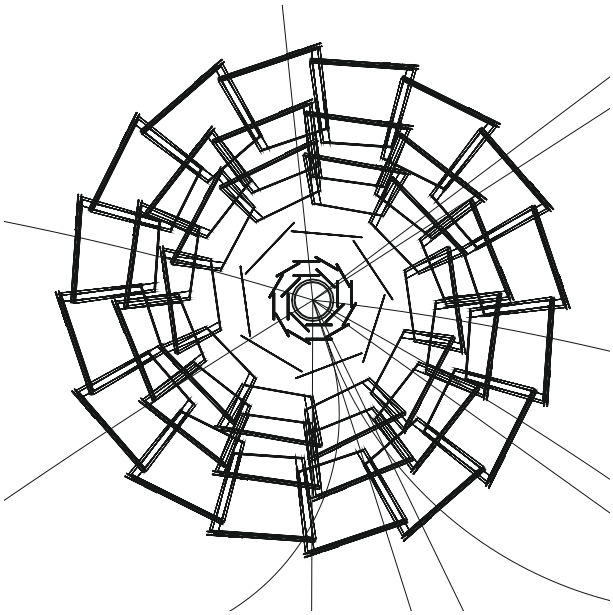

Università di Pisa

**Dipartimento di Fisica**

**Corso di Laurea Magistrale in Fisica**


Tesi di Laurea Magistrale in Fisica

# Development and performance of the track finder for the Belle II Vertex Detector




Candidato
**Valerio Bertacchi**

Relatore
**Prof. Francesco Forti**

Correlatore
**Prof. Eugenio Paoloni**


Anno Accademico 2016/2017

*E fu il calore d'un momento*
*Poi via di nuovo verso il vento*

# Contents

# Introduction

Belle II is the general purpose detector that will be operated by the second generation *B-Factory* SuperKEKB and it is currently at an advanced construction stage in KEK (Japan). SuperKEKB is a high luminosity asymmetric $e^+e^-$ collider (peak luminosity up to $8 \times 10^{35}$ cm$^{-2}$s$^{-1}$), with the center-of-mass energy tuned on the peak of the $\Upsilon(4S)$ resonance. It realizes an optimal environment to perform precision measurements in the heavy flavour sector and it will allow collecting an integrated luminosity of 50 ab$^{-1}$ of $\Upsilon(4S)$ events. The Belle II detector is designed to study a broad spectrum of heavy flavour physics, including the $\mathcal{CP}$ violation of $B$ and $D$ systems and the rare decay modes of $B$, $\tau$ and $D$. The improved precision that will be achieved, thanks to the increased statistic and to the high performance of the detector, may allow to observe physics Beyond the Standard Model at energy scales higher than those directly accessible. Several channels in the Belle II physics program already show tensions with Standard Model predictions that might be signatures of New Physics. In this thesis, the current heavy flavour physics status is described in chapter 1 together with a report on tensions with the Standard Model in the semitauonic sector. In chapter 2 instead, the SuperKEKB accelerator and the Belle II detector are described.

In order to perform the $B$ physics measurements with the required precision and to fully exploit the B-Factory environment, it is essential to reconstruct the complete $\Upsilon(4S)$ events including the momenta and the vertices of the charged particle tracks down to tens MeV/$c$ and the energy neutral particles. To fulfil this demanding task, Belle II adopts a high-performance silicon *Vertex Detector* (VXD), realized with two *Pixel Detector* layers (PXD) and four *Silicon Strip Detector* layers (SVD), and a *Drift Chamber* (CDC). The Belle II - Pisa group is involved in the production of the forward and backward regions SVD modules. This thesis work has been partially devoted performing electrical qualification test on the produced forward and backward modules, in the INFN Laboratories, while the main focus of the activity has been on the reconstruction software development, as described below. In chapter 3 after a description of the VXD and its operating principle, the assembly and test procedures performed during the thesis period are presented.

The Belle II tracking software is designed to perform the pattern recognition, vertexing and fitting procedures in the online (i.e. in the high level trigger of the experiment) and offline event reconstruction. One of the prominent features of the Belle II tracking software is its capability of efficiently reconstructing tracks down to $p_\perp \sim 50$ MeV/$c$ using a VXD standalone track finder (VXDTF), necessary for instance to reconstruct the low momentum pions from $D^*$ decays. In this momentum range the pattern recognition is particularly hard: only a few hits seen in VXD are available to reconstruct the tracks and the material effects, like energy losses and multiple scattering, significantly divert the trajectory from an ideal helix. To overcome this issue the VXDTF adopts the *Sector Map* concept, a lookup table called Sector Map, that contains the geometrical information of the patterns, is used to compare the measured track information with the expected ones. The pattern recognition begins by building segments (pairs of hits), then triplets (pairs of segments sharing the inner hit) and then four-hit track (pairs of triplets sharing

the inner segment). To reject the random combinations of hits, at each step geometrical and dynamical filters with the information provided by the Sector Map are applied. The Sector Map is realized collecting the information from a sample of simulated events, called *Training Sample*, which defines *a priori* the Sector Map that will be used in the actual data taking. After the reduction of the combinatorial burden performed by the Sector Map filters the tracks are identified and collected by a *cellular automaton*. Then, a best candidate selection is performed by a *Hopfield Neural Network* to remove the overlap between tracks which share hits.

During the period of this thesis, the Belle II tracking group has been involved in a redesign of VXDTF code, in order to improve its flexibility and maintainability. This redesign allows developing new strategies to improve the performance of the track finder, in term of efficiency, fake rate and time consumption. In chapter 4 the Belle II track finder is described in details as well as the features of the redesigned version. In addition, the comparison between the performances of the VXDTF and its redesigned version is discussed.

One of the critical points of the VXDTF is the definition of the Training Sample, that directly affects the Sector Map complexity and then the track finding efficiency and purity together with CPU time and memory footprint. In fact, if inside the Training Sample there are tracks which are very far from helical trajectories, and therefore not trackable, the Sector Map will allow the collection of patterns that will be discarded in fitting phase. Therefore these tracks increase the time consumption and the fake rate of the track finder, without a significant efficiency gain. One of the central points of this thesis is the development of the selection procedure of the Training Sample, with the definition of proper requirements on the simulated tracks in order identify and discard *a priori* the non trackable patterns, without discarding events of physical interest. One of the possible sources of this class of tracks are low the momentum multiple scattering effects which make the track depart from the ideal trajectory. This effect increases at low the momentum, coherently with the observed degradation of the performance of the track finder. A chance to identify these tracks comes from the analysis of the track parameters in their evolution along the trajectory: these parameters should be constant along the track, and a strong variation in a certain position inside the detector is a signature of an unusually large interaction. In chapter 5 the procedure of selection is presented together with the studies on track parameters needed for the definition of the cuts. Finally, the performance of the Sector Map trained with the selected sample are discussed with a comparison of track finder performances.

Part of the thesis works covers the application of the redesigned track finder on the real data, with the participation to the DESY Testbeam in February 2017, where a six-layer sector of VXD has been tested on an electron beam with an energy up to 6 GeV. The purpose of the testbeam was to test the SVD and PXD sensors with the complete readout and trigger chain and to test the tracking software after the substantial upgrades of the redesign. During the testbeam track finder performances studies has been performed together with the characterization of the SVD sensors in term of hit efficiency. These results on testbeam data and a description of the measurement strategy are presented in chapter 6.

In conclusion in my thesis work, I have covered some of the most relevant aspects of the silicon tracking detector operations, from the single module electrical tests, to the optimization of tracking software, to the test of the tracker performance on beam.

# Chapter 1

# The Standard Model and Heavy Flavour Physics

The Standard Model (SM) of elementary particles is the current theory that best describes the fundamental forces that rule the microscopic world. The SM places all the constituents of the matter and their interaction except gravity in a coherent theoretical framework. This theory is able to explain and predict all physics effects observed at a fundamental level, so it might seem that the microscopic world is well understood. However, at very different energy and spatial scales, astronomical and cosmological measurement reveals large effects non-explainable with Standard Model interactions. Some of these macroscopic effects could have a microscopical origin out of SM theory. Examples are the *baryon asymmetry* (asymmetry between baryonic matter and antibaryonic matter in the observable universe, not explainable with $\mathcal{CP}$ violation from SM source), or *dark matter* (anomalies in gravitational effects explainable with a not-electromagnetic matter). Besides, several SM measurements show some tensions with respect to theoretical expectation. These tensions are experimentally not significant enough to claim Beyond Standard Model (BSM) discoveries, but they might be hints to effects requiring further investigations. On the other hand the discovery of a new scalar resonance at $\sim$ 125 GeV [1] [2], identified as the Higgs Boson, experimentally confirmed the scheme described by the SM.

In this picture, experimental particle physics is trying to investigate the limits of the Standard Model looking for New Physics effects. This is pursued in two ways: the "energy frontier" and the "intensity frontier". In the first one, the purpose is to investigate higher energy scenarios to directly search for higher mass particles or new interactions. In the second one, precision measurements are performed to search for small effects showing deviations from the Standard Model scenario: these effects could be a portal to New Physics at higher energy scale, not directly accessible. The theoretical perspective, in both cases, is to reduce the Standard Model to an effective theory valid only below some energy scale.

In this approach of precision measurements, *B-Factories* have an important role for the intensity frontier approach [3]. These experiments, described in detail in chapter 2, allow achieving a large integrated luminosity of high purity events useful to study $B$, $D$ and $\tau$ physics. The first generation of B-Factories contributed to the experimental confirmation of the heavy flavour theory in the SM framework, while the next generation has the chance to reveal New Physics signatures.

## 1.1 Introduction to the Standard Model

The SM is a relativistic quantum field theory, in which the basic constituents are represented by spin-$\frac{1}{2}$ fields while interactions are mediated by spin-1 *gauge bosons*. All the fields (particle) in the SM acquire mass through the interaction with the spin-0 Higgs boson.

The strong force is mediated by gluons ($g$), massless bosons that exist in 8 color states, according to quantum chromodynamics (QCD), the theory of strong interactions. The strong charge which exists in three kinds, is called *color*, that so all colored particles interact strongly. Since gluons carry color charge they can interact with themselves.

The electromagnetic force is mediated by the photon ($\gamma$), a massless boson that exists in 2 states of polarization, and it interacts with all the particles with electric charge. This charge exists in one single kind and can assume positive or negative value. The photons do not carry electric charge.

The bosons mediating the weak force are $W^+$, $W^-$ and $Z^0$, three massive bosons, two electrically charged and one neutral. All the particles of the SM interact weakly, so a weak charge is assigned to all the particles. Electromagnetic and weak force arise from the same symmetry group of the theory and can be treated as a single force, called electroweak interaction.

In the SM matter is composed of fermions which are divided in two main categories: quarks and leptons. The quarks interact with the electroweak force and the strong force, instead leptons interact only with the electroweak force. Exist 6 kinds of quarks, classifiable in 3 generations of doublets. Each doublet contains an up-type quark with electric charge $\frac{2}{3}e$ and a down-type quark with electric charge $-\frac{1}{3}e$, where $e \simeq 1.602 \cdot 10^{-19}$ C is the absolute value of the charge of the electron. The kind of quark could be interpreted as a quantum number that each quark brings that identify the kind of quark. The doublets are:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix},$$

where the letters correspond to *u*p, *d*own, *c*harm, *s*trange, *t*op, *b*ottom. These quantum numbers are called *flavour*, and they are conserved in the strong and electromagnetic interaction, but not in the weak one. In this picture, 6 quantum numbers are assigned to each quark, with all values set to 0 except the one corresponding to its kind, that will be +1 for up quarks, -1 for down quarks (for instance, a charm quark has $U = D = S = T = B = 0$ and $C = 1$, instead a bottom quark has $U = D = C = S = T = 0$ and $B = -1$). These 3 generations have a completely specular picture of antiquarks (quarks are $\frac{1}{2}$-spin fermions and so they follow Dirac equation), with reversed flavour number, electric charge, and color. Isolated quarks are not observable because of color-confinement principle of QCD (all observable particles isolated must be colorless) a but bound state of one quark and one antiquark (meson) or a bound state of three quark (baryon) can be colorless and so observable isolated. The number of quarks is not conserved in the SM processes, but the difference $N_q - N_{\overline{q}}$ is constant, implying that the lightest baryon (the proton) must be stable in the SM and that the baryon number $N_B$ is a conserved quantity of SM.

The leptons are also classified in 3 generations of doublets:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}, \begin{pmatrix} \mu \\ \mu_u \end{pmatrix}, \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix},$$

where each doublet is composed by a charged and massive lepton (with electric charge $-e$) and a massless and neutral neutrino. Three more quantum numbers called *lepton family number* ($L_e$,

$L_\mu$, $L_\tau$) conserved in all interactions [1] are defined. Each lepton has the lepton family number of its generation equal to 1 and the numbers of the other generation equal to 0. Lepton number $L$ is defined as the sum of lepton family number. As quarks, leptons have corresponding antileptons (with reversed electric charge and lepton family number).

The Higgs Boson is a scalar, electrically neutral and massive. Through the *Higgs Mechanism* [5] it gives mass to the particles that couple to it. The result is a non-trivial hierarchy of mass between the particles of the SM: the massive bosons have masses in the scale of 100 GeV/$c^2$, the quarks cover from few MeV/$c^2$ to about 173 GeV/$c^2$ of quark $t$ and the massive leptons from about 0.5 MeV/$c^2$ to about 1.7 GeV/$c^2$ (in figure 1.1 precise values are shown ). Both in quarks and leptons the mass is bigger in higher generation, thus finally the three generations appear as three copy of the same structure that differ only in mass (and therefore in available decays and lifetime). The previous statement is exactly true for leptons and it is called *lepton universality*, instead for quarks generations there are differences in couplings that generate the phenomenology of heavy flavour physics, described in section 1.2. In figure 1.2 are shown the couplings between the particles of the SM previous described.

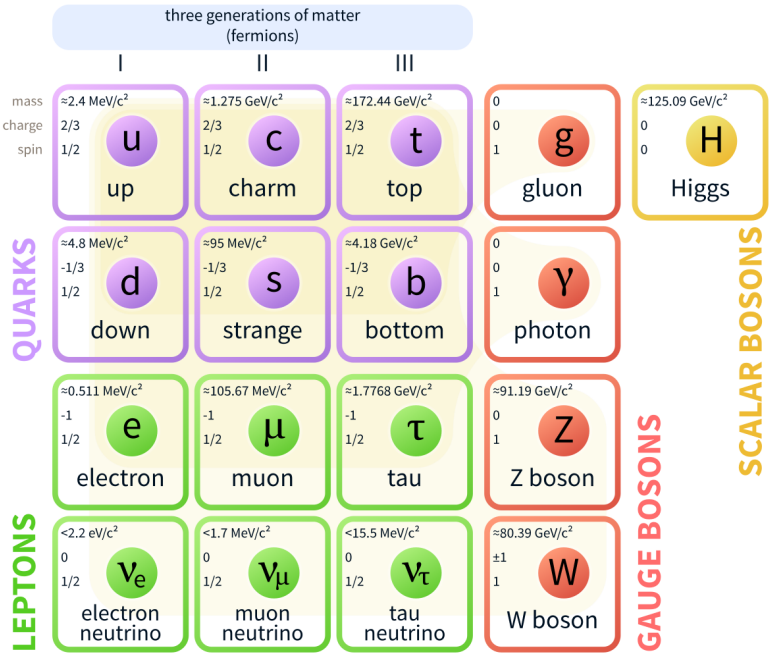



*Figure 1.1: Summary table of elmentary particles of Standard Model and their classification.*

### 1.1.1 Brief summary of the mathematical formulation of SM

The fields of the SM are the fermion fields $\psi_i$ (where $i$ is the kind of fermion), the gluons fields $G_a$ (where $a$ run from 1 to 8) the electroweak fields $W_1$, $W_2$, $W_3$ and $B$, the Higgs field $\varphi$. The SM is a gauge theory and its group of symmetry is:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y,$$

[1] the *neutrino oscillation* process violates the $L_{e,\mu,\tau}$ conservation and treats massive neutrinos, but is not included in the simplest version of SM. Anyway the neutrino oscillation phenomena can be included easily in an extended version of SM. Details about neutrino oscillation can be found in the review inside [4].

where $SU(3)_C$ is the gauge group of QCD and acts on $G_a$ (where the subscript $C$ is for "color"), $SU(2)_L \otimes U(1)_Y$ is the gauge group of electroweak theory, and $SU(2)_L$ acts on $W_{1,2,3}$ and $\varphi$, and $U(1)_Y$ acts on $B$ and $\varphi$ (where $L$ is for "left" and $Y$ is for "hypercharge").

