500 μm^2: square pixels ($50 \times 50\,\mu m^2$) and rectangular pixels ($100 \times 25\,\mu m^2$). Both options, shown in Figure 4.4, are compatible with the bump-bonding pattern of the RD53 readout chip, thanks to electrode routing in the sensor.

A Monte Carlo study using the CMSSW framework [114] was performed with high transverse momentum single muon tracks to compare the impact of the pixel geometry on the track resolution. The spatial resolution was determined using the known position of the particle from the Monte Carlo truth. The simulation was done with 150 µm-thick sensors with both pixel geometries and at two detection thresholds: 1000 e^- and 2000 e^-. The result is shown in Figure 4.5. The track resolution in the transverse plane simulated as a function of the pseudorapidity is shown in Figure 4.5(a) and the resolution along the beam axis as a function of the pseudorapidity is shown in Figure 4.5(b). The resolution in the transverse plane is in general better for rectangular pixels (shown in red) since their short pitch is oriented in the r-ϕ direction. The resolution along z is in general better for



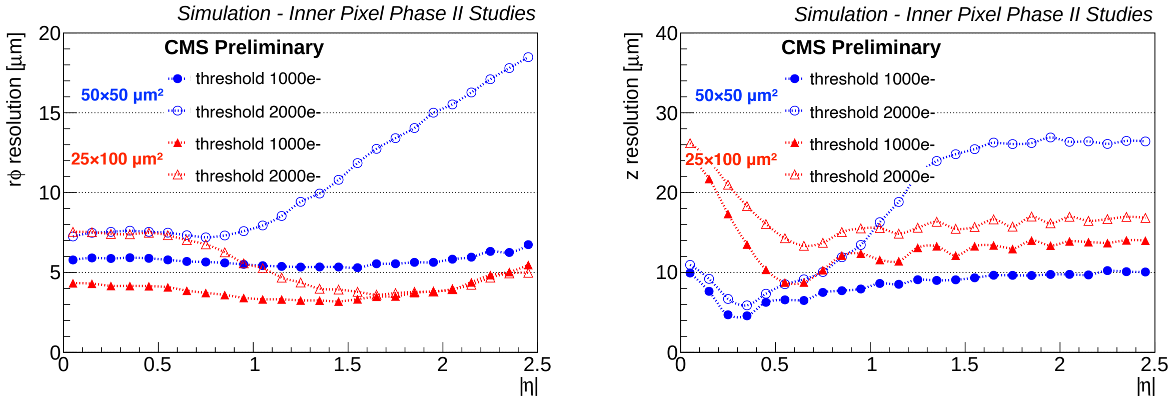
(a) Square pixels.

(b) Rectangular pixels.

Figure 4.4: Layout of two adjacent pixel cells for sensors from the HPK submission with $50 \times 50\,\mu\text{m}^2$ pixels (a) and $100 \times 25\,\mu\text{m}^2$ pixels (b). The n^+-implants are shown in green, the metal layers in blue, the p-stop isolation in red, the contacts in orange, and the bump bond pads in purple [46].

square pixels (shown in blue) given that they have a shorter pitch in this direction. However, the resolution of square pixels strongly deteriorates at large η at higher threshold. This is due to the breakage of long clusters due to the combination of large path lengths in the sensors at large η and a shorter pitch in the z direction in the square pixels. This study did not take into account the effects of radiation damage, but the higher threshold reduces the signal in a similar way as the charge trapping after irradiation. While strong quantitative statements can only be deduced when the reduced signal and increased noise after irradiation, as well as pileup and tracking in a dense environment are taken into account, this study supports the preference for rectangular pixels [130].

Another advantage of rectangular pixels is a reduced bandwidth, in particular in the forward part of the IT barrel, where the tracks arrive with a very shallow incidence angle,



(a)

(b)

Figure 4.5: Tracking resolution in $r - \phi$ (a) and in z (b) as a function of pseudorapidity, simulated with the square pixels (blue) and rectangular pixels (red) at two different thresholds: $1000\,\text{e}^-$ and $2000\,\text{e}^-$. Adapted from [130].

producing very long clusters. The number of pixels along the z direction in these long clusters is reduced to half with rectangular pixel thanks to their twice longer pitch in this direction. For these reasons, the rectangular pixels are currently the preferred option for the Phase-2 IT. However, the main drawback of the rectangular pixels is the observed cross-talk.

4.1.6 Cross-talk

Given the very small pixel size over a thin layer of silicon, the pixel capacitance becomes an important factor to consider. The pixel capacitance is important for two reasons. It represents a source of noise for the AFE electronics, but in particular it can cause cross-talk between neighbour pixels. The main contributions to the total pixel capacitance are:

- capacitance to the backside
- capacitance to the readout chip
- inter-pixel capacitance

Other contributions, as for instance the capacitance of the bump-bonds, are negligible and can be omitted. The capacitance to the backside and to the readout chip can both be calculated with the equation characterising a parallel plate capacitor:

$$C = \varepsilon_0 \varepsilon_{Si} \frac{A}{d} \tag{4.3}$$

where $\varepsilon_0 = 8.89 \times 10^{12}\,\mathrm{F/m}$ is the permittivity of free space, ε_{Si} is the relative dielectric constant of silicon, A is the pixel area and d is the distance to the back plane or to the readout chip [68, p. 57]. Given that the pixel size in the Phase-2 IT will be $2500\,\mu\mathrm{m}^2$ and the sensor thickness $150\,\mu\mathrm{m}$, the capacitance to the backside is about $1.7\,\mathrm{fF}$. Assuming a bump-bond height of $\mathcal{O}(20\,\mu\mathrm{m})$, the capacitance to the readout chip is of about $1.1\,\mathrm{fF}$.

The inter-pixel capacitance represents the dominant contribution to the total capacitance and it depends on the distance between pixel implants, as well as on the pixel perimeter. Hence, it is larger for rectangular pixels. For comparison, the perimeter of the square pixels considered by CMS is $200\,\mu\mathrm{m}$, while the perimeter of the rectangular ones is $250\,\mu\mathrm{m}$, which is 25% larger. The capacitance of rectangular pixels is dominated by the coupling between neighbours sharing the long edge.

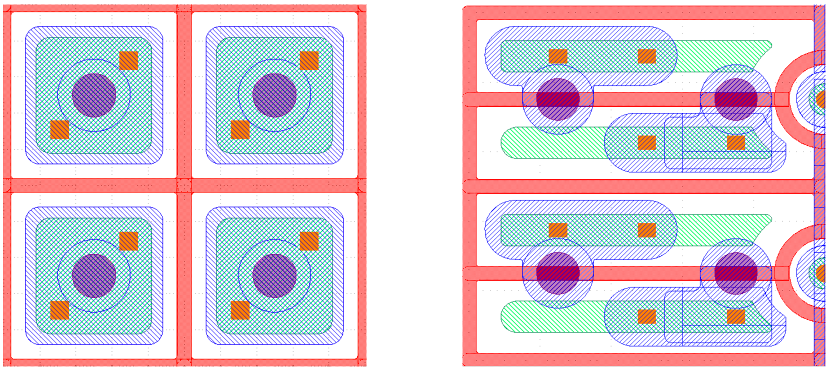
In addition to the pixel aspect ratio, the sensor layout plays an important role. An additional capacitive coupling is expected from the pairwise overlap between the metal collection electrodes shown in blue in Figure 4.4(b) and the pixel implants shown in green. This overlap is due to the routing of $100 \times 25\,\mu\mathrm{m}^2$ sensor pixels to the $50 \times 50\,\mu\mathrm{m}^2$ bump-bond pattern of the RD53A readout chip, and the required minimum bump-bond pad and implant sizes. As a consequence, a signal charge deposited in one pixel can induce signal on its neighbour(s). The worst case scenario would be the cross-talk with two neighbours, however since the overlap in the layout is present for pairs of pixels a cross-talk to only one neighbour is expected. A cross-talk to one neighbour of $\mathcal{O}(10\%)$ has been observed in the $100 \times 25\,\mu\text{m}^2$ pixel sensors bump bonded to RD53A chips [121]. Hence, the aim of this work was to develop methods to precisely measure cross-talk in different pixel sensors.

4.1.7 Sensors selected for cross-talk evaluation

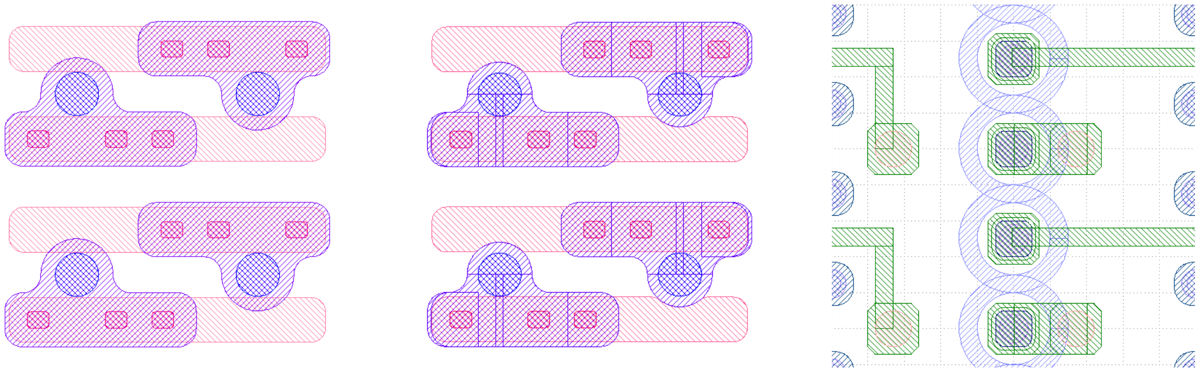
Five CMS sensors were selected for the cross-talk study and their layouts are presented in Figure 4.6. Figures 4.6(a) and 4.6(b) show two deep-diffused planar sensors fabricated at HPK. The pixel implants are represented in green, the metal collection electrodes in blue, the contacts in orange, the bump-bonds in violet and the *p*-stop insulation in red. Both HPK sensors are 200 µm thick with an active thickness of 170 µm. The HPK sensor shown in Figure 4.6(a) features square pixels and was used as a reference, since no asymmetric cross-talk related to the layout is expected in square pixels. The other HPK sensor shown in Figure 4.6(b) features rectangular pixels and a bias dot structure, placed in every other pixel column and row, visible on the right side of the layout.

Figures 4.6(c) and 4.6(d) show two other planar sensors fabricated at FBK. The pixel implants are represented in pink, the metal collection electrodes in violet, the contacts in red and the bump-bonds in blue. No pixel insulation or bias structure are present in these sensors. Both FBKs are 150 µm thick, but their active top side was directly bonded to a very low resistivity Czochralski wafer, which is known to affect the sensor doping and prevent the full depletion. The maximum thickness of the depleted volume is therefore smaller than 150 µm, resulting in a smaller signal than in the HPK sensors. The FBK sensor shown in Figure 4.6(c) has rectangular pixels and a standard layout. The other FBK sensor shown in Figure 4.6(d) was developed specifically to address the asymmetric cross-talk due to the overlap between the implant of one pixel and the collection electrode of the neighbouring pixel. The pixel implant of this sensor has a cut-out avoiding this overlap and is therefore referred to as the “bitten implant”.

A 3D sensor fabricated at CNM using the single-sided process was used for comparison and is shown in Figure 4.6(e). The CNM 3D sensor is 200 µm thick with an active thickness of 150 µm. The pixel implants are implemented as columns represented by the four blue squares in the centre of the layout, and the bias columns are represented as blue circles placed in each pixel corner. The columns have a diameter of 8 µm and a depth of 120 µm. The collection electrodes are shown in green and the bump-bonds are represented in orange.



(a) HPK square. (b) HPK rectangular.



(c) FBK standard. (d) FBK bitten. (e) CNM 3D.

Figure 4.6: Layouts of the CMS sensors selected for the cross-talk study showing four pixel cells.

4.2 Cross-talk measurement with injections

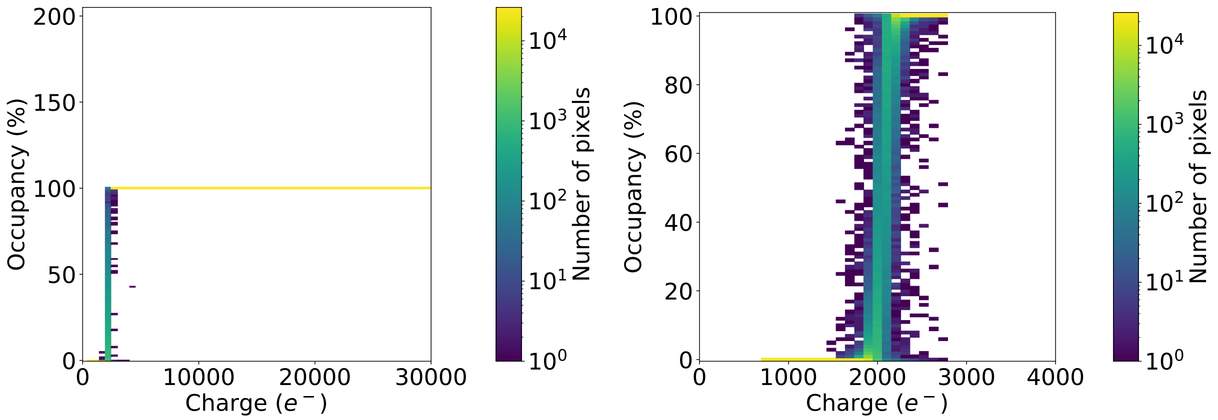
The first method to measure cross-talk was developed and integrated within the BDAQ53 test system (introduced in Chapter 3, Section 3.4.1). It uses the calibration injection circuit, implemented in the RD53A chip (presented in Chapter 3, Section 3.4.2). This method provides a quick estimation of the sensor cross-talk with no need for particles and therefore it can be performed easily in the lab.

4.2.1 Measurement method

The method is based on the S-curve measurement, i.e. the occupancy as a function of injected charge (explained in Chapter 3, Section 3.5.1). The full charge injection range of the RD53A can be used to verify if an injection of a very high charge triggers more than one pixel. Each charge is injected 100 times into each pixel of the pixel matrix. For every single injection all pixels are read out and the number of hits is counted. Without any cross-talk between neighbour pixels, each injection of a charge below threshold gives no hit and each charge above threshold gives exactly one hit, resulting in a standard S-curve plot. In presence of cross-talk, an injection of a high charge triggers the injected pixel plus its neighbour(s) and consequently the occupancy is higher than 100%.

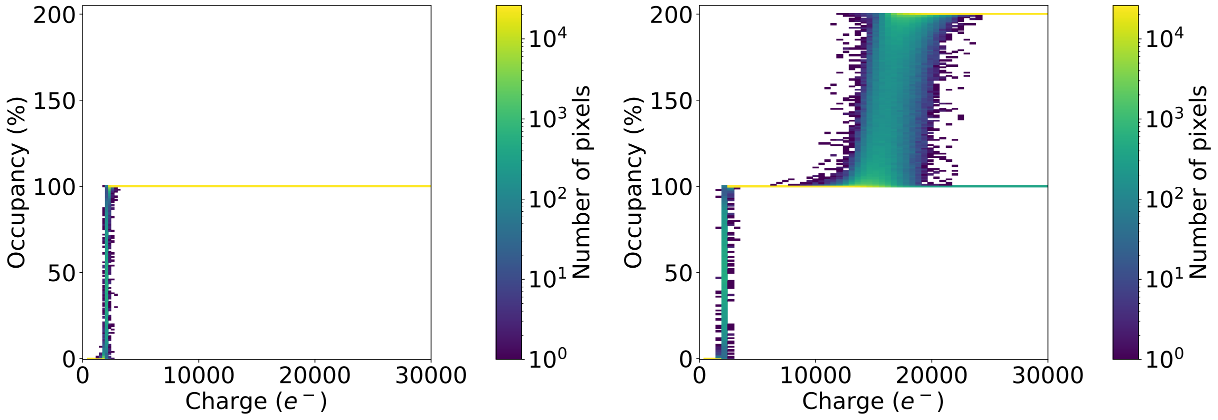
The cross-talk measurement with injections was performed with three different samples: a bare RD53A chip, i.e. with no connected sensor, an RD53A bump-bonded to an HLL sensor (used for the AFE evaluation presented in Chapter 3) with square pixels and another one connected to an HLL sensor with rectangular pixels. The result is shown in Figure 4.7, for charges up to 30 000 e$^-$. The measurement of the bare chip, presented in Figure 4.7(a), shows S-curves from 0 to 100% in the lower left part of the plot, as it is confirmed by the same plot with different scales shown in Figure 4.7(b). Once the threshold is reached, the occupancy stays at 100% for any higher charge. As expected any sign of cross-talk is observed in a bare chip. The same conclusion can be drawn from Figure 4.7(c), for a sensor with square pixels.

On the other hand, when the chip is connected to a sensor with rectangular pixels, double S-curves are observed, revealing the cross-talk in this type of sensors. In addition to the standard S-curves between 0 and 100%, a second set of S-curves appears between 100 and



(a) Bare chip.

(b) Bare chip zoom.



(c) Sensor with square pixels.

(d) Sensor with rectangular pixels.

Figure 4.7: Double S-curve measurement of cross-talk.

200%, which implies that each injection of a high charge triggers two pixels: the injected one and one neighbour. This demonstrates that the cross-talk in rectangular pixels is indeed asymmetric and occurring between pairs of sensor pixels, as expected from the sensor pixel layout, depicted in Figure 4.4(b).

The relative position of the lower and the upper S-curves can be used to calculate the amount of cross-talk. The mean charge at which 50% occupancy is reached, noted t in the following, corresponds to the threshold, as explained in Chapter 3, Section 3.5.1. The charge at which 150% occupancy is reached, noted s, represents the charge that needs to be injected to a pixel, such that the fraction of this charge transferred to the neighbour by cross-talk is above the neighbour's threshold. Thus, the cross-talk can be expressed as:

$$x = \frac{t}{t+s} \tag{4.4}$$

The pixel-to-pixel variation of the threshold induces a certain spread on the S-curves. The two variables t and s can be measured with a precision of Δt and Δs respectively, corresponding to the RMS. Making the assumption that the two measurements are uncorrelated, the standard uncertainty propagation results in an uncertainty on the cross-talk value given by:

$$\Delta x = \frac{\sqrt{(s\Delta t)^2 + (t\Delta s)^2}}{(t+s)^2} \tag{4.5}$$

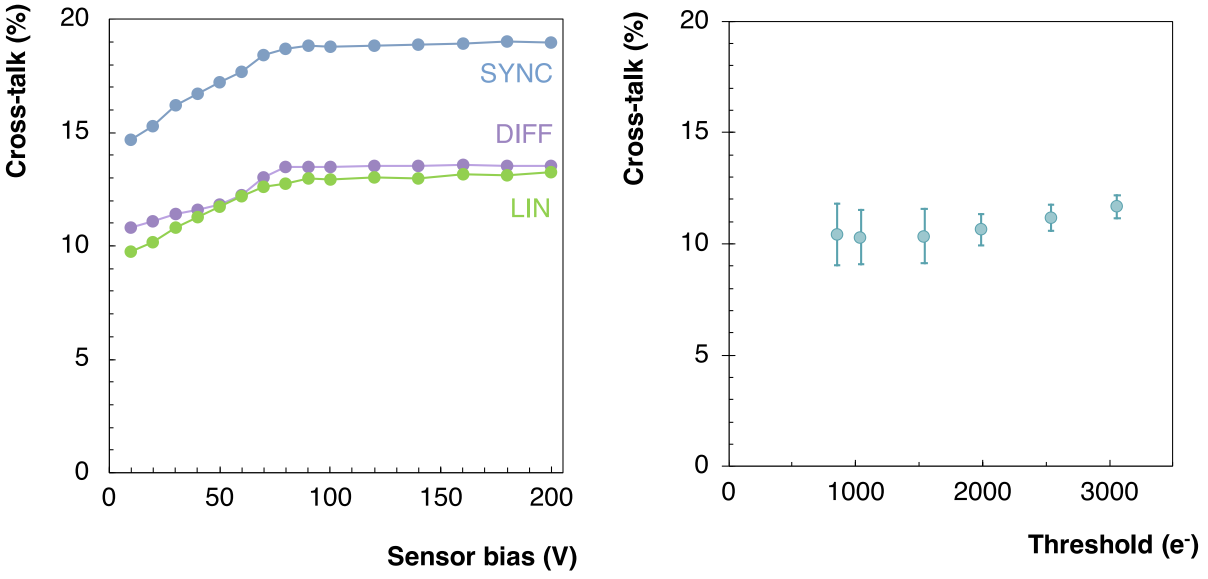
For each measurement with this method the latter formula was used for the error bars. The assembly of the RD53A chip with the HLL sensor with rectangular pixels was used to study the influence of different parameters on the asymmetric cross-talk and the results are presented in the following sections.

4.2.2 Influence of the sensor bias

First, the influence of the sensor bias voltage on the cross-talk was verified, given that it is the main operational parameter of a sensor. In the n-in-p type sensors, adopted by CMS for the Phase-2 Upgrade, the applied bias voltage is negative, but for simplicity the bias voltage is quoted in absolute values from now on. The cross-talk measurement with injections was performed for different sensor bias voltages, ranging from 10 V to 200 V and the result is shown in Figure 4.8(a).

The cross-talk shows a dependence on the sensor bias, pointing out that the charge transfer from one pixel to the neighbour depends on the size of the depletion zone. The cross-talk increases with the sensor bias until the full depletion is reached. Above the depletion voltage the cross-talk reaches a plateau. It is important to measure cross-talk with fully depleted sensors, to obtain comparable results. For all the following cross-talk measurements with the HLL sensor, the sensor bias was set to 120 V.

Moreover, the cross-talk as a function of the sensor bias measurement was done with each of the three AFEs implemented in the RD53A chip. The LIN and the DIFF AFE give very similar results, while the SYNC AFE shows about 4% higher cross-talk. Although the calibration injection circuit is the same for all three AFEs, their different design has an influence on the input impedance of the AFE PA and therefore on the cross-talk. All the following results presented in this chapter were obtained the LIN AFE, chosen for the CMS IT upgrade.



(a)

(b)

Figure 4.8: Cross-talk measured with injections as a function of the absolute value of the sensor bias voltage, obtained with each of the three AFEs in the RD53A chip (a) and cross-talk as a function of the threshold measured with the LIN AFE in the RD53A chip (b).

4.2.3 Influence of the threshold

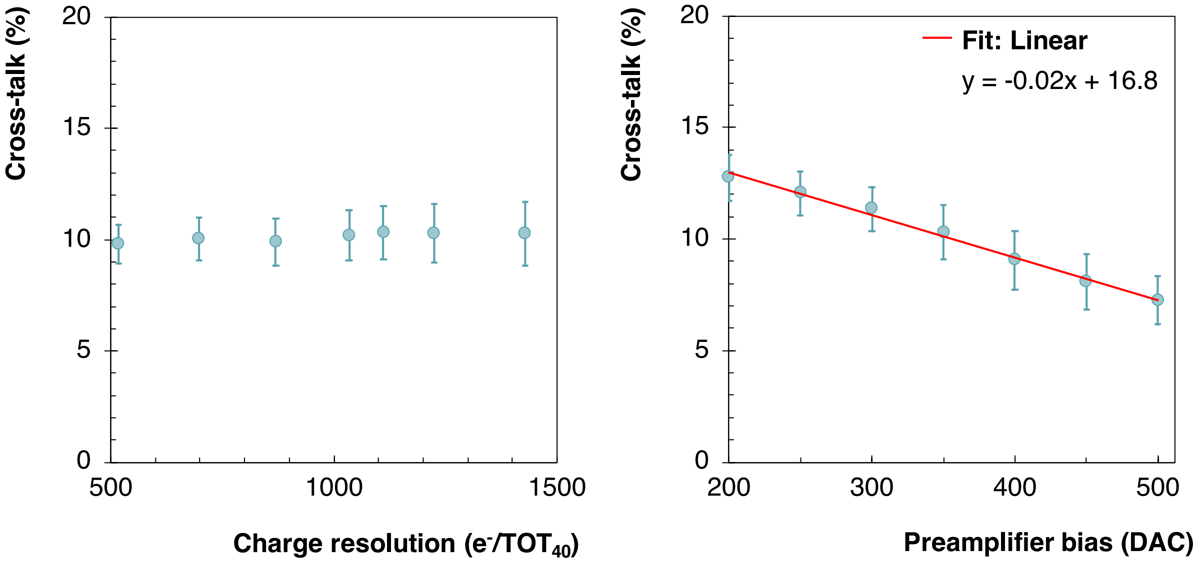
The cross-talk in a pixel detector manifest by inducing spurious hits, which enlarges the clusters and spoils the hit resolution. The detection threshold plays an important role in the observation of the cross-talk. At low threshold, the fraction of charge transferred to the neighbour pixel by cross-talk have a bigger probability to be above threshold and to cause a fake hit. At higher thresholds, this probability is reduced and the amount of fake hits is lower, although the cross-talk itself remains the same.

In principle, the amount of cross-talk measured with the injection method should not change with the threshold. In fact, setting the threshold to a higher value shifts both the lower and the upper S-curves to higher charges, but their ratio remains the same. Hence, the measured cross-talk should not change with threshold, as far as the upper

S-curve remains within the injection range of the chip, such that it can be evaluated. This assumption was verified, measuring the cross-talk for different thresholds, ranging from 800 to 3000 e^-. The result is shown in Figure 4.8(b). As expected, the change in cross-talk with increasing threshold is negligible and the cross-talk can be considered as constant within the error bars.

4.2.4 Influence of the charge resolution

Besides the threshold, another important detector parameter is the charge resolution, i.e. the calibration of the TOT. The charge resolution is expressed in electrons per TOT_{40} unit, corresponding to 25 ns. The baseline TOT calibration used in this thesis was set to 1100 e^-/TOT_{40}. More details about the charge calibration of the RD53A chip are given in Chapter 3, Section 3.7. The influence of the charge resolution on cross-talk was measured with injections and the result is shown in Figure 4.9(a). No dependence of cross-talk on the charge calibration was observed.



(a) (b)

Figure 4.9: Cross-talk measured with injections as a function of the charge resolution (a) and the preamplifier bias (b) obtained with the LIN AFE in the RD53A chip.

4.2.5 Influence of the preamplifier bias

Finally, the influence of the PA bias on cross-talk was investigated. The PA bias is a parameter of the AFE that sets the current of the PA and represents the main contribution to the AFE power consumption, as it was explained in Chapter 3, Section 3.2.2. In general, increasing the PA bias current has a positive impact on the performance, for instance it decreases the noise (Chapter 3, Section 3.6.5). The cross-talk measured with injections as a function of the PA bias is shown in Figure 4.9(b). The PA bias DAC ranges from 0 to 512 and the default value is 350. The study shows that the cross-talk decreases linearly with increasing PA bias. The linear fit of the data was performed to obtain the dependence equation, also shown in the plot. From this equation, it can be concluded that the cross-talk can be decreased by about 1% when the PA bias is increased by 50. Therefore, the PA bias represents a handle to reduce cross-talk, however, at a price of an increase in the chip power consumption. The relation between the PA bias and the power consumption was shown in Chapter 3, Section 3.4.5. Based on this study, an increase of the PA bias DAC by 50 from the default value, would imply an increase of the per pixel current consumption by about 9%. It was also observed that the dispersion of the upper S-curves increases with higher PA bias, however, it does not affect significantly the uncertainty of the cross-talk.

4.2.6 Cross-talk of the selected sensors

The cross-talk measurement method with injections, bench-marked with the HLL sensors because of their great availability, was applied to the five CMS sensors selected for the cross-talk study, described in Section 4.1.7. The operation parameters of the tested sensors are summarised in Table 4.1. All the sensors were operated at room temperature and at full depletion. The bias voltage of the planar sensors was set to 120 V, while for the 3D sensor 20 V were sufficient to reach the full depletion.

The HPK sensor with rectangular pixels and the CNM 3D sensor were tested with the default PA bias of 350, while the other three sensors were tested with a higher PA bias set to 410. Based on the equation obtained from the fit in Figure 4.9(b), and under the assumption that the cross-talk dependency on the PA bias is the same for all sensors, this increase of the PA bias from 350 to 410 should decrease the cross-talk by 1.2%. Hence, the FBK sensors are expected to show a lower cross-talk in the following measurements than the HPK sensor with rectangular pixels.

As far as the charge calibration is concerned, all the sensors with rectangular pixels were tested with the Krummenacher current (denoted KRUM and driving the TOT calibration) set to 29, while the sensor with square pixels was tested with a lower Krummenacher current of 20. The fact that the Krummenacher current is set to the same value does not imply the same charge calibration from one chip to another. However, as it was shown previously the charge calibration has a negligible impact on cross-talk, so these variations can be neglected.

The thresholds at which the cross-talk measurement with injections was performed for different sensors are also indicated in Table 4.1 together with the corresponding threshold dispersion. The five selected sensors were measured at different thresholds, however as

Table 4.1: Parameters used for the cross-talk measurement with injections.

Tested sensor	**Sensor bias (V)**	**PA bias (DAC)**	**KRUM (DAC)**	**Threshold (e$^-$)**	**Upper S-curve (e$^-$)**	**Cross-talk (%)**
HPK square	120	410	20	1020	> 37578	$< 2.6\%$
HPK rectangular	120	350	29	2259 ± 79	13390 ± 899	$(14.40 \pm 0.94)\%$
FBK standard	120	410	29	1344 ± 41	17239 ± 1572	$(7.20 \pm 0.65)\%$
FBK bitten	120	410	29	1243 ± 43	21231 ± 1969	$(5.50 \pm 0.52)\%$
CNM 3D	20	350	29	918 ± 94	18142 ± 1948	$(4.80 \pm 0.68)\%$

it was concluded previously the threshold has a negligible impact on the cross-talk. The mean charge at which 150% occupancy was reached, i.e. the mean position of the upper S-curves, is also given in Table 4.1 for all five selected sensors together with the RMS indicating the dispersion of the upper S-curves. The cross-talk calculated from the threshold and the mean position of the upper S-curves according to Equation (4.4) is given for all selected sensors in Table 4.1 and presented in Figure 4.10 for a better comparison.

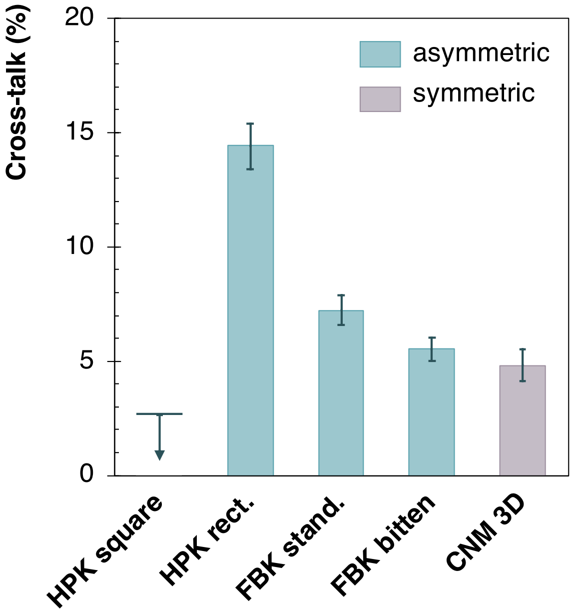


Figure 4.10: Cross-talk measurement with injections for the five selected sensors.

The HPK sensor with square pixels was measured at a threshold of $1020\,e^-$ and the threshold dispersion for this measurement was lost from the data. This sensor didn't show any upper S-curve within the injection range of the RD53A chip, i.e. almost up to $38\,000\,e^-$. This gives an upper limit on the value of the cross-talk in square pixels equal to 2.6%. This upper limit is represented in Figure 4.10 by a bar and an arrow. The HPK sensor with rectangular pixels shows the highest cross-talk among the selected sensors equal to 14.4%. The FBK sensors have a lower cross-talk than the HPK with rectangular pixels, which is partially due to the fact that they were measured with a higher PA bias.

The cross-talk of the FBK with bitten implant equal to 5.5% is lower than the cross-talk of the FBK with standard implant equal to 7.2%. This result shows that the bitten implant layout allows to reduce the cross-talk by almost 2% keeping the same settings.

And finally, given the different design of the CNM 3D sensor, the cross-talk was expected to be very low. At the threshold of 1000 e^- the upper S-curves were beyond the standard injection range of the RD53A chip and the cross-talk could not be evaluated. The threshold was lowered to about 900 e^- and the injection range was extended to the maximum. In this configuration, the cross-talk could be measured and the double S-curve measurement of the CNM 3D sensor is shown in Figure 4.11. The upper S-curves are between the occupancy of 100% and 300%, which indicates that a high charge injected into one pixel triggers the injected pixel plus two neighbours. This reveals that the cross-talk in the 3D sensor is actually symmetric and is therefore indicated in different colour in Figure 4.10.

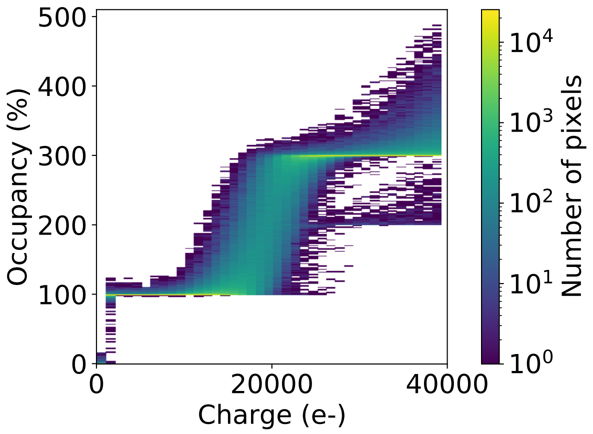


Figure 4.11: Double S-curve measurement of the CNM 3D sensor.

4.3 Cross-talk measurement with a particle beam

The second method to measure the sensor cross-talk is using particles provided by a test beam experiment. The evidence of cross-talk effect in pixel sensors was demonstrated in several previous test beam measurements, however these measurements did not allow to quantify the amount of cross-talk. Therefore, a new beam test method was established and put in place.

4.3.1 Measurement method

The method measures the asymmetric cross-talk occurring between pairs of sensor pixels. This effect is the dominant one, as it was inferred from the cross-talk measurement with injections detailed in the previous section. Let's consider the sensor pixel layout in Figure 4.12, showing one pixel column and four pixel rows. From the cross-talk point of view, this represents two pixel pairs: the pixel in row 0 is coupled to the pixel in row 1 and the pixel in row 2 is coupled to one in row 3. No measurable cross-talk is supposed to occur between rows 1 and 2. For each event, the pixel hit by the particle is denoted as *pixel 0*, its lower neighbour as *pixel* −1 and its upper neighbour as *pixel* +1, as indicated in Figure 4.12.

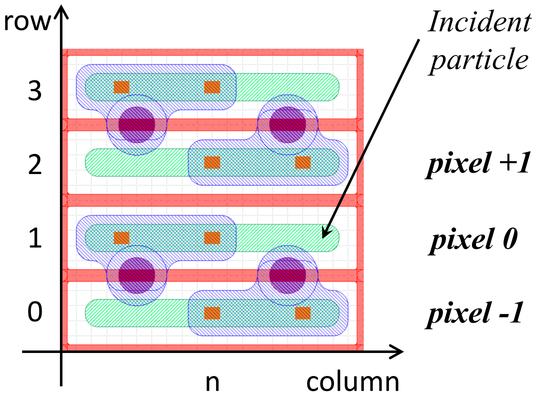
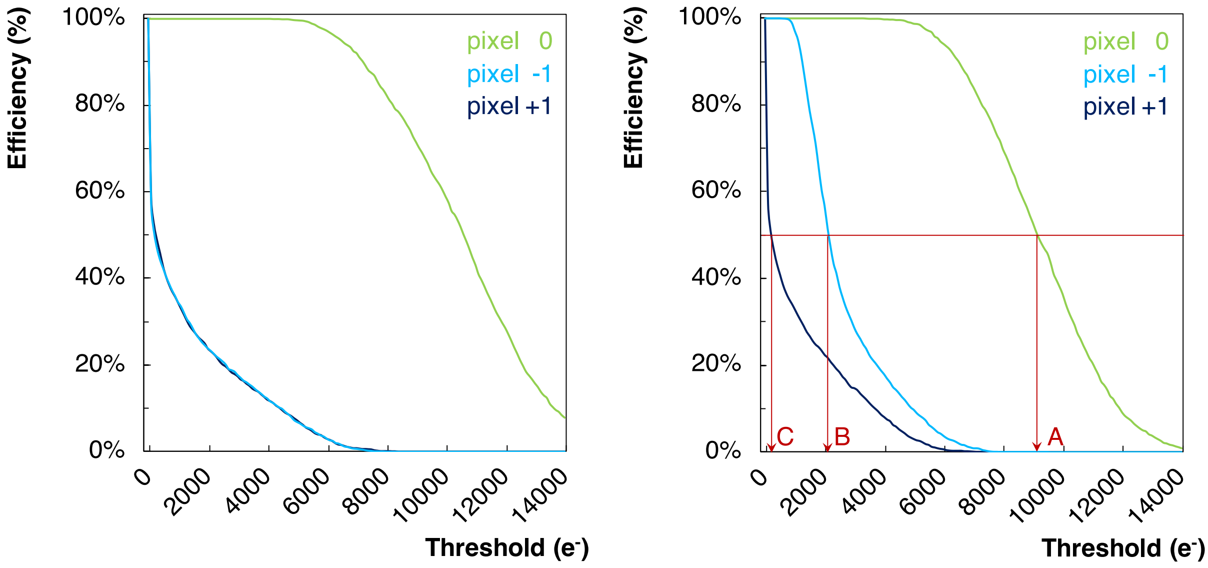


Figure 4.12: Illustration of the pixel assignment used for the cross-talk measurement method with the test beam.

The cross-talk measurement consists in performing a threshold scan in the particle beam. The goal is to extract the efficiency as a function of threshold, separately for pixel 0, +1 and −1. The corresponding plot, obtained from a simplified Monte Carlo simulation is shown in Figure 4.13. In this Monte Carlo, the deposited charge was distributed following a Landau distribution (introduced in Appendix D) with the MPV calculated based on the sensor thickness. A fixed amount of charge sharing and cross-talk were added to the simulation. The efficiency of the pixel 0, shown in green decreases from 100 to 0% with increasing threshold. If the charge transfer from pixel 0 to the neighbour pixels is only due to the charge diffusion or to a symmetric cross-talk, the charge transferred to both neighbours is on average the same resulting in the same efficiency curve, as it is shown in Figure 4.13(a). In presence of asymmetric cross-talk, pixel −1 receives additional charge from pixel 0. Thus pixel −1 collects more charge than pixel +1 and appears as more efficient. The efficiency curves of the two neighbours split, as it is shown in Figure 4.13(b), obtained from the simplified Monte Carlo simulation including 15% cross-talk.

The threshold at which a given efficiency is reached by each of the three pixels can be used to calculate the amount of cross-talk. For the sake of clarity, the variables introduced for this calculation are summarised in Table 4.2. The charge measured at 50% efficiency is denoted A, B and C for pixel 0, −1 and +1 respectively, as indicated in Figure 4.13(b). The charge deposited by the incident particle is denoted R_0 for pixel 0 and R_1 for each of the two neighbours, since the diffusion should be symmetric (on average). The charge



(a) No asymmetric cross-talk.

(b) With 15% asymmetric cross-talk.

Figure 4.13: Simplified Monte Carlo simulation demonstrating test beam measurement method of asymmetric cross-talk.

collected in each pixel is calculated from the deposited charge, taking into account charge transfers by cross-talk, noted x in the following.

Table 4.2: Variables used to calculate cross-talk from the measurement with a particle beam.

Considered pixel	Measured charge	Deposited charge	Collected charge
pixel 0	A	R_0	$(1-x)R_0 + xR_1$
pixel -1	B	R_1	$(1-x)R_1 + xR_0$
pixel +1	C	R_1	$(1-x)R_1$

The pixel 0 gives a fraction x of its charge R_0 to pixel -1, so it decreases its charge to $(1-x)R_0$. At the same time, it receives a fraction x of the charge in pixel -1, so its charge increases by xR_1. Hence the charge collected in pixel 0 is equal to $(1-x)R_0 + xR_1$. The pixel -1 transferred a fraction x of its charge R_1 to pixel 0, which decreased its charge to $(1-x)R_1$, but it also receives xR_0 charge from pixel 0. So the charge measured in pixel -1 is equal to $(1-x)R_1 + xR_0$. And finally the pixel +1 gives xR_1 fraction of its charge to its cross-talk pair pixel +2 and does not receive any charge back, since no charge was deposited in that pixel. It also does not receive any charge from pixel 0 given that they are not coupled. Hence the charge measured in pixel +1 is simply $(1-x)R_1$.

This leads to the establishment of a system of three equations, each one representing the measured charge in one pixel:

$$\begin{aligned} A &= (1-x)R_0 + xR_1 && (4.6) \\ B &= (1-x)R_1 + xR_0 && (4.7) \\ C &= (1-x)R_1 && (4.8) \end{aligned}$$

The three equations can be combined to eliminate the two unknown variables R_0 and R_1, and one equation of the second order in x is obtained as a function of the three measured quantities A, B and C. One of the two possible solutions to the second order equation would imply negative deposited charges and therefore the only physical solution is given by:

$$x = \frac{A + 2B - 2C - \sqrt{A^2 - 4BC + 4C^2}}{2(A+B)} \tag{4.9}$$

Inserting the charges A, B and C measured with the test beam into the Equation (4.9) allows to extract the amount of cross-talk in rectangular pixels.

4.3.2 Experimental set-up

DESY test beam facility

The data used to apply the above described method were taken at the *Deutsches Elektronen-Synchrotron* (DESY) [131] test beam facility in Hamburg in October 2019 and September 2020. The DESY II synchrotron, with a circumference of 292.8 m, supplies particle beam to three independent beam lines, called TB21, TB22 and TB24. The beams are not extracted directly from the accelerator, but are generated by a double conversion, as illustrated in Figure 4.14. The synchrotron is equipped with a movable carbon fibre target. When this primary target is placed in the beam, bremsstrahlung photons are generated. A secondary metal target is used to convert these photons into electron/positron pairs and a dipole magnet spreads the beam out horizontally. Then a set of collimators form either an electron or a positron beam with selectable momenta in the range from 1 to 6 GeV/c [132].

A reference detection device called a *telescope* is necessary to test a detector with a test beam. A pixel telescope is an experimental device, providing the position of the particles impinging on the detector, usually composed of several planes, equipped with pixel detectors and placed perpendicularly to the particle beam. The telescope allows to detect particle trajectories and provide the tracks. The DESY test beam facility is equipped with two pixel telescopes, which allow to track the beam particles. These high-precision pixel telescopes, developed within the EUDET project [134], offer an easy integration of the *device under test* (DUT) and a precise spatial resolution. Out of the seven EUDET-type beam telescopes existing worldwide, two are located at DESY: the DATURA telescope is installed on TB21 and the DURANTA on TB22.

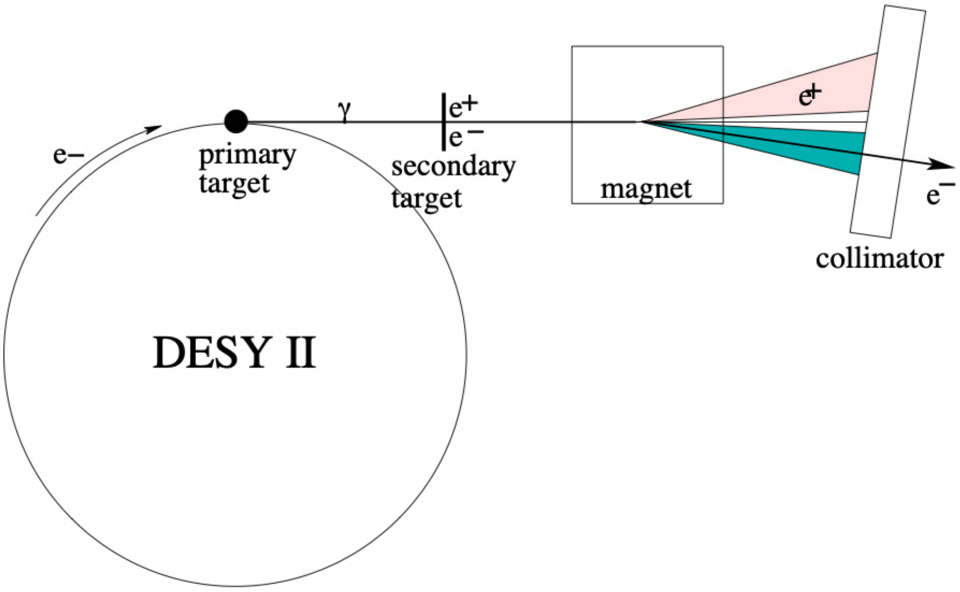


Figure 4.14: Test beam generation principle at DESY [133].

DATURA telescope

The results presented in this thesis were obtained from data taken with the *DESY Advanced Telescope Using Readout Acceleration* (DATURA) telescope. It is composed of two telescope arms with three detection planes each. The DUT is inserted between the two arms, as shown in Figure 4.15. The three telescope planes mounted before the DUT are referred to as the *upstream arm* and the three planes traversed by the beam only after the DUT are referred to as the *downstream arm*.

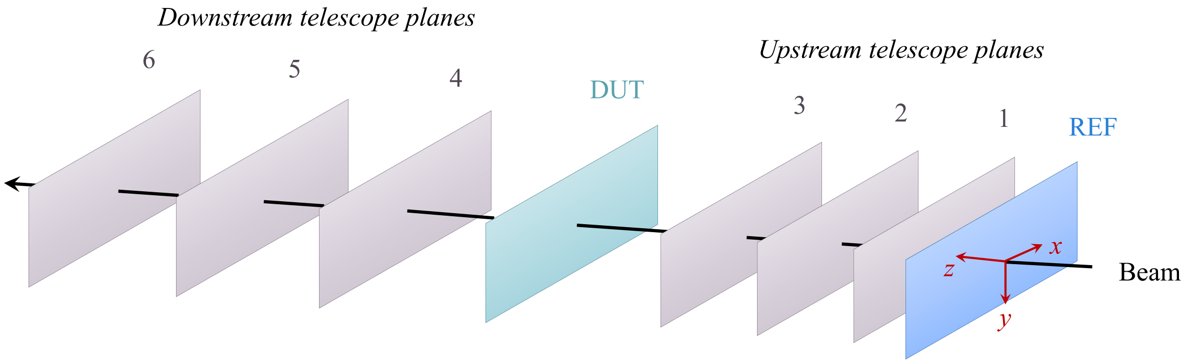
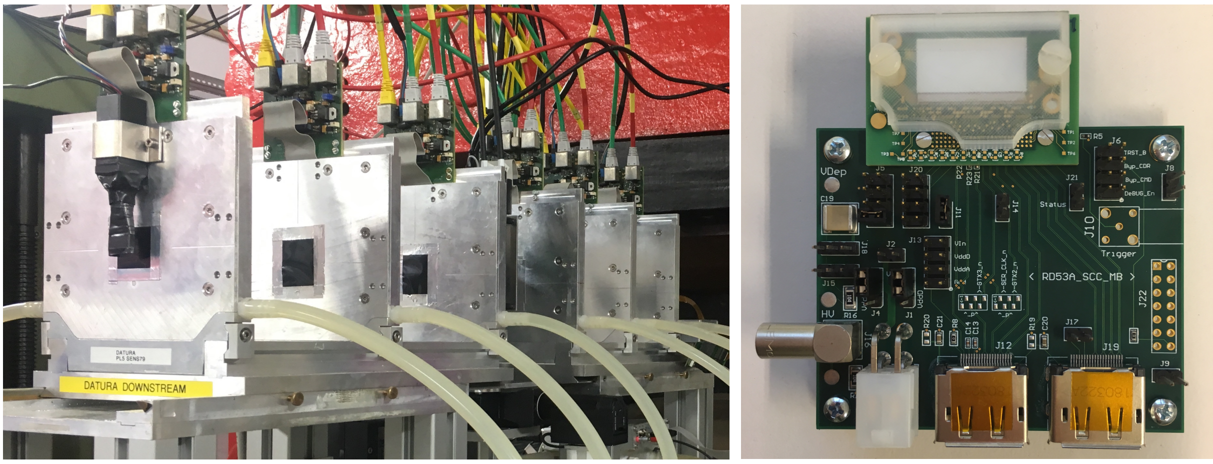


Figure 4.15: DATURA beam telescope configuration.

Each of the six telescope planes is equipped with a *minimum ionizing monolithic sensor array* (MIMOSA26) [135, 136] detector. The active area of the sensors is 21.2 mm × 10.6 mm, which is comparable to the size of the RD53A chip. The pixel matrix is composed of 1152 × 576 square pixels with a 18.4 µm pixel pitch. The MIMOSA26 sensor is 50 µm thick and is shielded from environmental light with a 25 µm-thick Kapton foil on each side. The resulting very low material budget of the telescope in combination with the fine pixel pitch allows to achieve an excellent track resolution, even at the rather low beam energies. The intrinsic resolution of a MIMOSA26 sensor was measured to be $(3.24 \pm 0.09)\,\mu\text{m}$ [137]. The track resolution depends on the beam momentum, the plane spacing and the material budget of the DUT. The best track resolution of the DATURA telescope without any DUT, using $5\,\text{GeV}/c$ beam and an equidistant plane spacing of 20 mm, was estimated to $(1.83 \pm 0.03)\,\mu\text{m}$ [137].