Inside the lagrangian of the SM a kinetic term of free field ($C_{\mu\nu}C^{\mu\nu}$, where $C_{\mu\nu}$ is the gauge tensor associated with the field) is present for $B$, $W$ and $G$ and then the dynamical terms of interaction between fermions and boson fields. i.e. an electroweak term ($\mathcal{L}_{EW}$, where enter $B$ and $W_j$) and a chromodynamic term ($\mathcal{L}_{QCD}$, where enter $G_a$) are present

A trivial mass term in the form $C^\mu C_\mu$ (where $C$ is a generic field) is forbidden to keep the theory gauge invariant, but with the introduction of a complex scalar field $\varphi$, it is possible to "give mass to vector bosons" [5]. This *spontaneous symmetry breaking mechanism* gives to the Higgs field a non-vanishing vacuum expectation value $\langle\varphi^0\rangle = v$ (VEV) that represent a mass-scale of the model. In conclusion, the Higgs Mechanism mix the four electroweak field to obtain three massive field, $W^\pm$ charged field and the $Z^0$ field, and the photon massless field $A$ (with own symmetry group $U(1)_{EM}$).

The Higgs Mechanism can be extended to give to the fermions their masses adding a Yukawa term (a term that couple a scalar and a fermion field, like $\overline{\psi}\varphi\psi$) in the Lagrangian. The mass scales of the fermions (the coupling constant of these terms) are free parameters of the SM.

The SM does not have any of the simpler discrete symmetry: the charge conjugation $\mathcal{C}$ which inverts all the internal quantum numbers of a particle transforming particle into its antiparticle, the parity operator $\mathcal{P}$ which reverses the spatial coordinate and momenta of a particle ant the time reversal $\mathcal{T}$ which inverts the arrow of time are not conserved in the SM processes. In addition, the electroweak interaction violates also the combination $\mathcal{CP}$. Instead the SM the combination of the three symmetries $\mathcal{CPT}$ is a symmetry of the SM.

For a more exhaustive description see, for instance, the reviews inside [4].

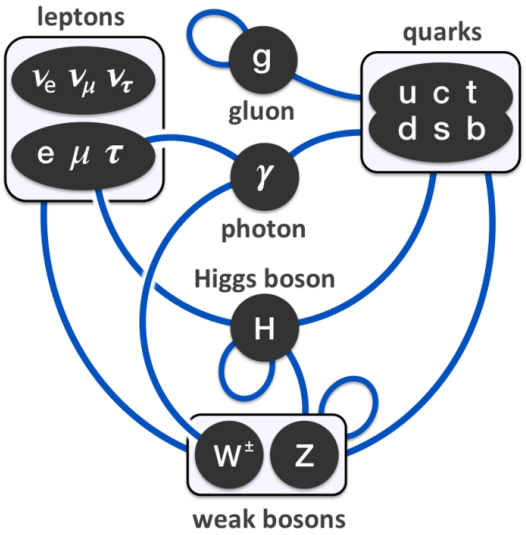



*Figure 1.2: Diagram of the coupling between the particles of the Standard Model.*

## 1.2 Heavy flavour and CKM matrix

In the previous section, it was mentioned that electroweak interactions do not conserve quark's flavour, in other words flavour transitions are allowed by the mediation of weak bosons. Only charged transitions are allowed, so from down to up quark or vice-versa, instead flavour changing neutral current (with $Z^0$ exchange) are forbidden in SM. Transition inside the same quark doublet (i.e. same generation) are preferred, instead transition between different generation are suppressed.

This is explained with the introduction of the $3 \times 3$ CKM matrix. This matrix "rotates" the mass eigenstates of down quarks $(d, s, b)$ to weak eigenstates $(d', s', b')$ that strictly couple with up quark of the correspondent generation. It is described by four free parameters: three mixing angles $\theta_{12}$, $\theta_{23}$ and $\theta_{13}$ and one irreducible complex phase $\delta$. This phase introduces in the SM a $\mathcal{CP}$ violation. The definition of matrix elements and the standard parametrization just described is shown below:

$$V_{CKM} \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{1.1}$$

where $s_{ij} \equiv \sin\theta_{ij}$, $\cos_{ij} \equiv \cos\theta_{ij}$ and $V_{ij}$ is the term involved in quark transition $q_i \to q_j$.

The suppression of the transitions between different generations can be noticed from the experimental values of the parameters: $s_{13} \ll s_{23} \ll s_{12} \ll 1$, and become apparent by using the Wolfenstein parametrization, an expansion of the CKM matrix elements in terms of small parameter $\lambda \equiv s_{12} = \sin\theta_C = 0.2272 \pm 0.0010$:

$$V_{CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & \simeq 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{1.2}$$

The link between the parametrization in 1.1 and the Wolfenstein parameter $A, \lambda, \rho, \eta$ is:

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \quad s_{23} = A\lambda^2 = \lambda\frac{|V_{cb}|}{|V_{us}|}, \quad s_{13}e^{-i\delta} = A\lambda^3(\rho - i\eta) = V_{ub}.$$

With this parametrization it is apparent that diagonal elements are close to one, while transitions between "far" generations have smaller couplings, in progressively increasing $\lambda$ powers. The corrections to higher $\lambda$ orders occur only at $\mathcal{O}(\lambda^7)$, and they are experimentally negligible.

An important tool in the analysis of CKM matrix elements comes from the unitarity of the matrix that gives the following set of relations between their elements:

$$\sum_{i=1}^{3} V_{ij}V_{ik} = 0, \quad \forall\ j, k \text{ with } j \neq k.$$

Each relation can be treated as the closed sum of three vectors in a plane, representing a triangle. They are called *unitarity triangle*, and their area is related to the amount of $\mathcal{CP}$ violation produced by the phase $\delta$. All the unitarity triangles have the same area but differ in term of sides and angles. The relation for which all sides are of the same order, in terms of Wolfenstein parameters $\mathcal{O}(\lambda^3)$, is the following:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \Leftrightarrow \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + 1 + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0. \tag{1.3}$$

The relation 1.3 resulting a triangle in the $\rho \times \eta$ plane, with one vertex in $(0,0)$, another in $(1,0)$ and the third in $(\rho(1-\lambda^2/2), \eta(1-\lambda^2/2))$. It is represented in figure 1.3, where the two useful definitions $\overline{\rho} \equiv \rho(1-\lambda^2/2)$, $\overline{\eta} \equiv \rho(1-\lambda^2/2)$ are used. The three angles can be expressed in term of $V_{ij}$ elements[2]:

$$\alpha = \phi_1 = \arg\left(\frac{V_{ud}V_{ub}^*}{V_{td}V_{tb}^*}\right), \quad \beta = \phi_2 = \arg\left(\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right), \quad \gamma = \phi_3 = \arg\left(\frac{V_{cd}V_{cb}^*}{V_{ud}V_{ub}^*}\right).$$

Since $\mathcal{CP}$ violation is related to the imaginary part of $V_{ij}$ elements i.e. to $\eta$, the measure of the angle of the unitarity triangle is a direct measurement of $\mathcal{CP}$ violation. The $\mathcal{CP}$-conserving case is instead equivalent to $\eta = 0$ i.e. vanishing triangle area.

The heavy flavour sector is an optimal framework to investigate small effects like $\mathcal{CP}$ violation and to test the limits of the Standard Model. In fact, in the processes which involve higher mass quarks, the non-perturbative QCD contribution are less important, therefore the theoretical previsions on quark transitions have better precisions, and the comparison with precise experimental measurements is possible. In addition, processes that involve $\tau$ quark, like $B \to \tau X$, allow to investigate the coupling between the heavy lepton and higher mass particles. In general, in the intensity frontier approach processes with very precise theoretical expectations inside the SM are investigated, to test the consistency of the model.

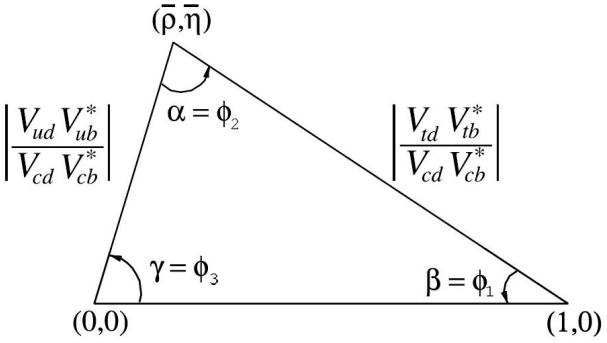



*Figure 1.3: The unitarity triangle with normalized base.*

## 1.3 Neutral mesons systems and $\mathcal{CP}$ violation

The neutral mesons play a central role in the study of $\mathcal{CP}$ violation because they offer the possibility to access several processes where the $\mathcal{CP}$ violation is manifested. In this section, a brief summary of neural mesons properties and oscillation phenomena is present, together with a description of the $\mathcal{CP}$ violation phenomenology. Further information can be found in [3].

### 1.3.1 Neutral Meson Oscillation

Let's consider a generic neutral meson $X^0$ and its $\mathcal{CP}$ conjugate $\overline{X}^0$. The production strong eigenstates, with a defined flavour, differ from mass eigenstates of time evolution, so a generic state at time $t_0$ of a produced $X^0$ is:

$$\psi(t_0) = a(t_0)\,|X^0\rangle + b(t_0)\,|\overline{X}^0\rangle + \sum_i c_i(t_0)\,|f_i\rangle\,,$$

[2] in the BaBar collaboration these angle are expressed as $\alpha, \beta, \gamma$; in Belle collaboration as $\phi_1, \phi_2, \phi_3$

where the $c_i$ represents the several decays amplitudes in $|f_i\rangle$ final states, while $a, b$ are the mixing amplitudes. Therefore, considering $t \gg t_{\text{production}}$ and restricting to the 2-state system $(X^0, \overline{X}^0)$ the evolution can be described by an effective Hamiltonian with the form

$$H_{\text{eff}} \equiv \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} + \frac{i}{2} \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}.$$

With the assumption $\mathcal{CPT}$ invariance, $\Gamma$ and $M$ matrices are hermitians, and the following eigenstates are obtained:

$$|X_1\rangle = \frac{1}{\sqrt{|p|^2 + |q|^2}} (p\,|X^0\rangle + q\,|\overline{X}^0\rangle \tag{1.4a}$$

$$|X_2\rangle = \frac{1}{\sqrt{|p|^2 + |q|^2}} (p\,|X^0\rangle - q\,|\overline{X}^0\rangle \tag{1.4b}$$

with the correspondent eigenvalues $\lambda_{1,2} = m_{1,2} - i\Gamma_{1,2}/2$ and the parameters:

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}}. \tag{1.5}$$

The physical parameters that can be extracted are the mass differences and the decay width of the system:

$$\Delta m = (m_1 - m_2) = \Re(\lambda_1 - \lambda_2) \tag{1.6a}$$

$$\Delta\Gamma = (\Gamma_1 - \Gamma_2) = -2\Im(\lambda_1 - \lambda_2), \tag{1.6b}$$

where $\lambda_{1,2}$ are the eigenvalues of the system.
Starting from a flavour-defined state it is possible to evaluate the probability to find a particle with the same flavour at time $t$ and evaluate the probability that mixing occurs. The calculation is omitted here, and only the final result is reported:

$$|X^0(t)\rangle = g_+(t)\,|X^0\rangle - \frac{q}{p} g_-(t)\,|\overline{X}^0\rangle$$

$$|\overline{X}^0(t)\rangle = g_+(t)\,|\overline{X}^0\rangle - \frac{q}{p} g_-(t)\,|X^0\rangle$$

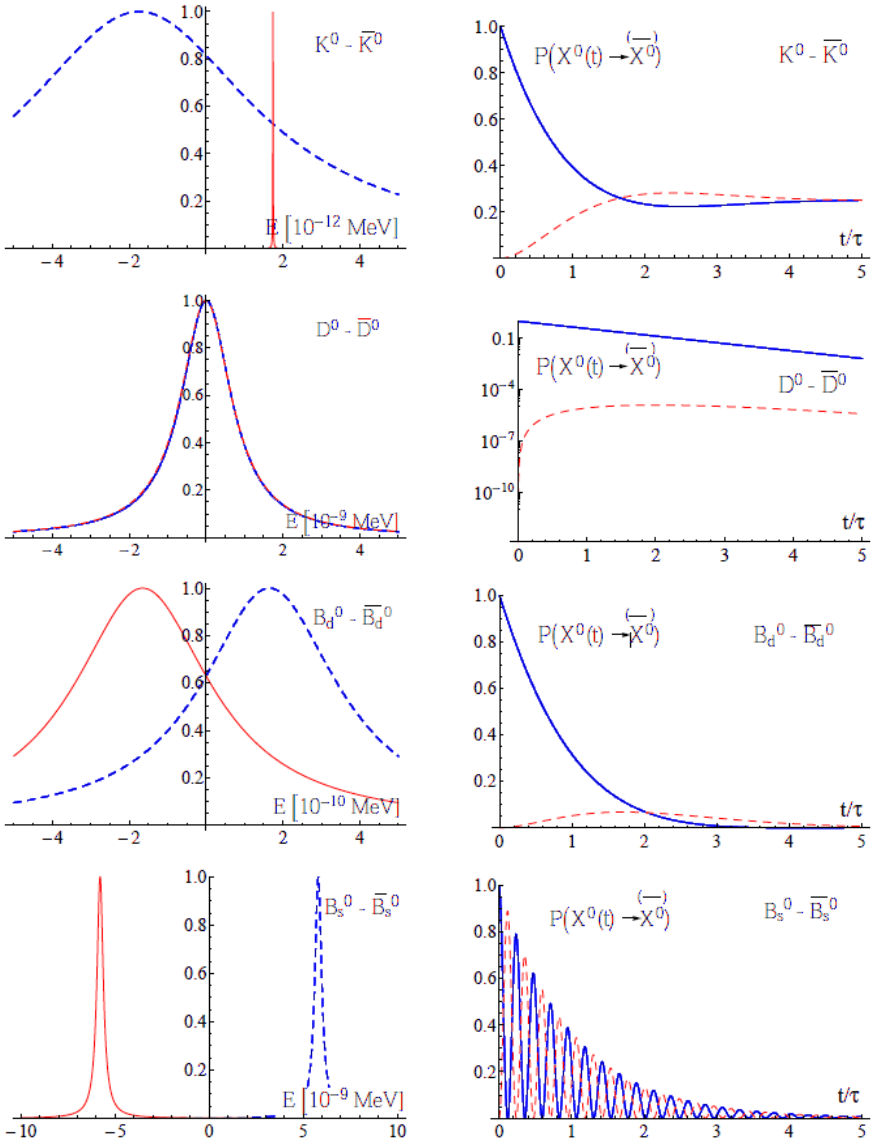
$$\text{where} \quad g_\pm \equiv \frac{1}{2}\left(e^{-m_1 t - \frac{1}{2}\Gamma_1 t} \pm e^{-m_2 t - \frac{1}{2}\Gamma_2 t}\right)$$

In this framework the probability $\mathscr{P}_{X^0 \to \overline{X}^0}(t) \propto \sin(\Delta m t)$, and so the period of oscillation $T_X = \frac{2\pi}{\Delta m_X}$.
In the SM there are four systems of neutral mesons which show an oscillatory behaviour: $K^0$, $D^0$, $B^0$, $B_s^0$. Experimentally the mixing has been observed in all the four systems, in table 1.1 the current measurements of $\Delta m$ and $\Delta\Gamma$ are shown.

| system | $\frac{\Delta m}{\Gamma}$ | $\frac{\Delta\Gamma}{2\Gamma}$ |
|---|---|---|
| $K^0 - \overline{K}^0$ | -0.95 | 0.99 |
| $D^0 - \overline{D}^0$ | $5 \cdot 10^{-3}$ | $5 \cdot 10^{-3}$ |
| $B^0 - \overline{B}^0$ | 0.77 | $5 \cdot 10^{-3}$ |
| $B_s^0 - \overline{B}_s^0$ | 26 | 0.15 |

*Table 1.1: Approximative value of two physical parameters of neutral mesons systems normalized to $\Gamma$ to allow comparisons (from [3]).*

Figure 1.4 points out the differences between various systems. The $K$-system shows large width differences and so two eigenstates with very different lifetimes ($K_L$, $K_S$), instead the quite small value of $\Delta m$ is manifest in long but observable period of oscillation. The $D$-system is critical: the $\Delta\Gamma$ and $\Delta m$ are very similar, in fact the mean values and the widths of masses are not distinguishable in the figure, while the period of oscillation is very long (not observable). The $B$-system has very small width differences and a small but observable $\Delta m$. The $B_s$-system instead has very small width differences but very large $\Delta m$, and it realizes two particles quickly oscillating. In conclusion the $K$ and $B$ system allow easier measurement, instead the $B_s$ and $D$ systems are the two harder to investigate because of their peculiar time dependence.




(a) *Mass plots of the propagation eigenstates (dashed blu and full red lines), normalized to 1 and with 0 mass average.*

(b) *Plot of the probability for a neutral meson to be found at time t in a particle (full blue line) or antiparticle state (dashed red line).*

*Figure 1.4: Features of the four neutral meson systems (from [3]).*

### 1.3.2 $\mathcal{CP}$ violation

$\mathcal{CP}$ violation occurs in three mechanisms, and two of these occur only in neutral meson systems:

- $\mathcal{CP}$ violation in decay (or direct violation)
- $\mathcal{CP}$ violation in mixing
- $\mathcal{CP}$ violation in the interference between decay and mixing.

In the next paragraphs the following notation is used to indicate the invariant transition amplitude between the meson state $|X\rangle$ and the multi-particle final state $|f\rangle$:

$$A_f = \langle f|H|X\rangle\,, \quad \overline{A}_f = \langle f|H|\overline{X}\rangle\,,$$
$$A_{\overline{f}} = \langle \overline{f}|H|X\rangle\,, \quad \overline{A}_{\overline{f}} = \langle \overline{f}|H|\overline{X}\rangle\,.$$

#### 1.3.2.1 $\mathcal{CP}$ violation in decay

This kind of $\mathcal{CP}$ violation occurs in neutral or charged meson decays when the two $\mathcal{CP}$ conjugate processes $X \to f$ and $\overline{X} \to \overline{f}$ shows different decay widths. The observable used to measure this violation is

$$\mathcal{A}_f = \frac{\Gamma_{X\to f} - \Gamma_{\overline{X}\to\overline{f}}}{\Gamma_{X\to f} + \Gamma_{\overline{X}\to\overline{f}}} = \frac{1 - |A_f/\overline{A}_{\overline{f}}|^2}{1 + |A_f/\overline{A}_{\overline{f}}|^2} \tag{1.7}$$

and so this $\mathcal{CP}$ violation is visible if and only if $|A_f/\overline{A}_{\overline{f}}| \neq 1$. This violation is the only one visible also for charged mesons.

#### 1.3.2.2 $\mathcal{CP}$ violation in mixing

This mechanism involves only neutral meson systems and occur only if the mass eigenstates differs from $\mathcal{CP}$ eigenstates. The $\mathcal{CP}$ eigenstates, for each neutral meson system, are

$$|X_\pm\rangle = \frac{|X_0\rangle \pm |\overline{X}_0\rangle}{\sqrt{2}},$$

from $\mathcal{CP}$ properties of physical states. This set of eigenstates and the propagation eigenstates in equation 1.4 are different only if

$$|\frac{q}{p}| \neq 1. \tag{1.8}$$

If the $\mathcal{CP}$ eigenstates and $H_{\text{eff}}$ are different $[\mathcal{CP}, H_{\text{eff}}] \neq 0$, and it means that $\mathcal{CP}$ is not conserved by evolution.

Conceptually the measure of a $\mathcal{CP}$ asymmetry in mixing is the measure of a different probability for a $X^0$ meson to oscillate in $\overline{X}^0$ with respect to a $\overline{X}^0$ to oscillate in $X^0$. With this idea, can be useful to select two different final states, accessible only for $X^0$ or $\overline{X}^0$. The most common example are semileptonic decays

$$\begin{cases} X^0 \to Y^- \ell^+ \nu_\ell & (\overline{d} \to W^+ \overline{u}) \\ \overline{X}^0 \to Y^+ \ell^- \overline{\nu}_\ell & (d \to W^- u) \end{cases}$$

were $u$ ($d$) means a generic up (down) quark and in the example shown $X^0$ is a $K^0$, $B^0$ or $B_s^0$ (the lepton charges are opposite in case of $D^0$). With the use of this class of decays, from the

lepton charge is possible to know the initial flavour of the particles and count the number of mesons that oscillate. The observable used in this case is:

$$\mathcal{A}_{\mathcal{CP}SL} = \frac{d\Gamma_{X\to\overline{f}}/dt - d\Gamma_{\overline{X}\to f}/dt}{d\Gamma_{X\to\overline{f}}/dt + d\Gamma_{\overline{X}\to f}/dt} \simeq \frac{1-|q/p|^4}{1+|q/p|^4}$$

#### 1.3.2.3 $\mathcal{CP}$ violation in the interference

This mechanism can occur in a system of neutral meson when both $X^0$ and $\overline{X}^0$ can decay to a final state $f$ eigenstate of $\mathcal{CP}$. In that case, there are two channels to reach the final state from an initial state $X^0$: direct decay $X^0 \to f$ or mixing and then decay $X^0 \to \overline{X}^0 \to f$. This kind of $\mathcal{CP}$ violation appear from the interference between these two channels, and can occur even in absence of direct and mixing violation too, in fact it occurs if and only if

$$\Im\lambda_f \neq 0 \quad \text{with} \quad \lambda_f = \frac{q}{p}\frac{\overline{A}_f}{A_f}, \tag{1.9}$$

which can be realized even it both $|q/p| = 1$ and $|\overline{\mathcal{A}_f}/\mathcal{A}_f| = 1$. The physical observable used to evaluate this $\mathcal{CP}$ violation is:

$$\mathcal{A}_{\mathcal{CP}f} = \frac{d\Gamma_{X^0\to f}/dt - d\Gamma_{\overline{X}^0\to f}/dt}{d\Gamma_{X^0\to f}/dt + d\Gamma_{\overline{X}^0\to f}/dt} \simeq S_{\mathcal{CP}f}\sin(\Delta m t) - C_{\mathcal{CP}f}\cos(\Delta m t), \tag{1.10}$$

with

$$C_{\mathcal{CP}f} = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}, \quad S_{\mathcal{CP}f} = \frac{2\Im\lambda_f}{1+|\lambda_f|^2}.$$

In case of $B^0$ system, where $\left|\frac{q}{p}\right| \simeq 1$ (not observable mixing violation) and $\Delta\Gamma \simeq 0$ the term in $C_{\mathcal{CP}f}$ vanishes and $\mathcal{S}_{\mathcal{CP}f} \simeq \Im\lambda_f \simeq \sin(2\beta)$, so the $\beta$ angle of unitarity triangle is directly accessible.

It is important to stress the time dependence of this $\mathcal{CP}$ violation: the evolution in time of physical states (with $H_{eff}$) include the factors $\frac{q}{p}$, so it is possible to measure the time evolution of $\mathcal{A}_{\mathcal{CP}f}$ if the oscillation can be observed. Anyway, it is possible to measure an "integrated" $\mathcal{CP}$ violation:

$$\mathcal{A}^{int}_{\mathcal{CP}f} = \frac{\int_0^\infty \Gamma_{X^0\to f}dt - \int_0^\infty \Gamma_{\overline{X}^0\to f}dt}{\int_0^\infty \Gamma_{X^0\to f}dt + \int_0^\infty \Gamma_{\overline{X}^0\to f}dt}$$

measuring the residual effect of asymmetry between $\mathcal{CP}$ states after the common exponential suppression of oscillation, but for some systems $\mathcal{A}_{\mathcal{CP}f}$ vanishes.

## 1.4 Heavy flavour experimental results

Historically, when only 3 quark $(u, d, s)$ were known, Cabibbo [6] gave a first explanation of the suppression of the transitions between different quark generation with the introduction of an angle $\theta_C$: with the prescription of multiplying by the factor $\cos\theta_C$ the transition amplitude in a process that involves $d \to u$ transition, or by $\sin\theta_C$ in case of $s \to u$ transition, the suppression was explained in most of the processes. However, some particular processes, like $K^0 \to \mu^+\mu^-$, were not observed thus a source of suppression had to be present. The solution was an extension of Cabibbo's hypothesis called the GIM (Glashow, Iliopoulus, Maiani) Mechanism [7]. This model has predicted the existence of a fourth quark of charge $q = \frac{2}{3}e$ (the quark charm), and

then $\theta_C$ was interpreted as a rotation angle between the mass quark eigenstates $(d, s)$ to the weak interaction eigenstates $(d', s')$:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}.$$

Therefore, only using the weak eigenstates the coupling is strictly $u \leftrightarrow d'$, $c \leftrightarrow s'$. Finally the discovery of the charm quark, in the $J/\psi$ resonance, gave to GIM Mechanism the experimental bases [8] [9].

This model includes the $\mathcal{P}$ symmetry violation of weak processes discovered in 1956 by Wu [10], but it is not able to explain the simultaneous violation of $\mathcal{C}$ and $\mathcal{P}$ discovered in kaons system in 1964 with the Cronin and Fitch experiment [11]. To give theoretical bases to the observed $\mathcal{CP}$ violation, Kobayashi and Maskawa introduced a third generation of quarks ($t$ and $b$) and a $3 \times 3$ extension of the Cabibbo matrix [12], which includes the $\delta$ phase that allows $\mathcal{CP}$-violating processes.

This model opened the door to a new class of experiments to investigate the heavy flavour and $\mathcal{CP}$ violation sector inside the SM framework. Two important experiment of the 80's are CLEO, situated at Cornell University (USA), that deeply studied the $B$-physics, and ARGUS experiment at DESY Laboratory (Germany) that made important discoveries in $B^0 - \overline{B}^0$ mixing [13] [14] [15]. Unfortunately, $\mathcal{CP}$ violation on the channels that they can investigate (direct and mixing channels) was too low to be observed. With the development of the B-Factory concept [16] realized in the BaBar experiment (1999-2008) at the PEP II accelerator (SLAC Laboratories, USA) and the Belle experiment (1999-2010) at the KEKB accelerator (KEK Laboratories, Japan), it has been possible to measure the $\mathcal{CP}$ violation in $B$ system in time-dependent (TDCPV) analysis. The theoretical and experimental idea of these experiments is described in chapter 2.

These two experiments built at the end of 90's reached fundamental result for flavour physics. Some of the most important are the first evidence of TDCPV in $B$ system, the measure of $\sin(2\beta)$, $\alpha$ and $\gamma$ angles of the unitarity triangle, the measurement of various elements of CKM matrix. Many these measurements have been combined in the so-called *CKM Fit*: a global fit to the CKM elements using all theoretical and experimental constraints, from which is possible to check the consistency with the theoretical model. The CKM mechanism and its complex phase as the dominant source of $\mathcal{CP}$ violation inside the SM was confirmed from this fit. The fit result consistent also subdividing the constraints arising from $\mathcal{CP}$-conserving variables ($\Delta m_{B_d}$, $\Delta m_{B_s}$, $|V_{ub}|$) and $\mathcal{CP}$-violating variables ($\varepsilon_K$, $\alpha$, $\beta$, $\gamma$). The CKM fit update on 2016 data is reported in figure 1.5

After the end of the first generation of B-Factories, Belle collaboration started a big upgrade both to the detector and to the B-Factory KEKB to improve the accelerator luminosity and the detector performances. The new B-Factory is called *SuperKEKB* and the upgraded detector *Belle II*, and they are detailed described in section 2.4 and 2.3.

The B-Factories are not the only experiments developed to study heavy flavour and $B$ physics. Several other experiments reached relevant results, like D0 and CDF at Tevatron (Fermilab, USA) or more recent CMS, ATLAS and LHCb at LHC (CERN, Switzerland). In particular, the last one is a detector specifically designed for heavy flavour physics and is the main competitor of the Belle II experiment.

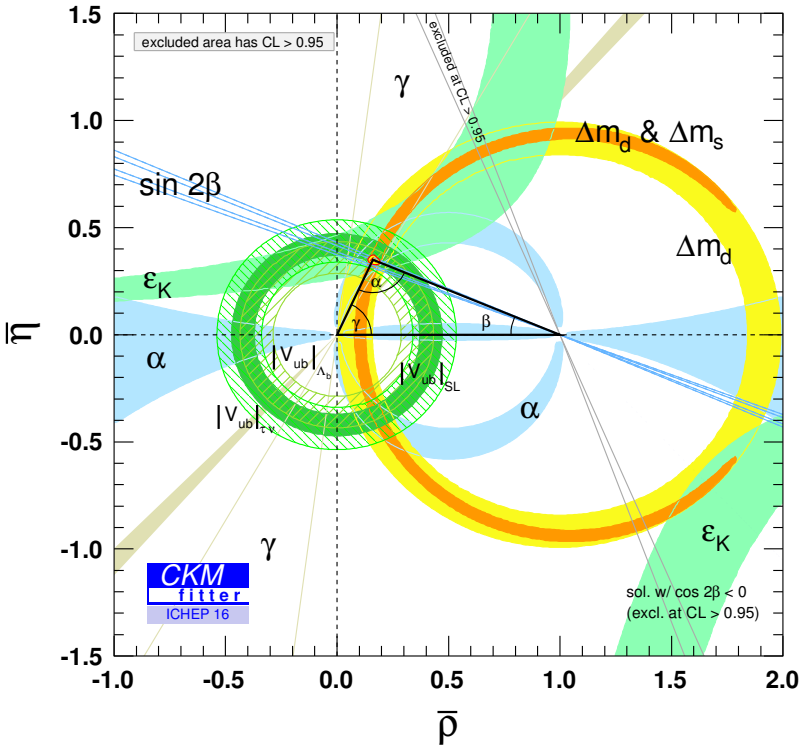



*Figure 1.5: CKM fit the $\bar{\rho} \times \bar{\eta}$ plane. In this plot constraints arising from both $\mathcal{CP}$-conserving ($\Delta m_{B_d}$, $\Delta m_{B_s}$, $|V_{ub}|$) and $\mathcal{CP}$ violating ($\varepsilon_K$, $\alpha$, $\beta$, $\gamma$) variables are shown. This plot is updated at summer 2016 analysis [17].*

## 1.5 $B \to D^{(*)}\tau\nu_\tau$ sector

In the intensity frontier approach, the channels used to look for New Physics are the ones with solid theoretical expectations i.e. reduced theoretical uncertainties. In these channels, the precise measurements can be compared with the SM calculation with the chance to find discrepancies which can be interpreted as New Physics signatures. In this picture, the channels with multi-hadron final state are less effective because hadron rescattering increases the uncertainties in theoretical calculations.