The DATURA beam telescope is shown in Figure 4.16(a). In this picture, the six detection planes are visible and the beam arrives from right to left. Each MIMOSA26 detector is embedded in an aluminium cooling jig, keeping the sensors at a constant temperature of about 18 °C. The zero-suppressed hit data from the sensors are transmitted over a ribbon cable to the auxiliary boards, visible on the top of the jigs, which establish the connection to the data concentrator board collecting the data from all six planes.



(a) The DATURA beam telescope.

(b) The DUT.

Figure 4.16: Photographs of the test beam set-up.

Device under test and time reference module

The DUT for the cross-talk measurement was composed of one of the sensors to be tested, presented in Section 4.1.7, bump-bonded to an RD53A chip, presented in detail in Chapter 3. The sensor size and pixel matrix were matching those of the RD53A chip. The assembly was mounted on a test card specially developed for test beam measurements. As it can be observed in Figure 4.16(b), the sensor is placed on a small and light-weight card with minimum material, called irradiation card, since it can be used for sensor irradiation. The irradiation card has a cut-out window behind the sensor to limit the material placed in the beam. A plastic cover, visible in the picture, protects the sensor and in particular the delicate wire-bonds. For the test beam measurements, a black opaque cover was used instead of the transparent one, to also shield the sensor from light. The irradiation card is connected to an adapter board with all necessary passive components and connectors.

The DUT was placed on a mechanical structure, positioned between the two telescope arms, equipped with a rotary stage with remote control. This enabled a micrometer precision positioning, as well as performing angular scans. The DATURA telescope offers not only an easy insertion of the DUT into the telescope, but also an easy integration of the custom DAQ system into the overall DAQ framework. The DUT was read out with the BDAQ53 test system, described in Chapter 3, Section 3.4.1.

The MIMOSA26 are read out at 80 MHz with a rolling-shutter digital architecture. Sixteen clock cycles are needed to read out one sensor double column and 576 double columns have to be read out before the first one is ready again. This results in an integration time of 115.2 µs per frame, causing event pileup. Since the RD53A chip is read out at 40 MHz, accepting triggers every 25 ns, an external time *reference module* (REF) with the same timing as the DUT was needed to correctly assign tracks to the DUT events. Hence, a CMS Phase-1 pixel module [138] was placed in front of the telescope and used as the time reference. The REF is made of a 285 µm-thick silicon sensor with $150 \times 100\,\mu\text{m}^2$ pixels and 16 PSI46dig [139] readout chips.

Trigger and data acquisition

The programmable *Trigger Logic Unit* (TLU) [140] provides a common trigger signal and a time stamp to all connected detectors. The trigger signal is based on the coincidence of two trigger devices, one placed in the front and the other at the back of the telescope. This way only triggers from tracks traversing the full telescope are accepted. Each trigger device is composed of a plastic scintillator, a light guide and a *photo multiplier tube* (PMT). The 3 mm-thick scintillators define a trigger acceptance window of $10 \times 10\,\text{mm}^2$, covering the central half of the MIMOSA26 sensor area, as well as of DUT area.

The hardware is integrated in the EUDAQ [141] data acquisition framework, which merges data streams of all components as event-based data. The EUDAQ framework is composed of different processes communicating together using the TCP/IP protocol. A diagram of the EUDAQ architecture is shown in Figure 4.17. A central Run Control software provides an interface for controlling the whole DAQ system. It sends commands and verifies the status of all the other processes. Each hardware producing data, i.e. telescope planes, DUT or TLU, is controlled with a software element, called a Producer. The Producer allows to initialise, configure, start and stop the hardware when a corresponding command is send by the Run Control. It also forwards the data to the Data Collector. The Data Collector receives data streams from all the Producers, and combines them into a single data stream that is written to disk. The Log Collector receives log messages from all other processes, allowing for debugging and monitoring and the Online Monitor reads the data file from disk and provides real-time data monitoring. The REF has a standalone

DAQ, synchronised with the rest of the hardware thanks to the TLU. The REF data are saved in a separate file.

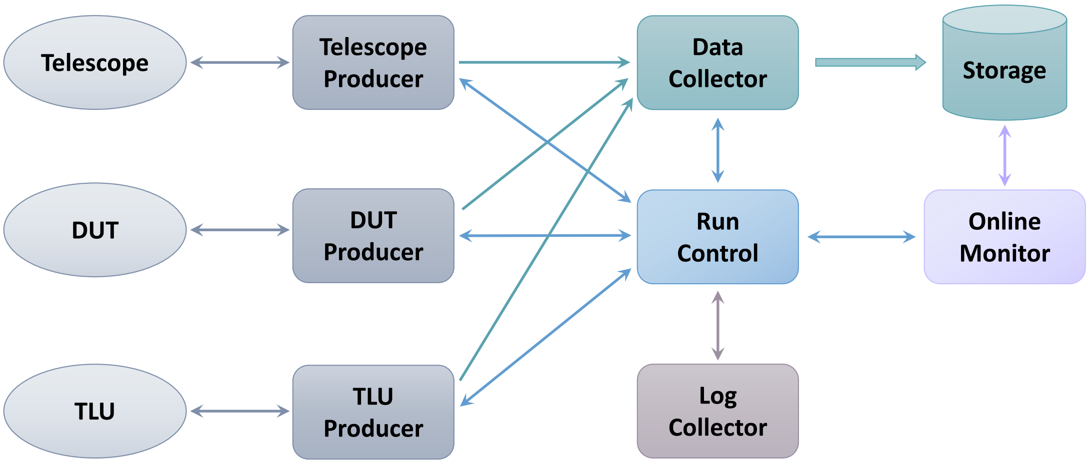


Figure 4.17: Overview of the EUDAQ framework architecture.

4.3.3 Data processing

Data decoding and clustering

The data taken with the previously described set-up are stored in a native "raw" binary format. The raw data are read by the offline analysis software and the EUDAQ converter is used to decode them to StandardEvent format. Each StandardEvent represents one data event from the whole telescope and the DUT, corresponding to one received trigger. The decoded pixel data is stored in an array of StandardPlanes, each representing one sensor plane of the telescope or the DUT. The REF data is loaded separately.

The charge deposited by a particle traversing the sensor can be shared among several adjacent pixels, due to either diffusion, Lorentz drift, incidence angle or cross-talk. The pixels triggered by the same particle are grouped into a cluster, using the *next-neighbour topological clustering*. This algorithm groups into the same cluster only direct neighbours sharing one pixel edge. The cluster represents a single hit in the sensor and the hit position is calculated. For clusters made of one pixel only, the hit position is simply the centre of the triggered pixel. For clusters composed of several pixels, a common position-finding method, called COG algorithm [142], is used. The cluster position is defined as:

$$x_{hit} = \frac{\sum_i x_i \, TOT_i}{\sum_i TOT_i}; \quad y_{hit} = \frac{\sum_i TOT_i \, y_i}{\sum_i TOT_i} \tag{4.10}$$

where $[x_i,\ y_i]$ are the coordinates of each pixel in the cluster and TOT_i its associated time-over-threshold (defined in Chapter 3, Section 3.1).

The fully binary readout of the MIMOSA26 sensors does not provide any information on the signal amplitude and therefore the hit position finding is simplified. In the RD53A chip of the DUT and the readout chips of the REF, the TOT information is available and the above mentioned weighted position calculation can be applied, which improves the spatial resolution of these detectors.

Track reconstruction

After clustering, the hits are transformed from the local sensor coordinates to the global three-dimensional space and are used for the track reconstruction. Since the test beam measurement was done with a collinear beam, with incidence direction normal to the sensor surface and no magnetic field, a simple straight-line tracking was sufficient for the analysis. Such tracking algorithm, based on the so-called *triplet method*, was developed in DESY and is explained in the following.

A track in the DATURA telescope is required to have exactly one hit in each of the six planes. The initial track candidates are built from triplets formed independently in the upstream and downstream telescope arms. Hit combinations in the first and third plane of each arm are fitted with straight lines. The angle between the linear fit and the beam axis is verified and track candidates too divergent from the beam axis are discarded. The accepted track candidates are interpolated to the middle plane. The distance between the interpolation and the hit position in the middle plane is called the triplet residual and denoted $\Delta x_{triplet}$, as indicated in Figure 4.18. The triplet residuals are used for further selection of track candidates: tracks with a hit within a certain distance are accepted and the others are considered as random hit combinations.

The upstream and downstream track candidates or half tracks are then extrapolated to the nominal position of the DUT. If the distance between the two extrapolated impinging points, denoted Δx_{track} in Figure 4.18, is within a certain range the two half tracks are accepted as a full track. The track position at the DUT corresponds to middle point between the two extrapolated half-tracks at the DUT plane.

Alignment

The mechanical alignment and the geometry description of the telescope are known with a much lower precision than the few micrometer telescope resolution. Hence, a fine alignment of the relative position of different sensors and their small rotations is necessary in the analysis. First, only the telescope planes are aligned, not taking into account the DUT and the REF. The telescope alignment is done in two steps and several iterations.

The first step is a pre-alignment, performed for the upstream and the downstream tele-

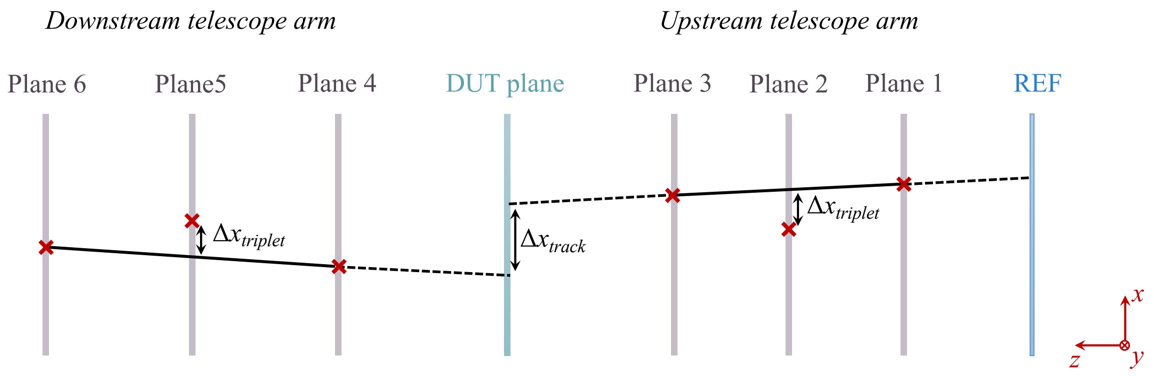


Figure 4.18: Track reconstruction.

scope arm separately. The middle plane of each arm, i.e. plane 2 and plane 5 are considered as fixed (using the plane numbering from Figure 4.18). The pre-alignment corrects the shifts in x and y and the rotation in x-y plane of the first and third plane of each arm, based on the cluster pair correlations. The clusters of each plane are correlated to those of the middle plane of the corresponding telescope arm, i.e. planes 1 and 3 are correlated to plane 2 and planes 4 and 6 are correlated to plane 5.

The second step consists in a precise alignment and requires full telescope tracks to be reconstructed. First, the z position of the first and third plane of each arm are adjusted with respect to the middle plane. Then, the whole downstream arm is aligned with respect to the upstream arm. The shifts and rotations are added simultaneously to all three downstream planes. In the end, the only plane still considered as fixed, with respect to which all the other planes were aligned, is plane 2. The full telescope alignment procedure is detailed in Appendix F. After the telescope is fully aligned the DUT and the REF are aligned using a similar procedure. Once the alignment is done, the track reconstruction is repeated to build the final tracks. The alignment procedure was also developed and implemented in DESY.

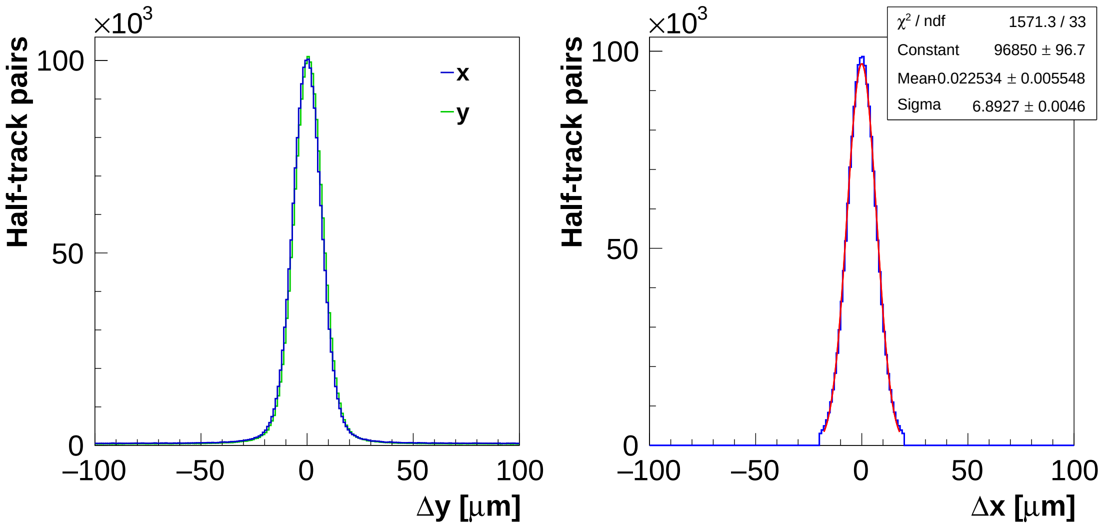
4.3.4 Resolution

Track resolution

The telescope alignment was verified based on the half-track matching. All combinations of the upstream and downstream half-tracks were considered. The half-tracks were extrapolated to the DUT plane and the distance between the impinging points, i.e. Δx_{track} in Figure 4.18, was is extracted. Figure 4.19(a) shows the distribution of the distances between half-track impinging points along the x axis in blue and y axis in green. They are very similar, given that the telescope sensors have square pixels and hence the resolution is the same along x and y. Both distributions are centred at zero, which means that the telescope translation alignment was done correctly.

The half-track matching distributions can also be used to extract the track resolution. They have long tails due to the long integration time of the MIMOSA26, spanning over several events, which results in half-track pairs not originating from the same particle. Only half-track pairs matching within 20 µm at the DUT are used as full tracks for the analysis. The half-track matching distribution along the x axis, after this selection, and therefore without the tails, is shown in Figure 4.19(b). The fit with a Gaussian distribution allows to extract the sigma, which divided by two corresponds to the track resolution.

The track resolutions obtained with this procedure for all five tested sensors are summarised in Table 4.3. They are similar for all sensors, because the same telescope was used. The slight differences in track resolution were introduced by moving the telescope planes closer or further away from the DUT when testing different sensors. The DUT thickness and the beam energy slightly different in each beam test also played a role. The best track resolution of 3.5 µm was obtained when testing the standard FBK sensor, because during this test the telescope planes were placed the closest to the DUT. It increased to 4.5 µm for the FBK with bitten implant and it was close to 6 µm for the HPK with rectangular pixels and the CNM 3D sensor, which were tested with the telescope planes further away from the DUT. The track resolution for the HPK sensor with square pixels could not be estimated with the same procedure. This sensor was placed inside a cold box during the beam test, which induced too much multiple scattering and hence larger deviations of the tracks after the DUT. The alignment was not converging correctly, because the distances between half tracks were too large resulting in a very broad half-track matching distribution. The final track positions were too far from the corresponding clusters, which



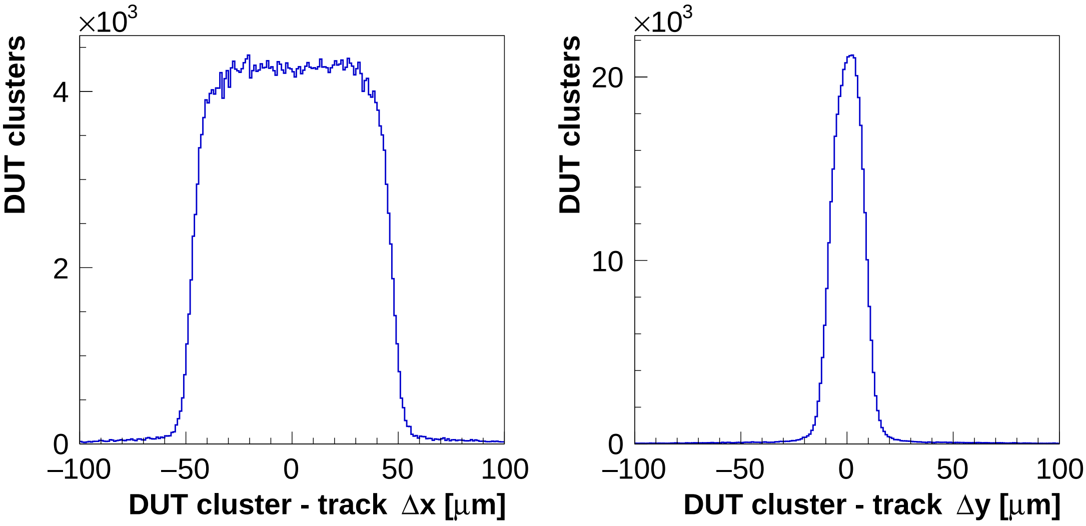
(a) (b)

Figure 4.19: Distribution of distances between impinging points of pairs of half-tracks (a) along x in blue and y in green and the same distribution along x selecting only half-track pairs matching within 20 µm fitted with a Gaussian distribution (b).

had an impact on the efficiency. Therefore, only the upstream half tracks, extrapolated as straight lines to the DUT, were used in the analysis of the results obtained with this sensor. The best and the worst track resolution obtained with the other four sensors were taken as the lower and the upper limit, respectively, to estimate approximately the track resolution of the HPK sensor with square pixels. Both were multiplied by $\sqrt{2}$, since only the upstream half-tracks were used, making the assumption that the resolution of the upstream and downstream half-tracks is equivalent.

Residuals

The DUT alignment with respect to the telescope was verified based on the residual distributions. The *residuals* are defined as the difference between the track position at the DUT and the corresponding cluster position. The residual distributions of the standard FBK sensor with rectangular pixels measured at a threshold of about 1100 e$^-$ are shown in Figure 4.20. The residuals along the x axis, i.e. along the long pixel pitch, is shown in Figure 4.20(a) and the residual distribution along the y axis, i.e. along the short pixel pitch, is shown in Figure 4.20(b). The shape of the residual distributions results from the convolution of a uniform distribution with a width of one pixel pitch and a Gaussian distribution representing the track resolution. This is particularly visible along the x axis, where the pixel pitch is larger than the width of the diffusion Gaussian. Along the y axis, the pixel pitch has a size comparable to the width of the Gaussian, which results in a Gaussian-shaped residual distribution. If both distributions are centred at zero, the DUT is correctly aligned with respect to the telescope.



(a) (b)

Figure 4.20: Residual distribution along the x (a) and y (b) axis obtained with the standard FBK sensor with rectangular pixels measured at a threshold of about 1100 e$^-$.

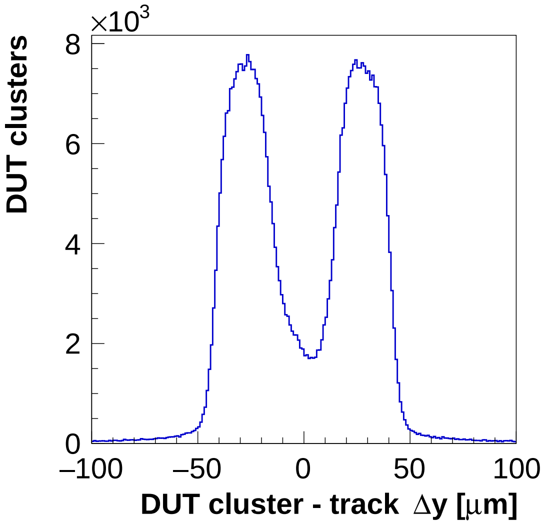


Figure 4.21: Residual distribution along y obtained with the standard FBK sensor with rectangular pixels measured at a threshold of about 1100 e$^-$ with a wrong pixel mapping.

The residual distributions can also be used to verify the pixel mapping in each DUT, i.e. the assignment of the sensor pixels to the pixels of the readout chip. For the sensor with square pixels, the sensor pixels have exactly the same geometry as the chip pixels and the correspondence is immediate. For the sensors with rectangular pixels, the mapping is more tricky and depends on the bump-bonding scheme of each DUT. The RD53A pixels in any pair of two adjacent chip columns, with an even column number on the left and an odd column number on the right, belong to the same sensor column, while the RD53A pixels in the same chip row belong to two adjacent sensor rows and can be connected in two different ways. An example of a residual distribution along y axis, i.e. along pixel rows, obtained with the standard FBK sensor with a wrong pixel row mapping is shown in Figure 4.21. The residual distribution is divided in two peaks, indicating that the mapping along the y axis is inverted. When the pixel row assignment is swapped the residual distribution along y looks like to one shown in Figure 4.20(b). It has a Gaussian shape and confirms the correct pixel mapping.

The residuals were extracted from the residual distributions along x and y as the sigma of the Gaussian fit for the short (25 μm) and square (50 μm) pitch and as the RMS of the distribution for the long pitch (100 μm). The residuals along the x and y axis obtained with all five sensors at a threshold of about 1100 e$^-$ are summarised in Table 4.3. The HPK sensor with square pixels has the same residuals along x and y, equal to about 15 μm. The other four sensors have rectangular pixels and therefore they have twice as large residuals of about 30 μm along the long pitch (x) and half smaller residuals of about 7.5 μm along the short one (y) compared with the square pixels.

DUT resolution

Finally, the previously obtained track resolutions σ_{track} and residuals σ_{res} allow to calculate the intrinsic DUT resolution given by:

$$\sigma_{DUT} = \sqrt{\sigma_{res}^2 - \sigma_{track}^2} \tag{4.11}$$

The DUT resolutions along x and y calculated for all five sensors are presented in Table 4.3. Since the track resolution of the HPK with square pixels was estimated only as a range between an upper and lower limit, the DUT resolution is also provided as a range. It is about 13–14 µm in both directions, which is slightly lower than $50\,\mu\text{m}/\sqrt{12} = 14.4\,\mu\text{m}$ thanks to the charge information from the readout chip. The DUT resolution for the rectangular pixels is about 30 µm along x and 6 µm along y. The obtained resolution along the short pitch is also lower than $25\,\mu\text{m}/\sqrt{12} = 7.2\,\mu\text{m}$, while the resolution along the long pitch is slightly higher than $100\,\mu\text{m}/\sqrt{12} = 28.9\,\mu\text{m}$, probably because the residuals along x were not extracted precisely from a fit, but taken as the RMS of the distribution enhanced by the tails. Among the sensors with rectangular pixels, the best DUT resolution was obtained with the FBK sensors and the worst with the CNM 3D sensor.

Table 4.3: Track resolution, residuals and DUT resolution measured in the test beam with the five selected sensors at a threshold of about 1100 e$^-$.

Tested sensor	**Track resolution**		**Residuals**		**DUT resolution**	
	x	y	x	y	x	y
HPK square	4.9–8.3	4.9–8.1	15.28 ± 0.03	15.37 ± 0.03	12.8–14.5	13.1–14.6
HPK rectangular	5.885 ± 0.006	5.715 ± 0.005	30.62 ± 0.07	8.52 ± 0.02	30.05 ± 0.07	6.31 ± 0.03
FBK standard	3.477 ± 0.003	3.464 ± 0.003	28.99 ± 0.05	6.60 ± 0.01	28.78 ± 0.05	5.62 ± 0.01
FBK bitten	4.523 ± 0.005	4.512 ± 0.005	29.16 ± 0.06	7.24 ± 0.01	28.80 ± 0.06	5.66 ± 0.02
CNM 3D	5.710 ± 0.004	5.580 ± 0.004	31.47 ± 0.05	8.97 ± 0.02	30.95 ± 0.06	7.03 ± 0.02

4.3.5 Efficiency

The efficiency is defined as:

$$\varepsilon = \frac{k}{n} \tag{4.12}$$

where n represents the number of selected tracks and k the number of tracks with a matching cluster in the DUT. Therefore, the event selection process is important to correctly extract the efficiency. The tracks selected for the efficiency normalisation (n) passed the following selection cuts:

- **Beam angle cut:** Only tracks not too divergent from the beam line were used. The angle between the upstream half-track and the beam axis was verified. A maximum divergence of 5 µm per millimetre of path along the z axis was allowed along both the x and y axis. This corresponds to a maximum incidence angle of 5 mrad.

- **Triplet cut:** The half-tracks should not be too “broken”, so the hit in the middle plane of each triplet had to be within ± 50 µm along both the x and y axis, for both the upstream and the downstream triplets.
- **Half-track matching cut:** The maximum allowed distance between the extrapolated upstream and downstream half-tracks in the DUT plane was set to ± 20 µm on both x and y axis, as it was mentioned previously.
- **Matching cluster in REF:** In addition to the geometrical considerations, a good track needs a correct timing. Therefore, only tracks with a matching cluster in the REF plane within ± 150 µm along the x and y axis were selected for the analysis. Moreover, the matching cluster was not allowed to be closer than 600 µm from another cluster to avoid any ambiguity.
- **Fiducial region:** The beam acceptance window defined by the overlap between the trigger scintillators was about $10 \times 10\,\text{mm}^2$. The beam particles crossing the DUT were only detected by the LIN AFE of the RD53A chip and the other two AFEs were disabled. The size of the LIN AFE is $6.8 \times 9.6\,\text{mm}^2$ and defines the so-called *fiducial region*. Only tracks pointing into the fiducial region had a chance to be detected by the DUT and were selected for the analysis. An example of the fiducial region is presented in Figure 4.22 showing the efficiency of the HPK sensor with square pixels as a function of the track position in the DUT plane. The green rectangle represents the contour of the RD53A pixel matrix. The red rectangle represents the contour of the LIN AFE defining the fiducial region for this sensor.

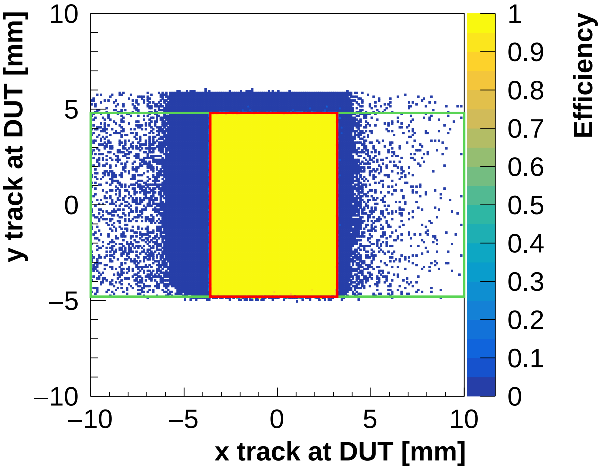


Figure 4.22: Efficiency as a function of the track position in the DUT plane measured with the HPK sensor with square pixels at a threshold of 1100 e^-.

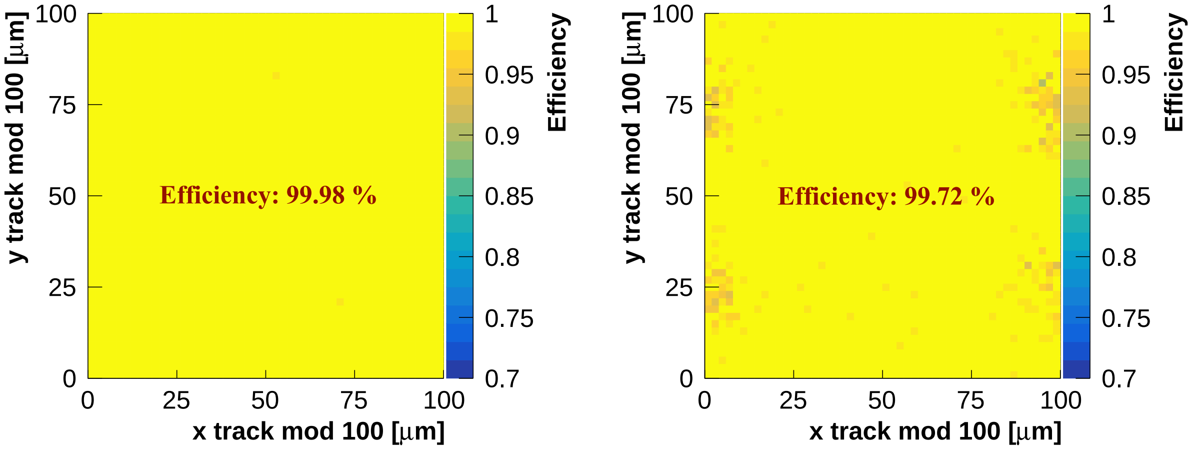
- **Noisy/dead pixels:** The tracks pointing to dead, faulty or noisy pixels that were disabled during data taking, were excluded from the analysis to preserve a high detection efficiency. In particular, in the CNM 3D sensor an island of faulty pixels plus some noisy pixels were excluded and in the HPK sensor with rectangular pixels the first core column of the LIN AFE, i.e. the first eight columns, were not 100% efficient and therefore, they were disabled during data taking and excluded from the fiducial region of this sensor.

After the track selection, the further event selection was based on the DUT clusters to account for k in the efficiency measurement. A cluster was considered to be matching with a selected track if its centre was within a distance of 500 µm from the track position in both x and y directions.

The efficiency as a function of the track position within four pixel cells measured with all five sensors at a threshold of about $1100\,e^-$ is presented in Figure 4.23. The efficiency of both HPK sensors, shown in Figures 4.23(a) and 4.23(b), and both FBK sensors, shown in Figures 4.23(c) and 4.23(d), is above 99%. The efficiency of the HPK sensor with square pixels and of both FBK sensors is very high, approaching 100%, while the efficiency of the HPK sensor with rectangular pixels is slightly lower. An efficiency drop visible in the pixel corners, in every other pixel row is due to the presence of the bias dot, as it can be observed in the sensor layout presented in Figure 4.6(b). The efficiency of the CNM 3D sensor, shown in Figure 4.23(e), is lower than for the other four sensors, because of an efficiency drop of about 30% in the pixel corners. The inefficiency is due to the tracks with normal incidence crossing the bias columns, located in every pixel corner, as it can be seen in the sensor layout in Figure 4.6(e). This effect is expected from a 3D sensor and is largely mitigated when the tracks arrive with a small non-zero incidence angle.

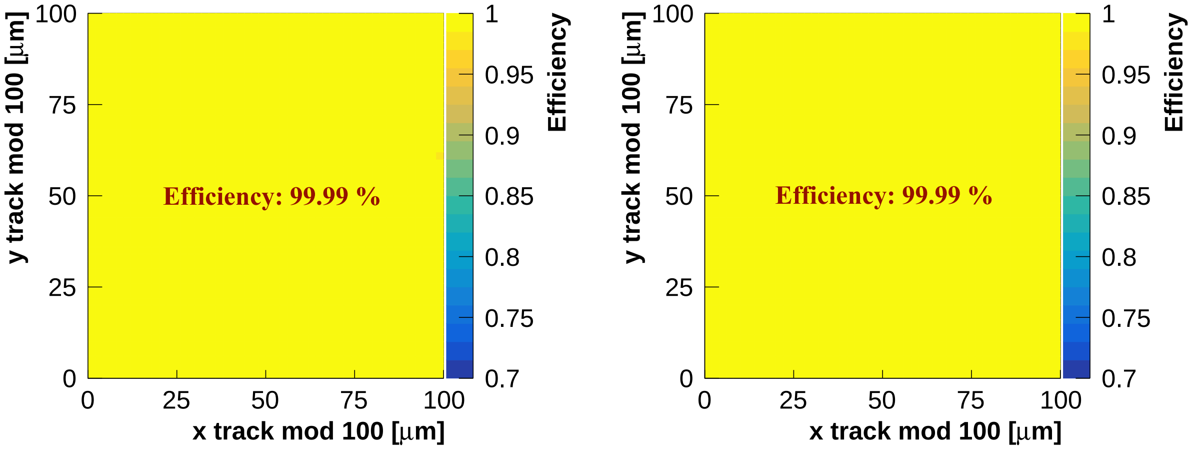
4.3.6 Cluster size

The average cluster size as a function of the track position within four pixel cells measured with all five sensors at a threshold of about $1100\,e^-$ is presented in Figure 4.24. These cluster-size maps reveal the pixel structure. The cluster-size map of the HPK sensor with square pixels is shown in Figure 4.24(a). The average cluster size in the pixel centre is close to one, it increases to about 1.5 when the track points close to the pixel edge and reaches almost two, when the track points close to the pixel corner. The higher cluster size at the pixel edges is due to the charge diffusion to the neighbouring pixel. The HPK sensor with rectangular pixels is shown in Figure 4.24(b). The overall cluster size in this sensor is between two and three, which is higher than in the HPK with square pixels. Moreover, the cluster map appears as two pixels, while four pixels are presented. The average cluster size anywhere within a pair of pixels is close to two and it increases up to three when the track points between two pixel pairs. This is evidence of a significant



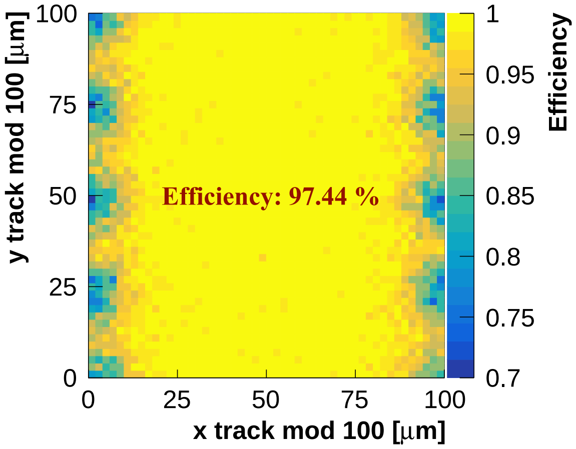
(a) HPK square.

(b) HPK rectangular.



(c) FBK standard.

(d) FBK bitten.



(e) CNM 3D.

Figure 4.23: Efficiency as a function of the track position within four pixel cells, measured in the test beam at a threshold of about 1100 e^- with the HPK sensor with square pixels (a), HPK sensor with rectangular pixels (b), FBK sensor with rectangular pixels and standard layout (c), FBK sensor with rectangular pixels and bitten implant (d) and CNM 3D sensor with rectangular pixels (e).

asymmetric cross-talk between pairs of pixels. The cluster-size maps of the two FBK sensors are shown in Figures 4.24(c) and 4.24(d). The average cluster size is equal to 1.2 in the pixel centre and is higher than two on the pixel edges for both sensors. The FBK sensor with the standard design shows a slightly higher cluster size at the pixel borders than the one with the bitten implant, which is a sign of a higher cross-talk. However, the two FBK sensors have a lower cluster size than the HPK sensor with rectangular pixels, which is consistent with the fact that they were tested with a higher front-end PA bias, known to reduce the cross-talk. Finally, the cluster-size map of the CNM 3D sensor is shown in Figure 4.24(e). The overall cluster size is comparable to the HPK sensor with square pixels. It is close to one in the pixel centre and lower than two close to the pixel borders.

The information about all clusters recorded at a threshold close to 1100 e^- by each of the tested sensors is summarised in the upper half of Table 4.4. For all sensors, the average number of clusters per event is very close to one. This means that the test beam was set to have only one electron passing through the telescope at a time, which ensures a very good detection efficiency even with a simple algorithm. The average number of clusters per event is just slightly above one, due to random noise hits in the DUT, which can be easily filtered out based on the track position. The lowest average cluster size of about

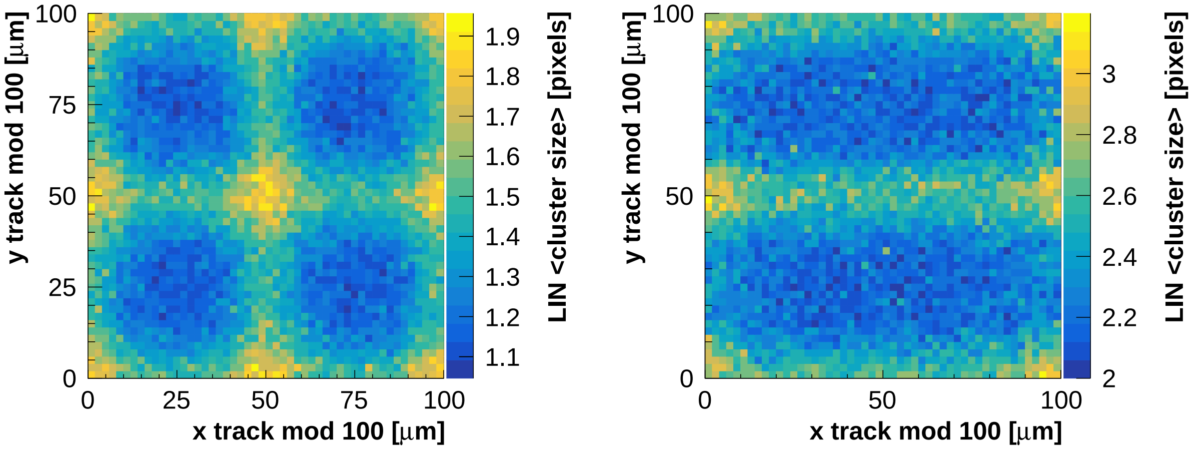
Table 4.4: Characteristics of all clusters and one-column clusters recorded in the test beam by the five selected sensors at a threshold of about 1100 e^-.

ALL CLUSTERS

Tested sensor	**Clusters /event**	**Cluster size**	**Columns /cluster**	**Rows /cluster**
HPK square	1.05	1.40	1.20	1.20
HPK rectangular	1.06	2.37	1.12	2.25
FBK standard	1.04	1.67	1.09	1.59
FBK bitten	1.04	1.58	1.10	1.48
CNM 3D	1.05	1.38	1.05	1.34

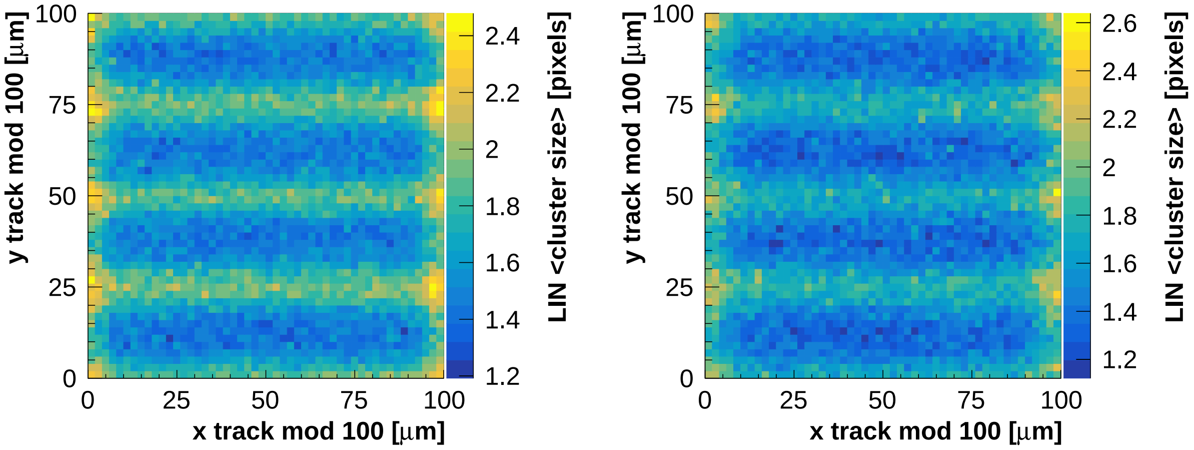
ONE-COLUMN CLUSTERS

Tested sensor	**Excluded clusters**	**Cluster size**	**Columns /cluster**	**Rows /cluster**
HPK square	24.4%	1.17	1.0	1.17
HPK rectangular	13.1%	2.21	1.0	2.21
FBK standard	9.0%	1.55	1.0	1.55
FBK bitten	10.4%	1.44	1.0	1.44
CNM 3D	7.3%	1.29	1.0	1.29



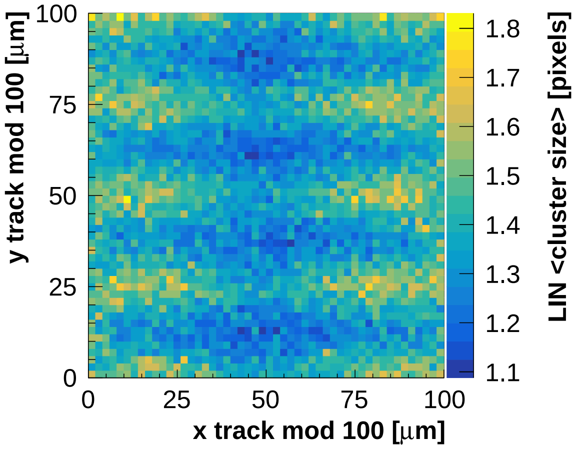
(a) HPK square.

(b) HPK rectangular.



(c) FBK standard.

(d) FBK bitten.



(e) CNM 3D.

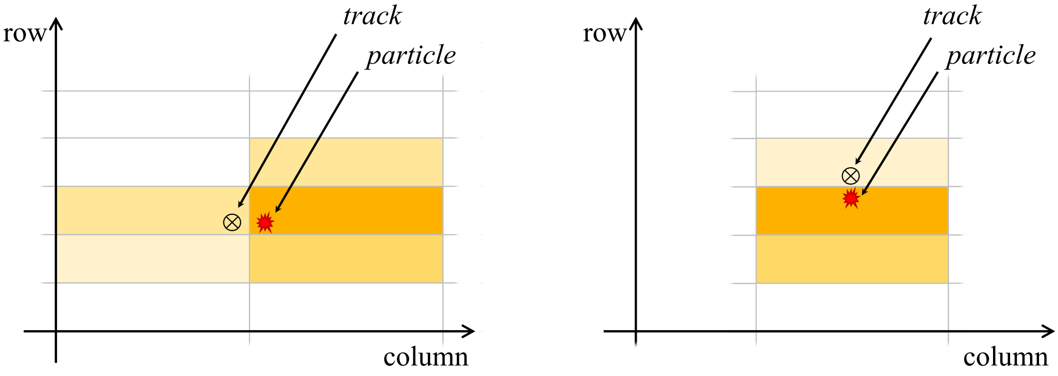
Figure 4.24: Cluster size as a function of the track position within four pixel cells, measured in the test beam at a threshold of about 1100 e^- with the HPK sensor with square pixels (a), HPK sensor with rectangular pixels (b), FBK sensor with rectangular pixels and standard layout (c), FBK sensor with rectangular pixels and bitten implant (d) and CNM 3D sensor with rectangular pixels (e).

1.4 was obtained with the square and the 3D pixels and the highest average cluster size reaching 2.4 was measured with the HPK with rectangular pixels, which is consistent with what was observed previously from the cluster-size maps. Between columns, only diffusion comes into play. Hence, the average number of columns per cluster is very similar for all sensors. It is slightly higher for square pixels, given that the pitch along the column direction (x) is half shorter (50 µm compared to 100 µm) favouring more diffusion. Moreover, with the square pixels the average number of columns and rows per cluster is the same, because the pitch is the same in both directions and this sensor type is not suffering from cross-talk. With the rectangular pixels, the average number of rows per cluster is higher than the number of columns, since the pitch along the row direction (y) is four times shorter (25 µm compared to 100 µm), increasing the diffusion and it is also enhanced by the cross-talk.

Since the asymmetric cross-talk to be measured is occurring only between rows, the analysis was simplified by selecting only one-column clusters. The events in which the matching cluster had more than one column were excluded from the analysis. The cluster information after the selection of one-column clusters is summarised in the lower half of Table 4.4. The fraction of excluded multi-column clusters was the highest for the square pixels, because the pixel pitch is half shorter along the column direction (50 µm compared to 100 µm) and hence the diffusion is more important and the probability of multi-column clusters is higher. For the rectangular pixels, the excluded multi-column clusters represented about 13% in the HPK sensor, 9 to 10% in the FBK sensors and about 7% in the CNM 3D. The overall cluster size, as well as the average number of rows per cluster, is slightly smaller in one-column clusters.

This event selection allowed to approximate the pixels to strips and to study the diffusion and cross-talk only in one direction. The key element of the analysis is the correct identification of the main pixel, i.e. the pixel effectively crossed by the particle in order to study its upper and lower neighbour. The multi-column clusters are due to tracks pointing close to the short pixel edge and the finite track resolution can cause the assignment of the main pixel to the wrong column. The misassignment of the main pixel in a multi-column cluster is illustrated in Figure 4.25(a) showing the triggered pixels of a multi-column cluster in shades of yellow, the colour saturation indicates the amount of signal in each triggered pixel, the impinging point of the incident particle is indicated by the red star and the track position by the black cross. The misassignment of the main pixel to a wrong column would affect not only the efficiency estimation of the latter but also of both neighbours, which would in turn affect the cross-talk analysis.

The same effect occurs also in the other direction and is even more important since the pitch is shorter and the tracks have a higher probability to point close to the pixel edge. This is illustrated in Figure 4.25(b) showing a one-column cluster in shades of yellow with the colour saturation proportional to the amount of signal in each triggered pixel. The particle impinging close to the pixel edge as indicated by the red star can be assigned to the neighbouring pixel pointed by the track as indicated by the black cross, which can fake the efficiency in any of the three pixels. In this direction, the effect cannot be removed from the analysis without sacrificing statistics as it will be discussed later in this chapter.



(a) Multi-column cluster.

(b) One-column cluster.

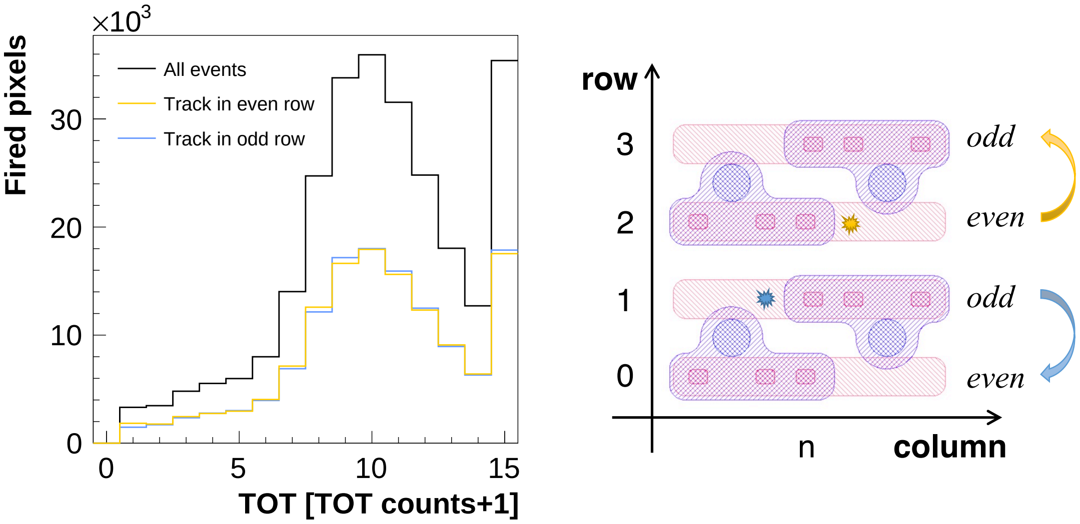
Figure 4.25: Illustration of the wrong assignment of the main pixel in a multi-column cluster (a) and in a one-column cluster (b) due to the finite resolution of the pixel telescope.

4.3.7 Pixel signal

The information about the amount of signal collected in different pixels of each cluster is provided by the RD53A chip as the TOT. The width of the signal pulse is expressed in TOT counts and one TOT count corresponds to one LHC clock cycle, i.e. 25 ns. The TOT distributions recorded in the triggered pixels of the selected clusters, detected with the standard FBK sensor at a threshold close to $1100\,e^-$, are presented in Figure 4.26.

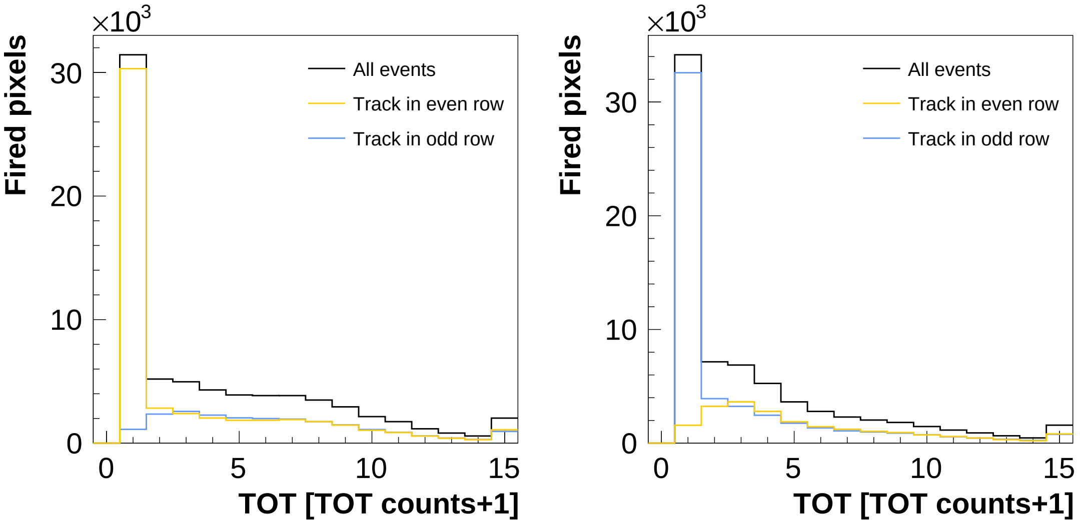
In each event, the pixel pointed by the track is denoted as the *main pixel*. The TOT distribution of main pixel in all selected events is shown in black in Figure 4.26(a). The distribution peaks at TOT = 10, which is given by the charge calibration of the RD53A chip, explained in Chapter 3, Section 3.7.1. The RD53A chip can count the TOT only up to 15 clock cycles and the TOT is usually calibrated for a test beam measurement such that the peak of the TOT distribution is close to the middle of the range. All signals larger than 15 clock cycles are assigned to TOT = 15, used as the overflow bin. Hence, the spike at TOT = 15 represents the right tail of the TOT distribution.