In the heavy flavour sector, various tensions with respect to SM theoretical expectations are currently present. One illustrative example are the semitauonic $B$ decays via $D^{(*)}\tau\nu_\tau$ channels, in fact both the measurement of the branching ratios and the $\tau$ and $D^*$ polarizations currently show a not negligible strain with respect to Standard Model prevision. Various experiments published measurement (BaBar [18], Belle [19][20][21][22], LHCb [23]) about these channels that reduce the chance that this strain could be a systematic bias not considered. Belle II has the opportunity to improve these measurements because of the larger statistic that it is going to collect and because of the reduced systematics that it can reach thanks to the increased performance of the detector and the very clean environment.

In the course of this thesis, some studies of the Belle II sensitivity in this class of decays have been carried out, although they are not complete enough to be presented here and only the current status of the measurement the Belle II prospects are discussed.

In the SM the decays $B \to D^{(*)}\tau\nu_\tau$ are described with a tree level diagram, in an electroweak transition $b \to c\tau\nu_\tau$ mediated by a virtual $W$ boson (the diagram is shown in figure 1.6). The semitauonic decays are useful to investigate processes Beyond Standard Model (BSM) because the high mass of the $\tau$ lepton is sensitive to additional couplings [24]. The best way to study the $B \to D^{(*)}\tau\nu_\tau$ decays is using the relative rates[3]:

$$R_{D^{(*)}} \equiv \frac{BR(B \to D^{(*)}\tau\nu_\tau)}{BR(B \to D^{(*)}\ell\nu_\ell)} \quad \ell = e, \mu. \tag{1.11}$$

In fact the quantities $R_{D^{(*)}}$ are independent of the CKM element $|V_{cb}|$ and largely of the hadronic matrix elements (form factors) [18]. The SM expectations of these ratios have small uncertainties [25], and so there are various BSM theories that can be tested with $R_{D^{(*)}}$ measurements.

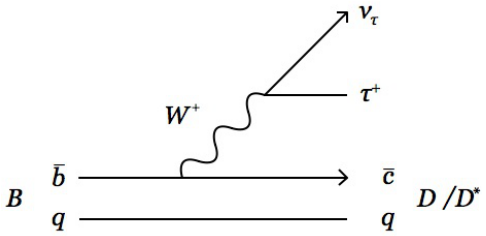



*Figure 1.6: Feynman diagram of $B \to D^{(*)}\tau\nu_\tau$ at tree level in SM prevision*

Another observable connected with semitauonic decays is the polarization of $\tau$ lepton and $D^*$ meson [26]. Their definitions are:

$$P_{\tau(D^{(*)})} \equiv \frac{\Gamma^+ - \Gamma^-}{\Gamma^+ + \Gamma^-},$$
$$P_{D^*} \equiv \frac{\Gamma_{//}}{\Gamma_{//} + \Gamma_\perp},$$

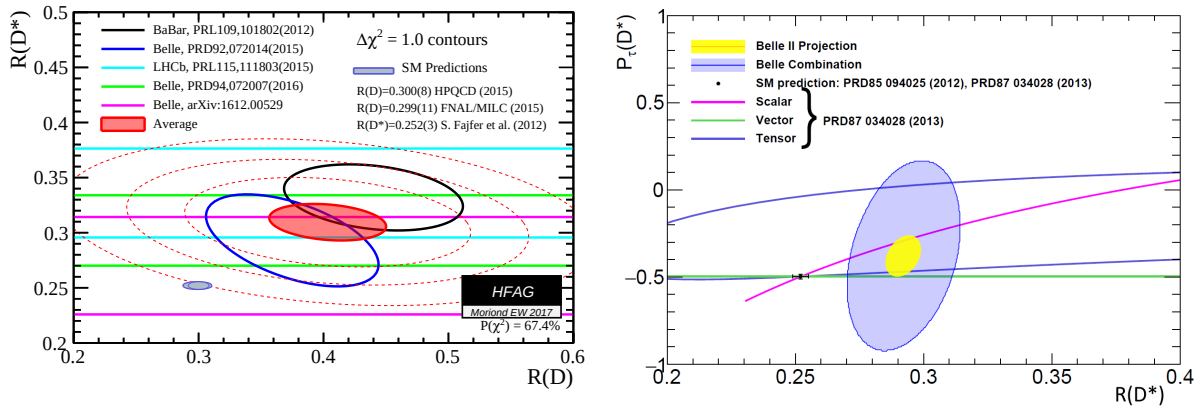
where $\Gamma^{+(-)}$ are the decay rate with the $\tau$ helicity $+\frac{1}{2}(-\frac{1}{2})$ and $\Gamma_{//(\perp)}$ with the longitudinally (transversely) polarized $D^*$.

The SM predictions come from the evaluation of the differential distributions of decay rate $d\Gamma_{D^{(*)}}/dq^2$ where $q^2 = (p_\tau + p_\nu)^2 = (p_B - p_{D^{(*)}})^2$ is the momentum transfer to the lepton pair. CKM matrix element $V_{cb}$ are extracted through the measurements of $B \to D^{(*)}\ell\nu_\ell$ for $\ell = e, \mu$, and then the form factors are evaluated using lattice QCD and HQET (Heavy Quark Effective Theory) In conclusion the SM predictions are [25] [27] [26] :

$$R_D^{SM} = 0.2999 \pm 0.003$$
$$R_{D^*}^{SM} = 0.257 \pm 0.003$$
$$P_{\tau(D)}^{SM} = 0.325 \pm 0.009$$
$$P_{\tau(D^*)}^{SM} = -0.497 \pm 0.013$$
$$P_{D^*}^{SM} = 0.458 \pm 0.009$$

The experimental measurements of $R_{D^{(*)}}$ performed by BABAR, Belle and LHCb in various channels are not consistent with SM predictions: the world average of measurements, taking care of correlation of $R_D$ and $R_{D^*}$ when are present both, is $3.9\sigma$ away from the SM value. The measurement of the polarization is also out of SM predictions, but at less than $2\sigma$. In figure 1.7 the SM values and a summary of experimental results and their uncertainties are shown.

[3] $BR$ symbol means *branching ratio*




(a) *Various measurement of $R_D$ and $R_{D^*}$ and the world average, with SM predictions $3.9\sigma$ away also shown (from [28].*

(b) *Measurement of $P_{\tau(D^*)}$ versus $R_{D^*}$. SM predictions and some theoretical expectation for extended Higgs sector are shown. Prospect of Belle II is also shown (from [27]).*

*Figure 1.7: Summary of measurement and SM prevision for the channels $B \to D^{(*)}\tau\nu_\tau$.*

### 1.5.1 New Physics models

In the presence of New Physics, the semitauonic $B \to D^{(*)}\tau\nu_\tau$ decays are described with a model-independent analysis, using the most general effective Lagrangian of the $b \to c\tau\nu_\tau$ transitions, that includes the 5 Wilson coefficients $C_X$ $(X = V_{1,2}, S_{1,2}, T)$ [29]:

$$\mathcal{L}_{eff} = -2\sqrt{2}G_F V_{cb}\big[(1 + C_{V_1})O_{V1} + C_{V_2}O_{V_2} + C_{S_1}O_{S_1} + C_{S_2}O_{S_2} + C_T O_T\big], \tag{1.12}$$

where the $O'_X s$ are 4-fermion operators defined by:

$$\begin{aligned} O_{V_1} &= \overline{c}_L\gamma^\mu b_L\overline{\tau}_L\gamma_\mu\nu_L \\ O_{V_2} &= \overline{c}_R\gamma^\mu b_R\overline{\tau}_L\gamma_\mu\nu_L \\ O_{S_1} &= \overline{c}_L b_R\overline{\tau}_R\nu_L \\ O_{S_2} &= \overline{c}_R b_L\overline{\tau}_R\nu_L \\ O_T &= \overline{c}_R\sigma^{\mu\nu}b_L\overline{\tau}_R\sigma_{\mu\nu}\nu_L. \end{aligned}$$

These coefficients should be all equal to 0 in SM, (i.e. the $\mathcal{L} \propto O_{V_1}$) but in the presence of New Physics one or more than one of them may turn out to be different from 0. The value of these coefficients can distinguish from various New Physics scenarios.

The easiest cases are the scenarios where one of the $C_X$'s is dominant or a scenario where two $C_X$ are not vanishing, $C_{S2} \neq 0$ and $C_T \neq 0$, like in some leptoquark models [30] [31]. The BaBar measurement [18] has excluded one of simplest extension of the Higgs Sector (2HDM Type II, where became relevant the channel with a charged Higgs that should have the role of $W$ boson of diagram 1.6), equivalent to a dominant $C_{S_1}$ in Wilson coefficients framework. In figure 1.8 the result of the analysis that allows to exclude the 2HDM Type II model is reported. Anyway because of tree level structure of the semitauonic $B$ decays and the magnitude of the experimental discrepancy the BSM physics is required to be close to weak scale to explain this discrepancy. With this hypothesis it is possible to identify two classes of models: a more complex extended Higgs sector [32] (not excluded by the previous measurement) or leptoquark models. The theoretical discussion is omitted here, but it possible to find details in [27] and [29].

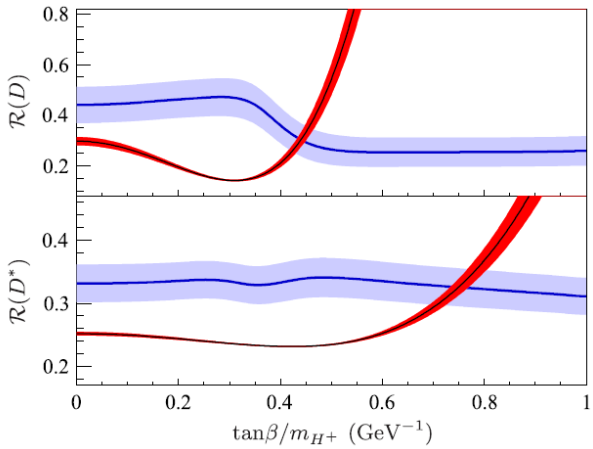



*Figure 1.8: Result of* BaBar *analysis [18]: comparison between experimental value of $R^{(*)}$ (blue band) and the 2HDM Type II prediction (red band) in function of $\tan\beta/m_{H^+}$, where $m_{H+}$ is the mass of the charged higgs and $\tan\beta$ is the ratio of the VEV of the two Higgs doublets. The widths of the bands represent the uncertainties at 68% (dark blue or red) and 95% (light blue or red). The SM prediction corresponds to $\tan\beta/m_{H^+} = 0$.*

In figure 1.9 the allowed values of the various $C_X$'s are shown considering the world average of the experimental measurement of $R_D$ and $R_{D^*}$. All the models with a single $C_X \neq 0$ and the two leptoquark model cited above are presented. The dominant $C_{S_1}$- scenario has been excluded (thus it is not shown), instead a dominant $C_{S_2}$ scenario remains possible (2HDM Type III) as the leptoquarks models which involve $C_T$ and $C_{V_{1,2}}$.

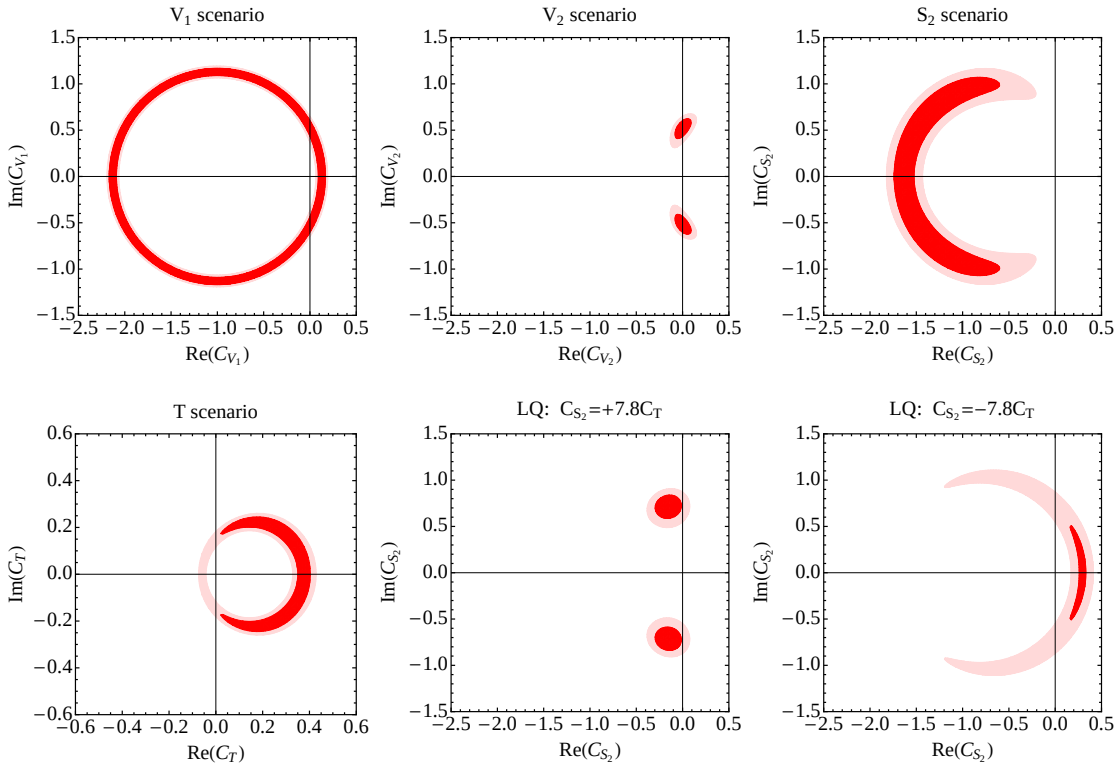



*Figure 1.9: Current constraints on various new physics scenarios based on $C_X$'s of lagrangian 1.12 from experimental measurement. In the complex plane of the $C_X$'s are shown the allowed region within the 95% (light red), 68% (red) of C.L.. The SM prevision is $C_X = 0$ for all the $C_X$'s (preliminary plots from [27]).*

### 1.5.2 Summary of previous measurements

The phenomenology of $B \to B^{(*)}\tau\nu_\tau$ decays is quite complicated and there are a lot of possible channels and decay chains which can be included in the reconstruction, and consequently different reconstruction strategies have been used. The unavoidable presence of at least 2 neutrinos, one from $\tau$ production and one for its decay (and one more for leptonic *tau* decays), complicate the reconstruction and prevent a full reconstruction of the event energy

Focusing on B-Factories analyses, independently from the signal decay chain, three different reconstruction techniques have been used called *hadronic tagging*, *semileptonic tagging* and *inclusive tagging*. In the B-Factory environment, the signal is produced from the $\Upsilon(4S) \to B\overline{B}$ decay, with the $B$ pair at rest in the $\Upsilon(4S)$ center-of-mass frame, but boosted in the laboratory frame. In the first two methods One of the two B's, called $B_{\text{tag}}$, is reconstructed first via well-known hadronic or semileptonic channels. The reconstruction the $B_{\text{tag}}$, and the knowledge of the initial $\Upsilon(4S)$ state allows inferring some important properties on the signal $B$ ($B_{\text{sig}}$), like charge, flavour and momentum. Then final state particles are associated with the required decay chain and finally $B_{\text{sig}}$ is reconstructed to obtain the signal. In the third method, the inclusive tagging, the signal tracks are selected first, and then the Rest Of Event is used to identify the $B_{\text{tag}}$ and obtain the $B_{\text{sig}}$ properties. The Hadronic tag has higher purity but lower efficiency than the semileptonic tag. In fact, semileptonic branching ratios are higher, increasing the efficiency, but they include decays with neutrinos that leave open the event, decreasing the purity. The inclusive tag has a higher efficiency than the semileptonic tag but a lower purity for the same reasons.

After the tagging, the signal is extracted using variables depending on the specific analysis, to better reduce the backgrounds. One typical B-Factory variable is the energy difference[4]

$$\Delta E = \frac{2p_B p_{e^+e^-} - s}{2\sqrt{s}} \overset{\text{c.m. frame}}{=} E^*_B - E^*_{\text{beam}}, \tag{1.13}$$

where all quantities with $^*$ symbol are estimated in the center-of-mass frame, $p_B$ and $p_{e^+e^-} = p_{e^+} + p_{e^-}$ are the four-momenta of the $B$ candidate and the electron-positron system and $\sqrt{s} = 2E^*_{\text{beam}}$ is the center-of-mass energy. If the $B$ is correctly reconstructed, the energy of its decay products is equal to half of the energy of the center-of-mass, i.e.:

$$\begin{cases} E^*_{\text{reco}} = E^*_{\text{beam}} = \sqrt{s}/2 \\ m_{\text{reco}} = m_B. \end{cases}$$

Therefore, if the decay products of the $B$ meson are correctly reconstructed, $\Delta E$ peaks at 0, otherwise, in case of misreconstructed particles, $\Delta E$ peaks at lower values. The uncertainty on $\Delta E$ arises from the beam energy spread and from the resolution on the $B$ energy measurement:

$$\sigma^2_{\Delta E} = \sigma^2_{E^*_{\text{beam}}} + \sigma^2_{E^*_B}.$$

A complementary variable with respect to $\Delta E$ is the beam-constrained mass[5]:

$$m_{bc} = \sqrt{\frac{(s/2 + \mathbf{p}_B \mathbf{p}_{e^+e^-})^2}{E^2_{e^+e^-}} - \mathbf{p}^2_B} \overset{\text{c.m. frame}}{=} \sqrt{E^{*2}_{\text{beam}} - \mathbf{p}^{*2}_B}, \tag{1.14}$$

which is mainly independent from the reconstructed mass of the $B$ and its decay products. The beam-constrained mass in case of correctly reconstructed $B$ peak at $m_B$, instead the events with

[4] in this paragraph has been set $c = 1$ to simplify the formulas.

[5] it is also called more precisely beam-energy substituted mass $m_{ES}$

some missing particles produce a shift toward lower $m_{bc}$ values. The uncertainty on this variable arises from the beam energy spread and the resolution in the $B$ momentum measurement:

$$\sigma^2_{m_{bc}} = \sigma^2_{E^*_{\text{beam}}} + \left(\frac{p^*_B}{m_B}\right)^2 \sigma^2_{p^*_B},$$

but because of the small momentum of the $B$ in the center-of-mass system, $\sigma^2_{m_{bc}}$ results dominated by the beam energy spread.

Both Belle [20] [22] [21] and BABAR [18] have used $m_{bc}$, $\Delta E$ and the unassociated photon energy in ECL $E^{\text{extra}}_{\text{ECL}}$ to identify the $B_{\text{tag}}$, and then missing mass $M^2_{\text{miss}}$ (Belle) or the $q^2$ (BABAR) to extract the signal. The measurement of $\tau$ polarization [19] has been performed only by Belle collaboration, which has fitted the $\tau$ helicity angle together with $E^{\text{extra}}_{\text{ECL}}$.

Most of the analysis reconstruct $\tau$ leptonically to reduce strongly the background. With this choice the most important background arises from $B \to D^{**}\ell\nu_\ell$ where $D^{**}$ is a class of excited states, with mass higher than $D^*(2010)^{\pm}$. In fact the $D^{**}$ events mimic signal if the $\pi^0 s$ from their decays are missing. The branching fraction and so the Monte Carlo simulation of this channel is not well known, and for example BABAR has used a control sample of $B \to D^*\pi^0\ell\nu_\ell$ to evaluate the background yield. On the other hand, it is possible to reconstruct $\tau$ from hadronic decays (Belle has been used $\tau \to \pi^-\nu_\tau$ or $\tau^- \to \rho^-\nu_\tau$). However, in this case the main background comes from hadronic $B$ decays and there is an increase of the systematic errors because of the uncertainties on $B$ branching fractions. The use of hadronic $\tau$ reconstruction makes the final state particle different in signal and normalization events, and it is necessary to correct

$$R_{D^*} \to R_{D^*} \times \frac{\varepsilon_{\text{norm}}}{\varepsilon_{sig}} \frac{1}{BR(\tau \to h\nu)}.$$

Anyway, both with leptonic or hadronic $\tau$ reconstruction the dominant source of the systematic uncertainty is the limited size of the by Monte Carlo sample, in term of branching fraction, reconstruction efficiencies, cross-feed estimation and knowledge of fit variables.

In all these analyses the $q^2$ spectrum has been measured because it could be a probe of New Physics but was always consistent with SM predictions.

LHCb collaboration has been quite a similar approach to these analysis [23], but a single decay chain has been reconstructed (only one $D^*$, $D$ and $\tau$ decay channel) because of the very much larger statistic of a hadronic collider which allows efficiency losses to gain purity. For the same reason the definition of $R_{D^*}$ is quite different too, in fact the normalization channel includes only $\mu$ decays. After background suppression with $M^2_{miss}$ and $q^2$, LHCb has reconstructed the signal with displaced vertex, mass constraints and isolation criteria.

The figure 1.7 shows that some measurement evaluated $R_{D*}$ only (LHCb measurement, and Belle's one with semileptonic tag and with hadronic $\tau$ reconstruction). In fact $R_D$ is more difficult to measure because it is not possible to use the variable $\Delta m_{DD*} = m_{D^*} - m_D$ to suppress the background and select the signal. Therefore the $B \to D\tau\nu$ channels have the worst resolution and have reduced purity, thus in the measurements that suffer for low purity (i.e. semileptonic tag, hadronic $\tau$ decays, open events from hadronic collider) the measurement of $R_D$ it is not performed with satisfactory performances.

At B-Factories no measurement has been done with $\tau$ reconstruction using decays in 3 charged particles, but the use of this channel can increase of more than 15% the available statistic of $B \to D^{(*)}\tau\nu_\tau$ decays. The main issue involved in this decay chain is the background from hadronic $B$ and $D$ decays that can mimic a 3-prong $\tau$ decay. The LHCb Collaboration overcome this issue with the use of a *tau* vertex displacement constraint, due to the large boost available

[33]. The increased performance of Belle II and in particular the increased vertex resolution has the chance to be able to include this channel in the Belle II $R_{D^{(*)}}$ analysis.

### 1.5.3 Belle II prospect

The 50 ab$^{-1}$ of integrated luminosity of Belle II are going to bring statistical uncertainties of $R_{D^{(*)}}$ below systematic uncertainties, so the goals of Belle II is to reduce these systematics. The increased resolution mass and vertex resolution, combined with higher tracking efficiency at low momentum and a better beam-background rejection help Belle II to improve the knowledge of $D^{**}$ particles and their decays, in term of kinematic and branching fractions. These improved performances can allow a better measurement of $B \to D^* \ell \nu + h^0$ (where $h^0$ are neutral hadrons) too. The new tagging and reconstruction methods of Belle II, called *Full Event Interpretation*, increase the tagging and reconstruction efficiency, reducing the corresponding systematics. Another central improvement arises from Monte Carlo: the produced sample allows to reduce all the related uncertainties, both for the larger size and for the better modelling of collision environment. At last the momentum resolution at Belle II provide $q^2$ distributions that can discriminate New Physics scenarios. In the context of model-independent analysis described in section 1.5.1 it is possible to connect the accessible New Physics mass-scale with the Wilson coefficients:

$$C_X \sim \frac{1}{2\sqrt{2} G_F V_{cb}} \frac{g g'}{M_{NP}^2} \tag{1.15}$$

where $g$ and $g'$ denote the general coupling of new particles to quark and lepton at New Physics mass scale $M_{NP}$. From equation 1.15 it is possible to evaluate, taking $g \simeq g' \simeq 1$ and $C_X$ of the expected order of magnitude, that the accessible mass scale at Belle II is $M_{NP} \sim 5 \div 10$ TeV/$c^2$.

# Chapter 2

# The Belle II experiment

The Belle II experiment at the SuperKEKB accelerator is the next generation of B-Factory, which is planning to collect an integrated luminosity of 50 $ab^{-1}$ at $\Upsilon(4S)$ resonance, with the goal to refine several measurements in heavy flavour sector of SM and searching for New Physics signatures. In this chapter, the main features of the B-Factories are described and the motivation for the Belle II experiment are discussed, followed by a high level description of the accelerator and the detector. The figures used in this chapter, unless otherwise specified, are taken from the Belle II Technical Design Report [34].

## 2.1 The B-Factory idea

From the legacy of the $B$-experiments of the '80s appeared that the $\mathcal{CP}$ violation in $B$ system is too low to be observed in direct or mixing channels. Then experiments designed to investigate TDCPV in the interference channels has been developed.

Let's consider a generic decay $B^0 \to f$, where $f$ is a $\mathcal{CP}$ eigenstate. In order to measure $\mathcal{A}_{\mathcal{CP}f}$, from equation 1.10, it is necessary to measure the yield of $f$ from both initial flavour states $B^0$, $\overline{B}^0$ and to know the time $\Delta t$ between the $B$ production and the decay in which the direct and the mixed channel could interfere. Therefore a procedure to *tag* the flavour of the initial state and the precise information of the production and decay vertices of the initial $B^0$ ($\overline{B}^0$) is needed. Finally, from the measurement of the different decay rates from $B^0$ and $\overline{B}^0$ initial states and so from $\mathcal{A}_{\mathcal{CP}f}$, it is possible to access to the $\beta$ angle of unitarity triangle.

In the B-Factory concept, the initial flavour is identified with the production of a $B\overline{B}$ pair from the decay of the $\Upsilon(4S)$, a strong resonance with mass $m_{\Upsilon(4S)} = 10.5794$ GeV/$c^2$ and $J^{\mathcal{PC}} = 1^{--}$ [4]. The quark composition of $\Upsilon(4S)$ is $b\bar{b}$, it is the lightest strong resonance with a mass sufficient to decay in $b$-flavoured mesons, and the branching fractions in $B\overline{B}$ pair is over 96%. The pair of $B$ meson results entangled in the decay and the $\Upsilon(4S)$ is unflavoured, so from the knowledge of the flavour of one $B$ it is possible to assign the flavour to the second $B$ too. The *flavour tagging* is performed reconstructing specific decay chains of one $B$ with a clean signature of the flavour of the mother meson $B_{\text{tag}}$. Then, if the signal decay $f$ is reconstructed from the other tracks of the events, the initial flavour state of the $B_{\text{sig}}$ is assigned.

To obtain the $\Delta t$ information the $\Upsilon(4S)$ is produced boosted with respect to the laboratory frame. In $\Upsilon(4S)$ frame the $B\overline{B}$ pair is produced almost at rest ($m_{Y(4s)} - 2mB^0 \simeq 19$ MeV), thus in the laboratory frame the flight direction of both $B$ is the almost same of the boost. After the decay and the reconstruction of the first $B$ the decay vertex position $z_1$ is evaluated. If the second $B$, decays in the signal $f$, it will be reconstructed. Then from the evaluation of the

second vertex position $z_2$, it is possible to obtain $\delta z = z_2 - z_1$ and thus $\Delta t$. Without the boosted center of mass will be measured $z_1 + z_2$ and so $t_1 + t_2$ only, without the chance to obtain the time interval in which the two channels can interfere.

In conclusion, the main features of a B-Factory can be summarized as follows:

**Completely known initial state.** The $\Upsilon(4S)$ initial state is produced from $e^-e^+$ collision. In that case the expected cross section is shown in table 2.1. The signal-to-noise ratio is about $\sigma_{b\bar{b}}/\sigma_{\text{tot}} \simeq 0.28$ and the charge track multiplicity from $\Upsilon(4S)$ is about 11.

**Boosted center-of-mass.** The boosted center-of-mass, necessary to evaluate $\Delta t$ is realized with an asymmetric collider. The boost is useful to increase the decay length of the $B\overline{B}$ pair at measurable level too. In the BaBar experiment the boost was $\beta\gamma = 0.56$, in Belle it was 0.42, while at Belle II it will be 0.28.

**High Luminosity.** Since branching fractions of interesting decays can be rather small, a machine that is able to produce a large sample of $B\overline{B}$ pair in needed. For an $e^+e^-$ machine this implies peak luminosities in excess of $\mathcal{L} \geq 10^{34}$ cm$^{-2}$s$^{-1}$, SuperKEKB is planning on $\mathcal{L} \geq 8 \cdot 10^{35}$ cm$^{-2}$s$^{-1}$.

**High-performance detector.** The detector must have a good vertex resolution ($< 10^2$ $\mu$m) to extrapolate the vertices position, an excellent particle identification capability to perform a high purity flavour tag, and a large coverage to exploit the high luminosity.

| $e^+e^- \rightarrow$ | Cross section [nb] |
|---|---|
| $\Upsilon(4S)$ | $1.05 \pm 0.10$ |
| $c\bar{c}$ | 1.30 |
| $s\bar{s}$ | 0.38 |
| $u\bar{u}$ | 1.61 |
| $d\bar{d}$ | 0.40 |
| $\tau^+\tau^-(\gamma)$ | 0.919 |
| $\mu^+\mu^-(\gamma)$ | 1.148 |
| $e^+e^-(\gamma)$ | $300 \pm 3$ |

*Table 2.1: Production cross section of $e^+e^-$ at $\sqrt{s} = m_{\Upsilon(4S)}c^2 = 10.58$ GeV (from [27], see it for uncertainties details).*

It is possible to find more details about the B-Factory theoretical concept and their experimental features in [3].

## 2.2 Physics motivations for a super B-Factory

The discovery of a SM Higgs particle at LHC experiments put strong constraints in various BSM theories: excluded several extended Higgs models and pushed up to TeV scale the chance to find new gauge bosons or fermions. In the intensity frontier approach, the New Physics that can be investigated scales with the inverse of the mass of BSM-particles, therefore the chance to found a New Physics signal has been strongly reduced.

However, a new generation of B-Factory, with a statistic increased of more than one order of magnitude and increased performances has the chance to investigate specific flavour coupling for which indirect searches can push the New Physics scale higher than direct searches [27]. The first generation of $B$-Factory has tested the CKM mechanism at the 10% level, but there are margins for of New Physics to exists below this upper limit. In case LHC finds New Physics at the energy frontier, precision measurement are essential to further understand the discoveries or,

in case LHC finds no evidence of New Physics, the B-Factories offer a unique way to probe for New Physics beyond TeV scale.

One of the main questions addressed by a new generation of B-Factory is the investigation of BSM sources of $\mathcal{CP}$ violation in quark sector: the SM $\mathcal{CP}$ violation is not sufficient to solve the baryon asymmetry and the first generation of B-factory and LHCb experiment measurements show several tensions with respect to SM expectation. With this purposes it is interesting to study TDCPV in $b \to s$ transitions, in which the SM $\mathcal{CP}$ violation is expected to be very small, and an observation can be interpreted as a BSM signal. With the same purpose, $\mathcal{CP}$ violation in charm mixing can be investigated. Charmless processes, like $B \to K\pi$, is another possible source of a large amount of direct $\mathcal{CP}$ violation.

Another interesting sector in flavour physics is the searches for flavour-changing neutral currents (FCNC) beyond the SM: the FCNC are strongly suppressed by the GIM mechanism, thus the measurement is very sensitive to New Physics contributions. Approaches include TDCPV searches in neutral channels or transitions with large missing energy like $b \to s\nu\overline{\nu}$. In this classes of decays is crucial the vertex resolution and the neutral reconstruction capability of a B-Factory.

The cross section of $e^+e^- \to \tau^+\tau^-$ at B-Factory center-of-mass energy is similar to the $B\overline{B}$ one, therefore the B-Factory results an effective "$\tau$-factory" too. The lepton flavour violation (LFV) has been measured in neutrino mixing phenomena only, but the study of the $\tau$ physics allows to investigate BSM sources of LFV. The $\tau$ physics program includes $\mathcal{CP}$ violation measurement, electric dipole moment measurement and $(g-2)_\tau$ measurement too.