Then, the events were divided into two categories: the events in which the track pointed into an even row and those in which the track pointed into an odd row. The pixel rows were counted from bottom to top, the bottom corresponding to the sensor edge close to the wire-bonds. The row numbering starts from zero as indicated in Figure 4.26(b) showing two pixel pairs. The lower pixel of each pixel pair is in an even row and the upper pixel of each pair is in an odd row. The TOT distribution of the main pixel in an even row is shown in yellow in Figure 4.26(a) and the one corresponding to the main pixel in



(a) Main pixel.

(b) Track in even/odd pixel.



(c) Upper neighbour.

(d) Lower neighbour.

Figure 4.26: Time over threshold distribution in the main pixel (a), the upper neighbour (c) and the lower neighbour (d) measured with the standard FBK sensor at a threshold of about 1100 e^-. And schematic showing the assignment of track to even/odd rows (b).

an odd row is shown in blue. Since the track has an equal probability to point to an even or odd row, these two distributions are equal and represent each half of the data sample.

The TOT distribution of the pixel directly above the main pixel, denoted as the *upper neighbour* is presented in Figure 4.26(c). The TOT distribution taking into account all selected events is shown in black. It is composed of a large spike at one clock cycle and a long tail of higher TOTs. If the track points to an odd row, as it is illustrated with a blue star in Figure 4.26(b), the upper neighbour is not part of the same pixel pair and therefore, it does not suffer the cross-talk and can receive charge from the main pixel only by diffusion. The TOT distribution of the upper neighbour when the track is pointing to an odd row, corresponding to the diffusion only is shown in blue in Figure 4.26(c). The distribution is broad since the amount of charge transferred by diffusion depends on the impinging point of the particle, i.e. a particle passing closer to the pixel edge induces more charge sharing to the upper neighbour and hence a higher TOT. Moreover, the much lower statistics in this distribution (blue), compared to the distribution of all events (black), demonstrates that the diffusion rarely induces enough charge to cross the threshold and to trigger the upper neighbour pixel.

If the track points to an even row, as it is illustrated with the yellow star in Figure 4.26(b), the upper neighbour is part of the same pixel pair and therefore, it can receive charge from the main pixel by diffusion and also by cross-talk. This is confirmed by the TOT distribution of the upper neighbour for tracks pointing to even rows, shown in yellow in Figure 4.26(c), comprising a broad tail equal to the one caused by the diffusion (blue) and a large spike at TOT = 1 caused by the cross-talk. The spike has almost the same statistics as the spike in the distribution if all events, demonstrating that the upper neighbour is triggered by cross-talk in every event in which the track points to an even row. The cross-talk is due to the capacitive coupling between the pixel implant (pink hashed area in Figure 4.26(b)) and the metal electrode (violet hashed area in Figure 4.26(b)) and the same fraction of the deposited charge is migrating to the paired pixel independently of the impinging point of the particle within the pitch. This fraction being constant and the threshold being homogeneous across the pixel matrix, the same amount of charge is transferred to the neighbour in each event, which results in the same TOT. Moreover, since this TOT equals to only one clock cycle at the threshold of 1100 e^-, the fraction of charge induced by cross-talk is close to the threshold and if the threshold is increased the charge will not be sufficient to cross the threshold and to trigger the pixel. Hence, the cross-talk induced spike in the TOT distribution will quickly disappear with increasing threshold.

Finally, the TOT distribution of the pixel directly below the main pixel, denoted as the *lower neighbour* is presented in Figure 4.26(d) for all events in black, for tracks in even rows in yellow and for tracks in odd rows in blue. When all the events are taken into account, the TOT distribution of the lower neighbour has a very similar shape as the distribution of the upper neighbour and the cross-talk induced spike appears in the lower neighbour when the track points to an odd row, as expected.

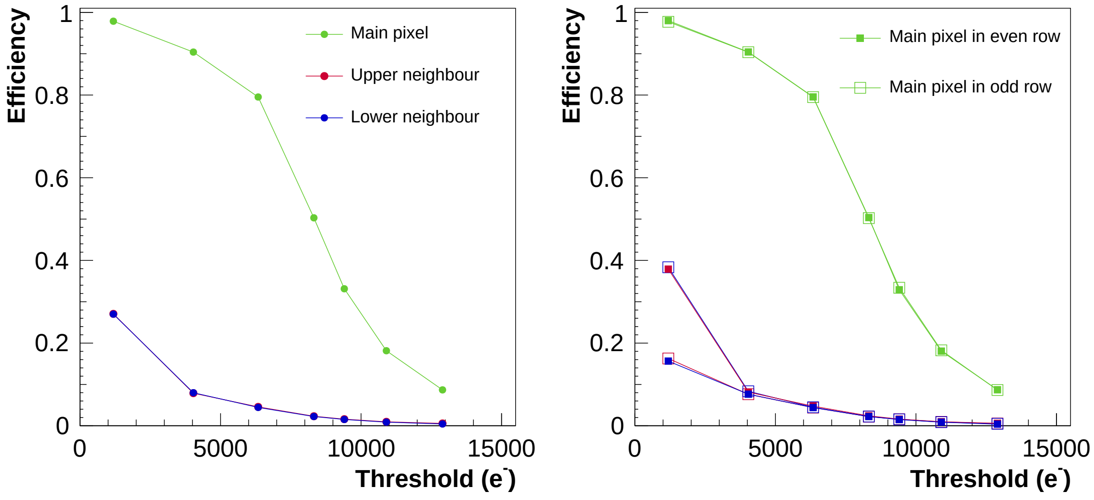
4.3.8 Cross-talk measurement

The efficiency as a function of threshold, measured with the standard FBK sensor, including all selected events is shown in Figure 4.27(a). The efficiency of the main pixel is shown in green. As expected, it is close to one at a threshold of 1100 e^- and it decreases with increasing threshold. The efficiency of the upper and lower neighbour, shown in red and blue respectively, is resulting from the combination of charge sharing and cross-talk. It is also higher at lower threshold. Moreover, the efficiency of the upper and lower neighbour are exactly equal when all events are used and the asymmetric cross-talk cannot be visualised. This confirms what was inferred in the previous section from the TOT distributions: when the particle impinges into an even row, represented by the yellow star in the sketch in Figure 4.26(b), the cross-talk induces signal in the upper neighbour, as indicated by the yellow arrow. On the other hand, when the particle impinges into an odd row, represented by the blue star in Figure 4.26(b), the cross-talk induces signal in the lower neighbour, as indicated by the blue arrow. Hence, if all the events are taken into account, the cross-talk asymmetry is compensated between the tracks in even and odd rows and the effect cancels out.

To estimate the cross-talk, the efficiency as a function of threshold was measured separately for the events with the main pixel in an even and odd row. The result obtained with the standard FBK sensor is shown in Figure 4.27(b). The main pixel is still shown in green, the upper neighbour in red and the lower neighbour in blue. The events with the main pixel in even rows are indicated with full square markers and those with the main pixel in odd rows are represented with open square markers. The efficiency of the main pixel is exactly the same whether it is in an even or odd row. At a threshold of $1100\,\mathrm{e}^-$, in the events with the main pixel in even rows, the efficiency of the upper neighbour (red full marker) is higher than the efficiency of the lower neighbour (blue full marker), while in the events with main pixel in odd rows it is the opposite. The difference between the two neighbours is proportional to the sensor cross-talk. As mentioned previously, the cross-talk contribution quickly decreases with increasing threshold and already at the threshold of about $4000\,\mathrm{e}^-$ the signal induced by cross-talk is not detected anymore and both neighbours have the same efficiency only due to the diffusion.

A new pixel mapping was introduced to merge the statistics in the rest of the cross-talk analysis. The main pixel is still the one pointed by the track. The *pixel with cross-talk* corresponds to the upper neighbour for tracks in even rows and to the lower neighbour for tracks in the odd rows. And on the contrary, the *pixel without cross-talk* corresponds to the lower neighbour for tracks in even rows and to the upper neighbour for tracks in the odd rows. The efficiency as a function of threshold with this pixel assignment is presented in Figure 4.28 for all five tested sensors. The result obtained with the HPK sensor with square pixels is shown in Figure 4.28(a), the HPK with rectangular pixels in Figure 4.28(b), the FBK with rectangular pixels and standard layout in Figure 4.28(c), the FBK with rectangular pixels and bitten implant in Figure 4.28(d) and the CNM 3D sensor in Figure 4.28(e). The main pixel is shown in green, the pixel with cross-talk in red and the pixel without cross-talk in blue.

The cross-talk measurement of the HPK sensor with square pixels shows the same efficiency in both neighbours, which confirms the expected absence of asymmetric cross-talk



(a) All events.

(b) Track in even/odd row.

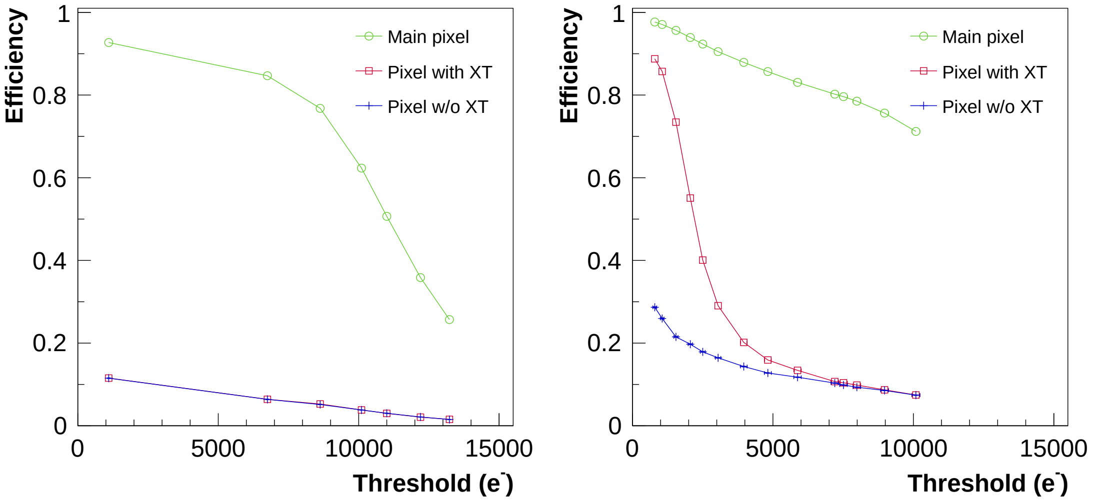
Figure 4.27: Efficiency as a function of threshold measured with the standard FBK sensor including all events (a) and dividing the events into even and odd rows (b).

in square pixels. This is also the case for the CNM 3D sensor, which is not expected to suffer any asymmetric cross-talk by design, as it was discussed in Section 4.2.6. The measurements of the HPK sensor with rectangular pixels, the FBK with standard implant and the FBK with bitten implant show a higher efficiency in the pixel with cross-talk than in the pixel without, which is an evidence of the asymmetric cross-talk in the planar sensors with rectangular pixels. In the HPK sensor with rectangular pixels, the efficiency difference between the two neighbours is about 60% at the threshold of 1100 e^-. In the FBK sensors this difference is much smaller, which indicates that the FBK sensors have less cross-talk. This result is consistent with the cross-talk measurement presented in Figure 4.10. Moreover, the efficiency difference between the two pixel neighbours in the FBKs is about 20% and 10%, for the standard layout and the bitten implant, respectively. This proves that the bitten implant indeed allows to reduce the cross-talk.

Threshold saturation

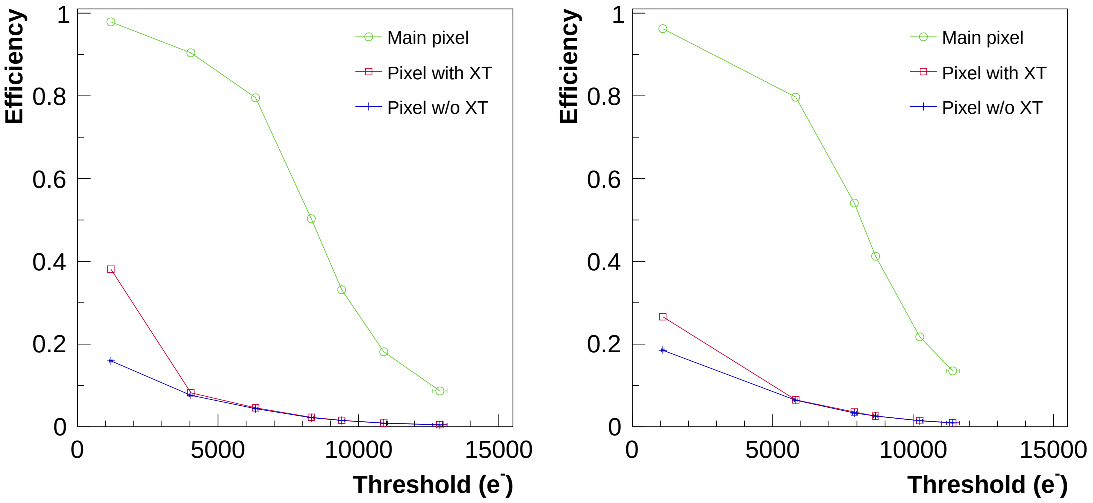
The cross-talk measurement with the test beam leads to the same conclusions as the measurement with injections and the two methods appear consistent. The next step consisted in the extraction of the amount of cross-talk from the test beam data to allow for comparison with the cross-talk values obtained with injections and to fully validate the two methods. There are three candidates for which the cross-talk could be evaluated from the test beam data: the HPK with rectangular pixels and both FBK sensors.

The method initially foreseen and described in Section 4.3.1 consisted in finding an efficiency reached by all three pixels in the efficiency versus threshold measurement in Fig-



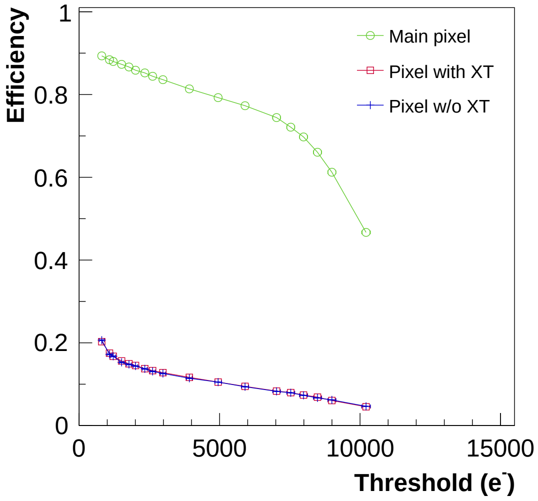
(a) HPK square.

(b) HPK rectangular.



(c) FBK standard.

(d) FBK bitten.



(e) CNM 3D.

Figure 4.28: Efficiency as a function of the threshold, showing the cross-talk (XT) in all five studied sensors.

ure 4.28 and to apply the formula in Equation (4.9). In the case of the HPK with rectangular pixels, not a single value of efficiency is simultaneously reached by all three pixels. One solution to this would be to measure efficiency for lower thresholds, which would allow to increase the efficiency in the pixel without cross-talk (blue curve in Figure 4.28(b)). However, the threshold tuning at lower thresholds is more complicated and the higher noise would affect the efficiency.

Another solution would be to the measure the efficiency at higher thresholds to decrease the efficiency in the main pixel (green curve in Figure 4.28(b)). This was not possible either, because it was found out that the threshold DAC in the RD53A chip saturates at very high thresholds. The mean threshold measured as a function of the threshold parameter of the LIN AFE (denoted Vthreshold_LIN) is presented in Figure 4.29. The threshold increases linearly with the threshold DAC up to 10 000 e$^-$ with the nominal analogue and digital supply voltage of 1.2 V, which is represented in dark blue. For higher values of the threshold DAC, the threshold saturates and cannot be set higher than 11 500 e$^-$. This behaviour was reproduced in circuit simulations and was confirmed not to occur in the improved version of the LIN AFE that will be implemented in the final chip.

It was also found out that the threshold saturation in the RD53A can be mitigated by increasing the analogue supply voltage (VDDA). The threshold as a function of the threshold parameter of the LIN AFE with the maximum allowed analogue supply voltage is shown in light blue in Figure 4.29. This workaround allowed to extend the maximum achievable threshold to up to 15 000 e$^-$ and was used in the test beam measurement of the FBK sensors. Indeed, the cross-talk measurement of the HPK sensor presented

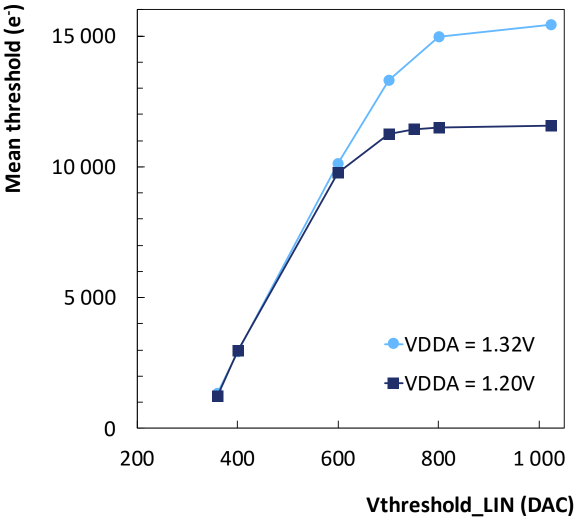


Figure 4.29: Mean threshold as a function of the parameter setting the threshold in the LIN AFE measured with an RD53A chip with the nominal analogue supply voltage of 1.2 V (dark blue) and the maximum analogue supply voltage of 1.32 V (light blue).

in Figure 4.28(b) was obtained with VDDA = 1.2 V, while the measurement of both FBK sensors presented in Figures 4.28(c) and 4.28(d) was performed with VDDA = 1.32 V. The higher analogue voltage in combination with a smaller active thickness of the FBK sensors, providing a smaller signal, allowed to increase the threshold enough to reach the efficiency of the main pixel below 20%.

Cross-talk multi-fit

Although the cross-talk measurement of the FBK sensors provides efficiencies reached by all three pixels, the small number of data points, in particular at low thresholds limits the accuracy of the cross-talk evaluation. A linear interpolation between data points does not represent well the threshold values at which a given efficiency is reached, which results in different cross-talk values when Equation (4.9) is applied at different efficiencies. Since no new data could be taken during this work because of the COVID-19 pandemic, another way to extract cross-talk was elaborated based on the data fit.

A Monte Carlo simulation generating 10^5 events was used to develop and validate the fit functions. The charge deposited in each event was randomly distributed according to a Landau distribution. The width and the MPV of the Landau were used as simulation parameters denoted p0 and p1, respectively. The track position along the short pixel pitch was randomly distributed following a uniform distribution. A fraction of the deposited charge could migrate to one of the two neighbours by diffusion depending on the track position. When the simulated track pointed close to the pixel centre, all the deposited charge remained in the main pixel and when the track pointed exactly at the pixel edge, half of the deposited charge was transferred to the neighbour. Between these two boundaries, the diffusion was modelled to increase linearly as the track position approaches the pixel edge. The simulation parameter representing the size of the diffusion

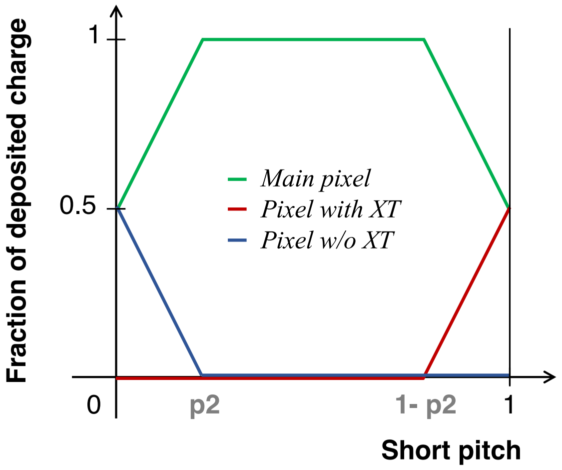


Figure 4.30: Linear diffusion model used in the simple Monte Carlo simulation.

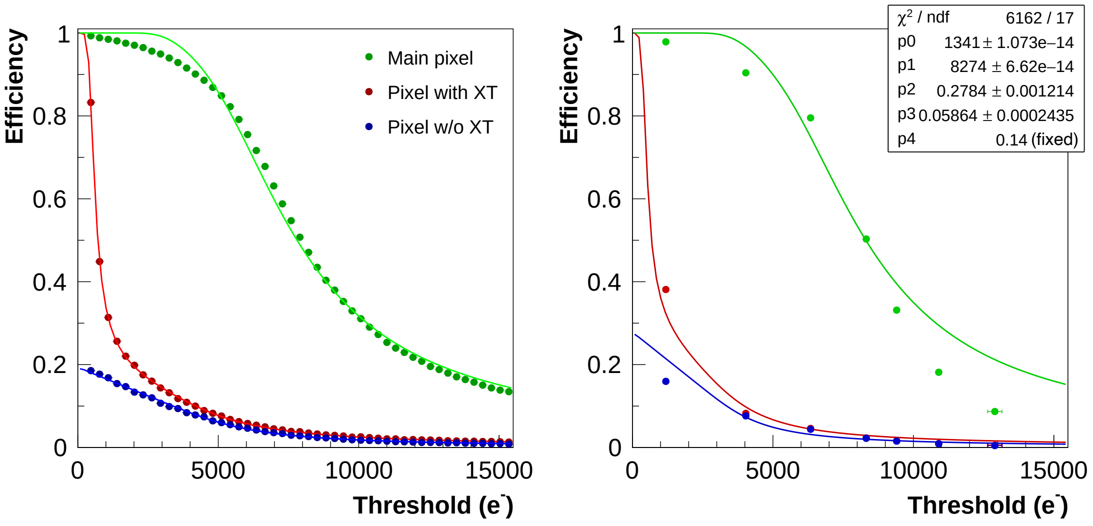
region was denoted p2 and expressed as a fraction on the short pitch, as it is illustrated in Figure 4.30. On the top of the diffusion, one of the two neighbours received charge by cross-talk. The cross-talk was simulated simply as a constant fraction of the charge in the main pixel transferred to the pixel with cross-talk and the same fraction of the charge in the pixel with cross-talk transferred back to the main pixel. The fraction of charge exchanged by cross-talk was used as another simulation parameter denoted p3. The last simulation parameter denoted p4 represented the track resolution expressed as a fraction of the short pitch.

The result of the Monte Carlo simulation with the five simulation parameters set to the values approximately expected from the measurement of the standard FBK sensor is shown in Figure 4.31(a). The simulation parameters were set to:

- Landau width: $\mathrm{p0} = 1200\,\mathrm{e}^-$,
- Landau MPV: $\mathrm{p1} = 8000\,\mathrm{e}^-$,
- Diffusion: $\mathrm{p2} = 5\,\mu\mathrm{m}/25\,\mu\mathrm{m} = 0.2$,
- Cross-talk: $\mathrm{p3} = 7\% = 0.07$,
- Track resolution: $\mathrm{p4} = 3.5\,\mu\mathrm{m}/25\,\mu\mathrm{m} = 0.14$.

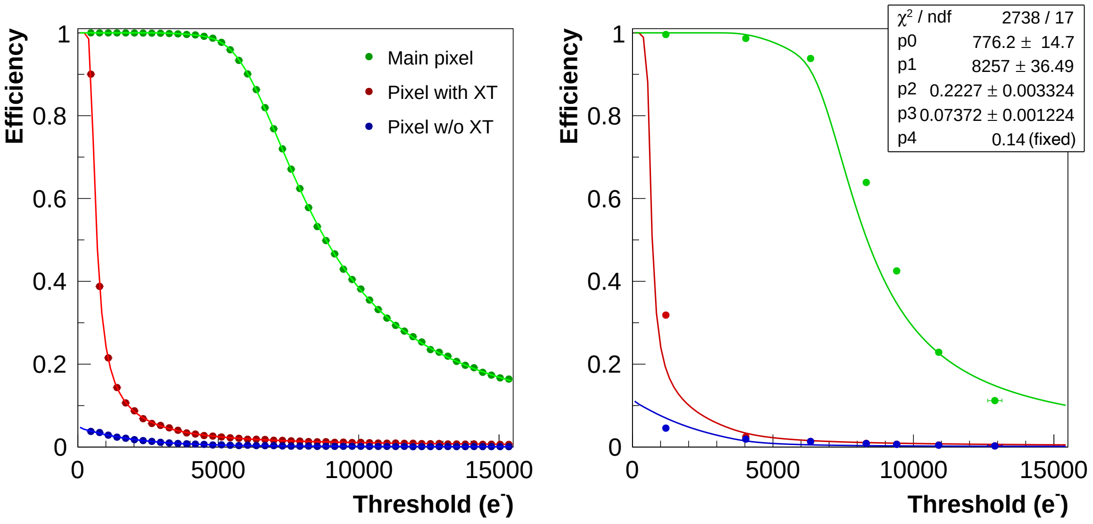
The Landau width and the MPV were inferred from the test beam measurement. Indeed, the efficiency of the main pixel as a function of the threshold corresponds to the integral of the Landau distribution from the threshold to infinity. Hence, the MPV corresponds to the threshold at which the derivative of the curve has the largest slope, i.e. around the 50% efficiency, which is reached at around $8000\,\mathrm{e}^-$. The diffusion was set arbitrarily to $5\,\mu\mathrm{m}$ from each pixel edge, the cross-talk was set based on the measurement with injections, presented in Section 4.2.6 and the track resolution was set based on the resolution measurements presented in Section 4.3.4. The Monte Carlo simulation with these parameters reproduces convincingly the cross-talk measurement of the standard FBK sensor shown again in Figure 4.31(b).

A fit function for each of the three efficiency curves was implemented based on the previous considerations. All three fit functions have the same five parameters p0 to p4. Each fit function was fitted separately to the corresponding efficiency curve obtained with the Monte Carlo and the result of the fit is also presented in Figure 4.31(a). All three efficiency curves in the cross-talk measurement originate from the same signal and therefore, their fit functions have to share the same fit parameters. A multi-fit was performed instead of a simple fit, i.e. the three efficiency curves were fitted simultaneously with their corresponding fit functions in order to minimise the global χ^2 and to optimise the five fit parameters simultaneously to all three data sets. The result of the multi-fit performed with data from the standard FBK is shown in Figure 4.31(b). The track resolution parameter was fixed to $\mathrm{p4} = 0.14$, since it was known from the measurement, leaving four out of five



(a) Monte Carlo all tracks.

(b) Test beam data all tracks.



(c) Monte Carlo tracks within 7.5 µm from the pixel edges excluded.

(d) Test beam data tracks within 7.5 µm from the pixel edges excluded.

Figure 4.31: Efficiency as a function of the threshold for all tracks obtained from the Monte Carlo (a) and from the test beam measurement (b) and only for tracks pointing further than 7.5 µm from the pixel edges obtained from the Monte Carlo (c) and from the test beam measurement (d). The fit functions are also represented.

fit parameters free. The total *number of degrees of freedom* (ndf) is equal to 17, which corresponds to three sets of seven measurement points minus four free parameters. The multi-fit resulted in a global $\chi^2 = 6162$, which corresponds to $\chi^2/ndf = 362$ indicating that the fit is not very good.

Indeed, the fit functions do not fit very well the data, in particular for the main pixel. This is also visible in the fit of the Monte Carlo data for the main pixel at low thresholds, which can be explained by the finite track resolution of the telescope, leading to the misassignment of the main pixel and lowering the efficiency, as it was explained at the end of Section 4.3.6 and illustrated in Figure 4.25(b). This effect is contained in the test beam data and is also accounted for in the Monte Carlo simulation, however it was not implemented in the fit functions of this simple fit model. Hence, the events in which the track was pointing close to the edge were excluded from the cross-talk measurement to mitigate the misassignment effect and to improve the fit. It was estimated empirically from the Monte Carlo that the size of the excluded region close to the pixel edges had be at least 1.5 times larger than the track resolution to sufficiently reduce the missasignment and to obtain a good fit of the Monte Carlo data. The track resolution being equal to 3.5 µm for the standard FBK, a margin was added and the tracks pointing within 7.5 µm from the pixel edges were excluded from the Monte Carlo, as well as from the test beam data, and the results are shown in Figures 4.31(c) and 4.31(d), respectively. The elimination of the tracks pointing close to the pixel edges increased significantly the efficiency of the main pixel (green) at low thresholds. It also notably reduced the diffusion (blue). The fit functions reproduce much better the Monte Carlo data and the quality of the multi-fit was also improved. The global χ^2 was more than twice smaller after the exclusion of tracks close to the edge, which resulted in a χ^2/ndf reduced from the initial value of 362 to about 160. The fit parameters resulting from this multi-fit indicate that the signal Landau had an MPV of about 8300 e^- and the width of about 780 e^-. The diffusion region spans across about 22% of the short pixel near each pixel edge, i.e. about 5.5 µm. The cross-talk measured in the standard FBK sensor with this method equals to about 7.4%.

Error bars

The cross-talk value obtained from the test beam data multi-fit matches the value obtained with injections. However, the uncertainties on the fit parameters obtained from the multi-fit cannot be trusted given the large χ^2/ndf. The cross-talk measurements presented in Figure 4.28 and in Figures 4.31(b) and 4.31(d) include error bars, even though they are not well visible, because they are smaller than the markers. The uncertainty on the threshold corresponds to the threshold dispersion across the pixel matrix. The threshold and the threshold dispersion were measured with the RD53A right before each run. The threshold was tuned at about 1100 e^-, as explained in Chapter 3, Section 3.5.2, and progressively increased without retuning. Hence, the threshold dispersion was increasing with the threshold and the horizontal error bars start to be visible at the highest threshold in the measurement of the standard FBK sensor. The statistical uncertainties on the efficiency were calculated based on the Bayes' Theorem as proposed by Ullrich and Xu [143]. If the number of tracks with a matching hit in the DUT is denoted k and the total number of tracks is denoted n, the efficiency probability density function can be defined as:

$$P(\varepsilon; k, n) = \frac{(n+1)!}{k!(n-k)!}\varepsilon^k(1-\varepsilon)^{n-k} \tag{4.13}$$

The efficiency is given by the most probable value: $\varepsilon_{m.p.} = \frac{k}{n}$ and the associated uncertainty can be calculated from the variance:

$$V(\varepsilon) = \sigma^2 = \frac{(k+1)(k+2)}{(n+2)(n+3)} - \frac{(k+1)^2}{(n+2)^2} \tag{4.14}$$

Since the efficiency estimate corresponds to the MPV of the efficiency probability density function, which is different from the mean: $\langle\varepsilon\rangle = \frac{k+1}{n+2}$, the statistical uncertainties are asymmetric, as it is illustrated in Figure 4.32. The resulting error bars being very small, the multi-fit was too constrained and could not reach a satisfactory χ^2. Based on the assumption that there could be some unknown systematics increasing the uncertainty of the cross-talk measurement, the fit errors can rescaled with the $\sqrt{\chi^2/ndf}$. The cross-talk extracted from the multi-fit with the rescaled uncertainty equals to $(7.37 \pm 0.02)\%$. This value is in agreement within error bars with the cross-talk measured with injections, presented in Table 4.1, equal to $(7.20 \pm 0.65)\%$.

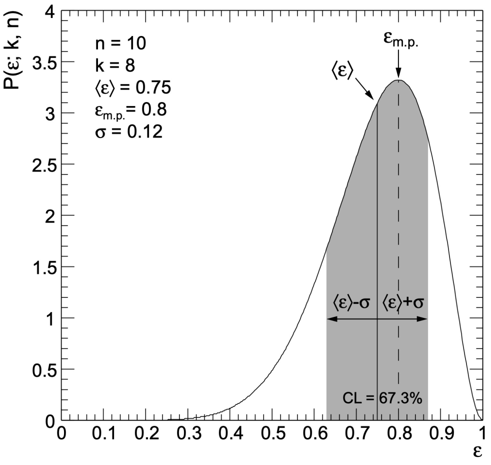


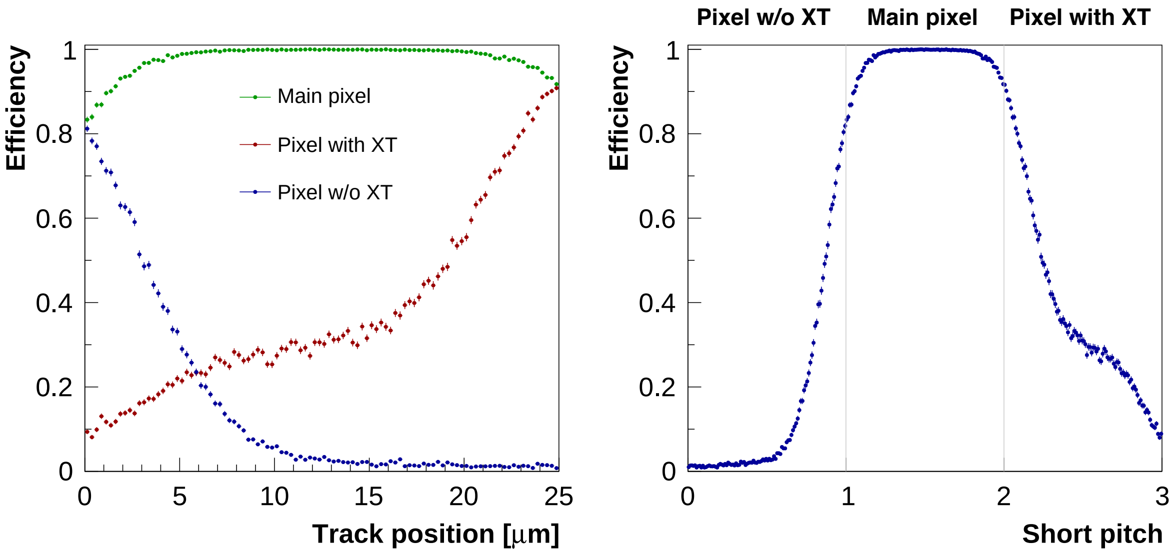
Figure 4.32: Example of the efficiency probability density function $P(\varepsilon; 8, 10)$ [143], showing the mean $\langle\varepsilon\rangle$, most probable value $\varepsilon_{m.p.}$ and the statistical uncertainty σ.

Efficiency profile

The efficiency of the three pixels measured with the standard FBK sensor at a threshold close to $1100\,e^-$ is presented in Figure 4.33(a) as a function of the track position along the short pitch of the main pixel. The efficiency of the main pixel shown in green is very close to one around the pixel centre and decreases when the track points close to the pixel edges, in particular within 5 µm from each edge. This indicates that the diffusion combined with the track resolution affects the edge region of about $5\,\mu m/25\,\mu m \approx 0.2$, which is in agreement with the diffusion parameter extracted from the test beam data multi-fit discussed in the previous section: p2 $= 0.22 \pm 0.04$. The efficiency of the pixel without cross-talk shown in blue is only enhanced by diffusion. It is close to zero when the track is far from the edge and increases as the track points closer to the edge. The efficiency in the pixel with cross-talk shown in red results from the combination of the diffusion and the cross-talk. The diffusion profile of this pixel, in principle the same as in the other neighbour, is convoluted with the cross-talk effect. The efficiency of this pixel is always higher than zero, even for tracks pointing far from the pixel edge, because the cross-talk does not depend on the track position. It can be noted that the average values of these three efficiency curves correspond to the three data points at the lowest threshold in the cross-talk measurement in Figure 4.31(b).

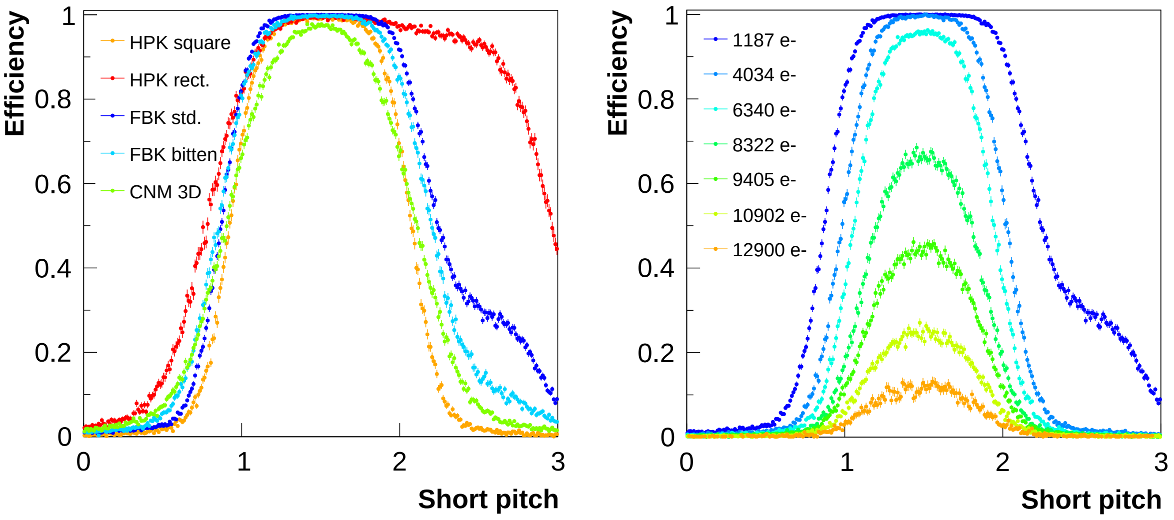
The efficiency of the three pixels as a function of the track position can be unfolded to obtain the continuous efficiency profile of the three pixels along the main pixel pitch shown in Figure 4.33(b). For this, the pixel pitch was normalised by 25 µm to unity, the efficiency profiles of the two neighbours were mirrored and placed on either side of the efficiency profile of the main pixel. The x axis of the resulting efficiency profile represents three times the short pitch with a size normalised to one. The middle pitch corresponds to the main pixel, the left pitch corresponds to the pixel without cross-talk and the right pitch to the one with cross-talk. The presence of the asymmetric cross-talk is clearly put in evidence in the efficiency profile.

The efficiency profiles of all five sensors tested at a threshold of about 1100 e$^-$ are presented



(a) (b)

Figure 4.33: Efficiency as a function of the track position along the short pixel pitch measured with the standard FBK sensor at a threshold close to 1100 e$^-$ (a) and efficiency as a function of the track position transformed to the efficiency profile (b).



(a) All sensors.

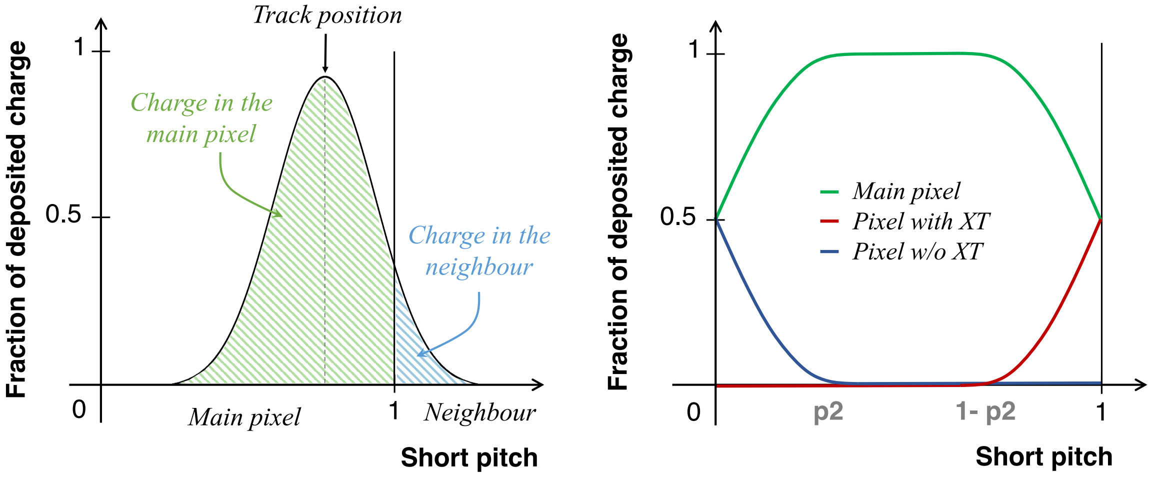
(b) FBK standard.

Figure 4.34: Efficiency profiles of all five selected sensors measured at a threshold of about 1100 e^- (a) and efficiency profile of the standard FBK sensor measured at different thresholds (b).

in Figure 4.34(a). The HPK sensor with square pixels is shown in orange, the HPK with rectangular pixels in red, the FBK with standard layout in dark blue, the FBK with bitten implant in light blue and the CNM 3D in green. The efficiency profiles put in evidence the asymmetric cross-talk in the three planar sensors with rectangular pixels. The HPK has the highest cross-talk, with the efficiency in the pixel on the right above 80% across about 2/3 of the pitch. The standard FBK sensor has a much lower cross-talk than the HPK, but still well visible in the efficiency profile. The FBK with bitten implant has the lowest cross-talk of the three almost not visible in the efficiency profile. However, the pixel on the right has a higher efficiency than the CNM 3D, which is not the case for the pixel on the left. Hence, a slight asymmetry is still present. The efficiency profile of the HPK with square pixels is symmetric showing only diffusion. The efficiency in the two neighbours is the lowest among all the sensors, because the pitch is twice as large for square pixels (50 µm compared to 25 µm) and therefore the diffusion region represents a smaller fraction of the pitch. The efficiency profile of the CNM 3D sensor is symmetric as well. The efficiency of the main pixel in the 3D sensor is lower than for the other three sensors, which is in agreement with the efficiency comparison discussed in Section 4.3.5. The efficiency profiles were also evaluated at different thresholds, as it is shown in Figure 4.34(b) for the standard FBK. As expected, the efficiency decreases with increasing threshold, and in particular, the cross-talk effect disappears quickly with increasing threshold.

Efficiency profile multi-fit

The efficiency profiles contain a more complete information about the diffusion and the cross-talk in different sensors and can therefore be used to elaborate an improved fit model to obtain a better fit and to extract the cross-talk from the test beam measurement with a better precision. For that, a Monte Carlo simulation, similar to the previous one, was used to develop and validate the fit function for the efficiency profile. The charge deposited in each event was again randomly distributed according to a Landau distribution with the width and MPV used as simulation parameters. The track position along the short pixel pitch was randomly distributed following a uniform distribution. A fraction of the deposited charge could migrate to one of the two neighbours by diffusion depending on the track position. The diffusion was simulated in this model with a Gaussian distribution centred at the track position. The fraction of charge transferred to the neighbour by diffusion corresponds to the integral of the Gaussian between the neighbour pixel's edges, as it is indicated in Figure 4.35(a) in blue. The fraction of charge remaining in the main pixel corresponds to the integral of the Gaussian between the edges of the main pixel, as indicated in Figure 4.35(a) in green. The integral of the Gaussian results in an S-curve shaped edges in the fraction of the deposited charge as a function of the track position, as illustrated in Figure 4.35(b), which matches better the efficiency profile than the linear diffusion used in the first simple fit model. The fraction of the short pitch corresponding to the diffusion region of the pixel was used as another simulation parameter.



(a) (b)

Figure 4.35: Gaussian diffusion model used in the improved Monte Carlo simulation.

On the top of the diffusion, one of the two neighbours could receive charge by cross-talk. The cross-talk was simulated as a constant fraction of the charge in the main pixel transferred to the pixel with cross-talk and the same fraction of the charge in the pixel with cross-talk transferred back to the main pixel. The fraction of charge exchanged by cross-talk was used as another simulation parameter and the last parameter represented the track resolution expressed as a fraction of the short pitch.

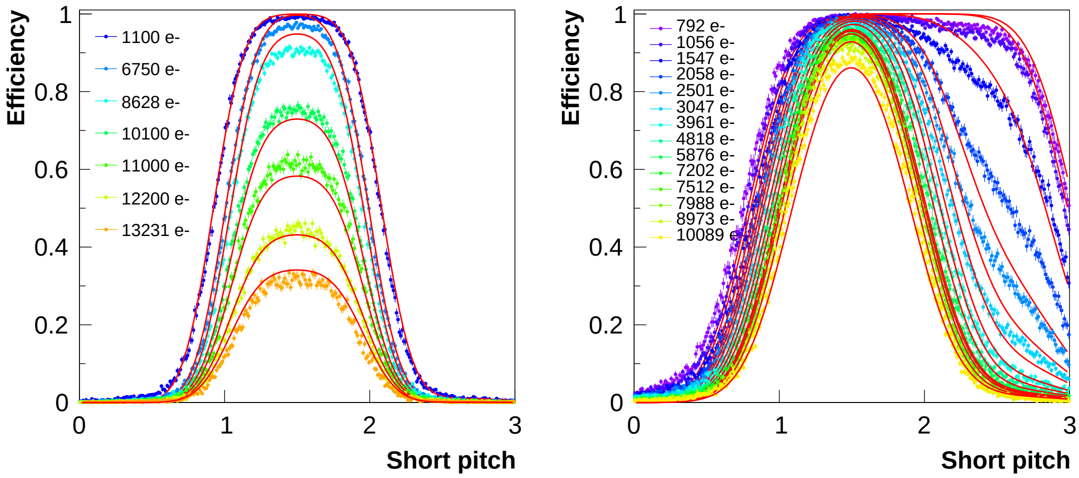
The function to fit the efficiency profile was implemented based on these considerations and had as free fit parameters the five parameters used in the simulation. The track mis-assignment due to the finite track resolution was also taken into account in this model. The track position was implemented as a Gaussian probability distribution centred at the impinging position of the particle. The efficiency profile fit function was implemented for each threshold, using the threshold as a known fixed parameter of the function. The efficiency profiles measured at different thresholds with the same sensor were fitted simultaneously with their corresponding fit functions, sharing the same five fit parameters to minimise the global χ^2. The result of this multi-fit applied on the efficiency profiles obtained with all five tested sensors at different thresholds is shown in Figure 4.36. The fit functions reproduce well the shape of the efficiency profiles from data.

The fit parameters obtained from the multi-fit applied to all tested sensors are summarised in Table 4.5. The Landau MPV is the largest in the HPK sensors, corresponding to about 10 000 e^- in the HPK with square pixels and about 13 000 e^- in the HPK with rectangular pixels. The Landau MPV is about 9600 e^- in the CNM 3D sensor, while in the two FBKs it is few hundreds of electrons lower. This is in agreement with the difference in the active thickness of the different sensors. Indeed, the two HPKs are the thickest ones with an active thickness of 170 µm, while the FBKs and the CNM have an active thickness of 150 µm. Moreover, the FBK sensors are expected not to reach the full depletion due to their fabrication process, as explained in Section 4.1.7. Hence, the depleted volume of the FBKs has the smallest thickness producing the smallest signal among the tested sensors. However, the measured Landau MPV divided by the sensor thickness results for all tested sensors in a signal of approximately 60 e^-/µm, except for the HPK with rectangular pixels, which produces almost 80 e^-/µm. Hence, the sensor thickness is not the only reason for the differences in the Landau MPV and other effects come into play.

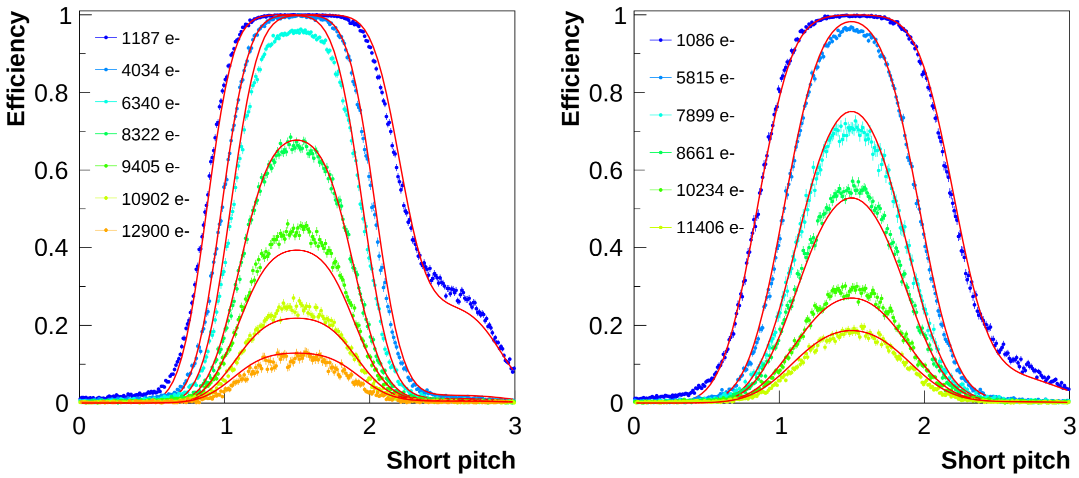
The Landau distribution was the largest in the HPK sensor with rectangular pixels and

Table 4.5: Fit parameters obtained from the efficiency profile multi-fit at different thresholds.

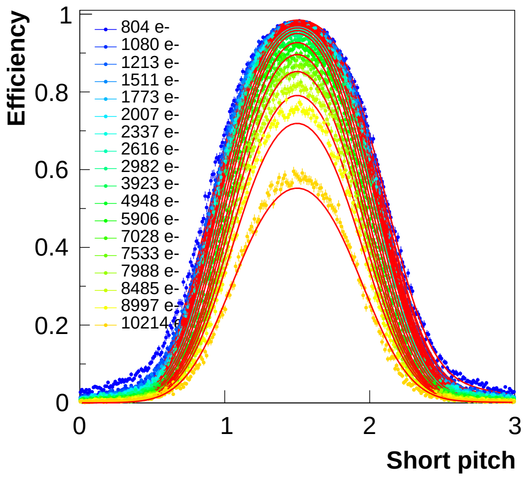
Tested sensor	**Landau MPV (e^-)**	**Landau width (e^-)**	**Diffusion region (μm)**	**Track resol. (μm)**
HPK square	$10\,289 \pm 6$	1053 ± 4	2.73 ± 0.01	8.19 ± 0.01
HPK rectangular	$13\,356 \pm 10$	1215 ± 6	3.33 ± 0.01	5.84 ± 0.01
FBK standard	9188 ± 5	570 ± 2	2.62 ± 0.01	$3.66 \pm 2 \times 10^{-8}$
FBK bitten	8730 ± 7	590 ± 2	3.11 ± 0.01	4.77 ± 0.01
CNM 3D	9602 ± 5	1192 ± 7	1.75 ± 0.01	$6.31 \pm 2 \times 10^{-6}$



(a) HPK square.
(b) HPK rectangular.



(c) FBK standard.
(d) FBK bitten.



(e) CNM 3D.

Figure 4.36: Efficiency profile multi-fit for all five tested sensors.

the CNM 3D with the width of about 1200 e⁻. The HPK with square pixels had a smaller width of about 1000 e⁻ and the two FBK sensors presented the narrowest Landau distributions with a width of about 600 e⁻. The size of the diffusion region is very similar for all planar sensors, being about 3 µm from the long edge, while it is slightly smaller for the 3D sensor in which it is below 2 µm.

The track resolution obtained from the multi-fit was the largest for the HPK with square pixels, which is consistent with the fact that for this sensor only the upstream half-tracks were used. It was slightly worse for the HPK with rectangular pixels and the CNM 3D sensor than for the two FBK sensors, because the two telescope arms were placed a bit further from the DUT during the beam test of the HPK and the CNM sensors. The track resolutions obtained for all sensors from the efficiency profile multi-fit are reported in Figure 4.37 in yellow and compared to the resolutions obtained from the half-track matching distributions presented in Section 4.3.4 and reported in Figure 4.37 in green. The track resolution of the HPK sensor with square pixels obtained from the half-track matching was estimated only as a range and therefore it is represented by an error bar delimiting the upper and the lower limit. Both methods give very similar results and the same trends are observed. The resolution extracted from the multi-fit is systematically slightly higher than the resolution extracted from the half-track matching, which is probably due to the long non-Gaussian tails excluded from the half-track matching distributions before the fit.

The χ^2 and the ndf associated to the efficiency profile multi-fit are provided in the upper half of Table 4.6. The ndf is large since it represents the number of bins in the efficiency

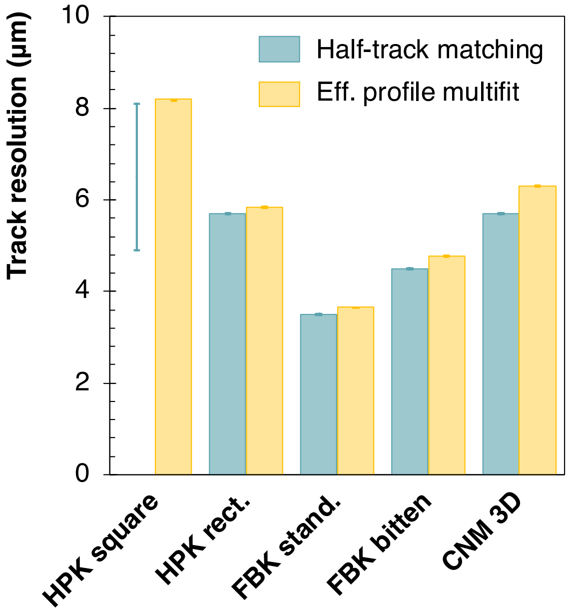


Figure 4.37: Comparison of the track resolutions obtained for all sensors from the half-track matching distributions (green) and from the efficiency profile multi-fit (yellow).

profile, which is equal to 300 for all sensors and all thresholds, multiplied by the number of thresholds fitted simultaneously for each sensor, to which were subtracted the five fit parameters. For instance, the efficiency profiles of the standard FBK sensor were measured at seven different thresholds, as it is shown is Figure 4.36(c). Hence, for this sensor the multi-fit had the $ndf = 7 \times 300 - 5 = 2095$. Given the large ndfs, all multi-fits were well constrained and very stable. The resulting χ^2/ndf is also given in the upper half of Table 4.6 and is below ten for all tested sensors. With such low χ^2/ndf, the statistical errors associated to the fit parameters reported in Table 4.5 can be trusted.

The cross-talk fit parameter obtained from this multi-fit is also presented in the upper half of Table 4.6 for all five sensors. The corresponding errors were rescaled by multiplying the fit errors by the $\sqrt{\chi^2/ndf}$. Out of the three sensors with expected asymmetric cross-talk, the highest cross-talk was measured with the HPK with rectangular pixels. The cross-talk measured with the multi-fit method with the square and 3D pixels is very low, however it is not zero, which is not in agreement with the previous observations. Indeed, it can be noticed that the fit convergence pushes the cross-talk parameter to higher values to try to reproduce some tails in the distribution even if they are symmetric. Therefore, another step was added to this analysis to correct for this. After the first multi-fit, all fit parameters except the cross-talk were fixed to the values obtained from the multi-fit and presented in Table 4.5. The asymmetry of the efficiency profile from data and the asymmetry of the fit function were compared based on the weighted mean along the x axis.

Table 4.6: Fit parameters obtained from the efficiency profile multi-fit at different thresholds before and after the asymmetry correction.

MULTI-FIT BEFORE THE ASYMMETRY CORRECTION

Tested sensor	χ^2	**ndf**	χ^2**/ndf**	**Cross-talk (%)**
HPK square	11 403	2395	4.8	0.80 ± 0.05
HPK rectangular	43 570	4195	10.4	12.32 ± 0.05
FBK standard	17 592	2095	8.4	10.12 ± 0.04
FBK bitten	6812	1795	3.8	6.15 ± 0.08
CNM 3D	32 914	5395	6.1	1.84 ± 0.05

MULTI-FIT AFTER THE ASYMMETRY CORRECTION

Tested sensor	χ^2	**ndf**	χ^2**/ndf**	**Cross-talk (%)**
HPK square	1	7	0.1	0.00 ± 0.20
HPK rectangular	3066	13	235.8	12.42 ± 0.50
FBK standard	795	6	132.5	9.76 ± 0.40
FBK bitten	154	5	30.9	5.43 ± 0.60
CNM 3D	11	17	0.6	0.04 ± 0.10

A second multi-fit was performed leaving only the cross-talk parameter free and aiming to minimise the asymmetry difference between the data and the multi-fit.

The χ^2 and the ndf of this multi-fit after the asymmetry correction as well as the resulting cross-talk are summarised in the lower part of Table 4.6. The ndf is much lower than in the first multi-fit, since it corresponds to the number of thresholds fitted simultaneously for each sensor minus one free fit parameter, i.e. the cross-talk. The cross-talk value obtained from this corrected multi-fit applied to the HPK sensor with square pixels and the CNM 3D sensor is much lower than after the first multi-fit and is very close to zero. The χ^2/ndf is also much lower for these two sensors. For the other three sensors with asymmetric cross-talk, the χ^2/ndf increased by at least an order of magnitude after the asymmetry correction and the cross-talk changed slightly. It can be also noted that the uncertainties on the cross-talk value, rescaled by $\sqrt{\chi^2/ndf}$, increased by an order of magnitude with respect to the first multi-fit.

The cross-talk exacted from the corrected multi-fit is reported in Figure 4.38 in yellow and compared to the asymmetric cross-talk measured with the injection method presented in Figure 4.10 and reported in Figure 4.38 in green. Both methods result in the asymmetric cross-talk in square and 3D pixels consistently equal to zero. The cross-talk measured in the FBK sensor with bitten implant with the two methods shows a very good agreement, while for the HPK with rectangular pixels and the FBK with standard implant the two methods give different results. However, both methods show the same overall trend, when comparing the cross-talk in these three sensors: the highest cross-talk in the HPK with

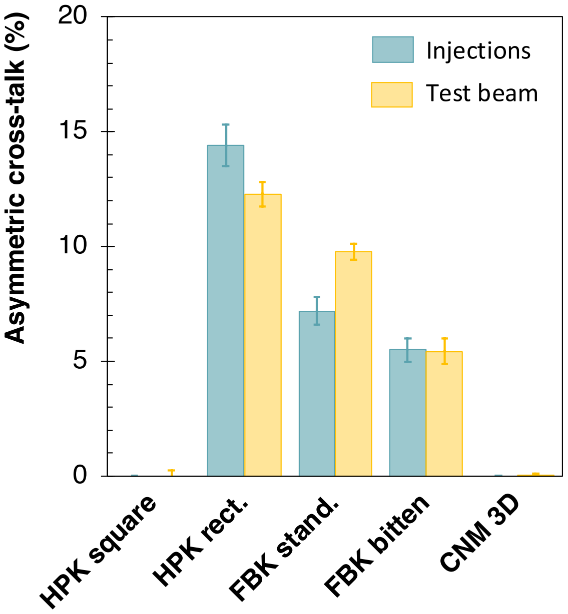


Figure 4.38: Comparison of the asymmetric cross-talk measured for all sensors with the injection method (green) and with the test beam method using the efficiency profile multi-fit (yellow).

rectangular pixels and the lowest in the FBK sensor with the bitten implant. The two methods are based on a different mechanism: one on the charge injection from the readout chip and the other on the charge collection from the sensor. These two mechanisms have a different timing of the signal, which can affect the cross-talk measurement and can contribute to the observed difference in the results. Moreover, only the statistical uncertainties were taken into account in this study and the systematic uncertainties of the test beam measurement could be studied with more accurate Monte Carlo simulations to obtain a more precise error estimate on the cross-talk measurement.

4.4 Summary and discussion

Two methods to measure cross-talk were developed in this work to help the choice of the sensor pixel geometry for the Phase-2 Upgrade of the CMS IT. The first method using the calibration injection circuit of the readout chip is based on the double S-curve measurement. It can be used to measure both the symmetric and asymmetric cross-talk and it can be performed quickly in the lab. Therefore, it is now widely used in the CMS pixel community to compare different sensor options. The second method based on the test beam measurement of the efficiency as a function of the threshold is sensitive only to the asymmetric cross-talk. Both methods were applied to five different sensors and lead to the same conclusions. No asymmetric cross-talk was observed between square or 3D pixels, while both methods provided evidence of the presence of asymmetric cross-talk between pairs of rectangular pixels in planar sensors. This cross-talk was confirmed to originate from the overlap between the pixel implant and the collection electrode of the neighbour pixel. It was also shown that the threshold and the charge calibration have a negligible impact on the cross-talk, while both methods confirmed a dependency of the cross-talk on the PA bias current.

Although the originally proposed cross-talk measurement method using the test beam could not be applied in this work, a multi-fit of the efficiency as a function of threshold measurement using a simple fit model allowed to extract the cross-talk in the standard FBK sensor equal to $(7.37 \pm 0.02)\%$. This value is consistent within error bars with the cross-talk measured with the injection method for the same sensor, which was equal to $(7.20 \pm 0.65)\%$. Moreover, a multi-fit of the efficiency profiles using an improved fit model allowed not only to extract the cross-talk value for all five sensors, but also to evaluate the characteristics of the Landau distribution and the diffusion, based only on a simple efficiency measurement at different thresholds, without using the TOT charge measurement. This method can therefore also be applied with pixel or strip detectors read out by binary readout chips not providing the charge information.

The two measurement methods showed consistent results and were hence fully validated and can now be used to evaluate the amount of cross-talk in different sensor options. In this work, it was shown that the cross-talk in rectangular pixels is of the order of 10% and the simulations showed that a 10% cross-talk has only a negligible impact on the tracking resolution. Moreover, the bitten implant topology was proven to reduce the cross-talk by about 2% in the FBK sensors with the same settings, making them promising candidates for the IT upgrade. The PA bias of the readout chip represents another handle to further reduce cross-talk at a price of a slight increase in the power consumption, which can be absorbed by the cooling system. Since the asymmetric cross-talk in rectangular pixels is studied in detail it could also be corrected offline to a certain extent. For all these reasons, the rectangular pixels remain the preferred option for the CMS Phase-2 Upgrade.

Finally, some limitations were encountered during the establishment of the test beam method and should serve as lessons learned in the future cross-talk studies when the test beams are available again. The test team measurement of the cross-talk should be made with the analogue supply voltage of the RD53A chip set to the maximum to reach higher thresholds. Thinner sensor providing smaller signal with the DESY test beam are preferable to be able to decrease enough the efficiency of the main pixel. More data points, in particular at low and high thresholds, would improve the accuracy of linear interpolations. With this precautions, the originally proposed method to measure cross-talk in the test beam could be applied and verified. To improve the fit method, the tracking resolution of the telescope should be reduced to the minimum, by placing the telescope arms as close as possible to the DUT and by removing any unnecessary material from the beam line. For instance, the thick plastic cover used for shielding of the DUT could be replaced by a thin Kapton foil. With better tracking resolution the multi-fit of the efficiency versus threshold would be improved without sacrificing the statistics. The systematic uncertainties of the measurement could also be investigated and included to the cross-talk measurement with test beam.

Chapter 5

E-links evaluation

One of the main challenges arising from the pixel detector operation at the HL-LHC is the very high bandwidth needed to read out the data. The older generation of output data links used in the current detector operating at 160 Mb/s [54] will not be sufficient to cope with the high trigger and data rates reached at the HL-LHC. Therefore, a new data link design offering multi-Gb/s read-out is necessary. The high-speed serial data link implementation for the IT readout system and the associated challenges are the topic of in this chapter.

The pixel modules, composed of readout chips, treated in Chapter 3 and silicon sensors addressed in Chapter 4, will be controlled and read out by ultra low-mass electrical links, also called *e-links*. The signal integrity in the e-links must be verified and validated to ensure the correct transmission of the hit data. Several e-link options are being envisaged by the CMS Collaboration. Different e-links were evaluated and compared and the results will contribute to the choice of the final data link implementation.

After the introduction of the IT system requirements driving the e-link design, different e-link prototypes are presented in Section 5.1. The electrical characterisation of the e-links as standalone electrical components is presented in Section 5.2 and their performance as a part of a readout system operating together with the readout chip is exposed in Section 5.3.

5.1 Inner Tracker electrical links

The architecture of the new IT readout system imposes demanding requirements on the readout links. After a short overview of the main signal integrity challenges in high speed serial links, the e-link system requirements are presented and the design of various e-link prototypes developed to meet these requirements is described in details.

5.1.1 Signal integrity challenges

The main signal integrity problems in high speed serial links are reflections, distortions and losses due to impedance mismatches along the transmission line. All interconnects for applications with clocks higher than 1 GHz should be designed as *uniform* transmission lines, i.e. they should have a constant cross section and controlled impedance in order to minimise reflections and optimise the signal quality. Non-uniform transmission lines will lead to signal-integrity problems and should be avoided, unless they are kept short enough [144].

The second main signal integrity issue in high speed serial links is the frequency-dependent loss from the conductor and the dielectric. In fact, the higher frequency signal components are more attenuated than the lower frequency components resulting in an increase of the rise time of the signal. When the rise time degradation approaches the period of a single bit, the digital information will be distorted, the bit is more likely to be misinterpreted by the receiver and the *bit error rate* (BER) will increase. This effect is called *inter-symbol interference* (ISI) and is a significant source of problems in high-speed serial links.

In addition to these two main signal integrity issues, other effects may occur in transmission lines. One of them is the skew, i.e. the time delay difference between two or more signal paths due to the difference in length or the variation in the dielectric constant. The skew between the two lines of a differential pair can cause signal distortions and contribute to the ISI. The other effect is the cross-talk, which is a capacitive and inductive coupling between two adjacent transmission lines that provides a path for unwanted noise from one line to the other and *electromagnetic interference* (EMI). Since the signal integrity issues scale with the data speed, it becomes an important concern for the new IT readout system.

5.1.2 System requirements

The Phase-2 IT readout system is schematically shown in Figure 5.1. The control of the pixel electronics will be ensured by the *downlink*, represented with grey arrows. The control signals issued by the DAQ will be sent at 2.56 Gb/s by the back-end DTC boards through optical links to the portcard, where they will be converted into electrical signals by the VTRx+. Then, the commands will be forwarded by the lpGBT chip at 160 Mb/s from the portcard through the e-links to the pixel modules. On the downlink path the lpGBT acts as a transmitter and the pixel modules as receivers.

The detector data readout will be ensured by the *uplink*, represented with red arrows. The pixel readout chips will send the hit data at 1.28 Gb/s from the pixel modules through the e-links to the portcard. At the portcard, the data from different e-links will be merged, serialised and transmitted by the lpGBT chip and converted to optical signals by the VTRx+. Then, the data will sent at 10.24 Gb/s through optical links to the back-end

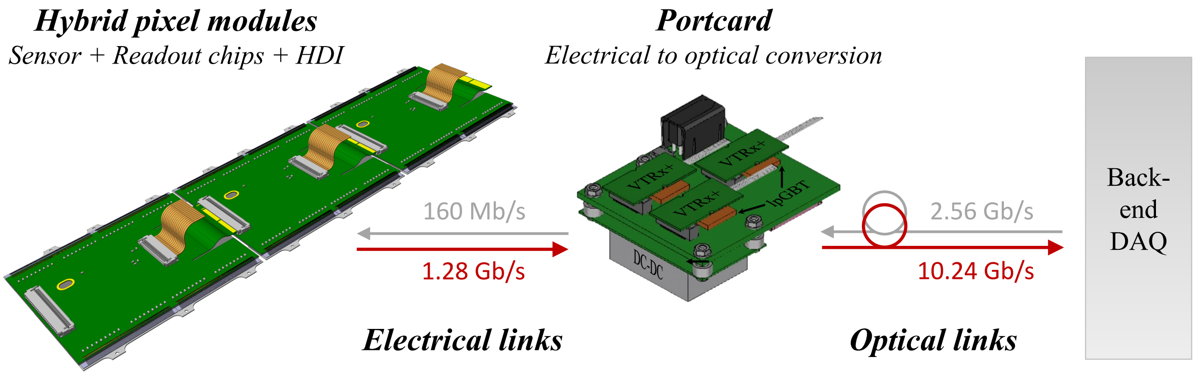


Figure 5.1: CMS Inner Tracker readout system architecture.

DAQ. On the uplink path the pixel modules act as transmitters and the lpGBT as a receiver.

For the entire IT readout system, 3892 e-links will be needed to control the pixel modules and 6908 e-links to read out the data. In total, 10 800 e-links will equip the detector. The minimum length of the e-links in the system will be 35 cm corresponding to the e-links reading out the inner rings of the TFPX and the maximum length will be 1.6 m needed for the readout of the TBPX ladders. Given the large amount of e-links needed, these cables have to be as low mass as possible to keep the material budget low and to limit the multiple scattering in the detector. However, the low mass will have an impact on the signal integrity, especially on the uplink with the data transmitted at high speed.

Another important aspect of the IT system is the serial powering (SP) of the pixel modules. Given that the pixel modules will be powered in series, different modules in an SP chain have different ground potentials and no common ground exist along the SP chain. Therefore, the e-links have to be AC-coupled to the pixel modules. For this reason, differential pairs with no return path will be used, which will also have an impact on the signal integrity. The data sent by the RD53 readout chips use a *non-return-to-zero* (NRZ) encoding and have no voltage reference.

The objective of the e-link development for the Phase-2 Upgrade is to minimise their mass, while keeping the signal amplitude and the jitter within acceptable levels for the lpGBT. The validation of the signal integrity is a crucial part of the pixel electronics system development. It is particularly relevant in the electrical part of the readout system, where the signals will be transmitted through ultra-low-mass cables. The scope of this chapter is therefore the e-links evaluation and validation of the signal integrity between the pixel modules and the portcard.

The desired system performance needs to be translated into an impedance requirement, for which the electrical links and interconnects will be designed. The value of impedance chosen for the system is not critical. For instance, a lower impedance will mean lower cross-talk and lower sensitivity to delays caused by connectors, components, and vias, but it also means higher power dissipation [144]. What is important is to have a controlled and constant impedance throughout the system. The target for the IT electrical readout system is the standard 100 Ω differential impedance. Therefore, the CMS e-links must have an impedance as close as possible to 100 Ω, to match with the impedance of the RD53 output transmitter of one side and with the lpGBT receiver on the other side.

5.1.3 Prototype E-links

The final e-links will be made of a maximum of seven pairs depending on the location in the detector, while the prototype cables for e-link evaluation and diverse system tests are made of 5 pairs, one pair for the downlink and four pairs to read out four chips of the prototype quad modules. Two e-link options are currently being explored by the CMS Collaboration: *twisted pairs* (TWPs) and *flexible flat cables* (FFCs).

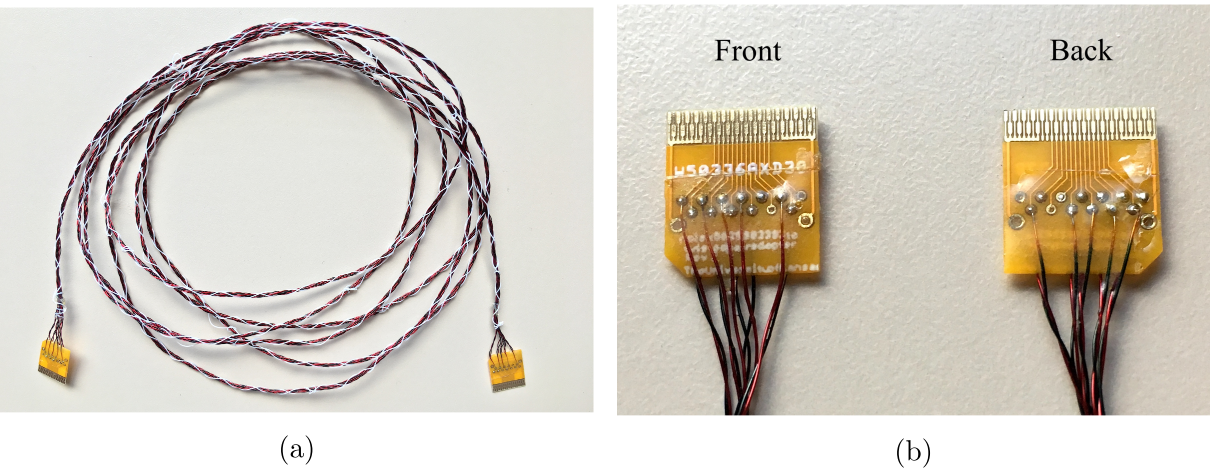
Twisted pairs

The TWP cables are the baseline solution for TBPX and TFPX e-links. The data links of the TEPX will be implemented in dedicated PCBs, currently under development and not addressed in this work. Along with their low mass, the TWPs offer an easy bending suitable for complex cable routing in the detector and particularly needed for the routing of the TBPX e-links from the barrel up to the TFPX service cylinder, where the corresponding portcards will be placed.

One TWP is made of two identical round wires twisted together. The spacing between the twisted wires is determined by the thickness of the wire insulation. One TWP corresponds to one differential line and several TWPs are twisted together to form a TWP cable. The TWP cables can be shielded or unshielded. The shield is used as the reference plane for the signal lines and it also provides an electrically conductive barrier to attenuate electromagnetic waves external to the shield to minimise the EMI issues.

The CMS TWP prototypes are unshielded to minimise their mass. Hence, the EMI and cable cross coupling effects may occur and cause signal cross-talk, deteriorating the signal quality. The prototypes are composed of five twisted pairs made of copper wires with double QML Kapton coating for insulation. A white plastic wire is braided around the TWPs to maintain them together and preserve the twisting. A photograph of a prototype TWP cable is shown in Figure 5.2(a).

The TWPs were produced in different lengths from 35 to 200 cm to validate their transmission performance and to demonstrate the feasibility of the new readout system. They were designed to be compatible with the prototype RD53A quad modules, on which a 33-pin Molex connector [145] is used for the readout link. Therefore, the TWPs are soldered on so-called *paddle boards*, shown in Figure 5.2(b), allowing their connection to this con-



(a) (b)

Figure 5.2: Photograph of a 2m-long twisted pair cable prototype (a) and zoom to the paddle boards (b).

nector. The size of the paddle boards is $10.2 \times 11.5\,\mathrm{mm}^2$ and they are about 0.2 mm thick. An epoxy coating is used to protect the wire solders on both sides of the paddle boards. The total weight of one paddle board equals to 0.45 g.

The wires of a TWP cable can have different thicknesses. The cross-sectional area of the wire determines the maximum current that a conductor can safely carry. The diameter of the TWPs is given by the *american wire gauge* (AWG). An increasing AWG indicates a decreasing wire diameter. Two different wire diameters are under evaluation for the CMS e-links: 34 AWG and 36 AWG, corresponding to a wire diameter of 160 µm and 127 µm, respectively. The cross section of both options, precising the wire diameter and the insulation thickness, is shown in Figure 5.3. The 34 AWG TWPs are slightly thicker and therefore, mechanically more robust. They have 6 twists per inch, while the 36 AWG TWPs have 4 twists per inch. The average mass per differential pair of the TWP prototypes is 0.39 g/m for the 34 AWG and 0.29 g/m for the 36 AWG.

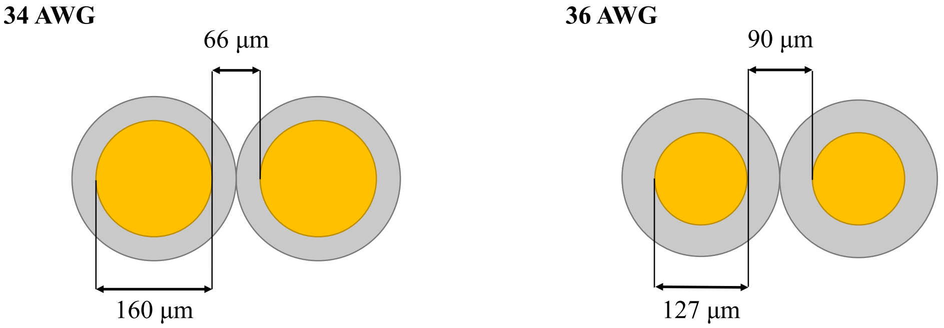


Figure 5.3: Cross section of a 34 AWG twisted pair cable (left) and a 36 AWG twisted pair (right). The copper wires are represented in yellow and the insulation in grey.

Flexible flat cables

An FFC is a type of ribbon cable consisting of a plastic film with multiple metallic coplanar lines. Each differential line is made of two narrow metal strips side by side and the spacing between them is referred to as the *pitch*. This type of cable usually offers a good EMI suppression and the wire coupling is very small, which limits the cross-talk issues. The FFCs can be designed to fit exactly the detector layout, and therefore they are good candidates for the readout of the TFPX discs. The advantage of FFCs over the TWPs is that the design can be very dense minimising the mass per differential pair. On the other hand, they are expected to attenuate more the signal amplitude and the high frequencies, in particular with increasing e-link length. Hence, they are not considered good candidates for the TBPX, where the e-links will be up to 160 cm long, compared to the maximum length of 50 cm in the TFPX.

The CMS FFC prototypes are made of Kapton polyimide layers and copper traces. They are terminated with integrated flat connectors that fit in the 33-pin Molex connector of the prototype modules. Three different FFC topologies were designed and produced for the e-link evaluation and the readout system demonstration and they are shown in Figure 5.4. The straight FFC was designed to read out one prototype module and therefore it features five differential lines: one for the downlink and four for the up-links, i.e. one data lane per readout chip. The straight FFC is 35 cm long and its mass per differential pair equals to 0.7 g/m. The copper traces are 50 µm thick and 152 µm wide. The distance between

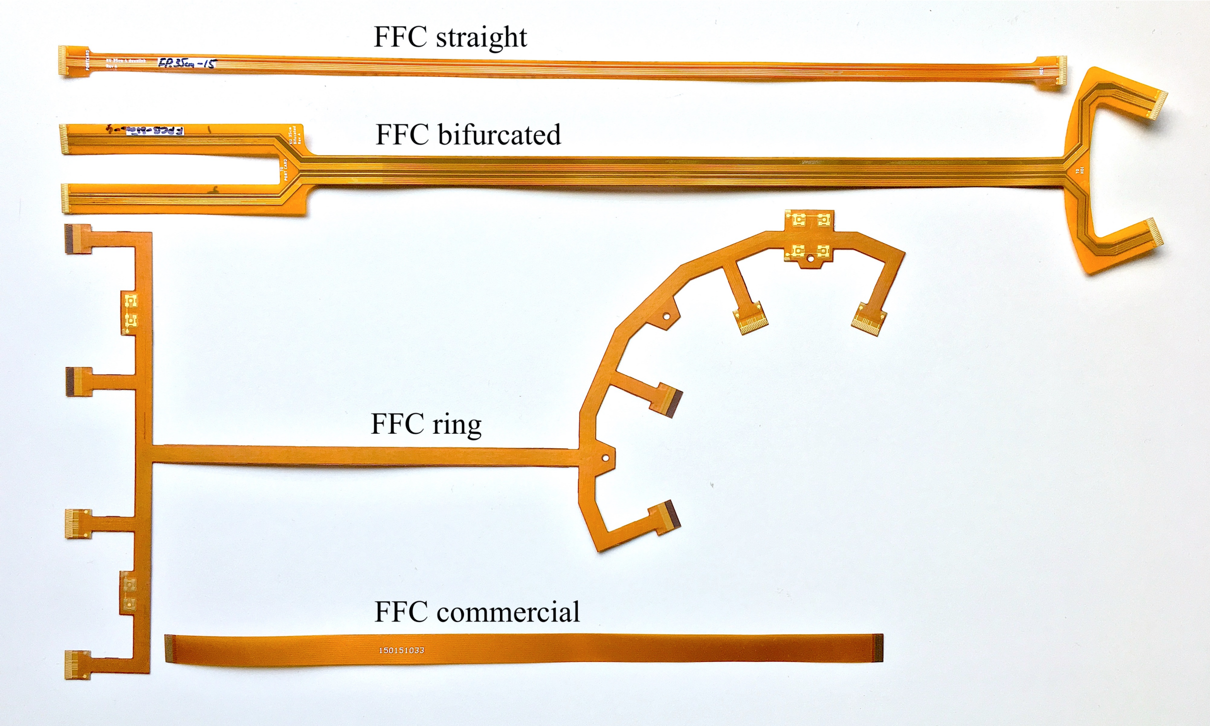


Figure 5.4: Photograph of the CMS flexible flat cable prototypes.

traces is 100 µm within a differential pair and 200 µm between differential pairs. The traces are implemented in one plane, while a copper mesh in the back plane is used as a reference plane. The bifurcated FFC was designed to read out two prototype quad modules arranged in a ring geometry. Hence, it contains ten differential lines. It is based on the same trace design. Therefore, it has the same mass per differential pair, but the signal traces are longer measuring about 40 cm.

The ring FFC was designed to readout up to four prototype quad modules arranged in a ring geometry. It contains 22 differential lines: 4 downlinks, $4 \times 4 = 16$ uplinks (one per readout chip) and two extra differential pairs for signal quality verification. The two extra lines are not routed to the 33-pin insertions for the module connectors, but to eight test points, where a micro-miniature coaxial connectors (MMCX series [146]) can be mounted for an easy connection to measurement instruments. The copper traces are 18 µm thick and 100 µm wide. The distance between traces is 50 µm within a differential pair and 170 µm between differential pairs, which makes this design more dense than the previous one. Twelve differential lines are implemented in one plane and ten in another plane. A 15 µm-thick aluminium layer is placed in between and is used as a reference plane. The layout of this FFC e-link results in four different trace lengths depending on the output: 26 cm, 31 cm, 36 cm and 38 cm. The total weight of this e-link is 2.25 g, which corresponds to 0.31 g/m per differential pair.

In addition to these CMS FFC prototypes, a commercially available FFC was also evaluated for comparison. The FFC commercialised by Molex [147], shown in Figure 5.4, was used since it was the longest available commercial FFC compatible with the 33-pix Molex connector on the modules. It is 25 cm long and it contains 33 single copper traces, i.e. up to 16 differential pairs. The traces are 18 µm thick and 200 µm wide and they are all

Table 5.1: Characteristics of the available e-link prototypes.

E-link type	**Diff. pairs**	**Trace size (μm)**	**Trace dist. (μm)**	**Ref. plane**	**Length (cm)**	**Mass per pair (g/m)**
TWP 36 AWG	5	⌀127 µm	90	none	35–200	0.29†
TWP 34 AWG	5	⌀160 µm	66	none	35–200	0.39†
FFC straight	5	55 × 152	100 (200)*	solid Cu	35	0.70
FFC bifurcated	10	55 × 152	100 (200)*	solid Cu	40	0.70
FFC ring	22	18 × 100	50 (170)*	15 µm Al	26–38	0.31
FFC commercial	16	18 × 200	100 (100)*	none	25	0.15

[*] The first number corresponds to the distance between traces of the same differential pair, while the number between parentheses corresponds to the distance between traces of two adjacent differential pairs.

[†] The mass of two paddle boards should be added per TWP cable for a fair comparison with the FFCs, which have the integrated connections already included in their mass.

equidistant with 100 µm between them. No reference plane is present in the commercial FFC and its total weight equals to 0.6 g, which corresponds to 0.15 g/m per differential pair. The characteristics of the above described e-links are summarised in Table 5.1.

5.2 Electrical characterisation of the e-links

The standard way to evaluate the electrical properties of a transmission line or an interconnect consists in sending a well-defined signal, measuring the received response and establishing the behavioural model. The behavioural model describes the way the transmission line affects the incident signal. It can be measured either in time domain, using a time-domain reflectometer, or in frequency domain, using a vector network analyzer. The second option was adopted for the e-link electrical characterisation addressed in this section.

5.2.1 The vector network analyzer setup

A *vector network analyzer* (VNA) is an instrument allowing to measure the interaction of a well-defined incident wave with a transmission line. It sends sine waves with different frequencies within a selected frequency range into the device under test (DUT) and it measures the reflected and transmitted waves. Hence, the VNA is composed of a signal generator and one or more signal receivers. The word "vector" in the instrument's name refers to the fact that both the magnitude and phase of the reflected and transmitted waves are measured [144, p. 558].

The Keysight N5225A PNA Microwave Network Analyzer [148] was used to characterise the CMS prototype e-links presented in the previous section. The test setup is shown in Figure 5.5. The instrument has four terminals or ports that can be used to either characterise two single-ended lines or one differential pair. Each port emits a sine-wave voltage at frequencies that can range from 10 MHz to 50 GHz. At each frequency, the amplitude and phase of the incident, reflected and transmitted waves are measured.

The long blue SMA cables with controlled 50 Ω impedance are a part of the instrument. They serve as the resistive load and termination and they allow to connect the DUT to the VNA. Prior to the measurement, the instrument has to be calibrated to compensate for the electrical properties of the connection cables and to exclude them from the measurement. The calibration is done using an electronic calibration module, also shown in Figure 5.5. The connection cables are connected two by two to the SMA connectors of the calibration module, hidden under the black covers on the photograph, and each pair is calibrated separately.

The e-links to be measured need an SMA connection to be connected to the instrument.

For that, an adapter board converting the 33-pin Molex connector to SMAs is used on each end of the e-link as it can be seen in Figure 5.5. The ends where signals enter or exit are called *ports*. The ports number 1 and 3 of the VNA are connected to one side of the e-link under test and the ports number 2 and 4 are connected to the other side. This way one differential pair of each e-link is tested. The measured response of the DUT describes the electrical properties of the e-link, but inevitably, it also includes a contribution from the two adapter boards that could not be removed in the calibration step.

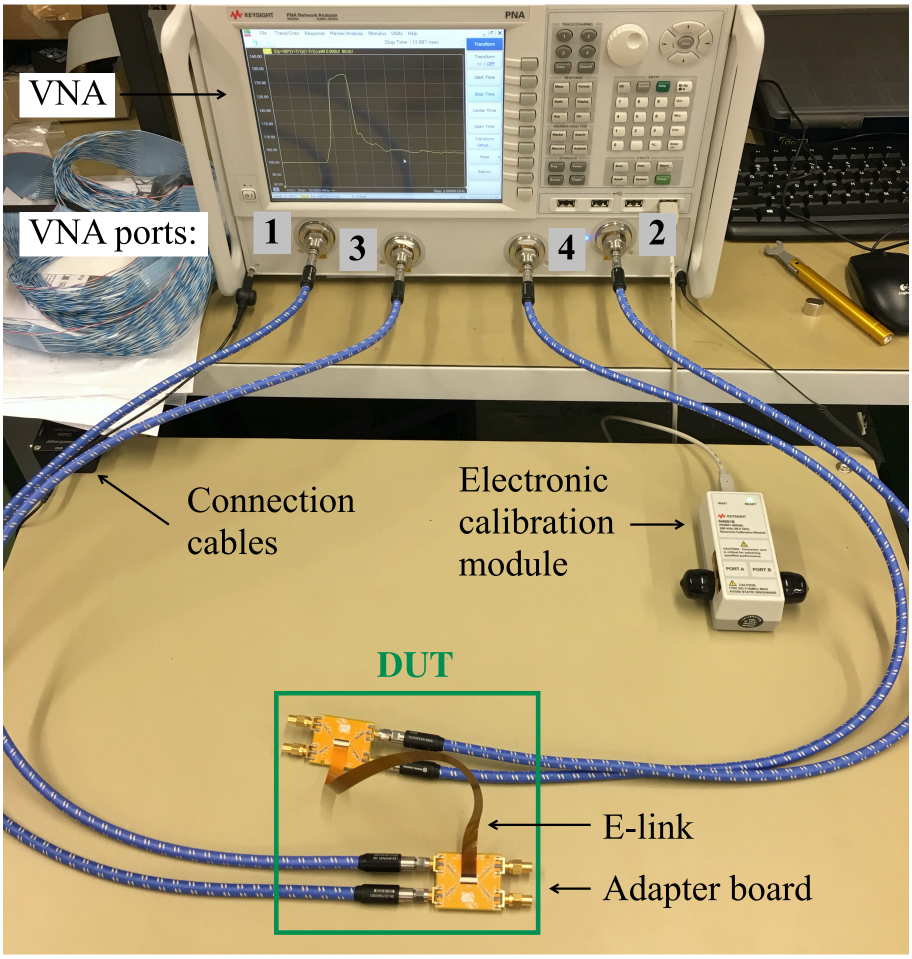


Figure 5.5: Photograph of the vector network analyzer setup.

5.2.2 S-parameters

In the frequency domain, the well-defined incident waveform is typically the sine wave and the behavioural model is described by the scattering parameters, also known as *S-parameters*. The S-parameters represent a formalism describing how the precision sine waveforms of different frequencies scatter from the ends of a transmission line. The electrical behaviour of any linear passive interconnect can be described with the S-parameters. They represent a complete description of the way the electrical signals interact with the interconnect and provide useful information about the impedance, the signal attenuation, the reflections or the amount of cross-talk.

Each S-parameter is the ratio of the output sine wave to the input sine wave. For all linear, passive elements, such as the e-links, the frequency of the output wave will be exactly the same as the frequency of the input wave, only the amplitude and phase of the wave will change. Hence, the ratio between the two sine waves is a complex number. The magnitude of an S-parameter can be obtained as the ratio of the amplitudes of the output to the input sine waves, in which case it is a number between 0 and 1. However, it is more often expressed in decibels as:

$$\text{mag(S) [dB]} = 20 \times \log\left(\frac{\text{amplitude(output sine wave)}}{\text{amplitude(input sine wave)}}\right) \tag{5.1}$$

The phase of the S-parameter corresponds to the phase difference between the output wave and the input wave expressed in degrees as:

$$\text{phase(S) } [^\circ] = \text{phase(output sine wave)} - \text{phase(input sine wave)} \tag{5.2}$$

The focus of this work is put on the magnitude of the S-parameters and from now on the phase is not further discussed.

Each S-parameter is the ratio of a sine wave scattered from the DUT at a specific port, to the sine wave incident to the DUT at a specific port. Therefore, they are denoted "Soi" where "o" indicates the output port number and "i" indicates the input port number. For example, the S-parameter for the sine wave going into port 1 and coming out of port 2 would be S21.

A DUT with only one port would have only one S-parameter, which is S11. A two-port device would have four possible S-parameters. When the incident wave is sent into port 1, the available S-parameters are S11 and S21. When instead the incident wave is sent into port 2, the S22 and S12 can be measured. For a four-port device, 16 S-parameters exist. The assignment of the S-parameters for a four-port device with the incident wave sent into port 1 is illustrated in Figure 5.6(a). In general, if the transmission line is not physically symmetrical, S11 will not equal S22. However, for all linear, passive devices, S21 will always be equal to S12. Therefore, a DUT with N different ports has N^2 S-parameters out of which $N(N+1)/2$ are unique terms.

The wave that scatters back to the source is called the *reflected wave* and the wave that scatters through the device is called the *transmitted wave*. The reflected and transmitted waves lead to the two most important S-parameters, which have specific names. The ratio of the reflected to the incident wave is called the *return loss*. It represents the sum of all the possible combinations of reflections occurring at all the impedance changes the signal encounters along the transmission line. The return loss is denoted S11 and is defined as [144]:

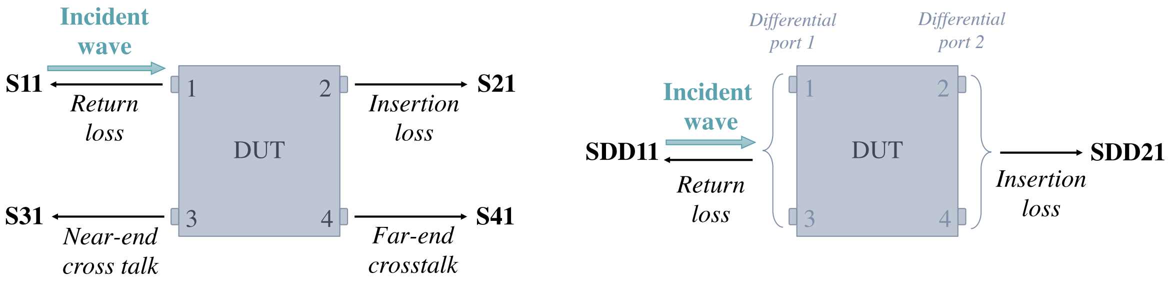
$$\mathrm{S11} = \frac{Z - Z_0}{Z + Z_0} \tag{5.3}$$

where Z is the instantaneous impedance at a given point of the transmission line and Z_0 is the characteristic impedance.

The ratio of the transmitted to the incident wave is called the *insertion loss*. It is a measure of the signal attenuation and it contains the information about losses, discontinuities, and couplings to other lines. The insertion loss is denoted S21 and it can be defined as [144]:

$$\mathrm{S21} = \sqrt{1 - \mathrm{S11}^2} \tag{5.4}$$

For any single frequency, the return loss or the insertion loss is a steady state value. They are typically measured as a function of frequency. The other two S-parameters indicated in Figure 5.6(a), i.e. S31 and S41, give an information about the cross-talk between adjacent transmission lines.



(a) Single-ended.

(b) Differential.

Figure 5.6: The assignment of the S-parameters when the incident wave is sent into port 1 of a four-port device in the single-ended configuration (a) and the differential configuration (b).

Two adjacent transmission lines can also be used as a differential pair and the associated S-parameters are called *differential S-parameters*. The definitions and notations given previously for the single-ended S-parameters are valid also for the differential ones, except that the “Soi” notation becomes “SDDoi” to indicate that the input and output waves are differential signals. A four-port DUT with two single-ended lines can also be described as one differential pair with two differential ports, as illustrated in Figure 5.6(b). In this case four pure differential S-parameters exist: SDD11, SDD12, SDD21 and SDD22.

The lab instruments, such as the VNA only measure the single-ended S-parameters, which can be mathematically transformed into the differential ones. The VNA was connected to the e-links in the same way as for a single-ended measurement, but it was configured to display and record the differential S-parameters. The differential return loss is obtained from the single-ended S-parameters as [144, p. 609]:

$$\mathrm{SDD11} = 0.5 \times (\mathrm{S11} + \mathrm{S33} - 2 \times \mathrm{S31}) \tag{5.5}$$

And the differential insertion loss is given by [144, p. 609]:

$$\text{SDD21} = 0.5 \times (\text{S21} + \text{S43} - \text{S41} - \text{S23}) \tag{5.6}$$

5.2.3 S-parameter measurement

The differential return loss and insertion loss were measured for the CMS e-links using the VNA setup described in Section 5.2.1 and shown in Figure 5.5. First, two standard 1 ns commercial SMA cables of 20 cm were measured and used as a reference. One of the cables was connecting the port 1 and 2 of the VNA and the other was connecting ports 3 and 4. The VNA was configured to take these two cables as a differential pair in the same configuration as illustrated in Figure 5.6(b). The measured differential S-parameters of the SMA cables are presented in Figure 5.7 as a function of frequency. The frequency range was scanned from 10 MHz to 2 GHz, with a step of 10 MHz. The return loss is shown in orange and the insertion loss in green. Both were measured with the incident wave sent into the differential port 1, as well as into the differential port 2, which results in four S-parameters.

As expected, the two insertion losses are exactly the same, i.e. SDD12 = SDD21 and they are very close to 0 dB over the measured frequency range. This implies that the impedance is perfectly matched to the impedance of the VNA ports and constant along the transmission line. The cable is transparent to the signal and the transmission is close to 100%. The two return losses are very similar, which indicates that the transmission line is symmetrical. In addition, the return loss is comprised between −50 dB at low frequency and −30 dB at higher frequency. The low return loss confirms a well controlled impedance in the commercial SMAs causing only small reflections. The reflections from

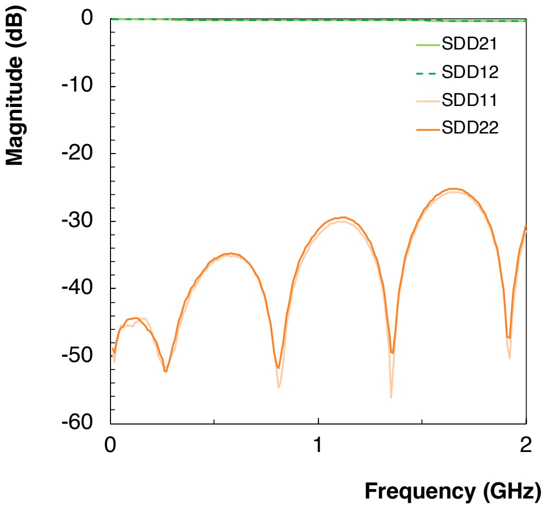


Figure 5.7: Differential S-parameters measured for a pair of commercial 1 ns SMA cables with the vector network analyzer.

the impedance mismatches in the connectors at both ends are visible as periodic oscillations in the return loss. The period with which the oscillations repeat is proportional to the length of the cable.

The same measurement was performed for all available e-link prototypes and the results are shown in Figure 5.8. The 35 cm TWPs of 36 AWG and 34 AWG are presented in Figures 5.8(a) and 5.8(b), respectively. The commercial FFC, the straight FFC, one output of the bifurcated FFC and the shortest output of the ring FFC are presented in Figures 5.8(c) to 5.8(f), respectively.