Despite the discovery of the SM Higgs, many extensions of Higgs sector are not excluded yet, and $B$ decays with $\tau$ production ($B \to \tau\nu$, $B \to D^{(*)}\tau\nu$) offer a not trivial way to investigate this sector of possible New Physics. Currently this class of decays shows tensions with SM, and the B-Factory environment has the correct properties (in term of vertex resolution, missing mass identification, luminosity) to study the tauonic and semitauonic $B$ decays. More in general, the decays which involve $\tau$ are challenging because of the large number of neutrinos involved, but accessible with a new generation B-Factory. The semitauonic decays and semileptonic $B \to K^*\mu^+\mu^-$ decays can be useful also to test the lepton universality.

The capability of a new generation of B-Factory to discover New Physics is not limited to the flavour sector. A B-Factory has a high sensitivity to dark matter via missing energy decays, i.e. via direct detection of new particles. Appropriate specific triggers can be developed for this purpose. In addition, the possibility to tune the center-of-mass of the collider on various strong resonance $\Upsilon(nS)$ allows studying a large family of quarkonia decays, to investigate low energy QCD with high level of precision. In this sector the detection capability for neutral channels of the B-Factory can be fully exploited. At last, a B-Factory with increased performance can measure with unprecedented precision the CKM observables increasing the knowledge of SM.

This physics program is going to be covered by the Belle II experiment at the SuperKEKB B-Factory starting from 2018. This detector is going to have a complementary role in flavour physics with respect to LHCb experiment: the latter has a higher statistic for $B_s$ and $B$ mesons, and its measurements dominate all-charged final states. On the other hand, Belle II measurements dominate $B$ channels with neutral final states (multiple neutrinos or photons).

## 2.3 SuperKEKB

The next generation of B-Factory is represented by SuperKEKB [35], the upgraded collider of KEKB, placed at KEK Laboratory in Tsukuba, Japan. The upgrade operations started in 2010 and are going to be finished at the end of 2017. The principal motivation for this substantial

upgrade is to increase the instantaneous luminosity of the machine from $2.1 \cdot 10^{34}$ cm$^{-2}$ s$^{-1}$ (KEKB) to $8 \cdot 10^{35}$ cm$^{-2}$ s$^{-1}$, to reach the statistic needed for the physics goals of the Belle II experimental program. This luminosity increase is obtained by using a larger beam current, and by a smaller beam dimension at the *interaction point* (IP), with the use of the nano-beam scheme, for which is crucial to keep the beam emittance ($\sim$ the phase-space volume of the bunch) as low as possible

SuperKEKB is an $e^+e^-$ asymmetric circular collider, the bunched electron beam with an energy of 7 GeV and the positron beam with an energy of 4 GeV. The center-of-mass energy is $\sqrt{s} \simeq \sqrt{4E_{e^+}E_{e^-}} = 10.58$ GeV.

The electrons are produced in a pre-injector by a pulsed laser directed on a cold cathode target, then they are accelerated by a linear accelerator (*Linac*) to 7 GeV and injected in the *High Energy Ring* (HER) of SuperKEKB. The positrons are produced by the collision of electrons with a tungsten target and then they are injected in a damping ring to reduce their emittance. When the positrons reach the required emittance they are accelerated to 4 GeV in the Linac and injected in *Low Energy Ring* (LER).

The two beams collide at the IP with the particular *nano-beams* geometry discussed in section 2.3.1, one of the major upgrade of SuperKEKB intended to improve the luminosity of the collider. The beam asymmetry produces a Lorentz boost between the frame of the center-of-mass of the colliding leptons and the detector rest frame (i.e. the laboratory frame) equal to:

$$\beta\gamma = \frac{|\mathbf{p}_{e^+} + \mathbf{p}_{e^+}|c}{\sqrt{s}/c^2} \simeq \frac{E_{e^-} - E_{e^+}}{\sqrt{4E_{e^+}E_{e^-}}} \simeq 0.28,$$

equivalent to a mean flight distance for the $B$ mesons of 130 $\mu$m. This distance is sufficient to track the displaced vertex of the $B$ mesons, but is reduced with respect to KEKB (it had a $\beta\gamma \simeq 0.42$). This is due to the exponential increase of power absorption which set a limit to the energy of the HER. On the other hand the beam geometry at IP and bunch shape (i.e. to luminosity of the accelerator), that need a strongly reduced dispersion of the bunches, set a lower limit to the LER energy: to obtain the same $\beta\gamma$ of KEKB the energy of the LER should be reduced to 3.5 GeV, but this implies higher beam losses due to Touscheck scattering, not sustainable for the luminosity requirements.

The luminosity requirement imposed several other modifications to the accelerator structure[27]: the electron injection and positron target are modified, the damping ring, the radiofrequency system, the optics, the beampipe and the vacuum system are renewed. In table 2.2 some relevant accelerator parameters are summarized.

The data taking of the B-Factory is subdivided in three main phases:

- **Phase I**: a preliminary phase without final focus and without collisions, used to evaluate beam background with a radiation detector only (Belle II detector was moved out of the beam line).
- **Phase II**: run with the final focus, but at low luminosity to calibrate and tune the accelerator and the detector response. A fraction of the data may be collected with a center-of-mass tuned on a different bottomominum resonance (currently under study which between $\Upsilon(3S), \Upsilon(5S), \Upsilon(6S)$). During Phase II the vertex detector of Belle II will not be installed completely, but only a sector of VXD will be present, to calibrate the detector and measure the radiation damage, thus the collected data missing the full resolution vertex information.
- **Phase III**: run at full luminosity with complete Belle II detector to perform the physics program of the collaboration. Most of the data are going to be collected at the $\Upsilon(4S)$ resonance.

| | SuperKEKB |
|---|---|
| $E$ (LER/HER) [GeV] | 4.0 / 7.0 |
| $\mathcal{L}_{\text{peak}}$ [$\text{cm}^{-2}\text{s}^{-1}$] | $8 \cdot 10^{35}$ |
| $\mathcal{L}_{\text{int}}^{\text{goal}}$ [$\text{ab}^{-1}$] | 50 |
| $\beta\gamma$ | 0.28 |
| Number of bunches | 2503 |
| Collision frequency [MHz] | 248.778 |
| Orbit lenght [m] | 3016.264 |
| Revolution frequency [kHz] | 99.39199 |
| Beampipe radius at IP [mm] | 10 |

*Table 2.2: SuperKEKB parameters (from [34]).*

### 2.3.1 Nano-beams scheme

The nano-beams scheme was initially designed for the SuperB collider [36] to reduce the beam size at IP and so increase the luminosity. The instantaneous luminosity of a collider is given by:

$$\mathcal{L} = \frac{\gamma_{\pm}}{2er_e}\left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right)\frac{I_{\pm}\xi_{y\pm}}{\beta_{y\pm}^*} \cdot \frac{R_L}{R_{\xi_y}},$$

where $\gamma$ is the relativistic Lorentz factor, $e$ is the absolute value of the electron charge, $r_e = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{m_e c^2}$ is the classical radius of electron, $\sigma_{x,y}^*$ are the width of the bunch at IP on the transversal plane, $I$ is the current of the beam, $\beta_y^*$ is the betatron function at IP ($\sim$ perturbation from the nominal trajectory), $\xi_y$ is the vertical beam-beam parameter, $R_L$ and $R_{\xi_y}$ are the reduction factors of luminosity and the vertical beam-beam parameter due to not-vanishing crossing angle and the $\pm$ sign is referred to the charge of the particles in the beam (with the insertion of the corresponding values $\mathcal{L}_+$ and $\mathcal{L}_-$ result obviously equivalent).

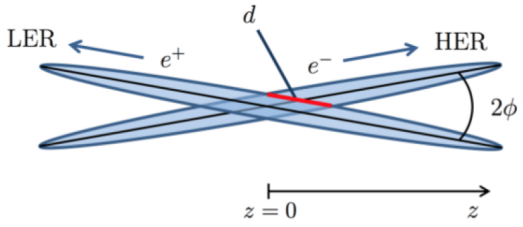



*Figure 2.1: The geometry of the IP in the nano-beams scheme. The effective longitudinal overlap size $d$ is highlighted in red.*

The idea of the nano-beams is to strongly reduce $\beta_y$ function with the minimization of the longitudinal size of the beam overlap at IP (in figure 2.1 the geometry is shown). The dimension of the effective overlap region is $d \simeq \frac{\sigma_x^*}{\phi}$, where $2\phi$ is the crossing angle of the beams. For that reason, the crossing angle has been chosen $2\phi = 83$ mrad (about four times KEKB crossing angle). In addition $\sigma_y^*$ is reduced to the size of tens of nm to reach a beam size at IP of 50 nm (from $\sim 1$ $\mu$m of KEKB). In conclusion with this scheme the betaron function $\beta_y^*$ is reduced by a factor 20 with respect of KEKB and since $\sigma_y^* \ll \sigma_x^*$, $R_L/R_{\xi_y} \simeq 1$, $\xi_y^{\text{SuperKEKB}} \simeq \xi_y^{\text{KEKB}}$, the current of the beams must be doubled to reach the required luminosity.

In table 2.3 are summarized the SuperKEKB IP parameters.

| | KEKB | | SuperKEKB | |
|---|---|---|---|---|
| | LER | HER | LER | HER |
| $E$ [GeV] | 3.5 | 8.0 | 4.0 | 7.0 |
| $\xi_y^*$ | 0.129 | 0.090 | 0.0881 | 0.0807 |
| $\beta_y^*$ [mm] | 5.9 | 5.9 | 0.27 | 0.30 |
| $\varepsilon_x$ [nm] | 18 | 24 | 3.2 | 4.6 |
| $I$ [A] | 1.64 | 1.19 | 3.6 | 2.6 |
| $\sigma_x$ [$\mu$m] | 80 | 80 | 7.75 | 10.2 |
| $\sigma_y$ [nm] | 1900 | 940 | 59 | 59 |
| $\sigma_z$ [mm] | 6 | 6 | 6 | 5 |
| $2\phi$ [mrad] | 22 | | 83 | |

*Table 2.3: SuperKEKB parameters with nano-beams scheme with the comparison with KEKB parameters with crab-crossing scheme (from [34]).*

### 2.3.2 Beam-induced background

A not-negligible background produced by SuperKEKB beam is expected in the Belle II detector. This background is made of particles produced by several physical processes of beam-material or beam-beam interaction, externally from the expected $e^+e^-$ collision at the IP. The background yields have been partially measured during Phase I, and precise predictions are also based on simulated SuperKEKB data.

The first background source is the Touscheck effect, an intra-bunch Coulomb scattering process which deviates the particles' energies from nominal values. The scattered particles propagate around the accelerating ring and finally are lost at the beam pipe inner wall, producing a shower. That shower might produce signals in the detector, if the loss position is close to it. To mitigate this effect, collimators and metal shields are located in the final section of SuperKEKB close to the detector area.

A second background source is the beam-gas scattering, i.e. the interaction between beam particles and residual gas molecules in the beam pipe, with Coulomb scattering or bremsstrahlung. These interactions might deviate the trajectories and energies from nominal values producing effects similar to Touscheck. The countermeasures used for Touscheck background are efficient also for beam-gas background.

Another source of background are the photons from radiative Bhabha scattering, which interact with the SuperKEKB magnets iron producing neutrons by giant photo-nuclear resonance mechanism. The effect is mitigated the over-bent along the ring of the scattered $e^+$ and $e^-$, because of they reduced energies. In addition, a neutron shielding is placed along the ring and close to the detector.

The last background source is the low momentum $e^+e^-$ pairs produced by two-photons QED process $e^+e^- \to e^+e^-e^+e^-$, which might spiral inside the detector.

An additional background source would be the synchrotron radiation of the beam (in particular the HER), which emits photons from few keV to tens of keV. However, the beam pipe shape is designed to avoid synchrotron radiation photons pass through the detector. In addition, the inner surface of beryllium beam pipe is gold-plated to absorb scattered photons. These precautions should completely suppress the synchrotron radiation background.

## 2.4 Belle II detector

Belle II is the detector designed for SuperKEKB collider, and it is a substantial upgrade of the Belle detector. It is a general purpose experiment, optimized to $\Upsilon(4S) \rightarrow B\overline{B}$ events reconstruction, with the capability to perform efficient tracking of charged particles, in particular in the low momentum range (down to 50 MeV/$c$). In addition, is present a neutral particles identification system, based on vertexing and precise $\gamma$ detection, and a multi-detector Particle Identification system. Anyhow an efficient reconstruction in $\tau$ sector and charmonium sector from the continuum production energies is also available. Belle II integrates a high efficiency and low bias hardware and software trigger to cope the high background of the SuperKEKB events, made possible by a low track multiplicity and low detector occupancy combined with an efficient online reconstruction. The high-resolution momentum and vertex reconstruction and the flavour tagging allow to study TDCPV and mixing parameters. In addition, the detector hermeticity and the knowledge of the initial state allow to perform missing mass analysis and use recoil techniques. Because of the low momentum range studied by the experiment a crucial feature of the detector is the material budget, kept as low as possible.

### 2.4.1 Detector overview

The structure of the detector, shown in figure 2.2, is as hermetic as possible with various sub-detector systems placed at various distances from the IP [34]. In the description is used the following coordinate system: the origin is set at IP, $r \in [0, \infty]$ is the distance from IP on the transverse plane, $z \in [-\infty, \infty]$ the distance from IP on the longitudinal plane with positive values for forward region, $\phi \in [-\pi\, \pi]$ is the azimuth angle where the 0 is set in the upward. Another useful variable is $\theta \in [0, 2\pi]$, the polar angle, there 0 is set to forward region. Is some cases a Cartesian system is used, with $z$ is set along the electron beam axis in the forward direction, $y$ upward and $x$ in the right direction to obtain a left-hand system.

The detector has an approximate cylindrical symmetry around the $z$-axis, while it has a significant *forward-backward* asymmetry to improve the solid angle acceptance in the boost direction (forward direction is the boost direction from IP).

From the innermost to the outermost sub-detector system Belle II is composed by:

**Pixel Detector (PXD):** 2 layers of pixel detector (DEPFET technology) which provide 2-dimensional position information. It is used to increase track and vertex resolution.

**Silicon Strip Detector (SVD):** 4 layers of Double Sided Silicon Strip Detector which provide 2-dimensional information. It is used to the tracking tasks for online and offline reconstruction. The SVD and the PXD form the Vertex Detector (VXD) of Belle II.

**Central Drift Chamber (CDC):** helium-ethane wire drift chamber, composed of 56 layers with stereo and longitudinal geometry, to obtain position information. It is used to trigger, tracking and particle ID tasks.

**Particle Identification System (TOP, ARICH):** Time-Of-Propagation counter for barrel region with a Čerenkov quartz radiator, and Ring-Imaging Čerenkov Detecor, with an areogel radiator for endcaps regions.

**Electromagnetic Calorimeter (ECL):** homogeneous calorimeter composed of CsI(T) Crystals that provide 16.1 $X_0$. It is used to measure the energy of photons, electrons and $K$, for trigger tasks and for luminosity measurements.

**Superconducting coil:** NbTi/Cu coil that provides a homogeneous magnetic field of 1.5 T parallel to the beam direction in the internal region. The iron structure of the KLM detector is used to return yoke of the field.

**$K_L$ and $\mu$ system (KLM):** alternated layers of Resistive Plate Chambers and iron plates in barrel regions and scintillator strips in endcap region. It provides 3.9 interaction lenght in the barrel region and it is used to detect $\mu$ and $K_L$ that escape from internal region.