Return loss

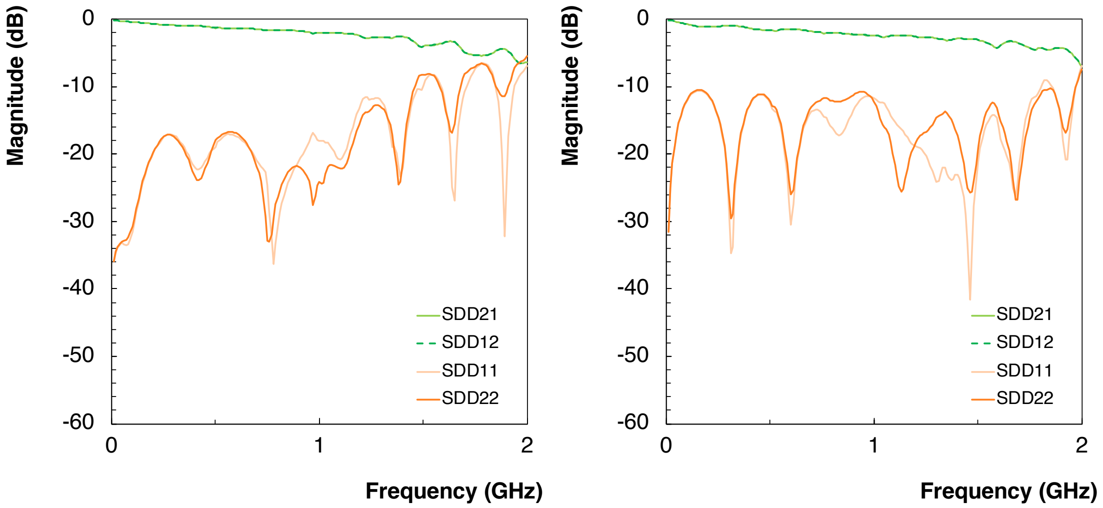
The general shape of the return loss shown in orange in Figure 5.8 is similar for all the e-links and is also similar to the return loss of the reference SMA cables. However, the return loss of the prototype e-links is on average comprised between -30 dB and -10 dB, which is higher than for the SMAs. Hence, the prototypes cause more signal reflections. In the case of the commercial FFC, for instance, the return loss becomes higher than the insertion loss close to 2 GHz, so more signal is reflected than transmitted at high frequency.

As mentioned previously, the oscillations in the return loss correspond to the signal reflections from the e-link ends and their period is proportional to the e-link length. The frequency of the oscillations increases with increasing length. The commercial FFC, which is the shortest e-link with its 25 cm shows only five oscillations, the short output of ring FFC measuring 26 cm shows seven oscillations and the longest of the measured e-links, which is the bifurcated FFC with 40 cm counts up to ten oscillations. Some irregularities in the return loss oscillations are observed, in particular with both flavours of TWPs and with the commercial FFC as well.

The two return losses SDD11 and SDD22 are very similar for the commercial and the straight FFCs because their straight flat shape makes them symmetrical. In the TWPs the connections to the paddle boards are not perfectly the same on both ends, which causes some asymmetry and results in a difference between SDD11 and SDD22. The two return losses are also different for the bifurcated and ring FFCs, where the routing topology is not the same on the module side and on the portcard side, as it can be observed in Figure 5.4.

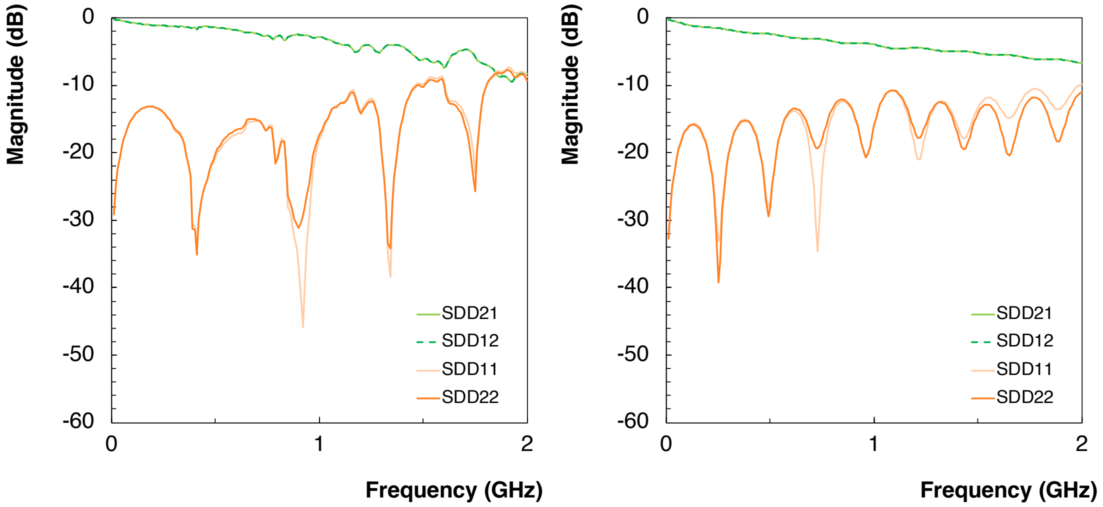
Insertion loss

The differential insertion loss describes the most important property of a differential pair. It is a measure of the transmitted signal and therefore, it shows the attenuation induced by the e-link. The two insertion losses SDD12 and SDD21, shown in green in Figure 5.8



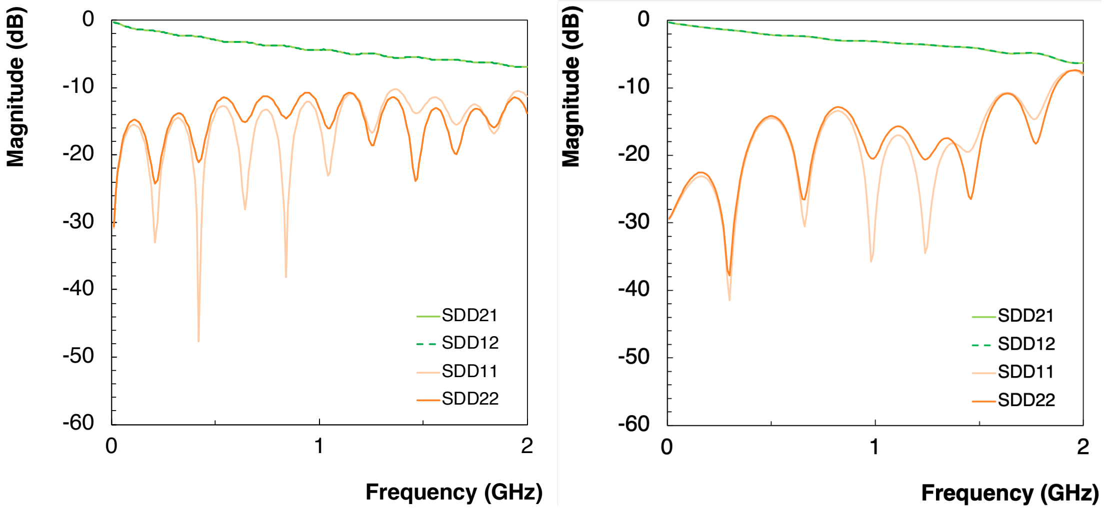
(a) TWP 36 AWG.

(b) TWP 34 AWG.



(c) FFC commercial.

(d) FFC straight.



(e) FFC bifurcated.

(f) FFC ring.

Figure 5.8: Differential S-parameters measured for all prototype e-links using the vector network analyzer.

are equal to each other in all the e-links. It is very close to 0 dB at low frequency and when the frequency increases the insertion loss decreases, so less and less signal is transmitted. This demonstrates a higher signal attenuation at higher frequency.

The general trend of the measured insertion loss is very similar for all the prototype e-links. For an easier comparison, the SDD21 of all the e-links was reported to Figure 5.9. The 35 cm TWPs are represented in blue; the 36 AWG in light blue and the 34 AWG in dark blue. The commercial FFC is represented in khaki colour, the straight FFC in yellow, the two outputs of the bifurcated FFC in orange and the four outputs of the ring FFC in pink. The SDD21 of the SMA cables is added for reference and is shown in green.

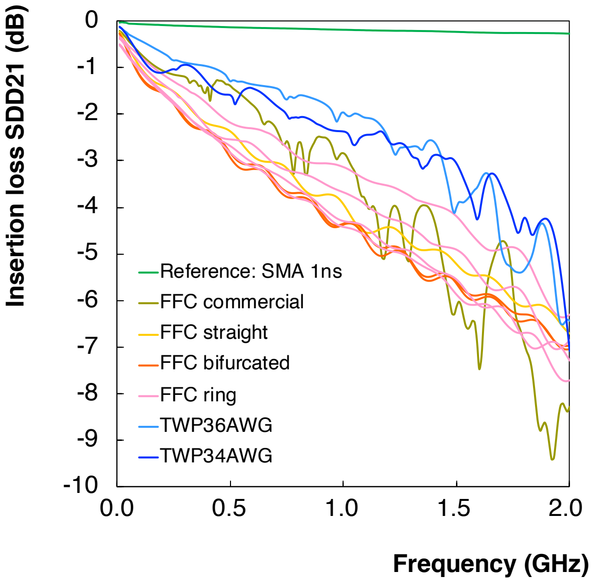


Figure 5.9: Differential insertion loss (SDD21) measured as a function of frequency for the prototype e-links using the vector network analyzer.

The insertion loss is mainly dominated by two factors. The conductor and dielectric losses give rise to a general monotonic decrease with increasing frequency. Hence, the slope of the insertion loss is related to the conductor and dielectric materials. A larger slope indicates a higher dissipation factor. The slope of insertion loss is similar for the two flavors of TWPs, since they are made of the same material. The different FFCs also have a similar slope since they are made of similar materials.

The second factor affecting the insertion loss is the impedance. When the impedance of the e-link is well matched to the impedance of the ports of the VNA, the transmission line is transparent to the signal, the signal is 100% transmitted and the insertion loss is very close to 0 dB, as it is the case for the SMA cables reaching only −2.6 dB at the highest measured frequency. The larger the impedance mismatch, the larger are reflections, and the less signal is transmitted, resulting in larger absolute value of the insertion loss in dB. Hence, there is a relationship between the insertion loss and the return loss that become particularly relevant at high frequency.

Other factors that can come into play in the insertion loss measurement are: the cross talk and the mode conversion occurring mainly when the two lines of a differential pair are not exactly of the same length. Overall, the TWPs demonstrate a higher insertion loss than the FFCs, which means that the signal is less attenuated in the TWPs. The 36 AWG TWP has a slightly higher insertion loss than the 34 AWG TWP up to 1.5 GHz. Among the FFCs, the two shorter outputs of the ring design attenuate less then the straight design and the bifurcated design attenuates the most, which is at least partially related to the e-link length.

5.2.4 Impedance measurement

The impedance of the e-link prototypes was also measured using the VNA setup described in Section 5.2.1 and shown in Figure 5.5. The port impedance of the instrument when used to measure single-ended S-parameters is 50 Ω. When two ports of the instrument drive a differential signal, the outputs are in series and the differential port impedance for the differential signal is 100 Ω. Therefore, the incident wave generated by the VNA sees an impedance of 100 Ω, when it propagates through the connection cables. Any impedance mismatch between the DUT and the connections cables will cause reflections and increase the differential return loss SDD11. This S-parameter also contains the information about impedance discontinuities of the DUT and can therefore be used to extract the impedance. Based on Equation (5.3) the impedance can be expressed as:

$$Z = Z_0 \frac{(1 + \mathrm{SDD11})}{(1 - \mathrm{SDD11})} \tag{5.7}$$

where $Z_0 = 100\,\Omega$ is the characteristic impedance of the VNA and its connection cables and SDD11 is the differential return loss expressed in linear scale, i.e. as a fraction between 0 and 1 and not in dB.

The impedance extracted from the return loss can be displayed in the frequency or time domain. In the frequency domain, it corresponds to the total, integrated impedance of the entire DUT measured for different frequencies. In the time domain, it corresponds to the instantaneous impedance the signal sees at each spatially distinct point of the transmission path through the DUT. The latter provides the full impedance profile along the transmission line, hence it was selected for the characterisation of the e-links. The differential return loss SDD11, measured by the VNA, was converted to time domain and transformed into the impedance time profile using Equation (5.7).

First of all, two standard 1 ns commercial SMA cables, with a length 20 cm and a controlled impedance of 50 Ω each, were measured and used as a reference. One of the cables was connecting the port 1 and 2 of the VNA and the other cable was connecting ports 3 and 4. The VNA was configured to take these two cables as a differential pair with a differential impedance supposed to be 100 Ω. The measured instantaneous impedance of the reference cables is presented in Figure 5.10 as a function of time.

The $t = 0$ represents the time when the incident signal leaves the connection cables and enters the DUT. Therefore, the impedance measured for $t < 0$ represents the controlled impedance of the connection cables of the VNA and is very close to 100 Ω. At $t = 0$, the impedance increases by about 1 Ω due to the connection between cables. The second impedance spike indicates the other connection at the far end. The impedance between the two spikes is the one of the reference cables and it is indeed very close to the expected 100 Ω. The time interval between the two connections visible in the impedance profile is 2 ns, which corresponds to twice the time delay of the SMA cables. Since the impedance is measured from the return loss, the signal launched into the line travels down the line to the termination, where it reflects and comes back to the source. The reflected signal will appear back at the source in a round trip time equal to twice the time needed to propagate from one end of the cable to the other.

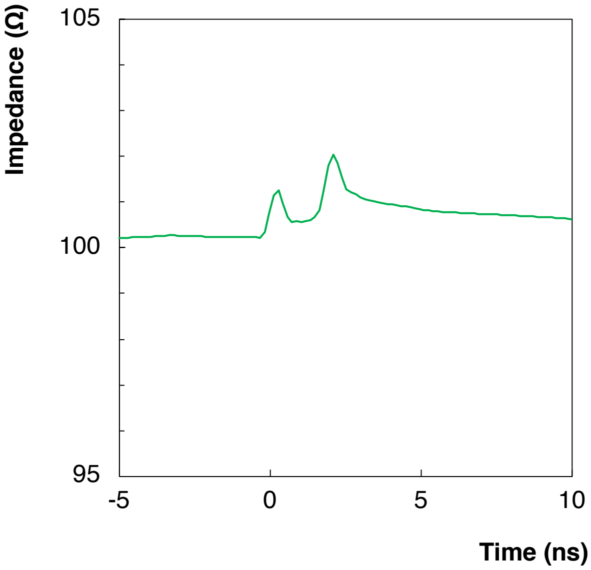
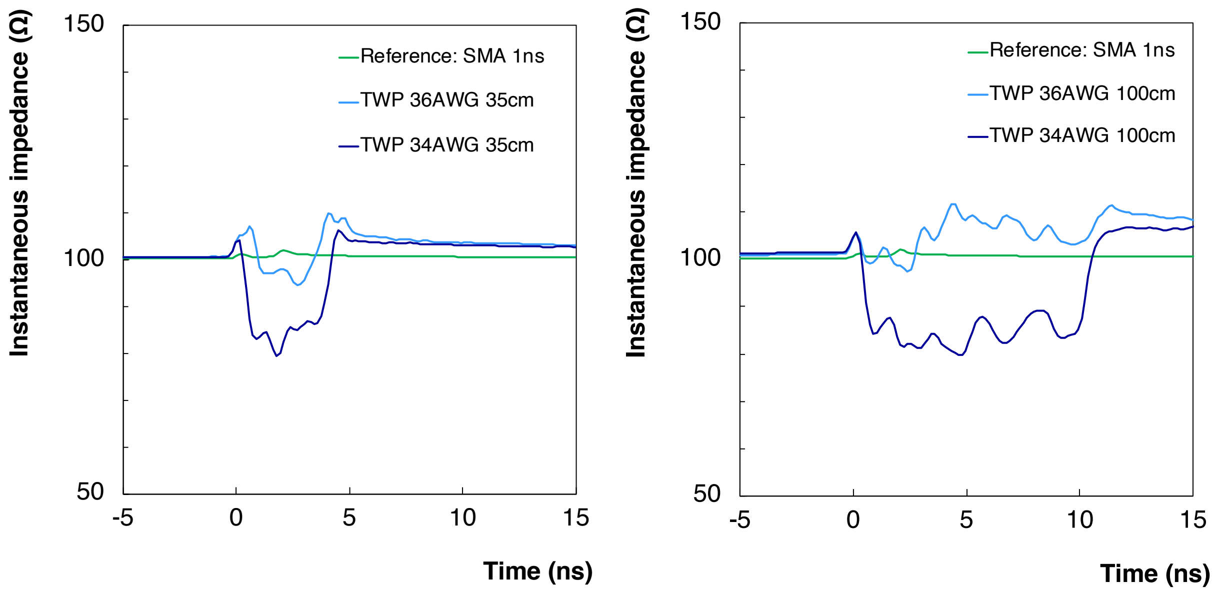


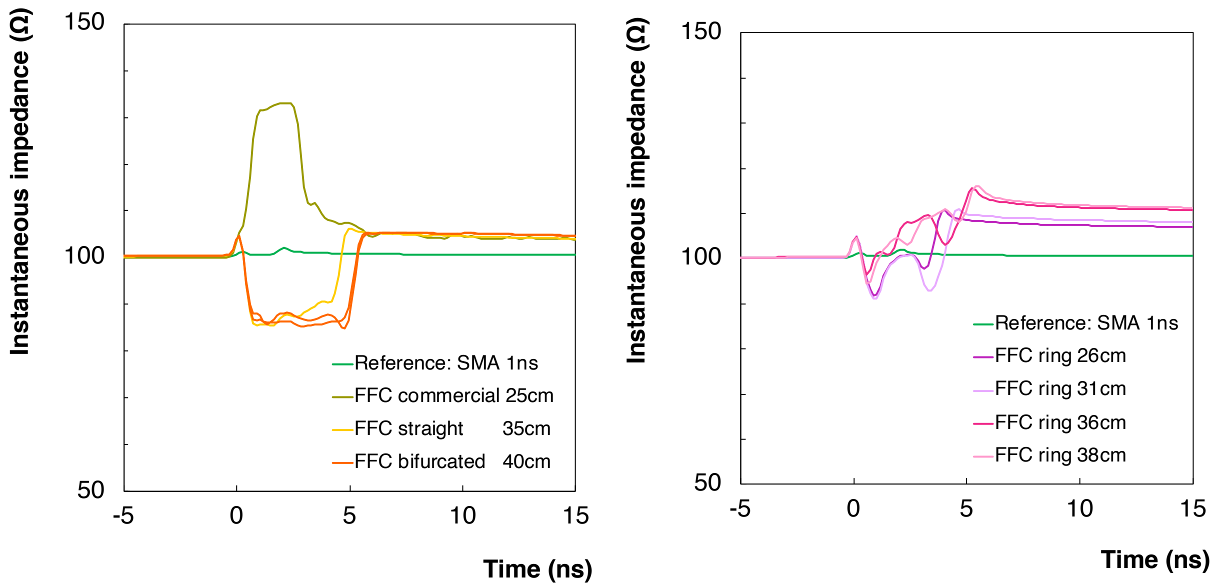
Figure 5.10: The impedance profile of a 1 ns commercial SMA cable measured with the vector network analyzer and taken as a reference.

The same measurement was done for all the available prototype e-links and the instantaneous impedance measured as a function of time is shown in Figure 5.11. The measured impedance of the 1 ns SMA cables is reported in green for reference. The 35 cm TWPs are presented in Figure 5.11(a). The measurement was done for both TWP flavours: the 36 AWG is presented in light blue and the 34 AWG in dark blue. When the signal enters the DUT, at $t = 0$, it first encounters the adapter board, which causes an impedance spike of about 105 Ω. Then the signal enters the TWP e-link and the impedance decreases and remains more or less constant during about 4 ns. A second spike indicates the adapter board on the far end, after what, the impedance returns slowly back to 100 Ω. The average impedance of the 36 AWG TWP is roughly 95 Ω, while for the 34 AWG it is around



(a) 35 cm TWPs.

(b) 100 cm TWPs.



(c) FFCs.

(d) FFC ring.

Figure 5.11: The instantaneous impedance as a function of time measured by the vector network analyzer for different CMS e-links: the 35 cm twisted pairs (a), 100 cm twisted pairs (b), the commercial, straight and both outputs of the bifurcated flexible flat cable (c) and the four outputs of the ring flexible flat cable (d).

80 Ω. The 36 AWG TWP is better matching the target 100 Ω impedance, which suggests a better signal integrity in the thinner TWPs.

The measurement was done also for the 100 cm TWPs and the result is presented in Figure 5.11(b). In this case, the time the signal needs to propagate through the e-link and back is about 12 ns. In fact, the signal propagates down the transmission line at the speed of light in the material surrounding the conductors. Hence, the signal speed primarily depends on the dielectric constant of the insulation and can be calculated as [144, p. 217]:

$$v = \frac{1}{\sqrt{\varepsilon_0 \varepsilon_r \mu_0 \mu_r}} = \frac{c}{\sqrt{\varepsilon_r \mu_r}} \simeq \frac{30}{\sqrt{\varepsilon_r}} [\text{cm/ns}] \tag{5.8}$$

where $\varepsilon_0 = 8.89 \times 10^{12}\,\text{F/m}$ is the permittivity of free space; ε_r is the relative dielectric constant of the material; $\mu_0 = 4\pi \times 10^{-7}\,\text{H/m}$ is the permeability of free space and μ_r is the relative permeability of the material. In general, for all interconnect materials, the relative permeability μ_r of dielectrics is close to one, and therefore it was neglected in Equation (5.8). The dielectric constant ε_r equals to one only in the air, hence the signal speed in the air is about 30 cm/ns. In all real interconnect materials the ε_r is always greater than one and the signal propagates slower than in the air [144]. The dielectric constant of the Kapton polyimide used as dielectric in the CMS e-links is $\varepsilon_r = 3.4$. Consequently, the signal speed in the e-links is approximately $v \simeq 30/\sqrt{3.4} \simeq$ 16.3 cm/ns. The round trip time of the signal in a 35 cm e-link is therefore: $t = 2 \times L/v \simeq 2 \times 35/16.3 \simeq 4\,\text{ns}$. While for a 100 cm e-link it is: $t = 2 \times L/v \simeq 2 \times 100/16.3 \simeq 12\,\text{ns}$.

In terms of impedance, the 100 cm 36 AWG TWP has an instantaneous impedance slightly lower than 100 Ω close to the near end and slightly higher than 100 Ω in the rest of the length. Overall, the measured impedance of the 100 cm 36 AWG TWPs matches the target 100 Ω impedance within ±10%, while the average impedance of the 34 AWG TWPs of the same length is around 85 Ω. This results are in agreement with the impedances measured for the 35 cm TWPs.

The impedance measurement of the FFCs is shown in Figure 5.11(c). The commercial FFC is represented in khaki colour, the straight FFC in yellow and the two outputs of the bifurcated FFC are both in orange. The 1 ns SMA is still presented in green for reference. The impedance of the commercial FFC is constant and equal to 130 Ω. Since this FFC is the shortest one, the round trip time of the signal is the shortest. The straight and bifurcated FFCs have very similar impedance profiles, because they are based on the same design. Their average impedance is about 85 Ω, which is 15% off with respect to the target 100 Ω impedance, while for the commercial FFC the mismatch is twice as big. However, the uniformity of the instantaneous impedance along the transmission line matters as well. Hence, the commercial FFC might show good signal integrity even though its integrated impedance is not well matched to the nominal 100 Ω impedance.

Finally, the impedance measurement of the four outputs of the ring FFC is presented separately in Figure 5.11(d) for the sake of visibility. The two shorter outputs of 26 cm and 31 cm are shown in dark and light violet, respectively, and the two longer outputs of 36 cm and 38 cm are shown in dark and light pink, respectively. Overall, the impedance of the ring FFC matches the target 100 Ω impedance within ±10%. The impedance of the shorter outputs is very close to 100 Ω and is constant along the e-link length. On the other hand, the impedance of the two longer outputs varies along length and is slightly higher.

This difference could be due to the fact that the two shorter outputs are implemented in one plane, while the two longer ones are implemented in another plane, where test structures are also present. In conclusion, only the 36 AWG TWP and the ring FFC e-link designs are close to the nominal impedance.

5.3 E-links performance with pre-emphasis

In addition to the electrical characterisation of the e-links as standalone components, it is also important to evaluate their performance as a part of a test system. In this section the effect of the e-links on the signal generated by the RD53A readout chip, running at the nominal data rate of 1.28 Gb/s is evaluated and the pre-emphasis functionality available in the readout chip is explored.

5.3.1 Test set-up

The signal sent through the e-links to evaluate their performance was generated by the RD53A chip. The chip can generate four different output signals: the serial data encoded with Aurora 64/66b protocol, a 640 MHz clock, a signal with only zeros or a test pattern [87]. The test pattern is a *pseudo random bit sequence* (PRBS) that can be used to model a well balanced stream of data to test the high-speed serial links. The PRBS repeats every N bits. Different PRBS signals exist and they are denoted as PRBSk, where the length of the repeated sequence is given by $N = 2^k - 1$. The sequence is composed of $2^{(k-1)}$ ones and $2^{(k-1)} - 1$ zeros, separated by $2^{(k-1)}$ transition edges [149]. The sequence contains all combinations of k bits except for all zeros. In particular, a unique sequence of k ones appears in the sequence and can be used as trigger.

The test pattern provided by the RD53A chip is a PRBS7 generated by a 7-bit *linear-feedback shift register* (LFSR) [87]. Therefore, each sequence is 127 bits long and is composed of 64 ones, 63 zeros and 64 edges. It contains exactly one sequence of seven ones. Two consecutive sequences of the PRBS7 from the RD53A chip displayed on the scope are shown in Figure 5.12.

The test set-up used for the e-link evaluation was based on the BDAQ53 test system and was similar to the one used for the AFE evaluation described in Chapter 3, Section 3.4.1. However, the e-links compatible with the 33-pin Molex connector, selected for the TBPX and TFPX module prototypes, cannot be directly connected neither to a single RD53A chip, nor to the scope. Hence, extra connections and cables were necessary and custom adapter boards were developed specifically for this test.

A schematic of the test set-up is presented in Figure 5.13. The PRBS7 signal from the RD53A chip was sent out of the SCC represented in blue as a differential signal via a 50 cm

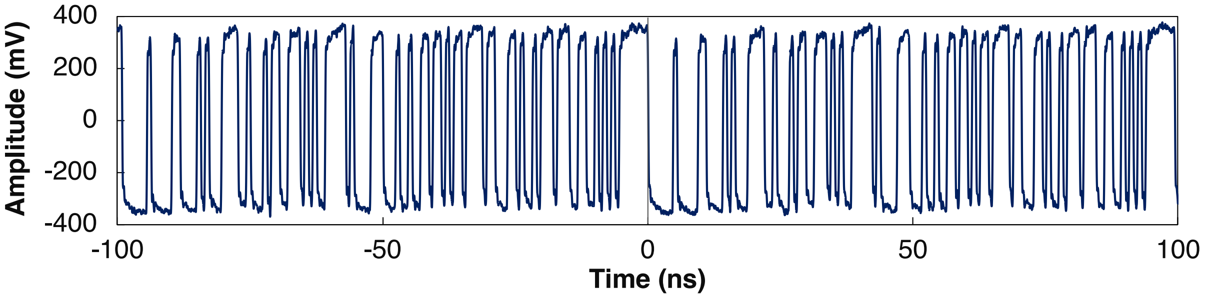


Figure 5.12: Two consecutive sequences of the PRBS7 test pattern generated by the RD53A chip.

long DP cable. A custom PCB depicted in green was used to split the differential lines of the DP cable into pairs of single ended lines routed with SMA cables. The downlink and one uplink sending the Aurora-encoded data were connected with pairs of 20 cm long SMA cables to the FPGA to maintain the communication with the DAQ system. The three remaining uplink lines were available for tests. One of them was routed with a pair of 20 cm SMA cables to an adapter PCB shown in yellow converting the SMA to the Molex connector where a prototype e-link could be plugged. On the other side of the e-link another adapter PCB converted the transmission lines back into SMAs. A pair of 30 cm SMA cables was used to transmit the signal to the oscilloscope. A Keysight DSO9254A scope with 2.5 GHz bandwidth and four analogue channels was used. The signal at the output of this transmission chain arrives to the scope as a pair of single ended lines and is displayed as a difference between them.

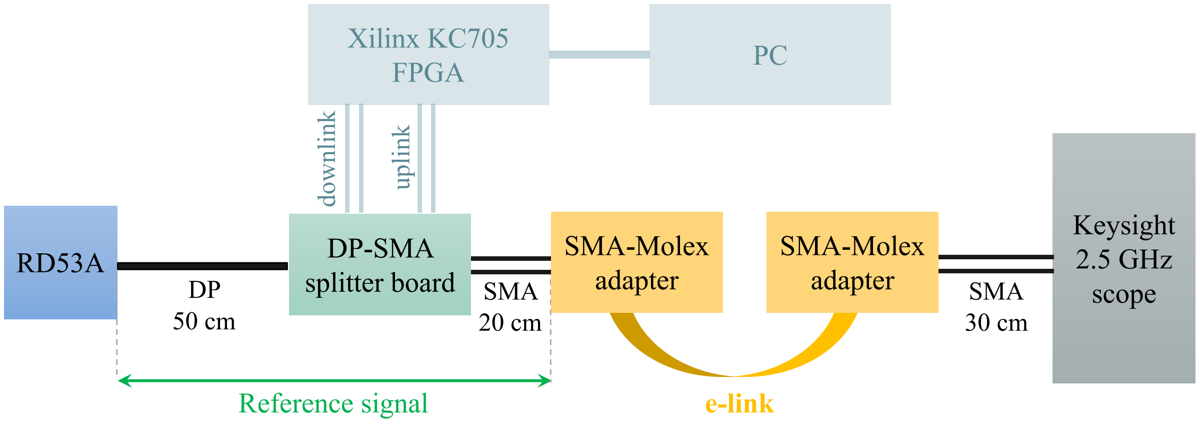


Figure 5.13: Schematic of the test set-up used for the e-links evaluation.

The signal passes through different cables and several connectors, where it encounters impedance discontinuities affecting the signal quality. Hence, this setup provides a pessimistic evaluation of the prototype e-links and their intrinsic performance can be only better. The signal at the output of the first pair of SMA cables was taken as a reference signal, as indicated in Figure 5.13. The first pair of SMAs was connected directly to the scope to evaluate the quality of the signal before its transmission through the e-link under test.

5.3.2 Measurement method and requirements

The quality of the transmitted signal is evaluated based on the *eye diagram* measurement. The eye diagram is a visual representation of a digital signal on the oscilloscope. It represents a statistical average of many superimposed samples of the waveform. Different bit sequences are aligned according to a certain reference point, e.g. using a clock or triggering on the rising edges of the signal, and the waveform samples are overlaid to form a single graph that looks like an eye. The signal amplitude is on the vertical axis, time is on the horizontal axis and the eye opening corresponds to one bit period. The formation of an eye diagram taking as an example a sequence of 3 bits is illustrated in Figure 5.14. Eight possible 3-bit sequences exist and are depicted in Figure 5.14 on the left. The superposition of all the sequences results in an eye diagram shown on the right.

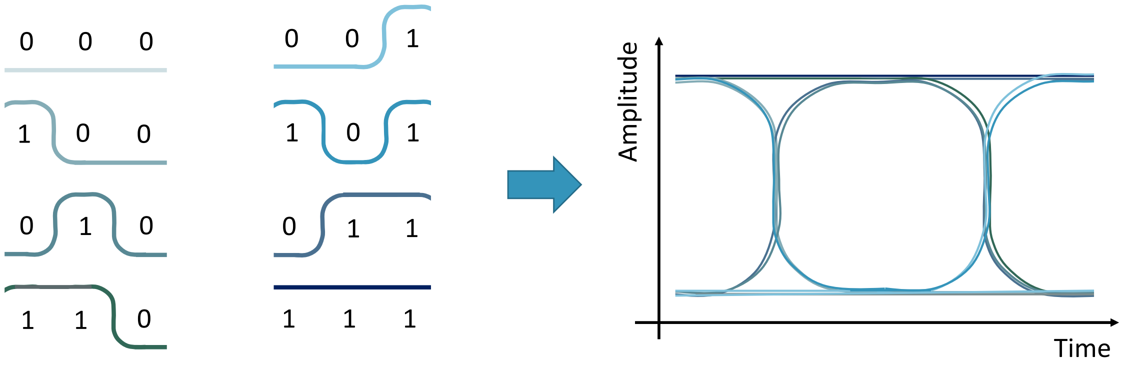


Figure 5.14: Illustration of the formation of an eye diagram.

Several parameters can be measured in an eye diagram to evaluate the quality of the transmitted signal. They are summarised in Figure 5.15 and explained in the following.

Time parameters:

- **Bit period**: horizontal opening of the eye between the mean crossing points. It corresponds to the inverse of the data rate. For the signal sent by the RD53A chip at 1.28 Gb/s the bit period equals to 781.25 ps.
- **Rise time**: mean transition time of the rising edge, typically measured between 20 and 80% of the slope. Alternatively it can be measured between 10 and 90% of the slope.

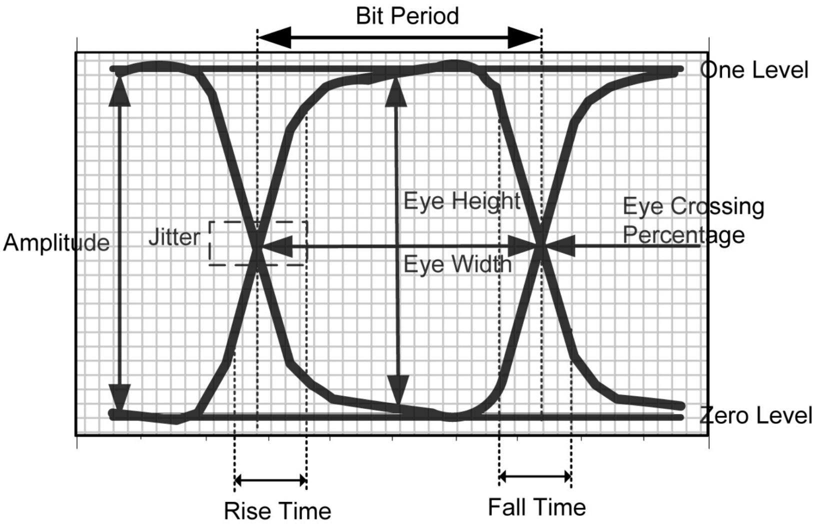


Figure 5.15: Eye diagram parameters [150].

- **Fall time**: mean transition time of the falling edge, typically measured between 20 and 80% of the slope or alternatively between 10 and 90%.
- **Eye width**: horizontal opening of the eye obtained from time histograms across the two crossing points, defined as:

$$\text{Eye width} = (\text{right crossing} - 3\sigma) - (\text{left crossing} + 3\sigma)$$

- **Eye jitter**: time deviation of the transitions from the expected crossing time obtained from a time histogram taken within a thin horizontal window centred at the crossing point.

Amplitude parameters:

- **One level**: mean value of the logical one evaluated from the amplitude histogram taken within the central 20% of the bit period.
- **Zero level**: mean value of the logical zero evaluated from the amplitude histogram taken within the central 20% of the bit period.
- **Eye amplitude**: vertical opening of the eye defined as the difference between the one and zero levels
- **Eye height**: vertical opening of the eye taking into account the eye closure due to noise. The eye height is measured from the amplitude histograms taken within the central 20% of the bit period and is defined as:

$$\text{Eye height} = (\text{one level} - 3\sigma) - (\text{zero level} + 3\sigma)$$

- **Crossing level**: amplitude of the crossing point of the rising and falling edges, measured as the mean of the amplitude histogram taken within a thin vertical window centred on the crossing point. It reveals pulse symmetry problems and amplitude distortions caused by differences in the one and zero level duration. The eye crossing percentage that should ideally be at 50% is calculated as:

$$\text{Eye crossing (\%)} = \frac{\text{crossing level} - \text{zero level}}{\text{one level} - \text{zero level}} * 100$$

Two of these parameters were selected as metrics for the e-link evaluation: the eye height as the amplitude parameter and eye jitter as the time parameter.

In an ideal system, the eye height would be equal to the eye amplitude. However, the noise in a real digital transmission system causes the eye to close. Hence, the eye height gives an estimate about the signal amplitude and at the same time it indicates the SNR. The eye height is an important amplitude parameter, since the data receiver, which is in our case the lpGBT, determines whether the data bit is a logical zero or one, based on the vertical opening of the eye. Any data bits scattered beyond the 3σ from the zero or one level into the open eye may lead to a possible error in the detection and increase the BER. The minimum amplitude needed by the lpGBT to ensure the correct treatment of the received data was estimated from the circuit simulations and taken as a CMS IT system requirement. This lpGBT amplitude requirement is:

$$\text{Differential amplitude} > 140\,\text{mV} \tag{5.9}$$

The jitter represents the time deviation from the ideal timing. The fluctuations can be random and/or deterministic. The jitter is evaluated from the time histogram taken at the crossing level. The standard deviation of the time histogram represents the jitter RMS and the full width of the histogram represents the jitter peak-to-peak. Given that the maximum jitter allowing a correct operation of the lpGBT was simulated as peak-to-peak jitter, the latter type of jitter was taken as requirement. The lpGBT jitter requirement is:

$$\text{Jitter peak-to-peak} < 300\,\text{ps} \tag{5.10}$$

It should be mentioned that even thought these two values were used as requirements for the e-link evaluation, they are only estimates and are not hard limits. This means that the lpGBT will not stop working exactly when the amplitude or jitter reach one of the two values. Instead, the performance of the lpGBT chip will progressively degrade around these limits resulting in a higher BER.

5.3.3 Pre-emphasis in RD53A chip

When the signal propagates down the transmission line, the higher frequencies are attenuated more than the low frequencies. The transmission line acts as a low pass filter, which results in a decrease in the bandwidth and in an increase of the signal rise time. These losses can severely degrade and attenuate the high-frequency content of the signal, making it difficult for the receiver to interpret it.

A dedicated circuit at the transmitter can add extra high-frequency components to the signal before its transmission to mitigate this effect. This way, by the time the signal reaches the receiver, the added higher frequencies have attenuated back and the original signal is preserved. This powerful wave shaping technique, commonly used to compensate for the signal attenuation and distortion in high-speed serial links is called *pre-emphasis*. The pre-emphasis distorts the signal in a controlled manner so that the effects of attenuation in the cables and interconnects are compensated for and the signal integrity is preserved. The pre-emphasis circuit plays a role of a high pass filter that in combination with the low pass filter effect of the transmission line results in a flat frequency response within the desired bandwidth, as it is shown in Figure 5.16.

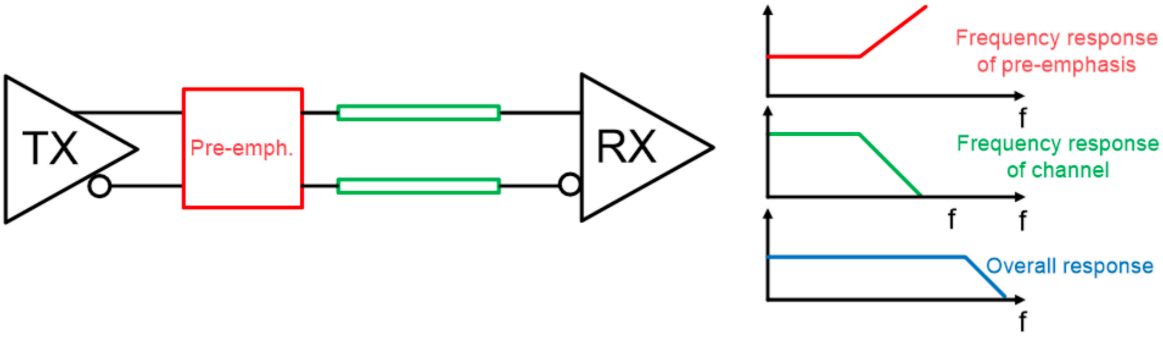


Figure 5.16: Illustration of the pre-emphasis.

The pre-emphasis can either amplify the high frequency component, attenuate the low frequency component or both. In digital transmission systems, the pre-emphasis usually reduces the amplitude of all the bits except for the first bit after a transition. The high frequency content due to the transition is emphasised compared to the low frequency content which is de-emphasised. The amount of pre-emphasis needed depends on the output signal strength and on the transmission line characteristics. It also requires that the distortions induced by the transmission line are predictable and reproducible.

In the RD53A chip, the data transmitter uses a clock recovered by the CDR circuit from the input command stream encoded with a custom DC-balanced protocol and running at 160 Mb/s [90]. The clock produced by the PLL is phase-aligned to the transitions of the command stream. When the PLL is locked, the internal VCO produces a 1.28 GHz clock provided to the high-speed serial transmitter, where the output data stream is formed from the encoded chip data. In addition to the nominal 1.28 GHz clock, the transmitter can also use lower speed clocks, either derived from this clock by the frequency divider, or provided externally [87, 91].

The output transmitter includes four parallel single serial links, also called *data lanes*, each of them sending data at 1.28 Gb/s. This results in total in a maximum bandwidth of 5.12 Gb/s to cope with the high hit rates expected in the innermost tracking layers of the detector. Each data lane consists of a serializer, receiving data from the data encoding logic and producing the output data stream driven off chip by the CML cable driver. The block diagram of the RD53A CML driver is shown in Figure 5.17. It features a common-source differential pair terminated with 50 Ω resistors to minimise the signal reflections. The effective output differential impedance is 100 Ω to match with the 100 Ω differential pairs of the e-links.

The CML driver features a programmable pre-emphasis with a three-tap configuration, as it is shown in Figure 5.17. This allows for a flexible pre-emphasis that can be configured for each individual data link to compensate for the signal attenuation and high-frequency loss introduced by the very low-mass e-links that will be used in the experiment.

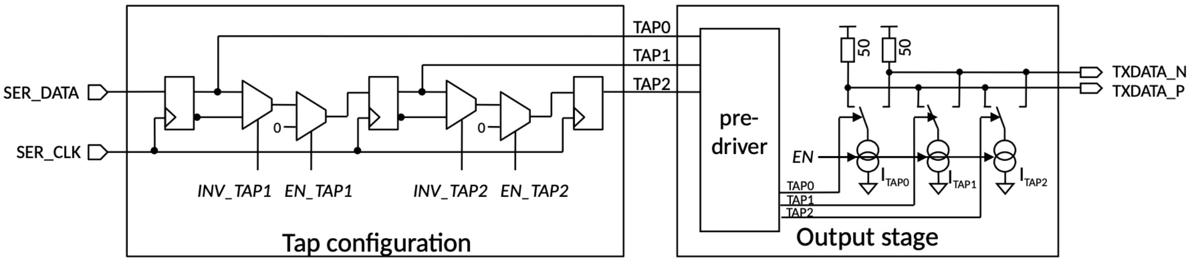
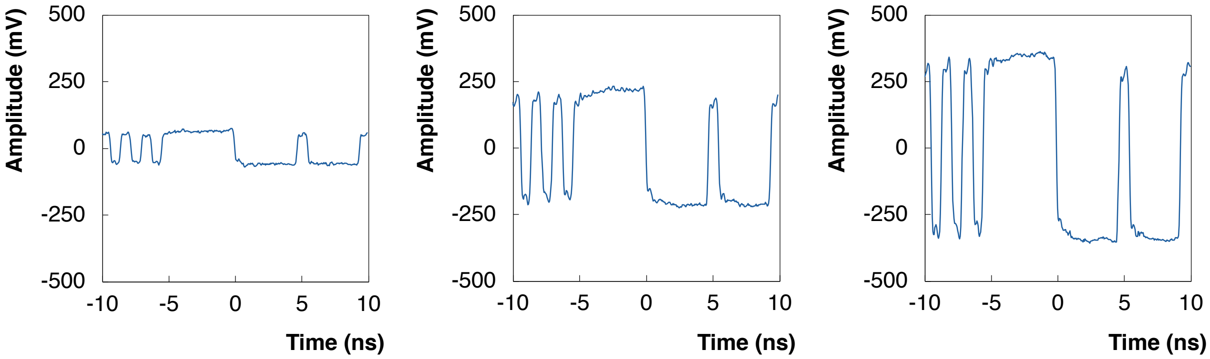


Figure 5.17: Block diagram of the RD53A CML cable driver with three tap configuration [87].

The effect of all the three pre-emphasis taps is not well visible in a sequence of alternating bits (010101). It is instead better visualised on longer sequences of zeros or ones. Therefore, the longest sequence of ones and the longest sequence of zeros in the PRBS7 generated by the RD53A chip are used in the following to explain and demonstrate the effect of each of the three pre-emphasis taps. The unique sequence of seven ones was used as a trigger on the scope and the waveform corresponding to the reference signal is presented in Figures 5.18, 5.19 and 5.21. The sequence of seven ones is visible between $t = -5$ ns and $t = 0$ ns, followed by a sequence of six zeros between $t = 0$ ns and $t = 5$ ns approximately.

TAP0. The main driver of each RD53A data lane is denoted TAP0. Its current can be set via a 10-bit global register. The TAP0 is always enabled and it drives the differential amplitude of the output signal. When the current of TAP0 increases, the signal amplitude increases proportionally, as it is shown if Figure 5.18. The reference signal waveform obtained with TAP0 = 100 is shown in Figure 5.18(a) and the average differential signal amplitude is about 120 mV. The waveform obtained with TAP0 = 500 is shown in Figure 5.18(b) and the signal amplitude is about 450 mV, while with TAP0 = 1000 shown



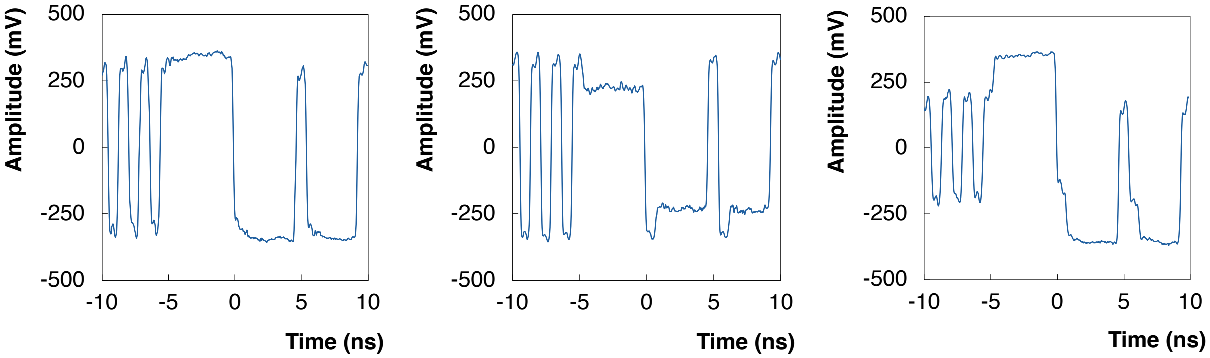
(a) TAP0 = 100.
(b) TAP0 = 500.
(c) TAP0 = 1000.

Figure 5.18: Influence of the TAP0 parameter in the RD53A chip on the amplitude of the reference signal waveform. TAP1 and TAP2 disabled.

in Figure 5.18(c) the amplitude reaches 510 mV.

TAP1. The first programmable pre-emphasis tap in the RD53A output driver is denoted TAP1. It can be enabled or disabled and it can be used with two polarities: inverted or non-inverted. The effect of the TAP1 pre-emphasis on the signal shape is shown in Figure 5.19. The reference signal waveform with TAP0 = 1000 and TAP1, and TAP2 disabled, i.e. with no pre-emphasis added to the signal, is shown in Figure 5.19(a). When the TAP1 is enabled and used with the inverted polarity the amplitude of all bits, except for the first one after each transition, is reduced, as it is shown in Figure 5.19(b). In particular, the three single bits of one on the left side of the plot are not affected. On the other hand, when the TAP1 is used with a non-inverted polarity the effect is the opposite: only the amplitude of the first bit after each transition is reduced, as it is shown in Figure 5.19(c). This is also confirmed by the reduced amplitude of the three single bits of one at the beginning of the sequence. Hence, the TAP1 pre-emphasis in the RD53A has to be used with the inverted polarity to boost the high frequencies.

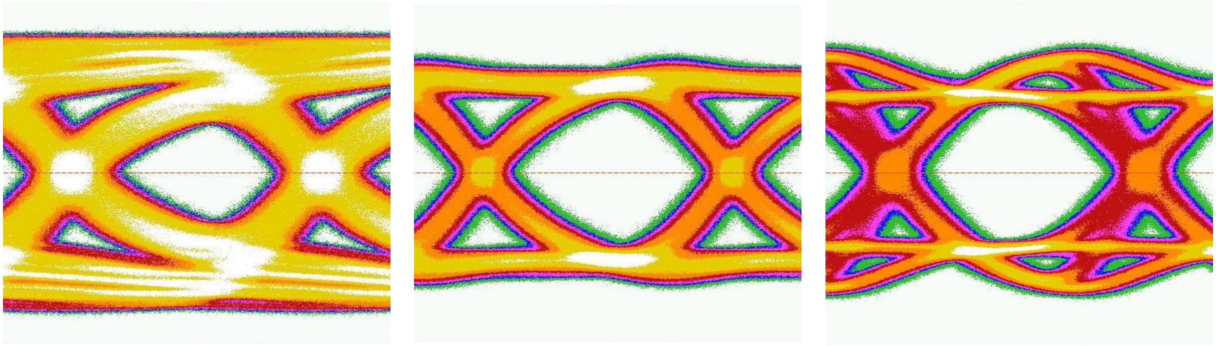
The amount of pre-emphasis added by TAP1 depends on the TAP1 current that can be adjusted individually by a 10-bit global register. If this current is set to zero, it is equivalent to disabling TAP1. When the TAP1 current increases, the amplitude difference between the first bit after transition and the other bits increases. The effect of different TAP1 currents on the signal eye diagram is shown in Figure 5.20. The longest e-link prototype was used to demonstrate the effect, since the it causes the biggest attenuation that can be compensated by the pre-emphasis. The eye diagrams were obtained with the 2 m long 36 AWG TWP. The TAP0 was set to 1000, the TAP1 was inverted and TAP2 was disabled. When no pre-emphasis is used with this long e-link, i.e. TAP1 = 0, the eye diagram presents a large dispersion and is not well opened, as it can be observed in Figure 5.20(a). Increasing TAP1 from 0 to 150, reduces the dispersion, increases the rise time and results in a better eye opening, as shown in Figure 5.20(b). When the TAP1



(a) TAP1 disabled. (b) Inverted TAP1. (c) Not inverted TAP1.