In table 2.4 are reported the technical details of the various sub-detectors of Belle II that are described in next sections.

| Purpose | Name | Component | Configuration | Readout channels | $\theta$ coverage |
|---|---|---|---|---|---|
| Beam pipe | Beryllium | | Cylindrical, inner radius 10 mm, 10 $\mu$m Au, 0.6 mm Be, 1 mm paraffin, 0.4 mm Be | | |
| Tracking | PXD | Silicon Pixel (DEPFET) | Sensor size: 15×(L1 136, L2 170) mm$^2$, Pixel size: 50×(L1a 50, L1b 60, L2a 75, L2b 85) $\mu$m$^2$; two layers at radii: 14, 22 mm | 10M | [17°;150°] |
| | SVD | Silicon Strip | Rectangular and trapezoidal, strip pitch: 50(p)/160(n) - 75(p)/240(n) $\mu$m, with one floating intermediate strip; four layers at radii: 38, 80, 115, 140 mm | 245k | [17°;150°] |
| | CDC | Drift Chamber | 56 wire layers subdivide in 9 superlayers, cell radius: 10 mm, 18.2 mm; | 14k | [17°;150°] |
| Particle ID | TOP | RICH with quartz radiator | 16 segments in $\phi$ at $r \sim 120$ cm, 275 cm long, 2cm thick quartz bars with 4×4 channel MCP PMTs | 8k | [31°;128°] |
| | ARICH | RICH with aerogel radiator | 2×2 cm thick focusing radiators with different $n$, HAPD photodetectors | 78k | [14°;30°] |
| Calorimetry | ECL | CsI(Tl) | Barrel: $r = 125 - 162$cm, end-cap: $z = -102 - +196$cm | 6624 (Barrel), 1152 (FWD), 960 (BWD) | [12.4°;31.4°], [32.2°;128.7°], [130.7°;155.1°] |
| Muon ID | KLM | barrel: RPCs and scintillator strips | 2 layers with scintillator strips and 12 layers with 2 RPCs | $\theta$ 16k, $\phi$ 16k | [40°;129°] |
| | KLM | end-cap: scintillator strips | 12 layers of (7-10)×40 mm$^2$ strips | 17k | [25°;40°], [129°;155°] |

*Table 2.4: Summary of the detector components (from [27]).*

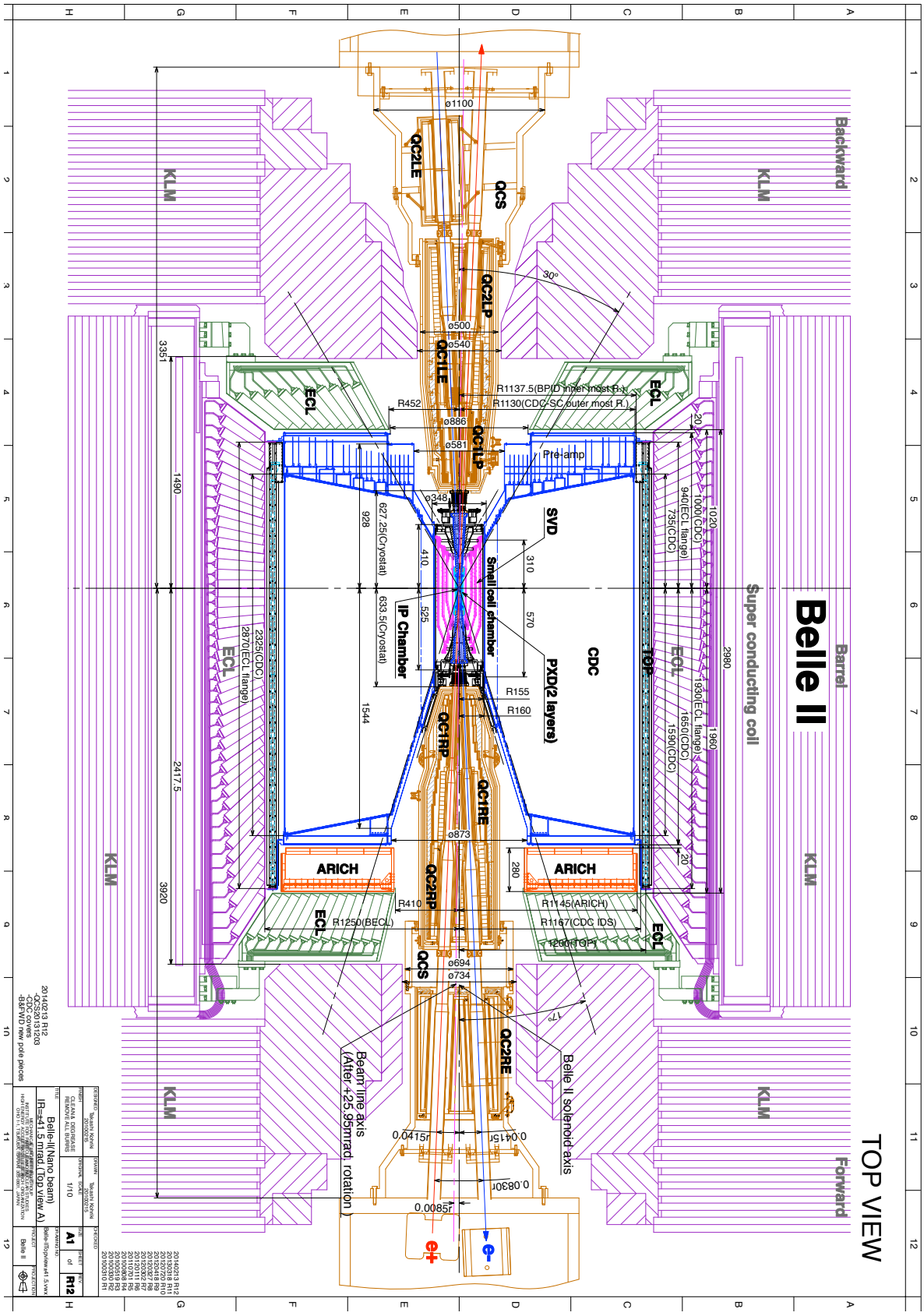


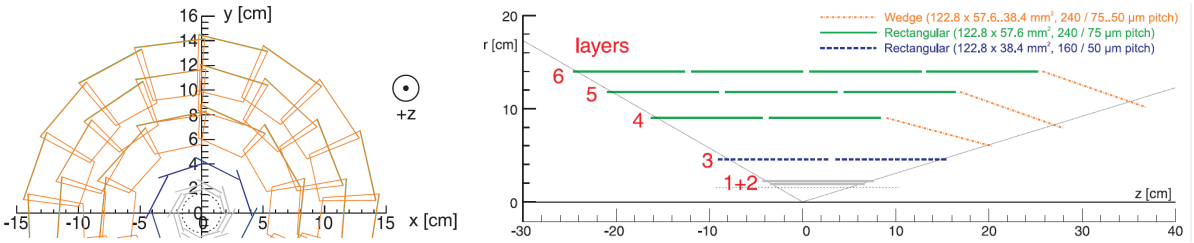

*Figure 2.2: Top view of Belle II, the beampipe at IP and final-focus magnets, all subdetectror are highlighted (from [27]).*

### 2.4.2 Vertex Detector (VXD)

The VXD is the innermost sub-detector of Belle II. It is composed of two devices, the silicon Pixel Detector (PXD) and the Silicon Vertex Detector (SVD), forming a 6-layer silicon vertex detector. It is a completely new device with respect of the SVD of Belle. In figure 2.3 the geometrical structure of VXD is shown. The VXD is described in detail in chapter 3 and here only a summary of the most important features is given.

The PXD is composed of two layers of pixelated sensors with DEPFET technology, to obtain a low material budget in the detection region. The layers are placed at 14 mm and 22 mm from IP while the beampipe radius is about 10 mm. The pixelated sensors have been chosen to sustain the higher hit rate (where a *hit* is the signal of a detector which return position information) due to the shorter distance from IP and the higher luminosity with respect to Belle. This solution allows keeping the occupancy of the detector, defined as the number of activated pixels in the same time over the total number of pixels, at about 3%. The amount of data provided from the PXD in a single event is nonetheless higher than the accepted Data Acquisition (DAQ) rate, therefore SVD+CDC-only tracks are reconstructed online and extrapolated onto the PXD sensors during the High Level Trigger (HLT) process. This extrapolation determines some Regions Of Interest (ROI) from which the PXD hits are selected for readout, allowing the system to remain within the DAQ bandwidth.

The SVD is composed of four layers of Double Sided Silicon Strip Detector, placed at 38 mm, 80 mm, 115 mm, and 140 mm from IP. SVD uses several types of sensors, described in chapter 3, with different shapes and strip pitch. It employs a slanted geometry for the forward sensors, to increase the acceptance reducing the material budget. An original feature of the SVD is the *Origami chip-on-sensor* concept, an innovative solution that uses a flexible fan-out to put all the readout chips on the same side of the modules in the detection region to reduce the connections and simplify the cooling system.




(a) $r - \phi$ *view (half detector)*

(b) $r - z$ *view (half detector)*

*Figure 2.3: Geometrical layer structure of VXD, in side and front view. The SVD is composed by external four layers, instead the PXD by the innermost two layers; different kind of sensor are shown in different colours. The beampipe is showed too with dashed line; dotted lines in side view highlight the angular acceptance of VXD (from* $\theta = 17°$ *to* $\theta = 150°$*). The trapezoidal shapes in front view are the overlapping slanted sensor in the forward region.*

### 2.4.3 Central Drift Chamber (CDC)

The CDC is a wire drift chamber with three main functions in Belle II: first it is the main tracking device to precisely measure momenta, second it provides particle identification (PID) information by measuring energy loss in gas volume, third it is used in hardware and software charged particles trigger.

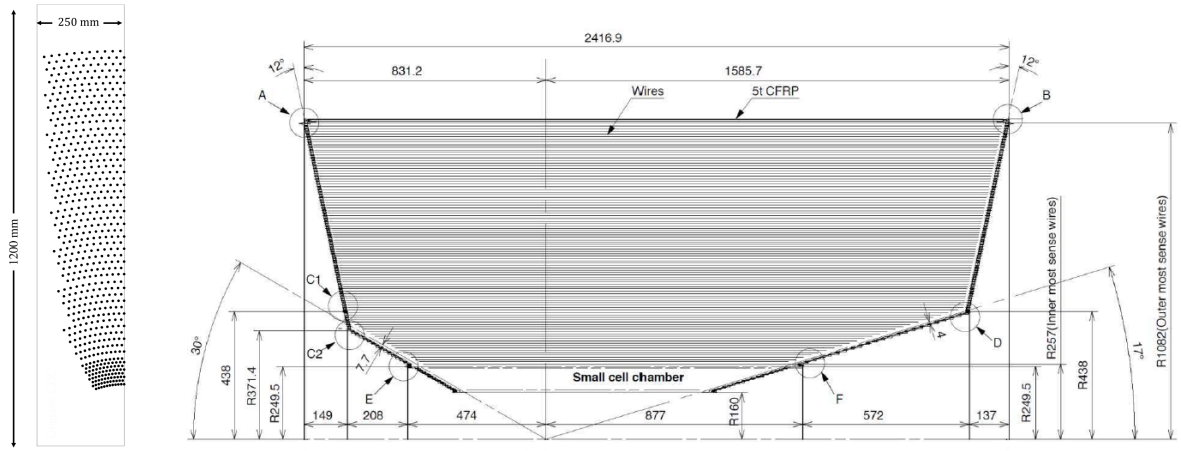
The chamber is composed of 8 superlayers formed by 6 layers of wires each, and an innermost

superlayer formed by 8 layers of wires. The chamber is filled with a mixture of Helium and Ethane (50% He, 50% $C_2H_6$), and the entire CDC is closed by two carbon cylinder and 2 aluminium endplates.

Two classes of wires are present: the field wires producing the accelerating electric field, and the sense wires collecting the released charge. The field wires are composed of aluminium and are thicker (126 $\mu$m of diameter) than the sense wires, which are made of gold-plated tungsten and have a diameter of 30 $\mu$m. The radial cell size is 10 mm for the innermost superlayer and 18.2 mm for the other superlayers. When a charged particle crosses the CDC ionize the gas mixture of the chamber producing $e^-$. The electrons are accelerated by the electric field and produce a charge avalanche that induces a signal on the sense wires, from which is possible to reconstruct the drift time and thus the initial particle position.

The front-end electronics is located near the backward endplate, and it uses an ASIC chip to amplify, shape and discriminate the signal. A TDC is used to measure the drift time and a FADC to measure the signal charge.

To obtain $z$ position information from the CDC half of the superlayers have a *stereo* wire configuration. It means that the wires are not strictly parallel to $z$-axis but present a small angle on $r$-$\phi$ plane. With the use of different inclinations it is possible to reconstruct the 3D position. There are 3 classes of superlayers, one axial ("A") and two stereo ("U", "V"). The configuration of the 9 superlayers is "AUAVAUAVA", chosen to optimize the trigger $z$ resolution.




(a) $r$-$\phi$ *view(sector)*

(b) $r-z$ *view (half detector)*

*Figure 2.4: Geometrical structure of CDC and wire distridibution.*

The innermost radius of CDC is 160 mm, larger than the Belle one (77 mm), because the higher expected background rates would make the chamber unusable at smaller radius. A complex endplate geometry (see figure 2.4) is employed to ensure good angular converge while limiting occupancy, especially from Bahbha scattering in the forward direction. The outermost radius is 1130 mm, larger than Belle one (880 mm) because the PID barrel device of Belle II is more compact than in Belle[1]. The angular acceptance is the same of VXD ($\theta \in [17°, 150°]$) to be able to merge all the tracks of two sub-detectors. The position resolution of CDC is about 100 $\mu$m, while the $dE/dx$ resolution is about 11.9% for particles with $\theta \simeq \pi/2$.

[1]the calorimeter of Belle is the same of Belle II so its inner radius is the upper limit for the available space for inner sub-detectors.

### 2.4.4 Particle identification devices (TOP, ARICH)

The main devices for particle identification in Belle II located outside the CDC are the TOP in the barrel region, the ARICH in the forward endcap region. Both systems detect Čerenkov light, but the operating principles are substantially different, thus they are described separately. The idea of Čerenkov detectors is to measure the $\theta_c$ angle of photons emitted by relativistic charged particles crossing a radiator material, obtaining $\beta$ of the particle with the relation:

$$\cos\theta_c = \frac{1}{n\beta},$$

where $n$ is the refractive index of the material. In Belle II, using the independent momentum measurement in the tracking system and the measurement of $\beta$ combined to the energy loss measured in the CDC the mass of the particles is determined. A PID likelihood is defined with the use of the measured momentum of the track, and various charged particle hypotheses are tested with a "PID selector" that compares the likelihood ratio between various mass hypothesis. The tested particles are $\pi^{\pm}$, $K^{\pm}$, $e^{\pm}$, $\mu^{\pm}$, $p$.

#### 2.4.4.1 Time of Propagation Counter (TOP)

The structure of the TOP is shown in figure 2.5: a single TOP module is made of a quartz bar with a focusing mirror in the forward region and an array of Photomultipliers (PMT) in the backward region.

The operating principle of the TOP detector is to obtain $\theta_c$ from the measurement of the time of arrival of Čerenkov $\gamma$'s from the emission point as a function of the angle of the Čerenkov cone $\Phi$ on the plane of the bar. From the combined information of the arrival time $t_{TOP}$, $\Phi$ and position and direction of the main particle provided from tracking it is possible to extract the $\theta_c$ information. The complete discussion is omitted here, for details see [37].

In the TOP detector the emitted photons are reflected internally in the quartz bar and reach the focusing mirror in the forward region. The mirror is built to preserve the $\Phi$ angle information and to reflect the photons to a specific PMT channel that measures the $t_{TOP}$. Thus the $\Phi$ angle is evaluated depending on the PMT activated channel. The photons emitted in the backward directions are first reflected by a mirror at the end of the quartz bar and directed to the focusing mirror.