Figure 5.19: Influence of the TAP1 parameter in the RD53A chip on the shape of the reference signal waveform. TAP0 = 1000, TAP2 disabled and TAP1 disabled (a), inverted and set to 250 (b) and not inverted and set to 250 (c).

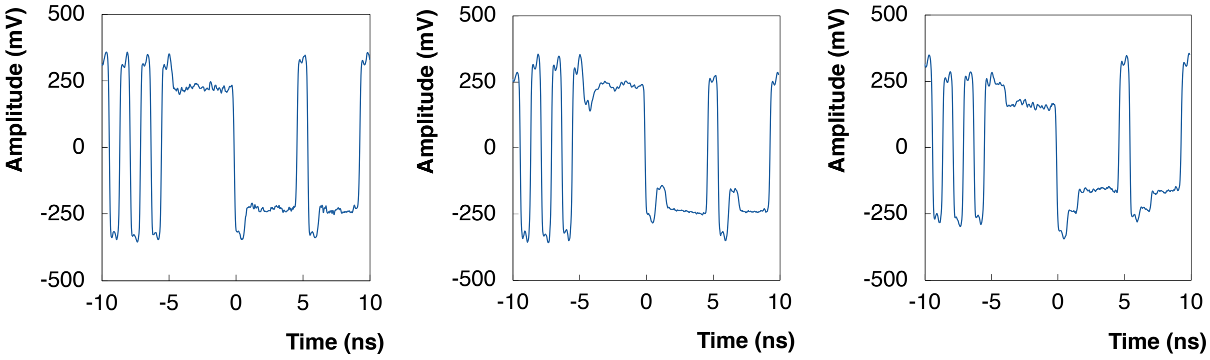
current is further increased up to 300, the eye diagram is distorted and the effective eye height is reduced, as shown in Figure 5.20(c). So there is an optimal TAP1 pre-emphasis setting, that maximises the eye opening and a further increase of TAP1 actually degrades the signal quality. The optimal TAP1 value depends on the transmission line characteristics (e-link design and length, connectors etc.), as well as on the TAP0 value. In fact, the higher the TAP0, the higher is the optimal TAP1.



(a) TAP1 = 0. (b) TAP1 = 150. (c) TAP1 = 300.

Figure 5.20: Eye diagrams obtained with the 2 m long 36 AWG TWP demonstrating the influence of the TAP1 pre-emphasis on the signal shape. The TAP0 was set to 1000, the TAP1 was enabled, inverted and set to 0 (a), 150 (b) and 300 (c) and TAP2 disabled.

TAP2 The second programmable pre-emphasis tap in the RD53A output driver is denoted TAP2. Similarly to TAP1, it can be enabled or disabled and it can be used with two polarities: inverted or non-inverted. The effect of the TAP2 pre-emphasis on the signal shape is shown in Figure 5.21. The TAP2 was designed in the RD53A to be used in addition to TAP1. The reference signal waveform with TAP0 = 1000, TAP1 = 250 inverted and TAP2 disabled is shown again in Figure 5.21(a) for comparison. When TAP2 is added to this signal with inverted polarity, the amplitude of the second bit after each transition is reduced, as it is shown in Figure 5.21(b). The effect is visible in the consecutive sequences of ones and zeros, while the single bits of one or zeros are not affected. On the other hand, when TAP2 is added to the signal with non-inverted polarity, the amplitude of all bits, except for the second bit after each transition, is reduced, as it is shown in Figure 5.21(c).



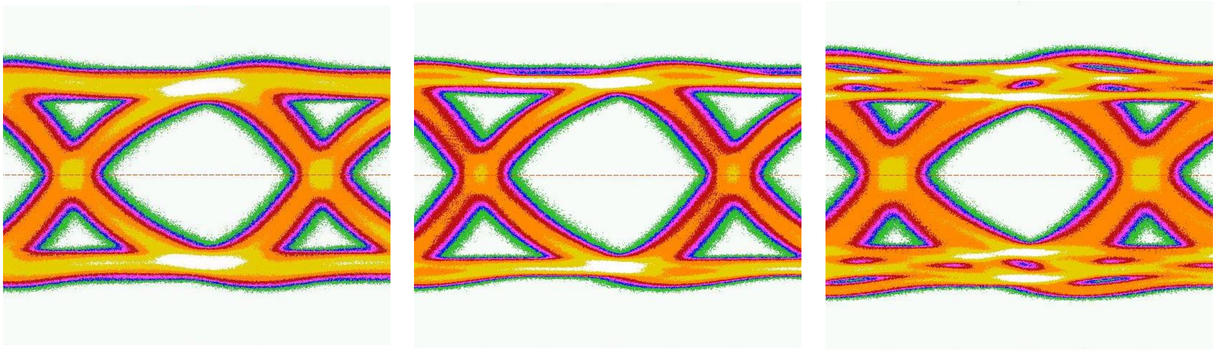
(a) TAP2 disabled. (b) Inverted TAP2. (c) Not inverted TAP2.

Figure 5.21: Influence of the TAP2 parameter in the RD53A chip on the shape of the reference signal waveform.

Similarly to TAP1, the amount of pre-emphasis affecting the amplitude of the second bit after each transition depends on the TAP2 current that can be adjusted individually by a 10-bit global register. If the TAP2 current is set to zero, it is equivalent to disabling it. The higher the TAP2 current and the bigger is the difference between the amplitude of the second bit after transition and the following bits. It can be also noticed that the first and second bit after transition do not have the same amplitude in Figures 5.21(b) and 5.21(c), which is due to the fact that the TAP1 was set to 250, while TAP2 was set to 100 only. If both taps were set to the same value the first and second bit would be emphasised in the same way.

The effect of the TAP2 on the eye diagram obtained with the 2 m long 36 AWG TWP is shown in Figure 5.22. In this measurement, the TAP0 = 1000, the TAP1 was enabled, inverted and set to 150 and TAP2 was enabled, non-inverted and its current was varied. Figure 5.22(a) shows the eye diagram for TAP2 = 0, i.e. only the first pre-emphasis tap is used. When the TAP2 is increased to 25, the eye diagram shown in Figure 5.22(b), has a narrower zero and one levels, the dispersion is smaller and the eye opening is slightly improved. When TAP2 is further increased to a value of 100, the zero and one levels are demultipied and the signal is distorted. The eye opening is slightly worse.

As for TAP1, an optimal value of TAP2 maximising the eye opening exists and a further increase degrades the signal. The optimal value of TAP2 depends on the TAP1 setting and is typically one order of magnitude lower. In this example, for TAP0 = 1000, the optimal TAP1 is 150, while the optimal TAP2 is 25. Moreover, the effect of TAP2 on the eye opening is much less significant than the effect of TAP1. The TAP1 is a first order effect and the TAP2 is a second order effect.



(a) TAP2 = 0. (b) TAP2 = 25. (c) TAP2 = 100.

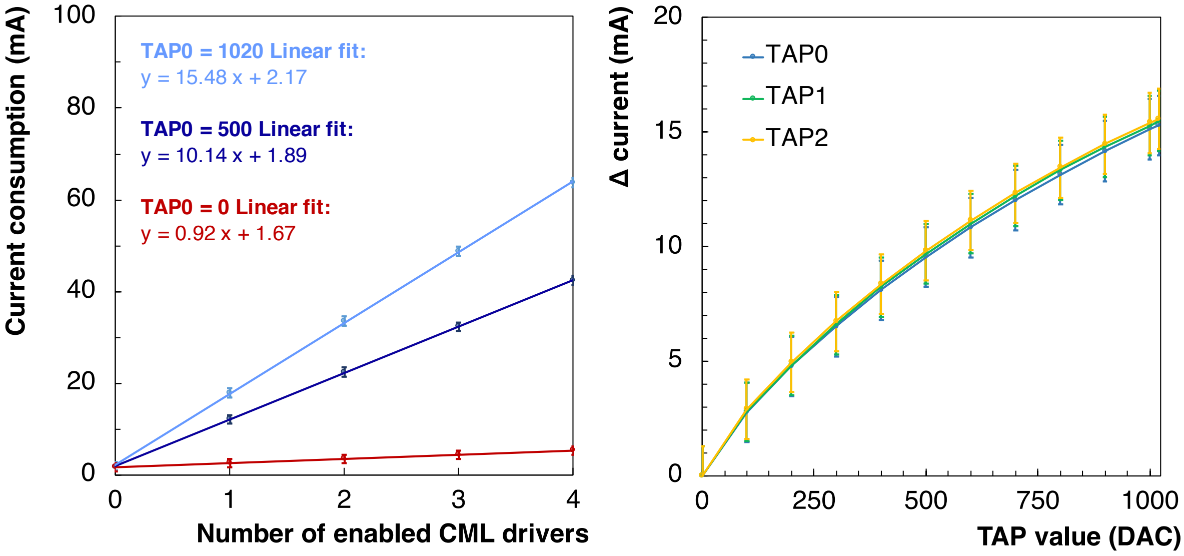
Figure 5.22: Eye diagrams obtained with the 2 m-long 36 AWG TWP demonstrating the influence of the TAP2 on the signal shape. The TAP0 was set to 1000, the TAP1 was enabled, inverted and set to 150 and the TAP2 was enabled, not inverted and set to 0 (a), 25 (b) and 100 (c).

5.3.4 Pre-emphasis and power consumption

The CML cable driver, as well as the three pre-emphasis taps, are current-based. The driver for each tap is identical and the tap currents sum up at the output node. Hence, the more pre-emphasis is added to the signal, the higher is the current consumption of the chip. Circuit simulations showed that the maximum current of each tap is expected to be 15 mA [91]. Since the power consumption of the pixel electronics is one of the main concerns for the IT system, the current consumption associated to the pre-emphasis functionality was also verified with a measurement.

The current consumption was measured for different number of enabled CML drivers with TAP0 = 0 and TAP1 and TAP2 disabled. The result is shown in red in Figure 5.23(a). When all drivers are disabled, the current consumption of the chip displayed, on the power supply, without sending any data out and without using the current-based output drivers is less than 2 mA. It increases linearly with increasing number of enabled drivers and each driver draws about 1 mA when TAP0 = 0. The current consumption as a function of the number of enabled drivers was also measured with TAP0 = 500 and TAP0 = 1020 and the results are reported in Figure 5.23(a) in dark blue and light blue, respectively. The current consumption also increases linearly with the number of enabled data lanes, as it is confirmed by the linear fits. Each enabled data lane draws about 10 mA when TAP0 = 500 and about 15.5 mA when the TAP0 is set close to the maximum (TAP0 = 1020).

The current consumption of each of the three taps was measured as a function of the TAP value. Only one data line was enabled for this measurement. First, the TAP0 was varied over the full range with TAP1 and TAP2 disabled. The current for TAP0 = 0 was taken as a reference and subtracted from all the measured currents and the current difference as a function of TAP0 is presented in blue in Figure 5.23(b). The current consumption increases with increasing TAP0 value and the maximum current drawn by this tap is equal to 15.3 mA. Then, the same measurement was done for TAP1 and TAP2 with TAP0 = 0 (since it cannot be disabled) and with the other tap disabled. The result is shown in Figure 5.23(b) in green and yellow for the TAP1 and TAP2, respectively, and is very similar to the result obtained for TAP0. The maximum current drawn by any of the pre-emphasis taps is about 15.5 mA. These results are consistent within error bars with the maximum current consumption of 15 mA per tap obtained from simulations.



(a) (b)

Figure 5.23: The RD53A chip current consumption associated to the output transmitter measured as a function of the number of enabled CML drivers (a) and as a function of the parameter driving the current of each pre-emphasis tap.

5.3.5 E-link comparison

The eye height and jitter peak-to-peak were measured with the scope at the output of the transmission chain shown in Figure 5.13. The measurement was done for all available prototype e-links to compare their performance with different pre-emphasis settings.

Influence of TAP0

First, the measurement was performed for different values of TAP0 ranging from 200 to 1023. For TAP0 values below 200, the eye diagram was often not open enough to be measured. The pre-emphasis TAP1 and TAP2 were disabled for this measurement.

The lpGBT requirements for both evaluation parameters (the amplitude and jitter) are indicated by a red line in all following figures. The reference signal indicated in Figure 5.13 is given for comparison and is always shown in green. It gives the maximum amplitude and minimum jitter that can be obtained with this set-up. Indeed, adding two adapter boards, an e-link under test and a pair of SMA cables can only worsen the signal quality.

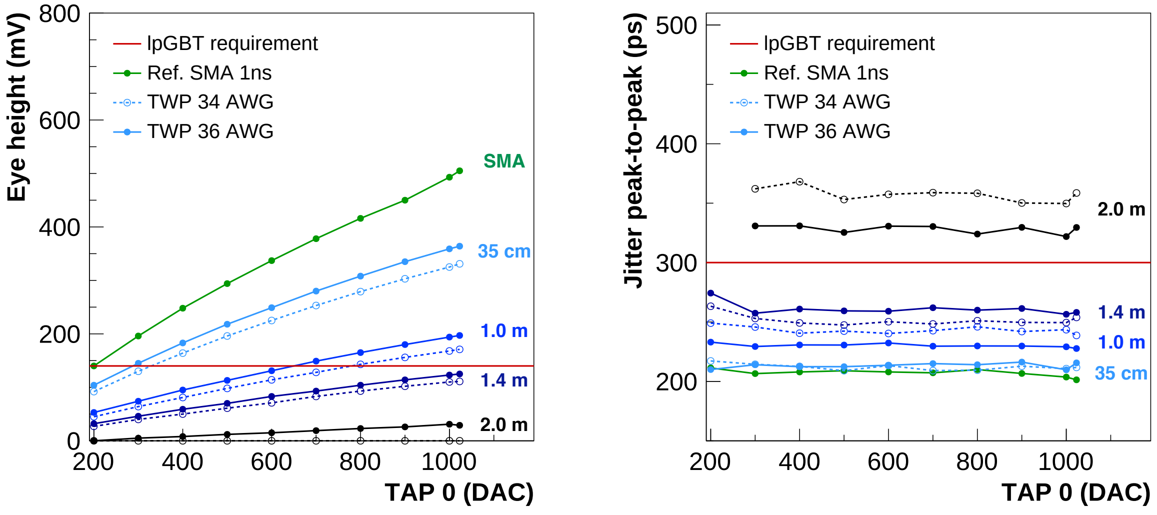
The measurement results comparing different TWP cables are presented in Figure 5.24. The two available flavours of TWPs were measured: the 36 AWG TWPs are represented with full markers and solid lines, while the 34 AWG TWPs are represented with open markers and dashed lines. Four different lengths of each TWP flavour were measured: 0.35 m, 1.0 m, 1.4 m and 2.0 m.

The eye height as a function of TAP0 is shown in Figure 5.24(a). As expected, the eye height increases linearly with increasing TAP0. For a given TAP0 setting, the eye height decreases when the e-link length increases, since the signal is more attenuated in longer cables. The 36 AWG TWPs demonstrate a higher amplitude than the 34 AWG. Hence, it is not worth to increase the material budget of the detector by using thicker TWPs, given that the signal quality is not better. The 0.35 m TWP reaches the lpGBT amplitude requirement already for a TAP0 of about 300. The 1 m TWP reaches the requirement as well, but a higher TAP0 is needed. The 1.4 m TWP is at the limit of the requirement when the TAP0 is set to the maximum. The signal is significantly attenuated in the 2 m TWP, the eye diagram is barely opened and the measured eye height is well below the requirement for any TAP0 value.

The jitter peak-to-peak as a function of TAP0 is shown in Figure 5.24(b). The flat curves indicate that the measured jitter is not sensitive to TAP0. When the length of the TWPs increases, the jitter increases because the signal quality is more affected in longer e-links. The 35 cm, 1.0 m and 1.4 m TWPs are within the lpGBT requirements, while the 2.0 m TWPs are outside. The jitter is even higher for the 2.0 m TWP with 34 AWG. The jitter of the 2.0 m TWPs could not be measured with TAP0 below 300. In fact, the signal amplitude was very small with these settings and after the attenuation introduced by the long e-link the eye diagram was closed and the scope was unable to measure its characteristics.

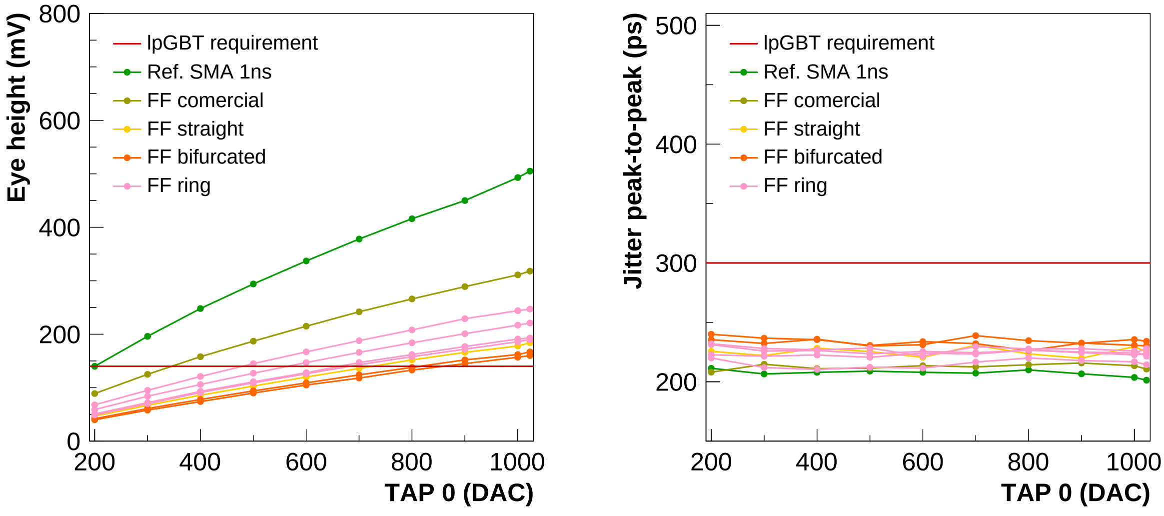
The same measurement was performed with the FFCs of similar lengths and the results are presented in Figure 5.25. The eye height as a function of TAP0 is shown in Figure 5.25(a) and the jitter as a function of TAP0 is shown in Figure 5.25(b). The commercial FFC is represented in khaki colour, the straight FFC in yellow, the two outputs of the bifurcated FFC in orange and the four outputs of the ring FFC are all represented in pink. Overall, all the FFCs can reach the amplitude requirement, if the TAP0 is correctly adjusted and all of them satisfy the jitter requirement with a margin of about 60 ps.

The commercial FFC reaches a higher amplitude than the prototypes. The ring design demonstrates the highest amplitude among the prototypes and the bifurcated design



(a) (b)

Figure 5.24: Eye height (a) and jitter peak-to-peak (b) as a function of TAP0, measured for 34 AWG and 36 AWG twisted pairs of 0.35 m, 1.0 m, 1.4 m and 2.0 m.



(a) (b)

Figure 5.25: Eye height (a) and jitter peak-to-peak (b) as a function of TAP0, measured for different flexible flat cables.

attenuates the most. These results can be correlated with the slight differences in the e-link length. In fact, the commercial FFC is the shortest flat e-link with only 25 cm, hence it provides the highest amplitude. The ring FFC features transmission lines with lengths between 26 cm and 38 cm. The straight FFC is 35 cm long, while the transmission traces in the bifurcated design are about 40 cm long, which corroborates the biggest attenuation.

Moreover, while the two outputs of the bifurcated FFC give very similar results, the four outputs of the ring FFC do not. This can be explained by the length of the transmission lines. The two sides of the bifurcated design are perfectly symmetric and have exactly the same length, as it can be observed in Figure 5.4. In the ring topology this is not the case. For an easier comparison, the results of all the e-links obtained with the maximum TAP0 and no pre-emphasis are presented as a function of the e-link length in Figure 5.26. The eye height is presented in Figure 5.26(a) and the jitter in Figure 5.26(b).

At comparable length, the TWPs show better performance than the FFCs, including the commercial one. The 35 cm TWPs reach an eye height of about 330 mV and 360 mV for the 34 AWG and 36 AWG, respectively, while the straight FFC with the same length reaches only about 180 mV. In addition, the eye height decreases more rapidly with the e-link length for the FFCs than for the TWPs. As a result, the 100 cm long TWPs show a similar eye height as the 35 cm FFCs. The same conclusions can be drawn for the jitter. At comparable length the TWPs have a lower jitter than the FFCs and the jitter increases more rapidly with length in FFCs. Hence, for the same performance we can afford longer e-links with the TWP design. The TWPs are therefore a more suitable option for the TBPX, where longer e-links will be needed, reaching a length of 1.6 m.

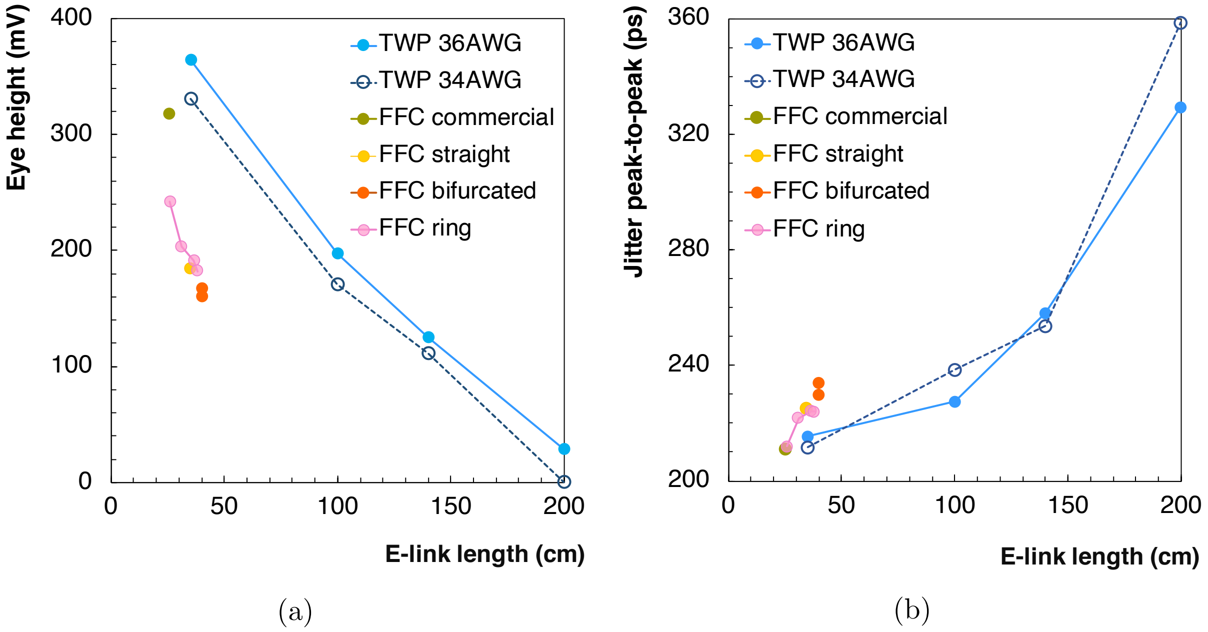


Figure 5.26: Eye height (a) and jitter peak-to-peak (b) as a function of the e-link length measured with TAP0 = 1023 and no pre-emphasis.

Influence of TAP1

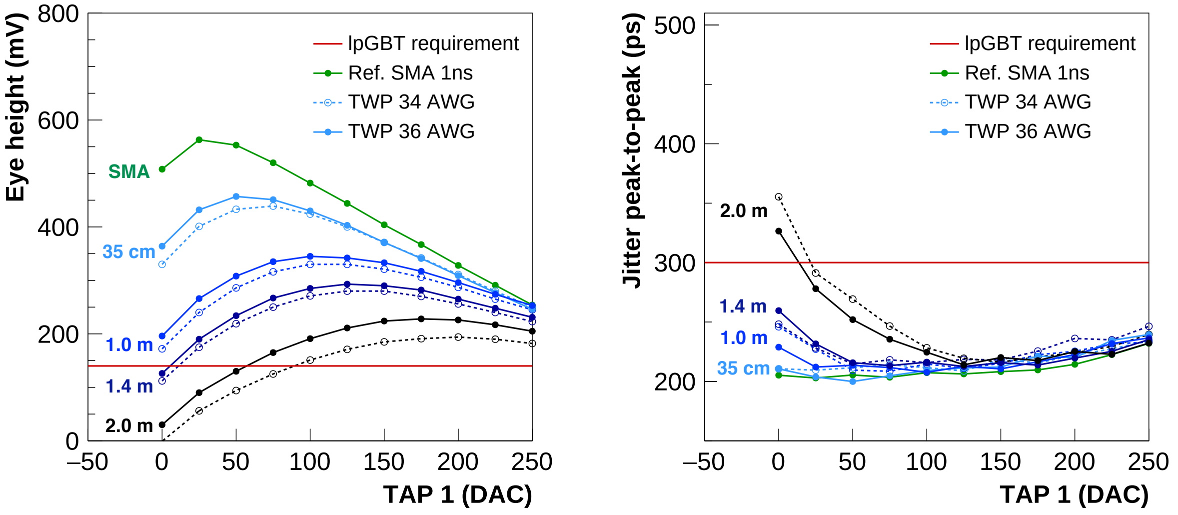
The eye height and jitter peak-to-peak were measured for all available prototype e-links as a function of TAP1 to evaluate the influence of the first pre-emphasis tap on the quality of the transmitted signal. The measurement was performed with TAP0 = 1023, i.e. the maximum signal amplitude. The TAP1 was scanned in the range from 0 to 250, which is a quarter of the full range. For higher TAP1 values, the eye diagram was too distorted and the performance deteriorated. The second pre-emphasis tap (TAP2) was disabled for this measurement. The results obtained with the TWPs are presented in Figure 5.27 and those obtained with FFCs in Figure 5.28. The lpGBT requirements for both evaluation parameters (amplitude and jitter) are again indicated by a red line and the reference signal is always shown in green.

When TAP1 is set to zero, no pre-emphasis is added to the signal and the measured eye height and jitter correspond to the ones presented in Figures 5.24 and 5.25 for the maximum TAP0 and to the values presented in Figure 5.26. When the TAP1 increases, the eye height increases, reaches a maximum and then decreases again because a too high pre-emphasis distorts the signal, as explained in Section 5.3.3. Similarly, the jitter decreases with increasing TAP1, reaches a minimum and then increases again. The optimal pre-emphasis setting is not the same for different e-links, however for all the available prototypes it is within the first quarter of the TAP1 range, even for the longest TWP. This leaves quite some margin to further increase the pre-emphasis when additional system effects will come into play.

This measurement highlights the same conclusions as the measurement with TAP0: with longer e-links the amplitude is smaller and the jitter is higher; the 36 AWG TWPs show a better signal quality than the 34 AWG TWPs, and for a comparable length the TWPs show better performance than the FFCs. Globally, with the optimal pre-emphasis, all evaluated e-links satisfy the lpGBT requirements. Both 34 AWG and 36 AWG TWPs reach the required amplitude and jitter, even the 2 m long e-link. This is a promising result, given that in the Phase-2 IT the maximum e-link length will be 1.6 m. The FFCs reach an amplitude between 300 mV and 400 mV, providing a margin of at least 150 mV, and the associated jitter is about 100 ps below the requirement.

Influence of TAP2

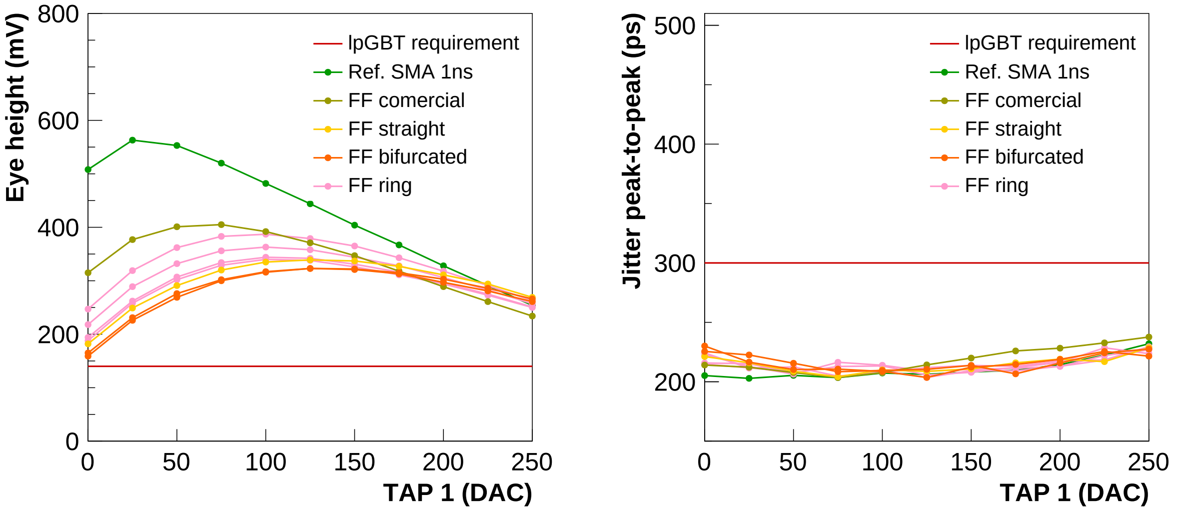
The second pre-emphasis tap can be used in addition to TAP1 to further improve the quality of the transmitted signal. The eye height and jitter peak-to-peak were also measured for all available prototype e-links as a function of TAP2 to evaluate its influence on the signal quality. The measurement was performed again with TAP0 = 1023, i.e. the maximum signal amplitude. The TAP1 was set to the value maximising the eye height, which was different for each e-link. The optimal TAP1 values were obtained from the previous measurement presented in Figures 5.27(a) and 5.28(a) and are also reported in Table 5.2. Once TAP0 and TAP1 were fixed, TAP2 was scanned in the range from 0 to 100, which is only 10% of the full range. For higher values, the eye diagrams were too distorted as it was explained in Section 5.3.3. The results obtained with the TWPs



(a)

(b)

Figure 5.27: Eye height (a) and jitter peak-to-peak (b) as a function of TAP1, measured for different twisted pairs.



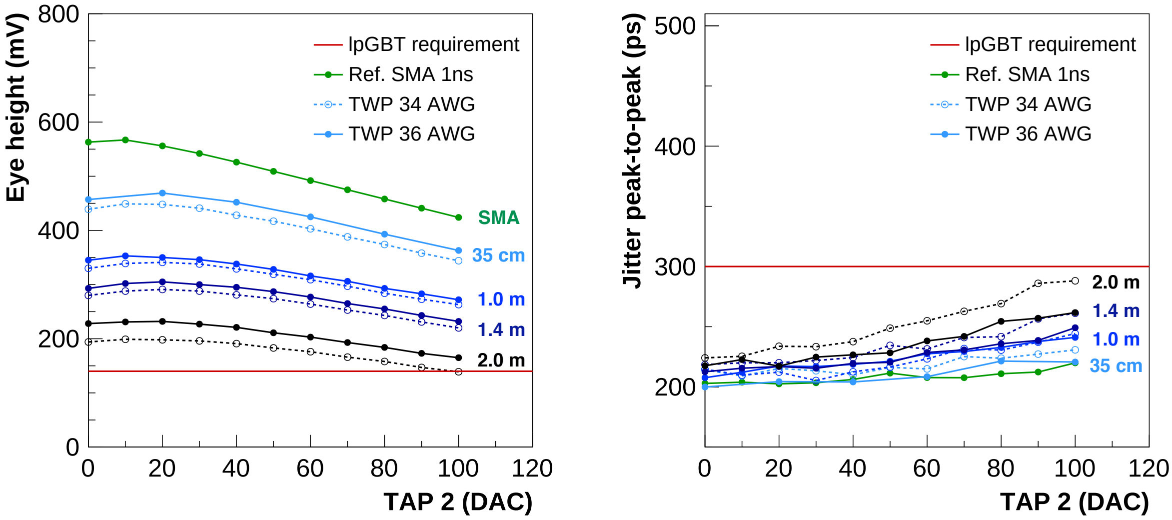
(a)

(b)

Figure 5.28: Eye height (a) and jitter peak-to-peak (b) as a function of TAP1, measured for different flexible flat cables.

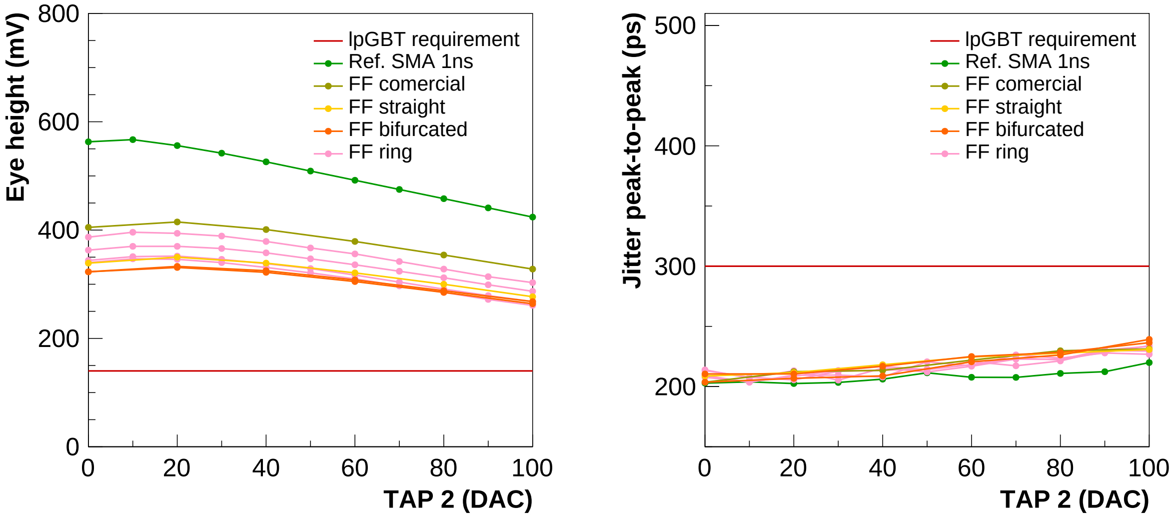
are presented in Figure 5.29 and those obtained with FFCs in Figure 5.30. The lpGBT requirements for both evaluation parameters are indicated by a red line and the reference signal is always shown in green.

The TAP2 demonstrates a behaviour similar to TAP1: when it increases, the amplitude first increases, reaches a maximum and then decreases. However, the TAP2 contribution to the eye height is much less significant than for TAP1 and the maximum is reached



(a) (b)

Figure 5.29: Eye height (a) and jitter peak-to-peak (b) as a function of TAP2 with TAP0 = 1023 and TAP1 optimised, measured for different twisted pairs.



(a) (b)

Figure 5.30: Eye height (a) and jitter peak-to-peak (b) as a function of TAP2 with TAP0 = 1023 and TAP1 optimised, measured for different flexible flat cables.

for a very low value of the parameter. As far as the jitter is concerned, it decreases only very slightly for a small TAP2 value before it starts increasing again when too much TAP2 pre-emphasis is added to the signal. In some cases, the jitter minimum is actually obtained for TAP2 = 0. This is due to the fact that the TAP1 was chosen in this study to optimise the amplitude rather than the jitter. Overall, the TAP2 measurement confirms the e-link comparison conclusions drawn from the TAP0 and TAP1 measurements exposed

Table 5.2: TAP1 and TAP2 pre-emphasis settings maximising the eye height measured with each prototype e-link and the induced change in eye height and peak-to-peak jitter.

E-link	**Length (cm)**	**Optimal setting**		**ΔEye height (mV)**		**ΔJitter (ps)**	
		TAP1	**TAP2**	**TAP1**	**TAP2**	**TAP1**	**TAP2**
Ref. SMA 1 ns	10	25	10	55	4	-2	1
FFC commercial	25	75	20	91	9	-8	-4
FFC straight	35	125	20	157	9	-20	6
FFC bifurcated 1	40	125	20	164	8	-15	0
FFC bifurcated 2	40	125	20	158	10	-27	3
FFC ring 1	36	100	10	150	7	-3	-5
FFC ring 2	38	100	10	152	7	-17	-4
FFC ring 3	26	100	10	140	9	-6	-6
FFC ring 4	31	100	10	145	7	-9	-7
TWP 34 AWG	35	75	10	109	10	4	-5
TWP 36 AWG	35	50	10	94	12	-23	1
TWP 34 AWG	100	100	20	158	11	-31	-3
TWP 36 AWG	100	100	10	149	8	-21	5
TWP 34 AWG	140	125	20	168	11	-30	2
TWP 36 AWG	140	125	20	167	12	-47	4
TWP 34 AWG	200	200	10	194	5	-131	1
TWP 36 AWG	200	175	20	198	4	-109	0

previously.

5.3.6 Pre-emphasis optimisation

All three pre-emphasis taps were optimised to evaluate the best performance of each e-link. First, the TAP0 was set to the maximum, i.e. TAP0 = 1023. Then, the optimal TAP1 maximising the eye height was found and fixed individually for each e-link, after what the optimal TAP2 also maximising the eye height was selected. The optimal TAP1 and TAP2 values maximising the signal amplitude are summarised for all e-links in Table 5.2. In this measurement, the average TAP1 is about 100 and TAP2 is 10–20. However, the optimal pre-emphasis settings depend on TAP0 and a rule of thumb is that the optimal value of a tap is about 10% of the previous tap, e.g. $\mathrm{TAP0} = 1000, \mathrm{TAP1} \approx 100, \mathrm{TAP2} \approx 10$ or $\mathrm{TAP0} = 500, \mathrm{TAP1} \approx 50, \mathrm{TAP2} \approx 5$.

The difference in eye height and jitter induced by the selected TAP1 and TAP2 values are also reported in Table 5.2. The reference signal needs only a small value of TAP1 and it increases the eye height by only 55 mV. This can be explained by the fact that the commercial electrical cables do not significantly degrade the signal integrity. For the

FFC e-links, the first pre-emphasis tap allows to increase the amplitude by about 150 mV. For the TWPs, it allows to gain about 100 mV in the shortest and up to 200 mV for the longest cable. In terms of jitter, the TAP1 pre-emphasis allows to decrease the jitter in the FFCs by about 10 to 20 ps and in the TWPs by about 20 ps in the shortest, and up to 130 ps in the longest one. It can be inferred from these results that the first pre-emphasis tap in the RD53A chip allows to gain $\mathcal{O}(100\,\mathrm{mV})$ in eye height and $\mathcal{O}(10\text{–}100\,\mathrm{ps})$ in jitter.

The TAP2 represents a second-order correction with respect to TAP1. It allows to gain an additional $\mathcal{O}(10\,\mathrm{mV})$ in amplitude and few picoseconds in jitter. In some cases, the jitter becomes actually few picoseconds worse. The TAP2 does not really improve the signal jitter in this measurement. The two pre-emphasis parameters could also be set to optimise the jitter if it was the limiting parameter of the system.

The maximum eye height obtained with different e-links after the pre-emphasis optimisation is shown in Figure 5.31 to better visualise the effect of the three taps. The left half of the histogram shows the FFCs and the right half the TWPs. The reference signal is shown on the very left. The eye height obtained with TAP0 only is represented in green, the additional amplitude gained with optimal TAP1 is represented in light blue and the amplitude gain induced by TAP2 is shown in dark blue. All the e-links with a length up to 1 m are within the lpGBT requirement even without pre-emphasis, while for the longer cables the pre-emphasis is necessary. As mentioned previously, the major improvement in eye height is obtained by optimising TAP1. The correction added by TAP2 is very small and can even be neglected. The difference between the maximum amplitude obtained with each e-link and the minimum amplitude required by the lpGBT, represents the available margin for further attenuation. All e-links are within the requirement, but

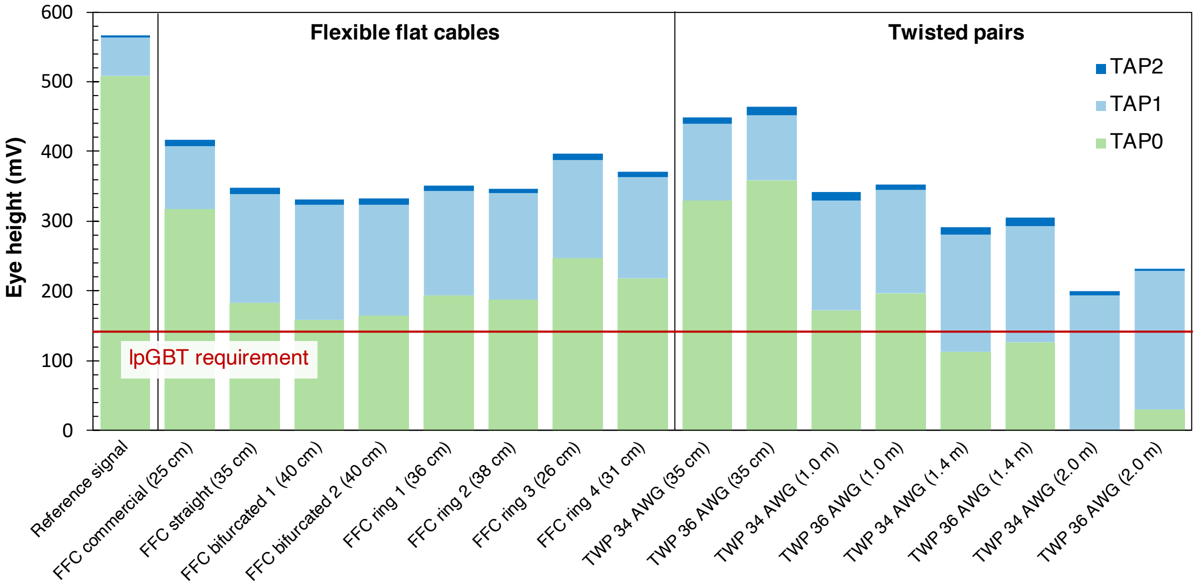


Figure 5.31: Maximum amplitude obtained with each e-link after pre-emphasis optimisation (TAP0 = 1023, TAP1 and TAP2 optimised).

the margin is reduced in longer e-links. The short TWPs have the biggest margin reaching an amplitude of about 450 mV. The 1 m-long TWPs reach an amplitude of about 350 mV, which is comparable to the maximum amplitude of the much shorter FFCs.

And finally, it is also important to determine the minimum tap settings allowing to satisfy the lpGBT requirements. The previous measurements showed that the jitter requirement is satisfied down to very low TAP0 values and the limiting factor was the amplitude. Hence, the goal of this measurement was to find the minimum TAP0 allowing to satisfy the amplitude requirement after TAP1 optimisation. The TAP0 was progressively decreased, for each value the TAP1 was optimised, the amplitude was measured and compared to the requirement. This was done until the amplitude reached the required 140 mV. Since the previous measurements showed that the contribution of TAP2 was negligible, it was not used in this measurement and TAP2 was disabled.

Figure 5.32 shows the TAP0 value for each evaluated e-link that leads to the amplitude exactly at the limit of the required 140 mV with optimised TAP1 reported in each bin. The y axis represents the full range of TAP0. The minimum TAP0 for the FFCs is between 300 and 400, leaving at least 60% of the TAP0 range available to compensate for additional signal attenuation occurring over time with irradiation or other unforeseen effects. The short TWPs could operate with the lowest TAP0 and leave almost 80% of the range available. The minimum TAP0 increases with increasing length. Nevertheless, even the longest e-link can operate with TAP0 set to about 650, which still leaves almost 40% of the range available for improvement of the signal.

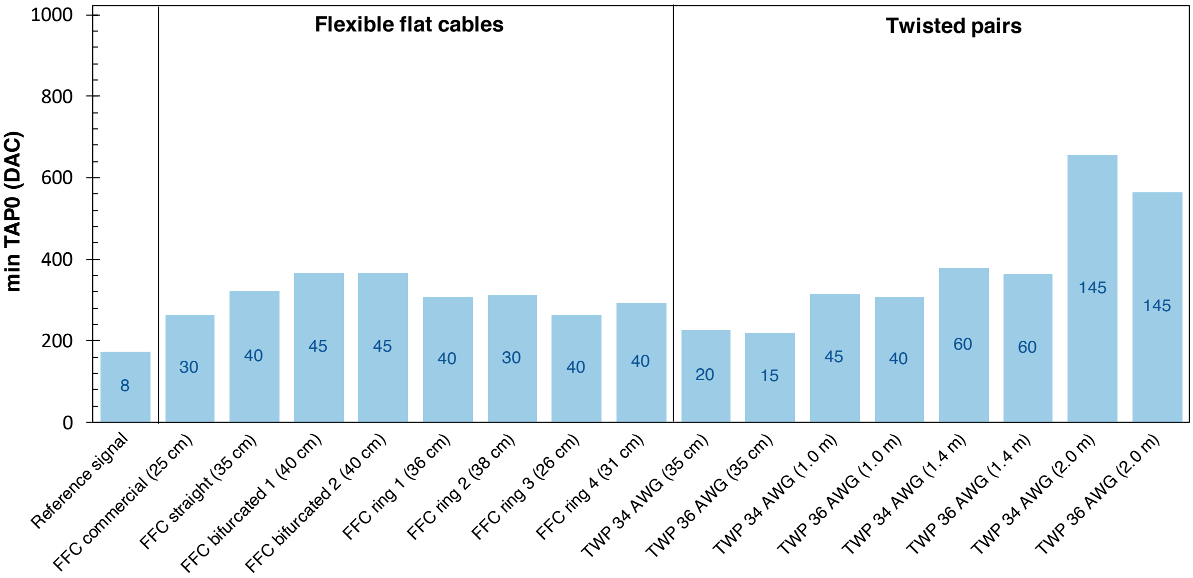


Figure 5.32: Minimum TAP0 needed for each e-link to reach the amplitude requirement. TAP1 optimised and TAP2 = 0.

5.4 Summary and discussion

The high hit and trigger rates expected at the HL-LHC are very challenging for the IT readout system and particularly for the electrical data readout. Hence, new electrical links supporting multi-gigabit readout, while having as low mass as possible without significantly affecting the signal integrity, must be developed and validated. Two options are considered by the CMS Collaboration: the twisted pairs or flexible flat cables. Different prototypes of both types were evaluated in this study. The VNA measurement showed that the e-link prototypes matching the best the impedance target of 100 Ω were the 36 AWG TWPs and the FFC with the ring design. The impedance of the other prototypes was about 85 Ω, which is 15% below the requirement.

The eye diagrams were used to evaluate the quality of the transmission of the RD53A data through the e-links at the nominal speed of 1.28 Gb/s. The measurements showed that the e-link length has an important impact on the signal amplitude and jitter and the signal quality deteriorates faster with increasing length in the FFCs than in the TWPs. At comparable length, the TWPs showed better performance than the FFCs. Between the two TWPs flavours, the 36 AWG preserves better the signal integrity then the 34 AWG. Among the FFC prototypes, the design showing the best performance is the one implemented in the ring topology, which is in agreement with impedance measurement.

Although the signal amplitude and jitter are indeed worse in the e-link prototypes than in the standard commercial cables, the readout chip offers a possibility to increase the signal amplitude and the pre-emphasis functionality allows to boost the high frequency component of the signal before its transmission through the low-mass e-links. With the maximum amplitude and optimised pre-emphasis, all the available e-link prototypes, even the 2 m-long ones, satisfied the amplitude and jitter requirements of the lpGBT, receiving the data and taking care of the data merging. All the prototypes were demonstrated to correctly operate with minimum amplitude and pre-emphasis settings, leaving sufficient margin to compensate for additional signal deterioration caused by the irradiation or unforeseen effects. The focus of this study was put on the quality of the signal arriving to the lpGBT receiver. Another possible handle that was not discussed in this work, but that can be used to improve the BER and ensure the correct data transmission thought the low mass e-links is the adjustment of the sampling point and the equalisation functionality of the lpGBT [78].

Following this study, the TWPs still remain the baseline choice for the TBPX, since they offer an easy bending facilitating the integration. Moreover, it was shown that the signal quality is better in TWPs than in the FFCs at comparable length and the signal integrity deteriorates less with increasing length in the TWP, which makes them better candidates for the barrel, where longer e-links are need. The signal integrity was validated in the currently available TWP prototypes for a length up to 2.0 m, which offers a safe margin. Moreover, the lighter 36 AWG TWPs show better performance and thus, they are the preferred option for the upgrade. A study of the twisting and other design and fabrication parameters is ongoing to further optimise the design of these e-links. For the TFPX, both e-link options are still envisaged. The TWPs show a better performance, but the design of FFCs can be made more dense and can offer a lower mass. Moreover, the design of the FFCs could be improved based on the presented results and a development of a flex combining the power distribution and readout links is also envisaged, which could reduce the material budget in the TFPX even further.

The next steps towards the final choice of the e-links for the detector would be to perform BER tests with the full readout chain composed of prototype modules, e-links, portcards and optical links, as well as to evaluate the cross-talk between neighbouring data lines.

Conclusion

The High Luminosity upgrade of the LHC will impose stringent requirements on the LHC particle physics experiments and challenging operating conditions for the particle detectors. The CMS Inner Tracker pixel detector will be exposed to extreme radiation levels and high particle density. Therefore, the current Inner Tracker will be replaced with a totally new pixel detector designed to withstand the harsh environment of the High Luminosity LHC, while maintaining its excellent detection performance. The goal of this work was to contribute to the development of the readout system for the new CMS Inner Tracker. The study was organised in three parts, each part consisting in the evaluation of prototypes of one system component towards one system choice.

The pixel readout chip is one of the most crucial components of the Inner Tracker readout system and in particular the analogue front-end part, which receives the particle signal from the sensor and processes it. A new generation readout chip is being developed in 65 nm CMOS technology by the RD53 Collaboration for the ATLAS and CMS experiments. A large scale demonstrator chip called RD53A was designed and produced. The particularity of this chip is that it contains three different analogue front-end designs, called the Synchronous, Linear and Differential, offering a choice to the two experiments. The selection of one of the three options for integration in the final CMS readout chip represented an important system choice. Therefore, a complete test program was devised and carried out to qualify, evaluate and compare the three analogue front-ends. A systematic test procedure was established and applied. The CMS requirements against which the three designs were evaluated were: the threshold of $1000\,e^-$, the noise occupancy below 10^{-6}, a maximum 1% dead time in the innermost layer of the detector and a radiation tolerance sufficient to aim for maximum one replacement of the innermost layer during the detector lifetime.

The outcome of the extensive evaluation program was that all three designs were fully functional and each one presented strong points, but also limitations. The Synchronous analogue front-end showed very good timing performance and had the advantage of an automatic threshold tuning. However, this analogue front-end also showed a higher noise and the origin of this noise was not fully understood and reproduced in simulations, which made this option risky for the detector. On the contrary, the Differential analogue front-end demonstrated very low noise, but a saturation in the preamplifier feedback loop was limiting the signal return to baseline, which had an impact on the dead time. Moreover, the threshold tuning of this analogue front-end became problematic after irradiation at cold temperature. A design improvement was proposed and demonstrated in simulation up to 500 Mrad. However, given the high radiation levels expected in the CMS Inner Tracker at High Luminosity LHC, the other front-ends that could potentially achieve a higher radiation tolerance were preferred. The Linear analogue front-end had all the studied performance parameters within requirements, but its main drawback was a larger time walk causing out-of-time pileup. This limitation was addressed and mitigated in an improved design, which was verified in simulation. In addition, the improved design of the Linear analogue front-end was also implemented in a test chip and validated also after irradiation up to 1 Grad. The analysis of the advantages and disadvantages of the three front-ends in the context of the CMS Inner Tracker operation requirements led to the selection of the Linear front-end with improved design for integration in the final CMS readout chip.

After the pixel readout chip and the selection of the analogue front-end, another important system choice concerned the pixel sensor. The CMS Collaboration is considering two options for the future sensor pixel geometry: either square pixels of $50 \times 50\,\mu\text{m}^2$ or rectangular pixels of $100 \times 25\,\mu\text{m}^2$. The rectangular pixels show a better tracking resolution in simulations and they also have the advantage of producing hits that require a lower bandwidth, namely in the forward part of the detector. On the other hand, the main drawback of the rectangular pixels is the observed asymmetric cross-talk arising from the routing of the rectangular pixels of the sensor to the square pixels of the readout chip. Two methods were developed to evaluate the amount of cross-talk in rectangular pixels and to contribute to the choice of the pixel geometry for the new CMS Inner Tracker.

One method was based on charge injections from the readout chip and the other was using physics signals from a test beam. Both methods were applied to five selected sensors and demonstrated the presence of asymmetric cross-talk between pairs of rectangular pixels in planar sensors, while no cross-talk was observed between square or 3D pixels. The injection method can be performed quickly and easily in the lab to measure both the symmetric and asymmetric cross-talk. The second method based on the test beam measurement is sensitive only to the asymmetric cross-talk. Although, the originally foreseen method the measure cross-talk with a test beam could not be verified in this work, the cross-talk could still be extracted from the test beam data in two different ways: either fitting the efficiency as a function of threshold or fitting the efficiency as a function of track position within the pixel pitch. The first option was applied to only one sensor, but resulted in a cross-talk in agreement within error bars with the cross-talk measured with injections. The second option could be applied to all five tested sensors and resulted in a moderately different cross-talk compared to the cross-talk measured with injections, however, the results were globally consistent and lead to the same conclusions.

The cross-talk study showed that the asymmetric cross-talk in rectangular pixels is of the order of 10% and depends on the sensor bias voltage, as the cross-talk reaches its maximum at full depletion. It does not depend neither on the threshold, nor on the charge calibration, while it does depend on the preamplifier bias current, which was identified as a handle to reduce cross-talk. This study also triggered an improvement in the CMS sensor design: sensors with the so-called bitten implant demonstrated to reduce the cross-talk by about 2%. Given that the simulations showed that a cross-talk of about 10% has only a negligible impact on the tracking resolution of the future detector, and as it can be further reduced with the preamplifier bias and an improved sensor design, the rectangular pixels are the preferred option for the new CMS Inner Tracker.

The third component of the Inner Tracker readout system addressed in this work were the electrical links, which carry the hit data from the pixel modules to the rest of the data acquisition system. Almost 11 000 electrical links will be needed to control and readout the pixel modules of the new CMS Inner Tracker. Given this large amount of data links the corresponding electrical cables should have the lowest possible mass in order to have the lowest disturbance on the particle trajectories. Moreover, the data will be transmitted at much higher speed than in the present detector which, together with the ultra low mass of the links, will affect the signal integrity. Two types of electrical links are envisaged by the CMS Collaboration: the twisted pairs or flexible flat cables. Different prototypes of both types were evaluated and compared. The signal quality after transmission through the prototypes was studied in this work to determine the choice of an optimal implementation of the data links.

First, different electrical links were characterised as standalone components by measuring their S-parameters and impedance. This measurement allowed to identify the prototype of each type matching the best the target impedance of 100 Ω. Then, the prototype electrical links were evaluated as a part of a test system. The quality of the signal from the RD53A chip, after its transmission through different electrical links at the nominal speed of 1.28 Gb/s, was evaluated based on eye diagrams. In particular, the influence of the pre-emphasis functionality on the signal quality was studied. The pre-emphasis implemented in the RD53A chip allows to boost the signal amplitude and high frequencies before its transmission. All the available prototypes satisfied the requirements in terms of amplitude and jitter, when the optimal pre-emphasis was used. Moreover, all the prototypes were also demonstrated to correctly operate with minimum amplitude and pre-emphasis settings, leaving a margin to compensate for additional system effects. Based on this study, the 36 AWG twisted pairs were selected as the baseline choice for the Inner Tracker barrel, because they offer an easy bending facilitating the integration and show the best performance, especially with higher length. The signal integrity was validated in the current twisted pair prototypes for a length up to 2 m. Both types of electrical links are still envisaged for the forward part of the detector. The twisted pairs show a better performance, but the flexible flat cables can be made more dense and offer a lower mass.

In conclusion, the work of this thesis contributed to the selection of the analogue front-end for the new CMS pixel readout chip. The test results were also used as valuable feedback for chip designers to improve the three designs. The results of the detailed cross-talk study supported the preference for rectangular pixels for the new Inner Tracker sensors. Finally, the evaluation of the signal integrity and the data transmission performance of different electrical links will be used as useful input for the development of the final electrical links for the detector.

As a next step, the recently submitted full size readout chip with the selected analogue front-end features many design improvements and will need to be fully characterised and validated. For the pixel sensors, the currently on-going task force is evaluating different sensor designs. In particular, the bitten implants have to be fully characterised in test beams towards to selection of the final sensor design. For the electrical links, the knowledge acquired in this study can be used to further improve the design of the prototypes. More readout tests, especially at system level, are necessary before the decision on the final links is taken. Moreover, different prototypes evaluated in this work were used to set up a first complete readout chain from pixel modules to the back-end, which allowed to demonstrate the feasibility of the new readout system. Yet, many interesting readout tests with the complete readout chain are still to be performed.

Appendices

Appendix A

CERN accelerator complex

At the begging of the CERN acceleration chain, a bottle of compressed hydrogen gas serves as source of protons, and vaporised lead is a source of heavy ions. An electric field is used to strip the source atoms of electrons to yield protons or lead ions, after what they progressively follow the accelerating chain:

- **Linear accelerators** are the first ones in the chain. They generate low energy particles using radiofrequency cavities to charge cylindrical conductors with alternating positive and negative charge. When the particles pass through the linear accelerator the conductors behind push them and the conductors ahead pull them, causing them to accelerate. In addition, small quadrupole magnets ensure that the particles remain in a tight beam.
  - **Linac2**: was accelerating protons to 50 MeV for injection into the *Proton Synchrotron Booster* (PSB). This linear accelerator was feeding the CERN accelerator complex with protons since 1978 until it was recently retired in 2018.
  - **Linac3**: accelerates heavy ions to 4.2 MeV for injection into the *Low Energy Ion Ring* (LEIR). About 500 mg of lead are used every two weeks of operation. During the acceleration through Linac3, all the electrons are stripped away from the Pb atoms, which are transformed into bare nuclei easier to accelerate than different ions.
  - **Linac4** [151]: is the successor of Linac2, inaugurated in 2017 and connected to the PSB in 2020. It accelerates negative hydrogen ions (H^-), consisting of a hydrogen atom with an additional electron, to 160 MeV before injecting them into the PSB.
- **Low Energy Ion Ring (LEIR)**: receives long pulses of lead ions from Linac3 and splits them into four shorter bunches, which are then accelerated from 4.2 MeV to

72 MeV in approximately 2.5 s before being injected into the *Proton Synchrotron* (PS).

- **Proton Synchrotron Booster (PSB)**: is composed of four superimposed synchrotron rings, which receive protons from the Linac4 and boost them from 160 MeV to 1.4 GeV before injecting them into the PS.

- **Proton Synchrotron (PS)**: accelerates either protons received from the PSB or heavy ions from the LEIR. Since the beginning of its operation in 1959, it has also accelerated alpha particles (helium nuclei), oxygen and sulphur nuclei, electrons, positrons and antiprotons. This 25 GeV accelerator, with a circumference of 628 m, made of 277 room-temperature electromagnets, accelerates particles for injection to the more powerful *Super Proton Synchrotron* (SPS). It also provides test beam lines for R&D studies, fixed-target experiments, as well as some irradiation facilities.

- **Super Proton Synchrotron (SPS)**: is the second-largest machine in the CERN accelerator complex with a circumference of 11 km, made of 1317 room-temperature electromagnets. It has handled many different kinds of particles during its operation, such as: sulphur and oxygen nuclei, electrons, positrons, protons and antiprotons. It accelerates particles to an energy of up to 450 GeV for injection to the LHC. The SPS also has its own beam lines for test beams and fixed-target experiments, namely NA61/Shine [152, 153], NA62 [154], COMPASS [155] and AWAKE [156].

- **Large Hadron Collider (LHC)**: is the largest circular accelerator of the CERN accelerator complex and is the last element of the accelerating chain. Particle beams are accelerated here up to a record energy of 6.5 TeV.

Apart from the main acceleration chain, the CERN accelerator complex also comprises the a standalone accelerator and two decelerators:

- **CERN Linear Electron Accelerator for Research (CLEAR)**: is a 20 m-long linear accelerator, producing bunched electron beams with an energy of about 220 MeV. It operates independently from the CERN's main acceleration chain and its primary focus is on general accelerator R&D, including prototyping and validation of accelerator components, beam-diagnostics development, radiation tests of electronic components and dosimetry for medical applications, in particular the cancer therapy.

- **Antiproton Decelerator (AD)** [157]: produces low-energy antiprotons and creates antiatoms for antimatter studies. A proton beam from the PS is fired into a block of metal, which creates many secondary particles, including a lot of antiprotons. The antiprotons, which emerge from the block at diverging angles, are focused before they reach the AD. Only a fraction of them have the right energy to be injected into the AD and transformed into a low-energy beam that can be used to produce antimatter.

- **Extra Low ENergy Antiproton (ELENA)** [158]: is a synchrotron, with a circumference of 30 m, coupled to the AD. It further decelerates antiprotons, reducing their energy from 5.3 MeV to only 0.1 MeV. Currently the AD and ELENA serve several experiments studying antimatter and its properties, namely: AEgIS, ALPHA, ASACUSA, BASE and GBAR.

Appendix B

Upgrade of the CMS subdetectors

Tracker upgrade. The silicon tracker will be entirely replaced during the Phase-2 Upgrade and the new tracker will feature increased radiation hardness, increased forward acceptance and compatibility with the higher data rates and a longer trigger latency. The granularity of both the IT and OT will be increased to maintain or improve the tracking and vertex reconstruction performance of the detector at much higher pileup. This will be achieved in the IT by using smaller pixels and in the OT by shortening the length of the silicon strips without significantly changing the pitch and by equipping a part of the detector with macro-pixels. A number of design improvements will also lead to a much lighter detector providing significantly improved transverse momentum resolution compared to the present detector [65]. In addition, the OT will provide tracking information to the L1 trigger at 40 MHz to identify tracks above a configurable transverse momentum threshold [65]. Currently, the L1 trigger uses only information from the calorimeters and the muon detectors and the track information is only available at the HLT. The participation of the tracker to the L1 trigger will ensure powerful background rejection at the earliest stage of the event selection and allow to keep the trigger rates at a sustainable level without sacrificing physics potential.

Calorimeter upgrade. The electromagnetic and hadronic endcap calorimeters will be entirely replaced with a new combined sampling *High Granularity Calorimeter* (HGCAL) [159]. It will count around 6.5 million channels divided into 52 layers. The first 28 layers with a total thickness of 34 cm will form the electromagnetic section made of high-density absorbers interleaved with hexagonal silicon pad sensors. The following 24 layers with a thickness approaching 160 cm will form the hadronic section. Similarly to the electromagnetic section, the detection layers in the hadronic regions close to the beam pipe will be made of silicon pad sensors. At larger distances from the beam line, plastic scintillator tiles read out by *silicon photomultipliers* (SiPMs) will be used. With silicon pad cell sizes of 0.5–1 cm^2 and the many sampling layers the HGCAL will have an unprecedented transverse and longitudinal granularity, leading to improved pileup rejection and particle identification [159, 160].

In the ECAL barrel, the existing lead tungstate crystals and *avalanche photodiodes* (APDs) will be preserved. However, after LS3 the crystals will be cooled down to a lower temperature, from the current 18 °C to 8 °C. This will allow to decrease the dark current in the APDs after irradiation and to limit the electronic noise in the readout. The ECAL front-end electronics will be improved to cope with the trigger latency and bandwidth requirements. The new front-end boards will also allow the exploitation of single-crystal information by the L1 trigger, while the in the present system, the L1 has only access to the information integrated over groups of 5×5 crystals. In the HCAL barrel, the *hybrid photodetectors* (HPDs) of the original detector were already replaced during the Phase-1 Upgrade with SiPMs offering a greater photon-detection efficiency. Hence, the photodetectors and readout electronics will not be further upgraded during LS3 and only the scintillation tiles closest to the beam line will be replaced [46, 161].

Muon system upgrade. The existing gaseous detectors equipping the muon system are expected to cope with the increased radiation and particle rates of the HL-LHC and will be conserved. However, the forward region of the muon system composed of CSC chambers lacks trigger redundancy, which will be improved during LS3 to maintain a good L1 trigger acceptance. The existing four CSC stations in this region will be enhanced with additional chambers using new detector technologies with higher rate capability. The first two stations are in a region where the magnetic field is still quite high. Hence, *gas electron multiplier* (GEM) chambers will be used for good muon position and momentum resolution and improved matching with tracks measured in the silicon tracker. The other two stations will be enhanced with improved low-resistivity RPCs with lower granularity, but good timing resolution to mitigate background effects. These new chambers will add trigger redundancy, improve the triggering and reconstruction performance, and increase the forward acceptance of the muon system. Moreover, the front-end electronics for the DT chambers and CSC chambers will be replaced with improved versions to increase radiation tolerance, readout speed, and performance [46, 162].

MIP Timing Detector. The new *MIP Timing Detector* (MTD) [163] will be inserted between the tracker and the electromagnetic calorimeter. The purpose of the MTD is to precisely measure the production time of MIPs with the timing resolution down to 30 ps. It will also contribute to a more precise charged hadron identification and the search for long-lived particles. The barrel part of the new detector, will be made of Cerium-doped Lutetium-Yttrium Oxyorthosilicate (LYSO:Ce) scintillating crystals readout by SiPMs. Due to radiation tolerance limitations, this technology can not be used for the MTD endcaps, which will be instrumented with *low gain avalanche detectors* (LGADs).

Appendix C

Physics and properties of silicon

Charged particles lose energy in matter via the different scattering processes described in the previous section. A part of this energy is converted into ionization, which can be detected as electrical signal and used to determine the particle position to reconstruct its trajectory. The majority of the current tracking detectors in HEP are based on semiconductor sensors. The advantage of semiconductors is their small ionization energy, i.e. about ten times smaller than for gas. Thus the same deposited energy produces an order of magnitude higher signal and a better energy resolution can be achieved. Moreover, because of their greater density, they have a greater stopping power and they allow to build compact detectors with very fast time response. The understanding of the detection principle of semiconductor sensors requires some knowledge of semiconductor physics and properties, summarised in this section. The emphasis is largely put on silicon, being the most commonly used semiconductor and the material that will be used for the new CMS IT.

Energy band structure

The energy levels of electrons in solid materials can be described by their electronic band structure. The adjacent energy levels are so close that they can be considered as a continuum and represented as energy bands. A solid has an infinite number of allowed energy bands, which may be separated by gaps or ranges of forbidden energies. The most important bands, those relevant for understanding semiconductors, are represented in a simplified electronic band structure in Figure C.1.

The lower band, called *valence band*, corresponds to the outer-shell electrons bound to the atoms of the crystal lattice and participating to the covalent bonds. The upper band, called *conduction band*, represents electrons free to move through the crystal. The electrons in this band contribute to the electrical conductivity of the material. The two bands are separated by a so-called *bandgap*, where no electronic states can exist. The energy gap is defined as the energy difference between the lowest level of the conduction band E_C and the highest level of the valence band E_V. The *Fermi level* E_F represents a hypothetical energy level that has 50% probability to be occupied at any given time [164].

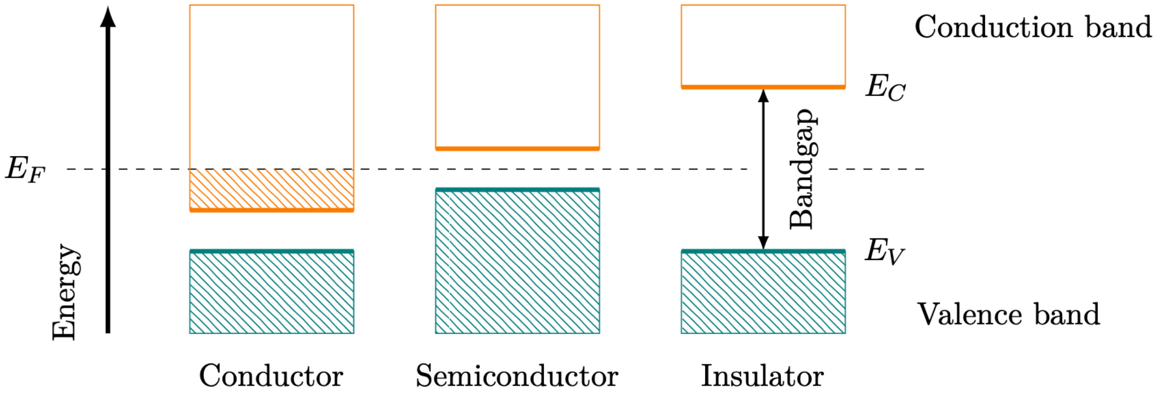


Figure C.1: Electronic band structure in solids [119].

Based on their band structure, solid materials are divided in three categories: conductors, semiconductors and insulators, as it is indicated in Figure C.1. In conductors, the conduction band is partially filled with electrons, i.e. the Fermi level is located inside the conduction band, or the valence and conduction bands may overlap, resulting in a high electrical conductivity. The insulators are not conductive, given that their bandgap is too large for electrons to be excited into the conduction band. A semiconductor has an intermediate-size, but non-zero, bandgap that behaves as an insulator at $T = 0\,\mathrm{K}$, but allows thermal excitation of electrons into the conduction band at higher temperatures. This results in an intermediate electrical conductivity, which increases with temperature.

Silicon as a semiconductor

At low temperature and in the absence of impurities, the valence band of a semiconductor is full and the conduction band is empty. Electrons of the crystal can be excited either receiving a part of the thermal energy or energy from a passing particle. If the received energy is larger than the bandgap, the electron is elevated to the conduction band leaving a hole in the valence band and an electron-hole pair is created. This causes an imbalance in the occupied states, induces a net charge displacement and hence conduction. In the absence of electric field, free charge carriers move by diffusion. If an electric field is applied, they move in the direction parallel to the applied field: electrons opposite to the field vector and holes in the same direction.

The semiconductor typically used in tracking detectors in HEP is silicon. This solid crystal, with a blue-grey metallic tinge, is a tetravalent element. It has the atomic number

14 and is member of the 14^{th} group of the periodic table. The band gap of silicon at $T = 300\,\text{K}$ is $1.124\,\text{eV}$. The main advantage of silicon as the detector material is its low ionization energy of only $3.6\,\text{eV}$. Compared to about $30\,\text{eV}$ needed to ionise a gas, the number of free charge carriers produced by ionization in silicon is about ten times larger. The average energy loss of a MIP in silicon is $390\,\text{eV}\,\mu\text{m}^{-1}$, creating 108 electron–hole pairs per µm. Silicon is present in most of the modern technology and, thanks to the electronics industry, it is the most intensively studied semiconductor. Therefore, its properties are well known and its processing is well developed. Moreover, it is available in large quantities at an affordable cost. Another advantage of silicon, largely exploited for the fabrication of silicon devices, is the existence of a natural oxide and the possibility of changing its properties by doping [165].

Intrinsic vs. doped silicon

A semiconductor is called *intrinsic* if the concentration of impurities is negligible and therefore the majority of free charge carriers is generated by thermal excitation. In this case the number of electrons in the conduction band must be exactly the same as the number of holes in the valence band, as electrical neutrality should be preserved.

The impurities in the crystal lattice of a semiconductor have an impact on its electrical properties. The conductivity may be enhanced by introducing intentionally a small fraction of other materials into the crystal structure. This process, called *doping*, creates additional energy states in the bandgap and increases the probability to excite the charge carriers. The elements used for doping are either from the 13^{th} group of the periodic table (e.g. boron) or from the 15^{th} group (e.g. phosphorus), so that they have either one valence electron less or one more than silicon.

Silicon is tetravalent, so it forms covalent bonds with four nearest silicon atoms. If we add a pentavalent element, such as phosphorus, it will create four covalent bonds with silicon, leaving one valence electron free, as it is illustrated in Figure C.2(a). The pentavalent impurities provide electrons to the semiconductor and therefore they are called *donors*. Doping with donors introduces a new energy level in the band gap very close to the conduction band, as indicated in Figure C.2(c), causing a shift of the Fermi level towards the conduction band. Since the extra electrons can be excited easily, at room temperature practically all donor atoms are ionised, the concentration of free electrons is equal to the concentration of donor atoms. The majority of charge carriers in the material are electrons and therefore silicon doped with donors is called *n-type* material.

A similar process happens when silicon is doped with trivalent impurities, such as boron. Boron has one valence electron less than silicon, so it will trap an electron from the valence band to establish four covalent bonds with the neighbouring atoms of silicon. Therefore trivalent impurities are called *acceptors* and they generate free holes in the semiconductor, as it is shown in Figure C.2(b). Doping with acceptors introduces an energy level in the band gap close to the valence band (Figure C.2(c)) and the Fermi level is shifted towards the valence band. At room temperature all acceptor atoms are ionised and the majority carriers are holes. Hence the acceptor-doped silicon is called *p-type* material.

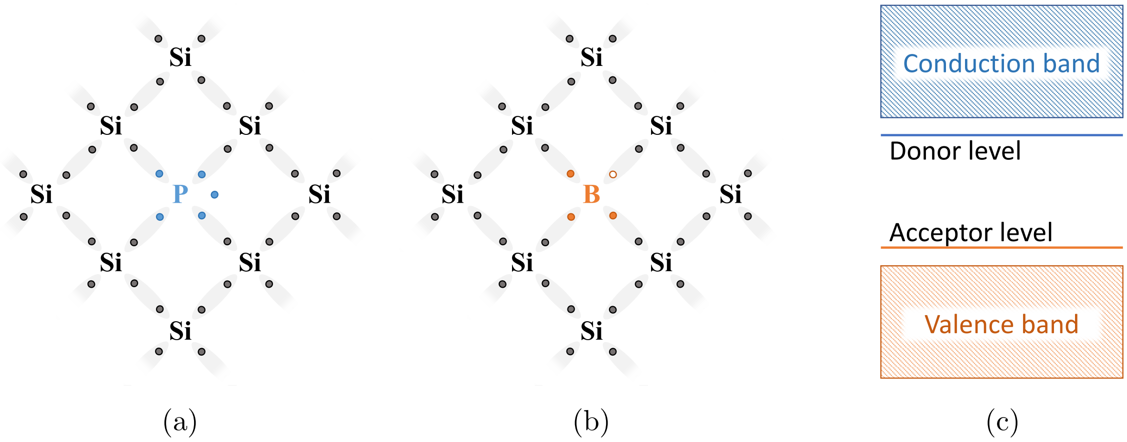


Figure C.2: Silicon doped with a donor atom of phosphorus providing a free electron (a), with an acceptor atom of boron providing a hole (b) and the additional energy levels created in the bandgap after doping (c).

PN junction

The doped silicon presents a much larger electrical conductivity than the intrinsic one, so it is favoured in radiation detectors. The conductivity can be further increased by doping with very high concentrations of impurities, in which case n^+ or p^+ notation is used. The concentration of charge carriers in an intrinsic silicon at room temperature is $\mathcal{O}(10^{10}\,\mathrm{cm}^{-3})$ and the typical doping levels are in the range between $\mathcal{O}(10^{12}\,\mathrm{cm}^{-3})$ for standard doping, up to $\mathcal{O}(10^{19}\,\mathrm{cm}^{-3})$ for high doping [165]. The signal created by a particle traversing silicon is $\mathcal{O}(10^4)$ electron-hole pairs, which is several orders of magnitude below the number of charge carriers already present in silicon and therefore the signal is difficult to distinguish. Hence the amount of free charge carriers has to be reduced to be able to detect the signal. This can be achieved in a so-called *PN junction.*

When two differently doped regions, one n-type and the other p-type, exist in the same crystal, a semiconductor PN junction is created, as it is depicted in Figure C.3. In the n-type region the density of conduction electrons is much higher than in the p-type region and the opposite applies for holes. This sharp gradient causes a diffusion of the majority charge carriers across the junction from the regions of high concentration to those of low concentration. The diffusing electrons and holes recombine near the junction and the region around becomes depleted of free charge carriers. In the *depletion zone* the donor and acceptor ions are left without their associated free charge carriers and the region is electrically charged, hence it is also referred to as the *space charge region.*

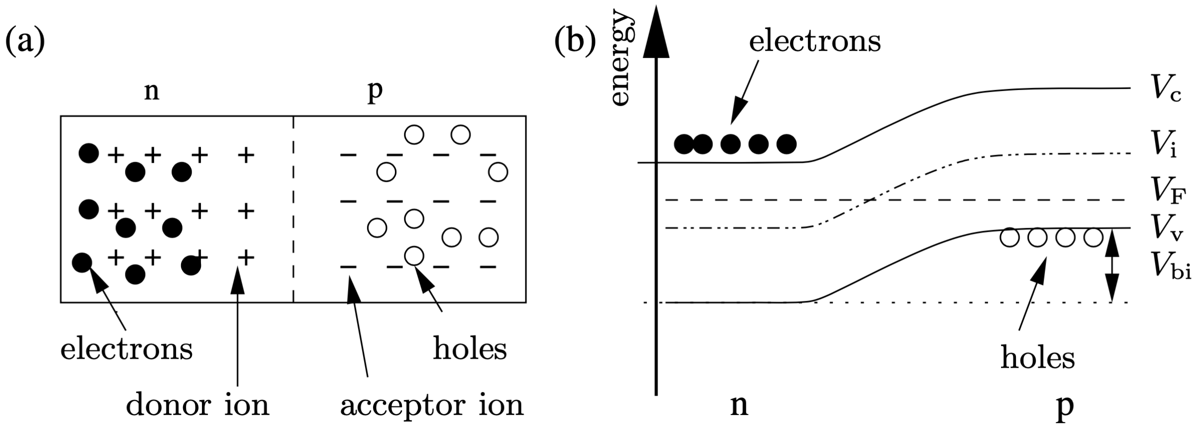


Figure C.3: Illustration if the PN junction (a) and its electronic band structure (b) [68, p. 41].

The electron diffusion results in a uniform positive space charge on the n-side of the junction and the hole diffusion results in a uniform negative space charge on the p-side. The total charge must be zero to preserve the electrical neutrality of the semiconductor. Therefore, if the concentration of donors on the n-side and acceptors on the p-side are equal, the depletion region extends to an equal distance on both sides of the junction, otherwise it extends more towards region with lower doping.

The accumulated space charge creates an electric potential difference across the PN junction, leading to the establishment of an electric field, preventing further diffusion. The equilibrium is reached when the Fermi levels of both materials are the same. The difference between the two Fermi levels by which the energy bands are bent is called the *build-in voltage* V_{bi}, as indicated in Figure C.3. The built-in voltage is usually of the order of few millivolts [119].

Reverse bias and leakage current

Any free charge carrier present in the depletion zone of the PN junction is swept out by the existing electric field: electrons towards the n-type and holes towards the p-type material. The charge carriers being continuously removed, the depletion zone is well suited for the detection of a particle-induced signal. However the width of the depleted zone in an unbiased PN junction is quite small, of the order of few microns [119].

By adding an electrode to each side of the junction and by applying an external voltage, the size of the depleted region can be increased or decreased, depending on the applied voltage polarity. When the p-side is put at higher potential than the n-side, the junction is in *forward bias* and the majority carriers flow freely through the junction. In this case, even a small voltage causes the junction to conduct large currents. On the other hand, when the n-side is at higher potential than the p-side, the junction is in *reverse bias*. The external voltage is applied in the same direction as the built-in voltage, which helps removing further majority carriers and extends the depletion region. Increasing the reverse bias voltage increases the thickness of the depletion zone. The latter has to be extended as much as possible to maximize the detection volume. The depletion zone can also be assimilated to a charged parallel plate capacitor and its capacitance is inversely proportional to the width of the depletion zone. Therefore, a larger depletion zone allows to minimize the capacitance and to lower electronic noise. The combination of these two effects of the reverse bias allows to increase the SNR of the detector.

In a reversely biased PN junction, a small current, called *leakage current* or *dark current* is still flowing. It is mainly caused by the thermal generation of electron-hole pairs within the depleted volume, which are swept away by the electric field. The generation rate increases with the size of the depletion zone, and therefore increases with the reverse bias voltage. Surface effects at the edges of the junction due to a relatively large voltage gradients may also contribute to the leakage current as well as impurities introduced during the processing. Moreover, the radiation damage to the crystal lattice increases the leakage current considerably, since the defects introduce additional energy levels in the band gap. The leakage current is an important parameter for silicon sensors since it strongly affects noise and power consumption of the detector. It can be reduced by appropriate cooling.

Appendix D

Particle interaction with matter

Depending on the type and energy of the particle and the type of the traversed material, different interactions with atomic constituents of matter take place and produce signal that can be detected. Given that the scope of this thesis is the development of the CMS IT that measures the trajectories of charged particles, this section exposes interactions of charged particles only. Charged particles passing through matter interact via two main electromagnetic processes:

- elastic scattering from nuclei of the material, causing the deflection of the particle from its incident trajectory,
- inelastic collisions with the atomic electrons of the material, inducing potentially large loss of energy.

Multiple scattering

The charged particles passing through matter undergo elastic Coulomb scattering interactions with nuclei, although they are much less probable than collisions with atomic electrons. Assuming that the nuclei of the material are much more massive than the incident particle, the energy transfer can be neglected and the main consequence of the nuclear collisions is the *multiple scattering*. The many small and independent angular deflections of the particle result in a change in direction of the particle trajectory, as illustrated in Figure D.1. The probability distribution of the net scattering angle in one plane, after many interactions, can be approximated with a centered Gaussian distribution with an RMS given by [166, 167]:

$$\theta_{RMS} = \frac{13.6\,\mathrm{MeV}}{\beta pc} z \sqrt{\frac{x}{X_0}} \left[1 + 0.038 \ln \frac{x}{X_0}\right] \tag{D.1}$$

where the scattering angle θ is expressed in radians, β is the particle velocity in units of speed of light c, p the particle momentum, z the electric charge in units of electron charge, denoted e and x the thickness of the traversed material. X_0 is called the *radiation length* and it is defined as the mean path length in the material over which the electron energy is reduced by a factor $1/e$ due to radiation loss only [168]. For instance the radiation length of silicon is 9.36 cm [68].

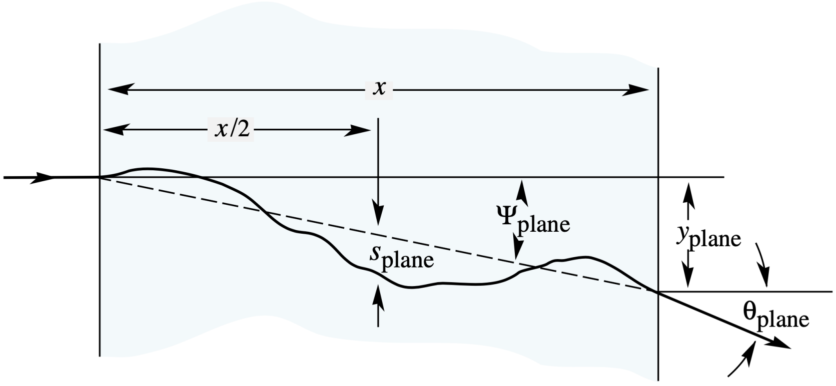


Figure D.1: Diagram illustrating the scattering of a particle through a material of thickness x resulting in a displacement of y and a scattering angle of θ in one plane [167].

Ionization energy loss

The dominant interaction of charged particles with matter is through inelastic collisions with the atomic electrons, resulting in a loss of energy of the incident particle, that cause ionization of the traversed material. The average energy loss per unit of path length of a charged particle, heavier than an electron at rest, passing through a layer of material, is given by the *Bethe-Bloch formula* [167, 168]:

$$-\langle\frac{dE}{dx}\rangle = K\frac{Z}{A}\frac{z^2}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_ec^2(\beta\gamma)^2T_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} - \frac{C}{Z}\right] \tag{D.2}$$

The energy loss depends on the path length of the particle in the material and is also called the *mass stopping power*, expressed in MeV g^{-1} cm^2. When the Bethe-Bloch formula is multiplied by the material density ρ the energy loss can be expressed as *linear stopping power* in MeV cm^{-1}.

For most materials the ratio is $Z/A \approx 1/2$, therefore, the energy loss is almost independent of the material type and is mainly dependent on the particle velocity β, that is to say on the particle energy. The energy loss calculated with the Bethe-Bloch formula as a function of kinetic energy for different type of particles is shown in Figure D.2.

K	$4\pi N_A r_e^2 m_e c^2 = 0.307\,075\,\mathrm{MeV\,cm}^2$,
Z	atomic number of the material (14 for silicon),
A	atomic mass of the material (28 for silicon),
z	electric charge of the particle in units of electron charge,
β	velocity of the traversing particle in units of the speed of light c,
γ	Lorentz factor $1/\sqrt{1-\beta^2}$,
$m_e c^2$	rest mass of the electron (0.511 MeV),
I	mean excitation energy (137 eV for silicon),
T_{max}	maximum kinetic energy transfer to an electron in a single collision,
$\delta(\beta\gamma)$	density effect correction,
C	shell correction.

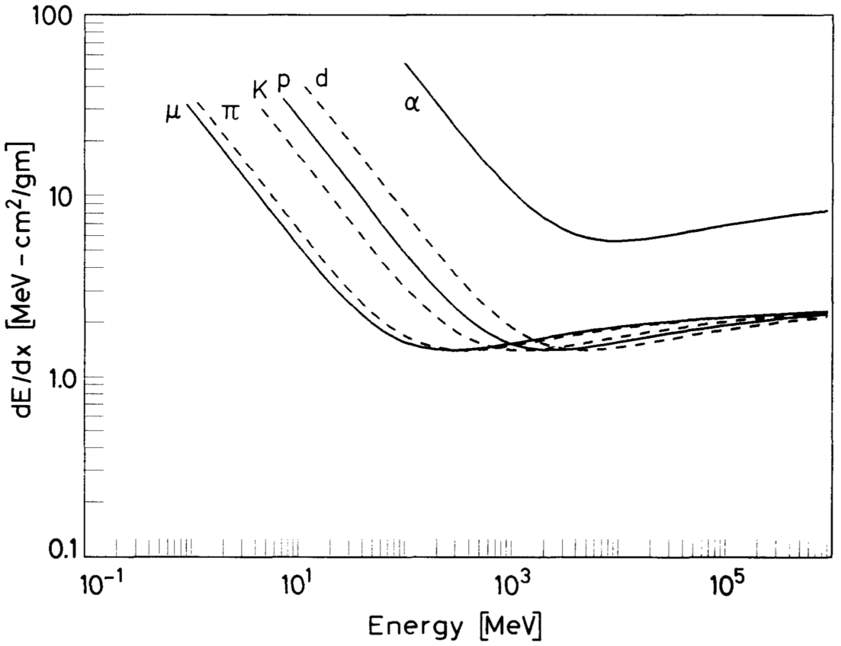


Figure D.2: Energy loss obtained with the Bethe-Bloch formula as a function of the kinetic energy for different type of particles [168]

At low energies the factor $1/\beta^2$ is dominant and the energy loss decreases with increasing velocity, until about $\beta = 0.96$ or $\beta\gamma = 3$ where a minimum is reached. A particle with an energy loss in the minimum of the Bethe–Bloch formula is called a *minimum ionizing particle (MIP)*. The value of the minimum depends on the square of the particle charge but very weakly on the particle mass. Therefore, the minimum energy loss is almost the same for all particles with the same charge, as it can be observed in Figure D.2. At higher energies the factor $1/\beta^2$ becomes almost constant and the logarithmic term leads to a slow rise again, which is finally cancelled by the density correction. Due to the flatness of the curve at higher energies, the expression MIP is often used for all particles with $\beta\gamma > 3$.

Two corrections can be added to the Bethe-Bloch formula: the density effect correction δ, and the shell correction C, which are important at high and low energies, respectively. The density effect arises from the fact that the electric field of a relativistic traversing particle tends to polarize the atoms along its path. The shell correction accounts for effects which arise when the velocity of the incident particle is comparable to the orbital velocity of the bound electrons [168]. The Bethe-Bloch formula with the two corrections is valid in the region of particle velocities between $0.1 < \beta\gamma < 1000$ with an accuracy of few percent [167]. At very low energies the particle velocity is comparable or even lower than the velocity of the electrons and at very high energies radiative effects become important.

Radiation energy loss

Like heavy charged particles, electrons (and positrons) also suffer an ionization energy loss when passing through matter, as described previously. The Bethe-Bloch formula is applicable, however some modifications are required to take into account their small mass and the fact that they interact with identical particles when traversing the material. This has an impact mainly on the maximum energy that can be transferred to the electron in a single collision T_{max}.

An additional energy loss mechanism comes into play for electrons: the emission of electromagnetic radiation arising from scattering in the electric field of a nucleus, called *bremsstrahlung*. When the electron is deviated from its trajectory by the electrical attraction of the nucleus, it causes a radiation energy loss.

The radiation probability is proportional to the inverse square of the particle mass. Therefore, for heavy charged particles the bremsstrahlung becomes relevant only for energies well beyond the range of validity of the Bethe-Bloch formula, while for electrons the bremsstrahlung contributes to the total energy loss because of their small mass. At energies below few MeV the radiative loss is still relatively small but it increases with energy and at few tens of MeV it is comparable to the ionization loss. If a critical energy is reached the bremsstrahlung dominates.

Unlike the ionization loss, which is quasi-continuous along the path, a big part of the radiation energy can be emitted in one or two photons, which causes large fluctuations in energy deposition by an electron beam.

Energy loss fluctuations

The energy loss is not a continuous, but a probabilistic process and therefore, it is subject to fluctuations. Two identical particles, with the same initial energy, traversing the same material, will not encounter the same number of collisions and hence the same energy loss.

The statistical fluctuation in the number of collisions and in the energy transferred in each collision, results in an energy loss distribution. The characteristics of the distribution depends on the thickness of the traversed material. For a very thick layer, the number of collisions is large, and the energy loss can be represented in a first approximation by a Gaussian distribution [168]. Less collisions are occurring in thinner layers and the statistical fluctuations are becoming more important and are better described by a *Landau distribution* [169].

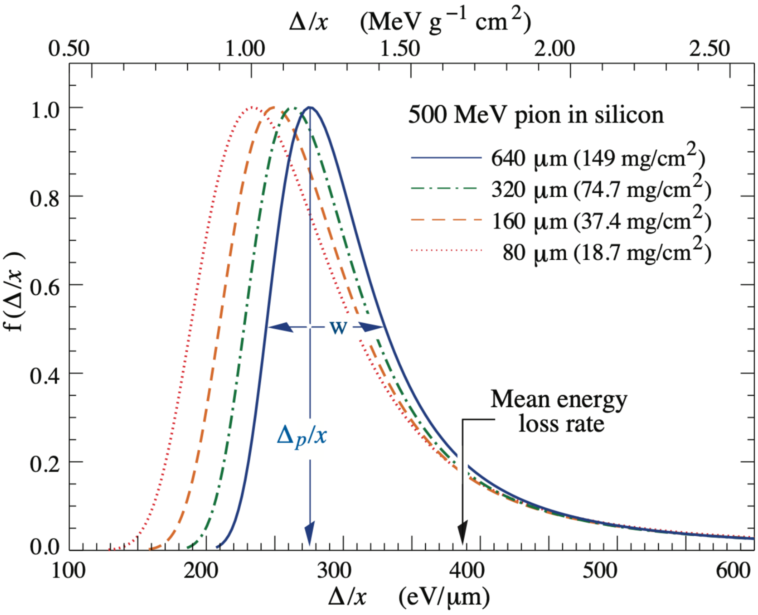


Figure D.3: Energy loss distribution for 500 MeV pions traversing silicon with four different thicknesses [167].

The energy loss distributions obtained from 500 MeV pions traversing silicon of different thicknesses are shown in Figure D.3. The Landau distribution has a long tail towards higher energies, which is partially due to possible large energy transfers in single collisions. However, the main cause of the Landau fluctuation are the rare, but highly ionizing *δ-electrons*, also called *knock-on electrons*. The inelastic collisions with atomic electrons of the material can be either soft, resulting in electron excitation, or hard in which case the electron obtains enough energy to become an ionizing particle itself and causes a secondary ionization. As a result of this tail, the mean energy loss given by the Bethe-Bloch formula is larger than the MPV of the Landau distribution.

In test beam experiments with thin silicon sensors, it was observed that while the general features of the Landau distribution are in agreement with measurements of the deposited energy, the width of the measured distribution is often broader than expected. The effect is particularly noticeable for very thin silicon layers of $\mathcal{O}(100\,\mu\text{m})$. The observed peak broadening in thin silicon layers compared to the simple Landau theory can be explained by the effects of binding of atomic electrons [170]. Many models address this topic [170, 171] but in practice the broadening can be well approximated by the convolution of a Landau with a Gaussian distribution, as it is shown in Figure D.4

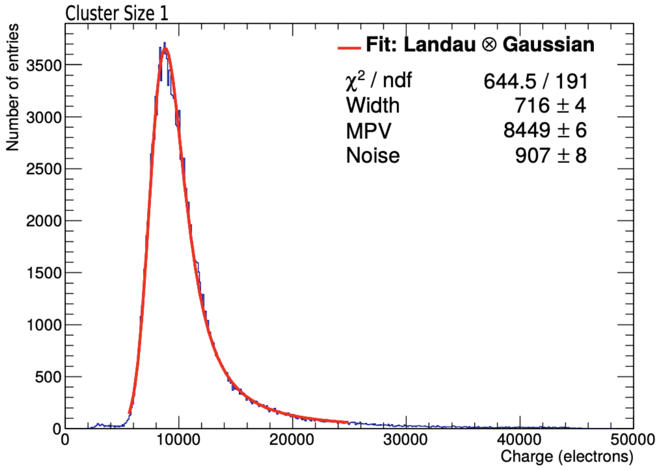


Figure D.4: Charge collected in a 130 µm silicon sensor with 120 GeV proton beam [172].

Ionization energy

The energy deposited by the charged particle in matter, via processes described previously, is absorbed by the atoms resulting in their ionization and excitation of their electrons. When high energetic charged particles penetrate a material, they cause a uniform ionization along their paths and produce many free electron-hole pairs. The average energy needed to produce one electron-hole pair, called *ionization energy* denoted ϵ, is a material property that is independent of the incident particle energy and type. The ionization energy is also inversely proportional to the temperature [164]. The average number of electron–hole pairs N, generated by a particle traversing the material, can be calculated by dividing the deposited energy E by the ionization energy ϵ:

$$N = \frac{E}{\epsilon} \tag{D.3}$$

A part of the energy deposited in the material is used for ionization and the rest of the deposited energy goes into the excitation of phonons, which dissipate thermally [68]. The fraction of deposited energy used for electron–hole pairs creation and phonon generation is subject to fluctuations, which translate in a fluctuation in the number of produced charge carriers that can be described by the *Fano factor* [173]. The Fano factor determines the best possible energy resolution of a detector. The experimental determination of this quantity is difficult but for most semiconductors the Fano factor is $\mathcal{O}(0.1)$ [68].

Appendix E

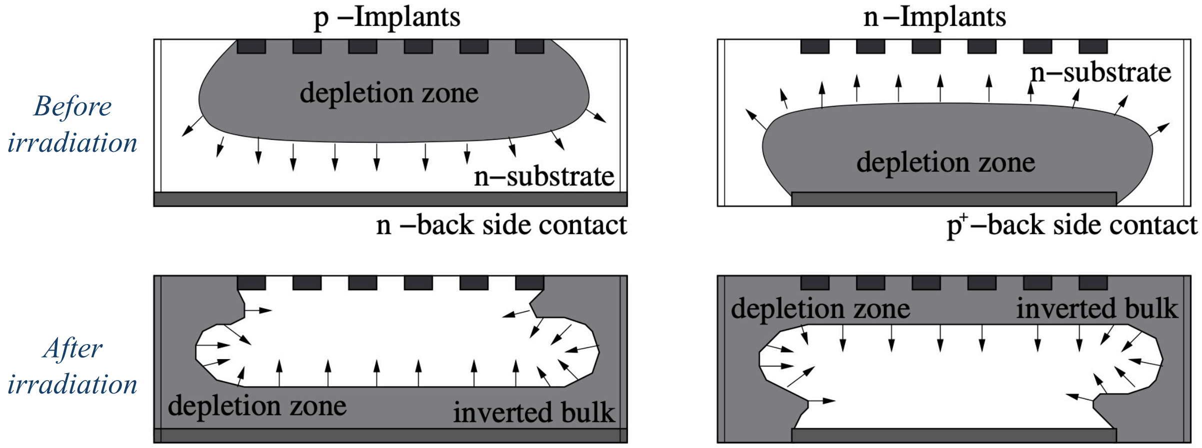
Sensor types

p-in-n sensor

Historically, the n-doped bulk was the first available in detector grade quality. The first silicon sensors were realised with p^+-implants in high resistivity n-substrate, as it was the case for the first CMS tracker. In these so-called p-in-n sensors, the p^+ side is segmented into pixels, while the ohmic backside is a continuous n^+-layer, as is shown in Figure E.1 (a). Therefore, no lithography is necessary on the backside and a single-sided process can be used. Moreover, the p^+-pixels are isolated from each other by design, so the main advantage of this type of sensors is the simplicity of the fabrication and consequently its low costs. In addition, they usually display a higher breakdown voltage. Before irradiation, the depletion zone of a p-in-n type sensor grows from the pixel side, as shown in Figure E.1(a) and therefore, it can be operated also partially depleted. After irradiation, the n-type bulk undergo the type inversion, the sensor becomes equivalent to p-in-p type and the depletion region grows from the junction at the backside. The full depletion is necessary after irradiation to avoid the p^+ implants to be shorted via the inverted undepleted bulk, which requires very high voltages.

n-in-n sensor

Since the radiation hardness of the p-in-n sensors is limited to about $\mathcal{O}(10^{14}\,\mathrm{n_{eq}/cm^2})$ [68], the CMS Phase-1 tracker adopted the n-in-n type sensors, as it is suitable for higher fluences. The doping of the implant and the bulk is of the same type and therefore, before irradiation the depletion zone grows from the PN junction located at the backside, as it is shown in Figure E.1 (b). The n^+-pixels are only isolated from each other at full depletion, which is not of particular concern in a non-irradiated sensor. Nevertheless, the electric field close to the pixels is weak and over depletion is desirable. After the type inversion, the depletion zone is growing from the pixel side and so the sensor can be operated even partially depleted after irradiation. The type inversion in this type of



(a) p-in-n type sensor (b) n-in-n type sensor

Figure E.1: Comparison of the p-in-n type sensor (a) and the n-in-n type sensor (b) with partial depletion before (up) and after irradiation (down) [68, p. 83].

sensors actually helps to improve the radiation tolerance, since even when the maximum applied bias voltage is insufficient to reach full depletion the area surrounding the readout electrodes is still depleted. The radiation tolerance is further improved by the fact that the n^+-pixels are collecting electrons, which have three times higher mobility than holes and are less prone to trapping.

The implants and the bulk having the same type of doping the PN junction is formed with the p^+-layer on the backside. The latter must not extend up to the sensor cutting edge, as it is the case in the p-in-n (Figure E.1 (a)). The mechanical damage caused by cutting of the individual sensors on wafers makes the edge highly conductive and this would short the junction. Therefore, a structured backside is used, the p^+-layer stops before the edge and a *guard ring* structures are implemented to establish a gradual voltage drop between the high bias voltage applied to backside and the cut edge, which is close to the ground potential. A multi guard-ring structure in an n-in-n sensor is illustrated in Figure E.2. Since the guard ring needs to be placed on the PN junction side, for sensors with the same implant and bulk doping, like n-in-n, the guard ring is placed on the backside, which requires the more difficult and costly double-sided fabrication process. In addition, the n^+-implants need pixel isolation. Hence, the major drawback of the n-in-n sensor is the cost and these sensors are at least a factor two more expensive than the p-in-n type [68].

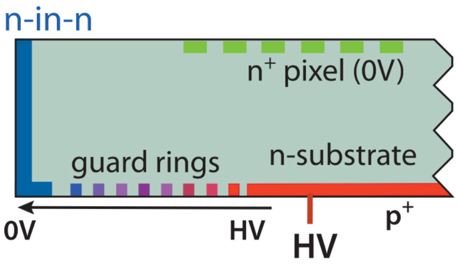


Figure E.2: Illustration of a multi-guard ring structure placed on back side of an n-in-n sensor type [69].

p-in-p sensor

More recently, the p-type substrates, which do not undergo the type inversion, became available in detector quality. Placing p-type implants in p-bulk would combine the disadvantage of expensive double-sided manufacturing of the n-in-n sensor and the hole collection of the p-in-n type, so it is not an attractive approach.

Appendix F

Telescope alignment

The telescope alignment procedure is done in two phases: the pre-alignment using cluster correlations and the alignment using tracks. The alignment procedure is done in several iterations. In each iteration all the previous alignment steps are repeated and a new step can be introduced. The telescope alignment procedure is explained in detail in the following and the telescope configuration and the plane numbering are reminded in Figure F.1.

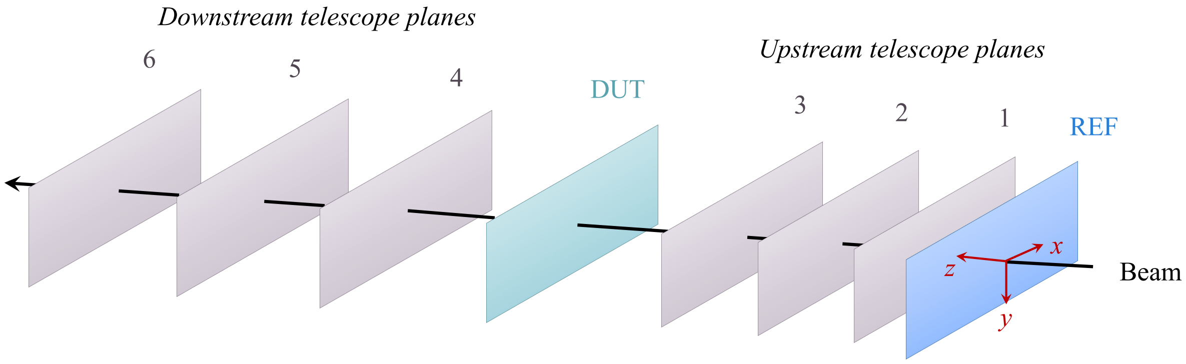


Figure F.1: Telescope plane numbering.

STEP 1: Cluster-based pre-alignment

- **Iteration 0 - Shift in X and Y :**
  The shifts along x and y are corrected, using the cluster pair correlations. For each telescope arm, the cluster positions in x and y of the first and the last plane are correlated to the cluster positions of middle plane considered as fixed. The histograms of the distances between cluster positions are built and the shift from zero of the mean of the distribution is used as the alignment correction of the corresponding plane.
- **Iteration - Rotation around the Z-axis**

Then the XY plane is rotated around the Z-axis. For each pair of planes, the distance between clusters in x versus the position on y is plotted and a linear fit is done. The slope of this plot is used as a correction of the X rotation. And the slope of the plot of the distances in y versus position on x is used as a rotation of Y.

- **Iteration 2 - No new step:**
  Repeat previous steps only.

STEP 2: Track-based alignment

After the pre-alignment step a finer alignment is achieved using tracking. The triplets found in each telescope arm are used.

- **Iteration 3 - First and third plane Z shift in Z:**
  For each valid triplet the residuals are compared to the triplet slope. The tracks with perfectly normal incidence are not sensitive to the Z position, it represents a weak mode. However, the beam is not perfectly collinear and some tracks are slightly divergent from the beam axis, which allows to correct the position along z. Indeed, if the triplet middle plane is misaligned in z, the inclined tracks will cause bigger residuals. The linear fit of the plot of triplet residuals versus the triplet slope is done for both telescope arms and the slope of the fit is used to to correct the shift in z. The same shift is applied to planes 1 and 3 and to planes 4 and 6. The middle planes 2 and 5 are still fixed.

- **Iteration 4 - Downstream planes X and Y shift:**
  From this point on only the plane 2 remains fixed and the three planes of the downstream arm are aligned all together with respect to the upstream planes, using the triplet-triplet matching at the DUT plane. The distribution of residuals between the upstream triplet and downstream triplets extrapolated to the DUT plane is used. The shift of the mean position of the residual distribution is applied to all three downstream planes to correct their x and y position.

- **Iteration 5 - Downstream planes X and Y rotation:**
  Then the rotations of the downstream planes with respect to the upstream ones are corrected, using the triplet-triplet matching on the DUT plane. The residuals along x versus the y position and residuals along y versus x position are fitted with a linear function and the slope is used as correction factor for all three downstream planes.

- **Iteration 6: Downstream arm shift in Z**
  Similarly to iteration 3 the residuals are compared to triplet slopes to correct the Z shifts relying on slightly inclined tracks. In this step however the triplet-triplet matching residuals are used and the same correction is applied to all three downstream planes.

After this no new alignment steps are introduced, but extra iterations can be performed, each time repeating all the steps, to reach a satisfactory alignment. The alignment shifts along x, y and z axes and the rotations around the x and y axes applied to each of the six telescope planes in 20 alignment iterations are shown in Figure Figures F.2 and F.3, respectively. These plots obtained during the telescope alignment procedure using the data from the HPK sensor with rectangular pixels show that the alignment constants remain stable after eight iterations. Therefore, eight iterations were used for the telescope alignment of the HPK with rectangular pixels and CNM data. The data from the HPK with square pixels and the two FBK sensors needed a more iterations because of a higher multiple scattering.

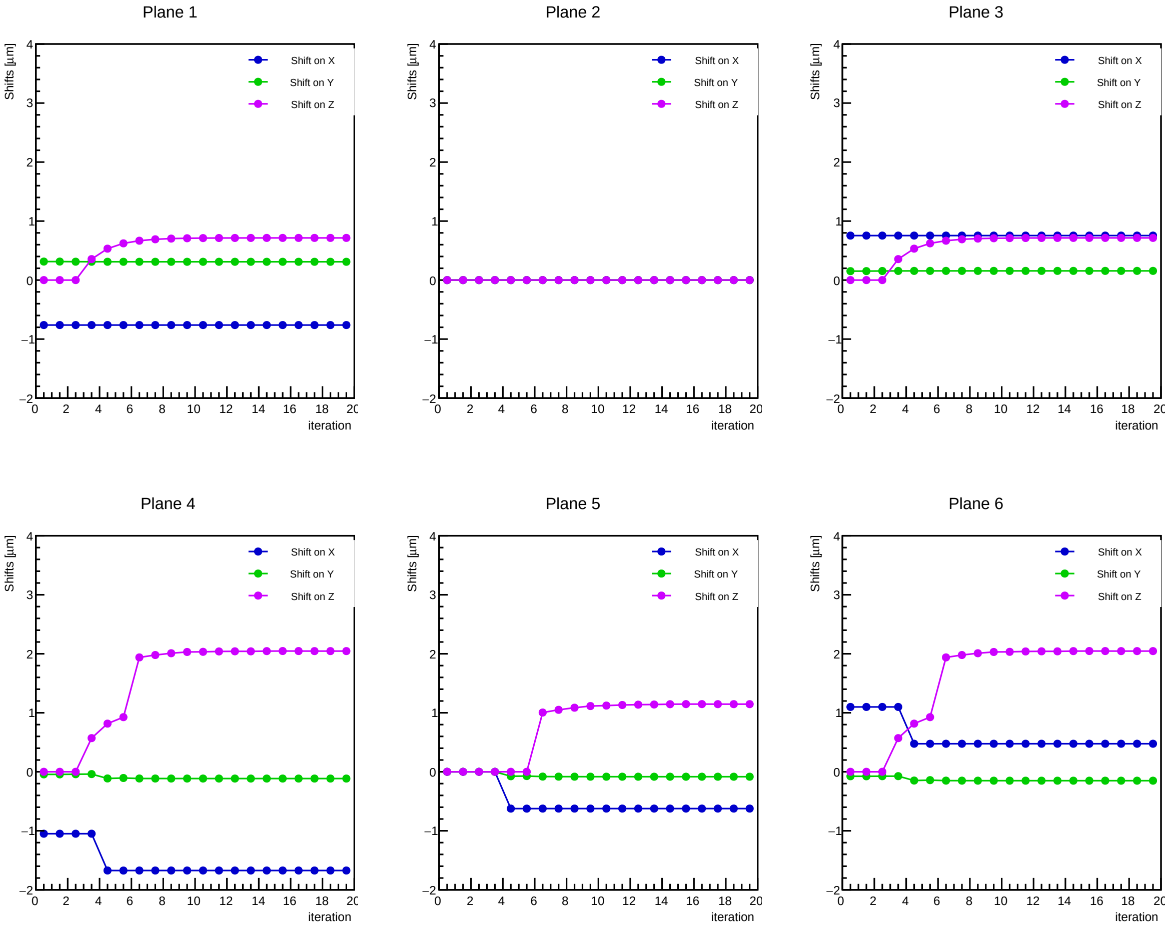


Figure F.2: Shifts along the x, y and z axis applied to each of the six telescope planes in each alignment iteration.

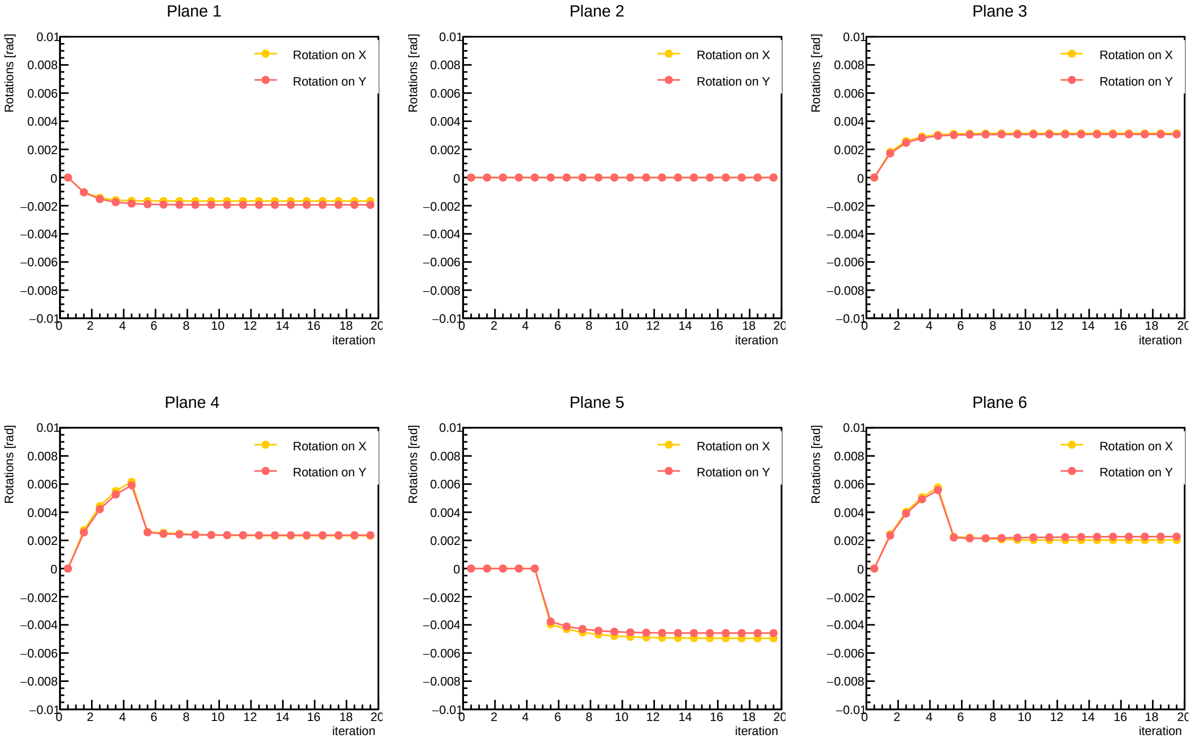


Figure F.3: Rotations around the x and y axis applied to each of the six telescope planes in each alignment iteration.

List of acronyms

ACB	analogue chip bottom
AD	Antiproton Decelerator
ADC	analogue-to-digital converter
AFE	analogue front-end
ALICE	A Large Ion Collider Experiment
APD	avalanche photodiode
ATCA	advanced telecom computer architecture
ATLAS	A Toroidal LHC ApparatuS
AWG	american wire gauge
BER	bit error rate
BRIL	Beam Radiation Instrumentation and Luminosity
BSM	beyond the standard model
BX	bunch crossing
CBA	central buffer architecture
CDR	clock data recovery
CERN	Conseil Européen pour la Recherche Nucléaire
CLEAR	CERN Linear Electron Accelerator for Research
CML	current mode logic
CMOS	complementary metal–oxide–semiconductor
CMS	Compact Muon Solenoid
CNM	Centro Nacional de Microelectrónica
COG	center-of-gravity

CSA	charge sensitive amplifier
CSC	cathode strip chamber
DA	differential amplifier
DAC	digital-to-analogue converter
DAQ	data acquisition
DATURA	DESY Advanced Telescope Using Readout Acceleration
DBA	distributed buffer architecture
DCB	digital chip bottom
DESY	Deutsches Elektronen-Synchrotron
DIFF	differential
DP	display port
DT	drift tube
DTC	data, trigger and control
DUT	device under test
ECAL	Electromagnetic Calorimeter
ELENA	Extra Low ENergy Antiproton
EMI	electromagnetic interference
ENC	equivalent noise charge
FBK	Fondazione Bruno Kessler
FE	front-end
FFC	flexible flat cable
FMC	FPGA mezzanine card
FPGA	field programmable gate array
GEM	gas electron multiplier
HCAL	Hadronic Calorimeter
HDI	high density interconnect
HEP	high energy physics
HGCAL	High Granularity Calorimeter
HL-LHC	High Luminosity LHC
HLT	High Level trigger

HPD	hybrid photodetector
HPK	Hamamatsu Photonics K.K.
IP	intellectual property
IP	interaction point
ISI	inter-symbol interference
IT	Inner Tracker
L1	Level-1 trigger
LCC	leakage current compensation
LDO	low-dropout
LEIR	Low Energy Ion Ring
LEP	Large Electron-Positron collider
LFSR	linear-feedback shift register
LGAD	low gain avalanche detector
LHC	Large Hadron Collider
LHCb	LHC beauty
LIN	linear
LIU	LHC Injectors Upgrade
lpGBT	low power gigabit transceiver
LS1	first long shutdown
LS2	second long shutdown
LS3	third long shutdown
LSB	least significant bit
MAPS	monolithic active pixel sensor
MIMOSA26	minimum ionizing monolithic sensor array
MIP	minimum ionizing particle
MPV	most probable value
MTD	MIP Timing Detector
ndf	number of degrees of freedom
NRZ	non-return-to-zero
NTC	negative temperature coefficient

OT	Outer Tracker
PA	preamplifier
PCB	printed circuit board
PLL	phase locked loop
PMT	photo multiplier tube
POR	power on reset
PRBS	pseudo random bit sequence
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
PU	pileup
QFT	quantum field theory
QGP	quark-gluon plasma
R&D	research and development
REF	reference module
RF	radio frequency
RMS	root mean square
RPC	resistive plate chamber
SCC	single chip card
ShLDO	Shunt low-dropout
SiPM	silicon photomultiplier
SM	standard model
SNR	signal-to-noise ratio
SP	serial powering
SPS	Super Proton Synchrotron
SYNC	synchronous
TBPX	Tracker Barrel Pixel detector
TEC	Tracker Endcap
TEPX	Tracker Endcap Pixel detector
TFPX	Tracker Forward Pixel detector
TIA	transimpedance amplifier

TIB	Tracker Inner Barrel
TID	Tracker Inner Disks
TID	total ionizing dose
TLU	Trigger Logic Unit
TOB	Tracker Outer Barrel
TOT	time-over-threshold
TWP	twisted pair
VCO	voltage controlled oscillator
VNA	vector network analyzer
XT	cross-talk

Bibliography

[1] S. Weinberg, 'The making of the Standard Model', Eur. Phys. J. C 34 (2004), 5, DOI: `10.1140/epjc/s2004-01761-1`.

[2] G. Hooft, 'The making of the standard model', Nature 448 (2007), 271, DOI: `10.1038/nature06074`.

[3] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013.

[4] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, 1995.

[5] Standard model classification, online picture.

[6] R. P. Feynman, 'Space-Time Approach to Quantum Electrodynamics', Phys. Rev. 76 (1949), 769, DOI: `10.1103/PhysRev.76.769`.

[7] F. Englert and R. Brout, 'Broken Symmetry and the Mass of Gauge Vector Mesons', Physical Review Letters 13 (1964), 321, DOI: `10.1103/PhysRevLett.13.321`.

[8] P. W. Higgs, 'Broken Symmetries and the Masses of Gauge Bosons', Physical Review Letters 13 (1964), 508, DOI: `10.1103/PhysRevLett.13.508`.

[9] J. Ellis, 'Limits of the Standard Model', 1st PSI Summer School on Condensed Matter Research: Magnetism, hep-ph/0211168, CERN-TH-2002-320 (2002).

[10] M. C. Gonzalez-Garcia and Y. Nir, 'Neutrino masses and mixing: evidence and implications', Rev. Mod. Phys. 75 (2003), 345, DOI: `10.1103/RevModPhys.75.345`.

[11] J. Ellis, 'Outstanding questions: physics beyond the Standard Model', Phil. Trans. R. Soc. A 370 (2012), 818, DOI: `10.1098/rsta.2011.0452`.

[12] CERN website: `https://home.cern`.

[13] F. J. Hasert et al., 'Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment', Physics Letters B 46 (1973), 138, DOI: `10.1016/0370-2693(73)90499-1`.

[14] R. Cashmore, L. Maiani and J.-P. Revol, *Prestigious Discoveries at CERN, 1973 Neutral Currents 1983 W & Z Bosons*, Springer, Berlin, Heidelberg, 2003, DOI: `10.1007/978-3-662-12779-7`.

[15] UA1 Collaboration, 'Experimental observation of isolated large transverse energy electrons with associated missing energy at $s = 540\,\text{GeV}$', Physics Letters B 122 (1983), 103, DOI: `10.1016/0370-2693(83)91177-2`.

[16] UA1 Collaboration, 'Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider', Physics Letters B 126 (1983), 398, DOI: `10.1016/0370-2693(83)90188-0`.

[17] UA2 Collaboration, 'Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\bar{p}p$ collider', Physics Letters B 122 (1983), 476, DOI: `10.1016/0370-2693(83)91605-2`.

[18] UA2 Collaboration, 'Evidence for Z^0→ e$^+$e$^-$ at the CERN $\bar{p}p$ collider', Physics Letters B 129 (1983), 130, DOI: `10.1016/0370-2693(83)90744-X`.

[19] G. Baur et al., 'Production of antihydrogen', Physics Letters B 368 (1996), 251, DOI: `10.1016/0370-2693(96)00005-6`.

[20] ALPHA Collaboration, 'Confinement of antihydrogen for 1,000 seconds', Nature 7 (2011), 558, DOI: `10.1038/nphys2025`.

[21] ATLAS Collaboration, 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC', Physics Letters B 716 (2012), 1, DOI: `10.1016/j.physletb.2012.08.020`.

[22] CMS Collaboration, 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC', Physics Letters B 716 (2012), 30, DOI: `10.1016/j.physletb.2012.08.021`.

[23] T. Berners-Lee et al., 'The World-Wide Web', Communications of the ACM, Vol.37, No.8 (1994).

[24] D. M. Sendall, 'The World-Wide Web past present and future, and its application to medicine', presented at the International Symposium on Hadrontherapy, CERN-OPEN-97-005 (1997).

[25] E. Mobs, 'The CERN accelerator complex', CERN-GRAPHICS-2019-002 (2019).

[26] L. Evans and P. Bryant, 'LHC Machine', JINST 3 (2008), S08001, DOI: `10.1088/1748-0221/3/08/S08001`.

[27] S. Myers, 'The LEP Collider, from design to approval and commissioning', John Adams' memorial lecture (1991), DOI: `10.5170/CERN-1991-008`.

[28] HiLumi in a nutshell: `https://project-hl-lhc-industry.web.cern.ch/content/hilumi-nutshell`.

[29] M. Schaumann et al., 'First Xenon-Xenon Collisions in the LHC', *9th International Particle Accelerator Conference*, 2018, DOI: `10.18429/JACoW-IPAC2018-MOPMF039`.

[30] ALICE Collaboration, 'The ALICE experiment at the CERN LHC', JINST 3 (2008), S08002, DOI: `10.1088/1748-0221/3/08/s08002`.

[31] ATLAS Collaboration, 'The ATLAS Experiment at the CERN Large Hadron Collider', JINST 3 (2008), S08003, DOI: `10.1088/1748-0221/3/08/s08003`.

[32] CMS Collaboration, 'The CMS experiment at the CERN LHC', JINST 3 (2008), S08004, DOI: `10.1088/1748-0221/3/08/S08004`.

[33] LHCb Collaboration, 'The LHCb Detector at the LHC', JINST 3 (2008), S08005, DOI: `10.1088/1748-0221/3/08/S08005`.

[34] G. Brumfiel, 'Eight-month delay for LHC, Broken magnets put particle collider in limbo', Nature 455 (2008), 1015, DOI: `10.1038/4551015a`.

[35] CERN releases analysis of LHC incident, `https://home.cern/news/press-release/cern/cern-releases-analysis-lhc-incident`.

[36] The Large Hadron Collider timeline: `https://timeline.web.cern.ch/timeline-header/93`.

[37] W. Herr and B. Muratori, 'Concept of luminosity', CERN Accelerator School: Intermediate Accelerator Physics (2006), DOI: `10.5170/CERN-2006-002.361`.

[38] CMS Luminosity public results: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[39] G. Apollinari et al., 'High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report', CERN-2015-005 (2015), DOI: `10.5170/CERN-2015-005`.

[40] O. Brüning and L. Rossi, *The High Luminosity Large Hadron Collider, the new machine for illuminating the mysteries of Universe*, World Scientific, 2015, DOI: `10.1142/9581`, eprint: `https://www.worldscientific.com/doi/pdf/10.1142/9581`.

[41] HL-LHC website: `https://hilumilhc.web.cern.ch/content/hl-lhc-project`.

[42] H. Damerau et al., 'LHC Injectors Upgrade, Technical Design Report, Volume I: protons', CERN-ACC-2014-0337 (2014).

[43] J. Coupard et al., 'LHC Injectors Upgrade, Technical Design Report, Volume II: ions', CERN-ACC-2016-0041 (2016).

[44] A. Sopczak, 'Precision Measurements in the Higgs Sector at ATLAS and CMS' (2020), arXiv: `2001.05927`.

[45] CMS Collaboration, 'Measurement of $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel in $41.5\,\mathrm{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$', CMS-PAS-HIG-18-030 (2019).

[46] CMS Collaboration, 'Technical Proposal for the Phase-2 Upgrade of the CMS Detector', CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02 (2015).

[47] CMS website: `https://cms.cern/collaboration/people-statistics`.

[48] CMS website: `https://cms.cern/news/cms-detector-design`.

[49] CMS Collaboration, 'Interactive Slice of the CMS detector', CMS-OUTREACH-2016-027 (2016).

[50] CMS Collaboration, 'The CMS magnet project: Technical Design Report', CERN-LHCC-97-010, CMS-TDR-1 (1997).

[51] CMS Collaboration, 'The CMS tracker system project: Technical Design Report', CERN-LHCC-98-006, CMS-TDR-5 (1997).

[52] CMS Collaboration, 'The CMS tracker: addendum to the Technical Design Report', CERN-LHCC-2000-016, CMS-TDR-5-add-1 (2000).

[53] CMS Collaboration, 'CMS Technical Design Report for the Pixel Detector Upgrade', CERN-LHCC-2012-016, CMS-TDR-11 (2012).

[54] The Tracker Group of the CMS Collaboration, 'The CMS Phase-1 Pixel Detector Upgrade', JINST 16 (2021), P02027, DOI: `10.1088/1748-0221/16/02/P02027`.

[55] CMS Collaboration, 'The CMS electromagnetic calorimeter project: Technical Design Report', CERN-LHCC-97-033 (1997).

[56] CMS Collaboration, 'The CMS hadron calorimeter project: Technical Design Report', CERN-LHCC-97-031 (1997).

[57] CMS Collaboration, 'The CMS muon project: Technical Design Report', CERN-LHCC-97-032 (1997).

[58] C. Lippmann, 'Particle identification', NIM A 666 (2012), 148, DOI: `10.1016/j.nima.2011.03.009`.

[59] CMS Collaboration, 'The CMS trigger system', JINST 12 (2017), P01020, DOI: `10.1088/1748-0221/12/01/P01020`.

[60] CMS Collaboration, 'CMS TriDAS project: Technical Design Report, Volume 1, The Trigger Systems', CERN-LHCC-2000-038, CMS-TDR-6-1 (2000).

[61] CMS Collaboration, 'CMS TriDAS project: Technical Design Report, Volume 2, Data Acquisition and High-Level Trigger', CERN-LHCC-2002-026, CMS-TDR-6 (2002).

[62] CMS Collaboration, 'The Phase-2 Upgrade of the CMS Level-1 Trigger', CERN-LHCC-2020-004, CMS-TDR-021 (2020).

[63] A. Ferrari et al., 'FLUKA: A Multi-Particle Transport Code' (2005), DOI: `10.2172/877507`.

[64] T. Böhlen et al., 'The FLUKA Code: Developments and Challenges for High Energy and Medical Applications', Nuclear Data Sheets 120 (2014), 211, DOI: `10.1016/j.nds.2014.07.049`.

[65] CMS Collaboration, ‘The Phase-2 Upgrade of the CMS Tracker’, CERN-LHCC-2017-009, CMS-TDR-014 (2017).

[66] CMS Collaboration, ‘Pileup mitigation at CMS in 13 TeV data’, JINST 15 (2020), P09018, DOI: `10.1088/1748-0221/15/09/P09018`.

[67] M. Backhaus, ‘The Upgrade of the CMS Inner Tracker for HL-LHC’, JINST 14 (2019), C10017, DOI: `10.1088/1748-0221/14/10/c10017`.

[68] L. Rossi et al., *Pixel detectors: from fundamentals to applications*, Particle Acceleration and Detection, Berlin: Springer, 2006, DOI: `10.1007/3-540-28333-1`.

[69] M. Garcia-Sciveres and N. Wermes, ‘A review of advances in pixel detectors for experiments with high rate and radiation’, Rep. Prog. Phys. 81 (2018), 066101, DOI: `10.1088/1361-6633/aab064`.

[70] L. Tlustos, ‘Performance and limitations of high granularity single photon processing X-ray imaging detectors’, PhD thesis, Vienna Technical University, 2005.

[71] W. Snoeys, ‘CMOS monolithic active pixel sensors for high energy physics’, NIM A 765 (2014), 167, DOI: `10.1016/j.nima.2014.07.017`.

[72] W. Snoeys, ‘Monolithic CMOS sensors for high energy physics’, NIM A 924 (2019), 51, DOI: `10.1016/j.nima.2018.06.034`.

[73] CMS Collaboration, ‘The Phase-2 Upgrade of the CMS Beam Radiation, Instrumentation, and Luminosity Detectors: Conceptual Design’, CMS-NOTE-2019-008, CERN-CMS-NOTE-2019-008 (2020).

[74] D. Koukola, ‘Serial Powering for the Tracker Phase-2 Upgrade’, PoS (Vertex2019), 039, DOI: `10.22323/1.373.0039`.

[75] V. Perovic, ‘Serial powering in four-chip prototype RD53A modules for Phase 2 upgrade of the CMS pixel detector’, NIM A 978 (2020), 164436, DOI: `https://doi.org/10.1016/j.nima.2020.164436`.

[76] S. Orfanelli, ‘The Phase 2 Upgrade of the CMS Inner Tracker’, NIM A 980 (2020), 164396, DOI: `10.1016/j.nima.2020.164396`.

[77] M. Karagounis et al., ‘An integrated Shunt-LDO regulator for serial powered systems’, Proceedings of ESSCIRC (2009), 276, DOI: `10.1109/ESSCIRC.2009.5325974`.

[78] LpGBT online manual: `https://lpgbt.web.cern.ch/lpgbt/`.

[79] C. Soós et al., ‘Versatile Link PLUS transceiver development’, JINST 12 (2017), C03068, DOI: `10.1088/1748-0221/12/03/C03068`.

[80] F. Faccio et al., ‘The bPOL12V DCDC converter for HL-LHC trackers: towards production readiness’, PoS (TWEPP2019), 070, DOI: `10.22323/1.370.0070`.

[81] G. Ripamonti et al., ‘2.5V step-down DCDCs: a radiation-hard solution for power conversion’, PoS (TWEPP2019), 071, DOI: `10.22323/1.370.0071`.

[82] Xilinx website, Aurora 64b/66b overview, `https://www.xilinx.com/products/intellectual-property/aurora64b66b.html`.

[83] E. S. Hazen et al., 'The APOLLO ATCA Platform', PoS (TWEPP2019), 120, DOI: `10.22323/1.370.0120`.

[84] P.-R. Luukka, 'CMS Inner Tracker Upgrade', PoS (Vertex2019), 008, DOI: `10.22323/1.373.0008`.

[85] Taiwan Semiconductor Manufacturing Company (TSMC) website: `https://www.tsmc.com/`.

[86] RD53 Collaboration, J. Christiansen and M. Garcia-Sciveres, 'RD Collaboration Proposal: Development of pixel readout integrated circuits for extreme rate and radiation', CERN-LHCC-2013-008, LHCC-P-006 (2013).

[87] RD53 Collaboration, M. Garcia-Sciveres, 'The RD53A Integrated Circuit', CERN-RD53-PUB-17-001 (2017).

[88] RD53 Collaboration, M. Garcia-Sciveres, 'RD53A Integrated Circuit Specifications', CERN-RD53-PUB-15-001 (2015).

[89] RD53 Collaboration, M. Garcia-Sciveres, F. Loddo and J. Christiansen, 'RD53B Manual', CERN-RD53-PUB-19-002 (2019).

[90] K. Moustakas et al., 'A Clock and Data Recovery Circuit for the ATLAS/CMS HL-LHC Pixel Front End Chip in 65 nm CMOS Technology', PoS (TWEPP2019), 046, DOI: `10.22323/1.370.0046`.

[91] T. Wang et al., 'A high speed transmitter circuit for the ATLAS/CMS HL-LHC pixel readout chip', PoS (TWEPP2018), 098, DOI: `10.22323/1.343.0098`.

[92] Xilinx Aurora 64b/66b Protocol Specifications, `https://www.xilinx.com/support/documentation/ip_documentation/aurora_64b66b_protocol_spec_sp011.pdf`.

[93] A. Paternò et al., 'New development on digital architecture for efficient pixel readout ASIC at extreme hit rate for HEP detectors at HL-LHC', Proceedings of IEEE NSS/MIC/RTSD (2016), 1, DOI: `10.1109/NSSMIC.2016.8069855`.

[94] P. Grybos, *Front-end Electronics for Multichannel Semiconductor Detector Systems*, vol. 8, EuCARD Editorial Series on Accelerator Science and Technology EuCARD-BOO-2010-004, 2010.

[95] F. Krummenacher, 'Pixel detectors with local intelligence: an IC designer point of view', NIM A 305.3 (1991), 527, DOI: `10.1016/0168-9002(91)90152-G`.

[96] E. Monteil et al., 'A synchronous analog very front-end in 65 nm CMOS with local fast ToT encoding for pixel detectors at HL-LHC', JINST 12 (2017), C03066, DOI: `10.1088/1748-0221/12/03/C03066`.

[97] L. Gaioni et al., '65 nm CMOS analog front-end for pixel detectors at the HL-LHC', JINST 11 (2016), C02049, DOI: `10.1088/1748-0221/11/02/C02049`.

[98] YARR repository on CERN Gitlab: `https://gitlab.cern.ch/YARR`.

[99] Lawrence Berkeley National Laboratory webpage: `https://www.lbl.gov`.

[100] University of Bonn webpage: `https://www.uni-bonn.de/en`.

[101] M. Daas et al., 'BDAQ53, a versatile pixel detector readout and test system for the ATLAS and CMS HL-LHC upgrades', NIM A 986 (2021), 164721, DOI: `10.1016/j.nima.2020.164721`.

[102] BDAQ53 repository on CERN Gitlab: `https://gitlab.cern.ch/silab/bdaq53`.

[103] Xilinx Kintex-7 FPGA KC705 Evaluation Kit: `https://www.xilinx.com/products/boards-and-kits/ek-k7-kc705-g.html`.

[104] J.-C. Beyer, 'Optimisation of pixel modules for the ATLAS inner tracker at the high-luminosity LHC', PhD thesis, Ludwig-Maximilians-Universität München, 2019.

[105] Semiconductor Laboratory of the Max Planck Society webpage: `https://www.mpg.de/8735784/hll`.

[106] G. Giakoustidis, 'Charge Calibration and Crosstalk Measurements with RD53A Assemblies', Master thesis, University of Bonn, 2019.

[107] E. Monteil, *Synchronous Front-end Testing specifications*, `https://twiki.cern.ch/twiki/pub/RD53/RD53ATesting/SYNC_FE_TESTING_SPECIFICATIONS_updated.pdf`, 2018.

[108] L. Gaioni, *Linear Analog Front-end Guidelines and Recommendations*, `https://twiki.cern.ch/twiki/pub/RD53/RD53ATesting/LIN_AFE_guidelines.pdf`, 2018.

[109] M. Standke et al., *Guide to the RD53A Differential Front End*, `https://twiki.cern.ch/twiki/pub/RD53/RD53ATesting/Diff_userguide.pdf`, 2018.

[110] L. Gaioni et al., 'Design and test of current-mode DACs for threshold tuning of front-end channels for the High Luminosity LHC', PoS (TWEPP2019), 002, DOI: `10.22323/1.370.0002`.

[111] V. Cindro, 'World Irradiation Facilities for Silicon Detectors', PoS (Vertex2014), 026, DOI: `10.22323/1.227.0026`.

[112] C. de La Taille, 'Automated system for equivalent noise charge measurements from 10 ns to 10 µs', Nuclear Physics B - Proceedings Supplements 32 (1993), 449, DOI: `10.1016/0920-5632(93)90059-F`.

[113] V. Radeka, 'Signal, Noise and Resolution in Position-Sensitive Detectors', IEEE Transactions on Nuclear Science 21.1 (1974), 51, DOI: `10.1109/TNS.1974.4327444`.

[114] D. Lange, 'The CMS Reconstruction Software', Journal of Physics: Conference Series 331 (2011), 032020, DOI: `10.1088/1742-6596/331/3/032020`.

[115] RD53 Collaboration, L. Gaioni and F. Loddo, 'CMS analog front-end: simulations and measurements', CERN-RD53-PUB-20-002 (2020).

[116] W. Shockley, 'Currents to Conductors Induced by a Moving Point Charge', Journal of Applied Physics 9 (1938), 635, DOI: `10.1063/1.1710367`.

[117] S. Ramo, 'Currents Induced by Electron Motion', Proceedings of the IRE 27.9 (1939), 584, DOI: `10.1109/JRPROC.1939.228757`.

[118] W. Riegler, 'An application of extensions of the Ramo–Shockley theorem to signals in silicon sensors', NIM A 940 (2019), 453, DOI: `10.1016/j.nima.2019.06.056`.

[119] S. Spannagel, 'Test Beam Measurements for the Upgrade of the CMS Pixel Detector and Measurement of the Top Quark Mass from Differential Cross Sections', DESY-THESIS-2016-010, CMS-TS-2016-010, CERN-THESIS-2016-059, PhD thesis, University of Hamburg, 2016, DOI: `10.3204/DESY-THESIS-2016-010`.

[120] J. Schwandt, 'CMS Pixel detector development for the HL-LHC', NIM A 924 (2019), 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors, 59, DOI: `10.1016/j.nima.2018.08.121`.

[121] G. Steinbrück, 'Development of planar pixel sensors for the CMS Inner Tracker at the High-Luminosity LHC', NIM A 978 (2020), 164438, DOI: `10.1016/j.nima.2020.164438`.

[122] M. Meschini et al., 'Radiation resistant innovative 3D pixel sensors for the CMS upgrade at the High Luminosity LHC', NIM A 978 (2020), 164429, DOI: `10.1016/j.nima.2020.164429`.

[123] D. Forshaw, 'Development of Radiation Hard Planar Silicon Tracking Detectors for the ATLAS Experiment at the HL-LHC', PhD thesis, University of Liverpool, 2014.

[124] Hamamatsu Photonics K.K. website., `https://www.hamamatsu.com/`.

[125] Fondazione Bruno Kessler website., `https://www.fbk.eu/`.

[126] S. Parker, C. Kenney and J. Segal, '3D — A proposed new architecture for solid-state radiation detectors', NIM A 395.3 (1997), Proceedings of the Third International Workshop on Semiconductor Pixel Detectors for Particles and X-rays, 328, DOI: `10.1016/S0168-9002(97)00694-3`.

[127] CNM website., `http://www.imb-cnm.csic.es/en?lang=en`.

[128] J. Duarte-Campderros et al., 'Results on proton-irradiated 3D pixel sensors interconnected to RD53A readout ASIC', NIM A 944 (2019), 162625, DOI: `10.1016/j.nima.2019.162625`.

[129] A. G. Alonso et al., 'Test beam characterization of irradiated 3D pixel sensors', Journal of Instrumentation 15.03 (2020), C03017, DOI: `10.1088/1748-0221/15/03/c03017`.

[130] G. Steinbrueck, 'The CMS Silicon Pixel detector for HL-LHC', *Proceedings of The 25th International workshop on vertex detectors — PoS(Vertex 2016)*, vol. 287, 2017, 016, DOI: `10.22323/1.287.0016`.

[131] Official DESY web page, `https://www.desy.de`.

[132] R. Diener et al., 'The DESY II test beam facility', NIM A 922 (2019), 265, DOI: `10.1016/j.nima.2018.11.133`.

[133] T. Behnke and al., 'Test beams at DESY', EUDET-Memo-2007-11 (2007).

[134] Official EUDET project web page, `https://www.eudet.org`.

[135] C. Hu-Guo et al., 'First reticule size MAPS with digital output and integrated zero suppression for the EUDET-JRA1 beam telescope', NIM A 623.1 (2010), 1st International Conference on Technology and Instrumentation in Particle Physics, 480, DOI: `10.1016/j.nima.2010.03.043`.

[136] J. Baudot et al., 'First test results Of MIMOSA-26, a fast CMOS sensor with integrated zero suppression and digitized output', *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2009, 1169, DOI: `10.1109/NSSMIC.2009.5402399`.

[137] H. Jansen et al., 'Performance of the EUDET-type beam telescopes', EPJ Techniques and Instrumentation 3.7 (2016), DOI: `10.1140/epjti/s40485-016-0033-2`.

[138] S. Spannagel, 'Status of the CMS Phase I pixel detector upgrade', NIM A 831 (2016), Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors, 71, DOI: `10.1016/j.nima.2016.03.028`.

[139] H. C. Kästli, 'Frontend electronics development for the CMS pixel detector upgrade', NIM A 731 (2013), PIXEL 2012, 88, DOI: `10.1016/j.nima.2013.05.056`.

[140] D. Cussans, 'Description of the JRA1 Trigger Logic Unit (TLU), v0.2c', EUDET-Memo-2009-4 (2009).

[141] E. D. Team, 'EUDAQ User Manual' (2016).

[142] R. Turchetta, 'Spatial resolution of silicon microstrip detectors', NIM A 335.1 (1993), 44, DOI: `10.1016/0168-9002(93)90255-G`.

[143] T. Ullrich and Z. Xu, 'Treatment of Errors in Efficiency Calculations' (2007), arXiv: `physics/0701199v1`.

[144] E. Bogatin, *Signal and power integrity, simplified; 3rd ed.* Pearson modern semiconductor design series, Boston, MA: Pearson, 2018.

[145] 33-pin Molex connector, online datasheet: `https://www.molex.com/webdocs/datasheets/pdf/en-us/5025983393_FFC_FPC_CONNECTORS.pdf`.

[146] Samtec, MMCX connector.

[147] Commercial flexible flat cable data sheet: `https://www.molex.com/webdocs/datasheets/pdf/en-us/0150151033_CABLE.pdf`.

[148] Keysight N5225A PNA Microwave Network Analyser, online information: `https://www.keysight.com/fr/en/product/N5225A/pna-microwave-network-analyzer-50-ghz.html#`.

[149] ADVANTEST Corporation, *PRBS (Pseudo Random Binary Sequence)*, Hideo Okawara's Mixed Signal Lecture Series, 2013.

[150] ON semicondutor, *Understanding Data Eye Diagram Methodology for Analyzing High Speed Digital Signals*, Application note AND9075/D, 2015.

[151] M. Vretenar et al., 'Linac4 design report', CERN Yellow Reports: Monographs 6 (2020), DOI: `10.23731/CYRM-2020-006`.

[152] NA61/SHINE Collaboration, 'NA61/SHINE facility at the CERN SPS: beams and detector system', JINST 9 (2014), P06005, DOI: `10.1088/1748-0221/9/06/p06005`.

[153] D. Tefelski, 'Upgrade of the NA61/SHINE Detector', *The XVIII International Conference on Strangeness in Quark Matter (SQM 2019)*, ed. by D. Elia et al., Cham: Springer International Publishing, 2020, 473.

[154] NA62 Collaboration, 'The beam and detector of the NA62 experiment at CERN', JINST 12 (2017), P05025, DOI: `10.1088/1748-0221/12/05/P05025`.

[155] P. Abbon et al., 'The COMPASS experiment at CERN', NIM A 577.3 (2007), 455, DOI: `10.1016/j.nima.2007.03.026`.

[156] AWAKE Collaboration, 'AWAKE Design Report: A Proton-Driven Plasma Wakefield Acceleration Experiment at CERN', CERN-SPSC-2013-013, SPSC-TDR-003 (2013).

[157] S. A. Baird et al., 'Design study of the antiproton decelerator: AD', CERN-PS-96-043-AR, CERN/PS 96-43 (AR) (1996).

[158] W. Bartmann et al., 'Extra Low Energy Antiproton ring ELENA: from the conception to the implementation phase', 5th International Particle Accelerator Conference (2014), DOI: `10.18429/JACoW-IPAC2014-TUOAA03`.

[159] CMS Collaboration, 'The Phase-2 Upgrade of the CMS Endcap Calorimeter', CERN-LHCC-2017-023, CMS-TDR-019 (2017).

[160] M. Valentan, 'The CMS high granularity calorimeter for the high luminosity LHC', NIM A 936 (2019), Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors, 102, DOI: `https://doi.org/10.1016/j.nima.2018.10.131`.

[161] CMS Collaboration, 'The Phase-2 Upgrade of the CMS Barrel Calorimeters', CERN-LHCC-2017-011, CMS-TDR-015 (2017).

[162] CMS Collaboration, 'The Phase-2 Upgrade of the CMS Muon Detectors', CERN-LHCC-2017-012, CMS-TDR-016 (2017).

[163] CMS Collaboration, 'A MIP Timing Detector for the CMS Phase-2 Upgrade', CERN-LHCC-2019-003, CMS-TDR-020 (2019).

[164] G. F. Knoll, *Radiation detection and measurement; 4th ed.* New York, NY: Wiley, 2010.

[165] F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*, vol. 275, Springer Tracts in Modern Physics (STMP), Springer, 2017, DOI: `10.1007/978-3-319-64436-3`.

[166] 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 58.1 (1991), 6, DOI: `10.1016/0168-583X(91)95671-Y`.

[167] Particle Data Group, 'Review of Particle Physics', Progress of Theoretical and Experimental Physics (PTEP) 2020.8 (2020), 083C01, DOI: `10.1093/ptep/ptaa104`, eprint: `https://academic.oup.com/ptep/article-pdf/2020/8/083C01/33653179/ptaa104.pdf`.

[168] W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments, A How-to Approach*, Springer, Berlin, Heidelberg, 1994, DOI: `10.1007/978-3-642-57920-2`.

[169] L. D. Landau, 'On the energy loss of fast particles by ionization', J. Phys. 8.4 (1944), 201.

[170] G. Hall, 'Ionisation energy losses of highly relativistic charged particles in thin silicon layers', Nuclear Instruments and Methods in Physics Research 220.2 (1984), 356, DOI: `10.1016/0167-5087(84)90296-5`.

[171] H. Bichsel, 'Straggling in thin silicon detectors', Rev. Mod. Phys. 60 (3 1988), 663, DOI: `10.1103/RevModPhys.60.663`.

[172] I. Zoi et al., *Beam Test Results of Thin n-in-p 3D and Planar Pixel Sensors for the High Luminosity LHC Tracker Upgrade at CMS*, tech. rep. CMS-CR-2017-325, 2017, DOI: `10.22323/1.314.0809`.

[173] U. Fano, 'Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions', Phys. Rev. 72 (1947), 26, DOI: `10.1103/PhysRev.72.26`.

UNIVERSITÉ DE STRASBOURG

IPHC
Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

CERN

CMS

École doctorale
Physique, chimie-physique
ED 182
Université de Strasbourg

Natalia EMRISKOVA

Développement et validation du système de trajectométrie interne de CMS pour LHC à haute luminosité

Résumé

Le détecteur à pixels de CMS, appelé le trajectomètre interne, va être entièrement remplacé pour LHC à haute luminosité. Le développement du système d'acquisition du nouveau détecteur repose sur plusieurs choix, chacun d'entre eux nécessitant une étude dédiée. La puce de lecture RD53A a été testée et ses trois électroniques analogiques frontales ont été comparées, ce qui a permis de choisir l'électronique Linéaire avec un circuit amélioré pour l'intégration dans la nouvelle puce de lecture de CMS. Deux méthodes de mesure de diaphonie dans les capteurs à pixels ont été proposées et vérifiées sur cinq capteurs dans le but de contribuer au choix entre les pixels carrés et les pixels rectangulaires. La diaphonie asymétrique de l'ordre de 10% a été mise en évidence dans les pixels rectangulaires et deux moyens de la réduire identifiés : la conception des pixels et la polarisation du préamplificateur. Finalement, différents prototypes de câbles électriques ont été évalués et comparés pour implémenter la liaison de données du détecteur assurant une intégrité des signaux électriques suffisante pour être correctement réceptionnés par la puce d'agrégation lpGBT. L'influence de préaccentuation implémentée dans RD53A sur la qualité du signal a aussi été étudiée.

Mots clés : détecteur à pixels au silicium, trajectomètre interne de CMS, amélioration, développement du système, puce de lecture RD53A, électronique frontale, électronique analogique, capteur à pixels au silicium, diaphonie, liaison de données, câbles électriques légers, préaccentuation.

Résumé en anglais

The CMS Inner Tracker pixel detector will be entirely replaced for the High Luminosity LHC. The development of the readout system for the new detector relies on several system choices each of which requires dedicated studies. The RD53A readout chip was tested and its three analogue front-ends were compared, which allowed to choose the Linear front-end with an improved design for integration into the new CMS readout chip. Two methods to measure cross-talk in pixel sensors were proposed and verified on five sensors to contribute to the choice between square and rectangular pixels. Asymmetric cross-talk of the order of 10% was put in evidence in rectangular pixels and two handles to reduce it were identified: the pixel design and the preamplifier bias. Finally, different prototypes of electrical links were evaluated and compared towards the implementation of the data link ensuring a sufficient signal integrity for a correct reception of the data by the lpGBT chip. The influence of pre-emphasis implemented in RD53A on the signal quality was also studied.

Keywords: silicon pixel detector, CMS Inner Tracker, upgrade, system development, RD53A readout chip, front-end electronics, analogue front-end, silicon pixel sensor, cross-talk, data link, low mass electrical links, pre-emphasis.