The entire TOP detector is made of 16 modules set around the CDC at 1.2 meter of radius from IP, with an angular acceptance of $\theta \in [31°, 128°]$. The bar quartz dimension is 0.45 m $\times$2 cm $\times$2.75 m. The TOP has a single photon time resolution of about 100 ps, achieved with 16 channel micro-channel plate (MCP) PMTs. Instead the production time of the main particle is known with the precision of about 50 ps [27][34].

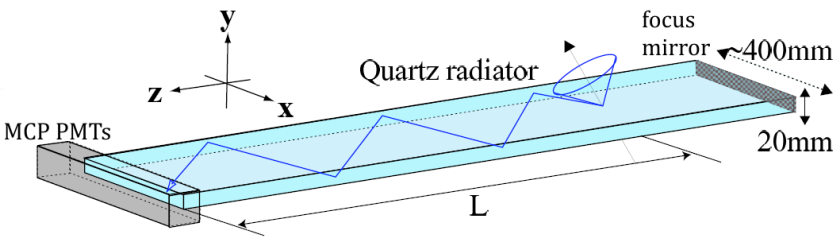



*Figure 2.5: Scheme of a single module of the TOP detector*

#### 2.4.4.2 Areogel Ring-Imaging Čerenkov counter (ARICH)

The ARICH is a proximity focusing Ring-Imaging detector, which uses as radiator a ring of areogel. An expansion volume of 20 cm divides the radiator from a ring of hybrid avalanche photon detectors (HAPD), and allows the Čerenkov photons to enlarge into rings. The structure is shown in figure 2.6.

The performance of RICH detectors depends on the number of detected photons $N_\gamma$ and the single photon resolution on the Čerenkov angle $\sigma_{\theta_c}$. $N_\gamma$ increases with the thickness of the radiator and the resolution per track improves as $\sigma_{\theta_c}/\sqrt{N_\gamma}$, but $\sigma_{\theta_c}$ degrades due to the uncertainty of the emission point. In the Belle II ARICH a peculiar solution is adopted to optimize the performance: two layers of areogel with different refractive indices ($n = 1.045$ upstream, $n = 1.055$ downstream) and 2 cm thickness are used, so that the two produced rings are overlapped on the detection surface, giving the $N_\gamma$ equivalent to a double radiator thickness.

The reached resolution is $\sigma_{\theta_c} \simeq 13$ mrad, optimized for charged tracks with momentum larger than 3.5 GeV/$c$, but the $\sigma_{\theta_c}$ doesn't show significative degradation also for lower momentum tracks. With $N_\gamma \simeq 10$ per ring, the resolution of a single track is about $\sigma_{\theta_c}^{\text{track}} \simeq 3$ mrad. The angular acceptance is $\theta \in [14°, 30°]$.

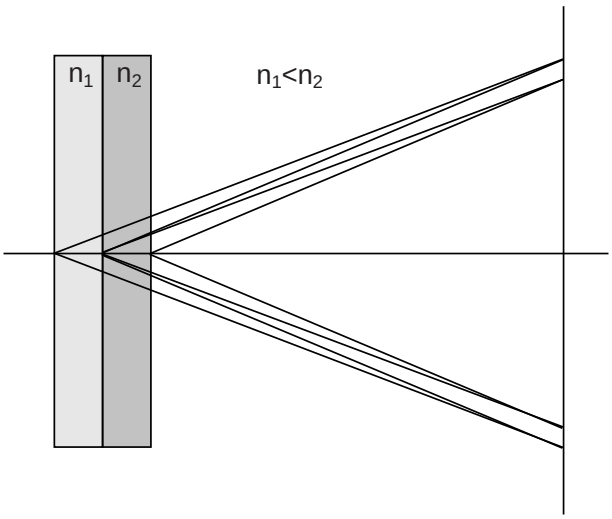



*Figure 2.6: Scheme of the ARICH structure: the two different radiators are shown with different ring expansions on the detection surface.*

### 2.4.5 Electromagnetic Calorimeter (ECL)

The ECL has several central roles in Belle II. First, it detects photons with high efficiency and measures their energy and angle. Secondly, it identifies electron and contributes to the $K_L^0$ detection with KLM sub-detector. Third, it generates a hardware and software photon trigger. In addition, the ECL is used to monitor online and offline the luminosity of SuperKEKB.

The Belle II ECL uses the same crystals of Belle's calorimeter, but a complete upgrade of the readout electronics was needed to cope the SuperKEKB increased luminosity. The calorimeter is subdivided into three regions, the barrel region, the forward and the backward region, and they collectively cover 90% of the solid angle in the center-of-mass system (with an angular acceptance of $\theta = [12.4°, 155.1°]$). The barrel region is extended for 3 meters and has an inner radius of 1.25 m. The annular endcap regions have the internal base at $z = 1.96$ m (forward) and $z = -1.02$

m (backward) from IP. There are two gaps of 1° between the barrel and the endcaps regions to allow the passage of the cables of internal sub-detectors.

The ECL is a homogeneous highly segmented calorimeter, composed by 8736 crystals of CSI(Tl) (caesium iodide thallium-doped). The crystals have a shape of a truncated pyramid with a length of 30 cm and a $6 \times 6$ cm$^2$ base, equivalent to 16.1 radiation lengths ($X_0$). The crystals are assembled in 8 cells separated by 0.5 mm thick aluminium septum walls and closed by two cylinders. Each cell provides the optimal operating environment for the crystals in term of humidity and temperature by a dry air flushing and a water cooling system.

At the external bases of the crystals $10 \times 20$ mm$^2$ Hamamatsu Photonics s2744-08 photodiodes are glued with a 1-mm plexiglass plate collecting light from the scintillating material. Each photodiode has a LED to inject light pulses into the crystal volume to monitor the optical stability.

The relatively long decay time of scintillations in CsI(T) ($1\mu$s), in the presence of elevated background level expected in Belle II, produces a not negligible overlapping of pulses from neighbouring background events. Therefore the new readout electronics samples the photodiodes' signals in 16 points and then fits the signal shape with a predefined proper function[2].

The energy resolution of ECL, from a prototype test, is given by:

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt[4]{E}}\right)^2 + (1.34\%)^2}$$

where $E$ is the energy in GeV. For instance it means $\sigma_E/E(100\text{ MeV}) \simeq 2\%$ and $\sigma_E/E(4\text{ GeV}) \simeq 1.4\%$. In Belle [3] the angular resolution of ECL is $\sigma_\theta \simeq 13$ mrad at low energy and $\sigma_\theta \simeq 3$ mrad at high energies, while the $\pi^0$ mass resolution is 4.5 MeV/$c^2$. Despite the higher background level, because of the new electronics, the performances are expected to be similar in Belle II.

In the forward region only, due to the very higher background level, it is under study the possibility to replace the CSI(T) crystals with pure CsI ones, because the latter are more radiation tolerant and have faster scintillation decay time. On the other hand, the signal is 10 times lower than in the thallium-doped crystals.

### 2.4.6 Magnet

A superconducting coil produces a 1.5 T homogeneous **B** field parallel to the beam direction. The coil is made of NbTi/Cu, and the internal volume is a cylinder of $2r = 3.4$ m and a length of 4.4 m. It operates with a 4400 A current and a liquid helium cryogenic system. The iron structure of KLM provides the return yoke of the magnetic field, therefore in the region of KLM outside the coil the direction of **B** (i.e. the curvature of the tracks) is inverted.

The magnetic field has been mapped with commercial Hall sensors with a precision of 0.1%. In figure 2.7 the field map used in Belle II software is shown. Notice that the inhomogeneities of the field are due mainly to edge effects and to the presence of final focus system magnets of SuperKEKB.

[2] $F(t) = Af(t - t_0)$, where the amplitude $A$ and the signal time $t_0$ are the fit parameters and $f$ a fixed function obtained from previous measurements.

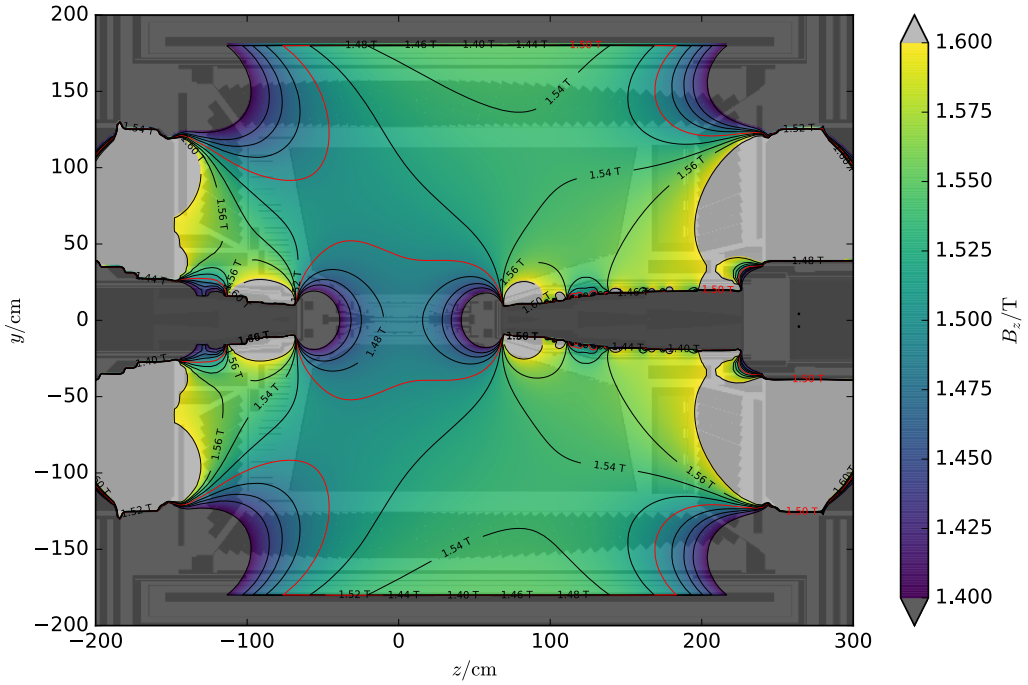



*Figure 2.7: Map of the $B_z$ field used in Belle II software (`basf2`)*

### 2.4.7 Kaons and muons system (KLM)

The KLM is located outside the superconducting coil, and it is build by alternating iron plates and active material detectors. The barrel region covers the polar angle range from $\theta = 45°$ to $\theta = 125°$, and the endcaps extend the range from $\theta = 20°$ to $\theta = 155°$. In the barrel region there are 15 detector layers and 15 iron plates, while in each endcap are present 14 detector layers and 14 iron plates.

The iron plates are 4.7 cm thick each and serve as the magnetic flux return for the superconducting solenoid and providing 3.9 interaction lengths ($\lambda_0$) in addition to the $0.8\lambda_0$ of the ECL, in which $K_L$ can shower hadronically.

The task of the KLM detector is to identify the muon tracks by measuring their penetration depth in the iron and to reconstruct neutral long-lived kaons with the use of the combined information of ECL and the hadronic KLM showers.

The barrel detector layers of KLM are Resistive Plate Chambers (RPC): a proportional gas chamber used in streamer mode with a dielectric plate between the electrodes to prevent the propagation of sparks and so increase the spatial resolution. The signal is read with metallic strips on one side of the chamber. Each KLM module is made of two coupled RPC, with independent power supply and orthogonal strips configuration, this pair of RPC is called *superlayer*. In figure 2.8 the structure of a RPC superlayer is shown.

Both the detector layer and the iron structure of the barrel region are exactly the same one used in Belle experiment, because the events rate results sustainable despite the increased luminosity. Instead, in the endcaps region and in the innermost barrel superlayer of the KLM, the RPCs have been replaced by two orthogonal layers of scintillator strips coupled with silicon photomultiplier (SiPM). Because the RPCs have a too long dead time to sustain the background rate of this region.

The muons are identified starting from CDC tracks: each track is extrapolated to KLM region with a $\pi$ mass hypothesis, and if a KLM hit is present near the extrapolation region it is assigned to that track. A threshold of $p > 0.6$ GeV/$c$ is set, the minimum momentum to reach the KLM innermost layer[3]. Then a *predicted range* is defined as the outermost layer crossed by the extrapolated track, taking care of the interaction between the particle and the material, and an *actual range* by the outermost layer with an associated hit. If the two ranges are substantially different the track is classified as a hadron track and rejected, otherwise the extrapolation restarts with a $\mu$ hypothesis and the track is fitted. In this second case the difference between the predicted and actual range and the quality of the fit are used in a likelihood ratio to test the $\mu$ hypothesis.

The muon detection efficiency plateaus at 89% above 1 GeV/$c$ with a hadron fake rate of about 1.3% due to pions that decay in flight in softer muons.

To reconstructs $K_L^0$ all the KLM hits within a 5° opening angle cone from IP are clustered, then a charged track veto is applied with the use of a CDC track matching. If the remaining neutral KLM clusters are aligned within a cone of 15° with an ECL cluster the two showers are associated. Finally a cluster size cut is applied: each candidate must have at least a hit on two different KLM modules or one hit in ECL and one hit in KLM.

The $K_L^0$ detection efficiency rises linearly from 0 at 0 GeV/$c$ to a 80% plateau at 3 GeV/$c$. The angular resolution is about 3° for KLM-only candidates and 1.5° for KLM+ECL candidates. The SiPMs offer an excellent time resolution of $\sigma_t \simeq 0.7$ ns, that allows to measure also the time of flight of $K_L^0$.

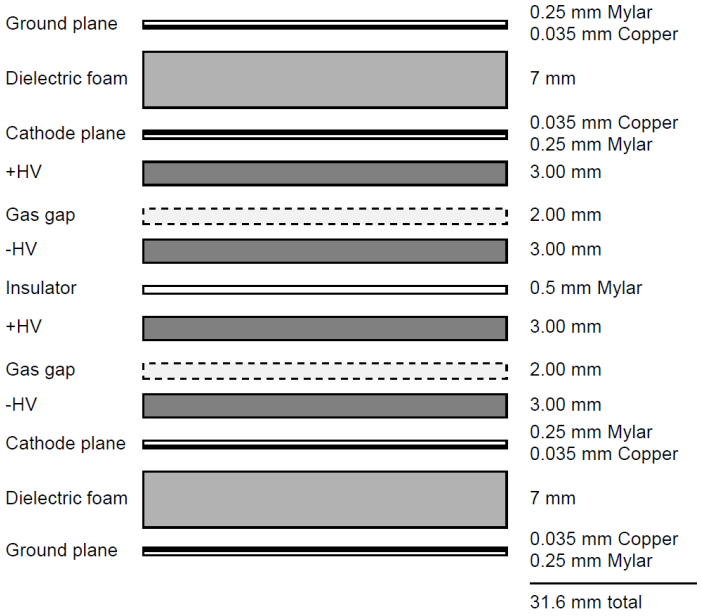



*Figure 2.8: Exploded cross section of an RPC superlayer.*

### 2.4.8 Trigger

The bunch crossing frequency of SuperKEKB is about 250 MHz. Since the bunch crossing time of 4 ns is much faster than the detectors signal collection time for all practical purposes the beam

[3]instead the threshold to escape the KLM is $p = 1.5$ GeV/$c$.

can be considered continuous. Anyway, at full luminosity, the expected event rate is about 50 kHz, and over than 90% of these events are Bhabha scattering or $2\gamma$ QED processes.

A trigger system is therefore required to select events from beam background and identify interesting ones. Despite $B\overline{B}$ events are catheterized by a higher charged track multiplicity with respect to others events, this variable can not be used in the trigger because $\tau$ and low multiplicity events would be discarded too. The required trigger must have instead an efficiency $\sim 100\%$ for $B\overline{B}$ events and a high efficiency for $\tau$ and low multiplicity events too. Some efficiency degradations are allowed to suppress the Bhabha and $2\gamma$ QED backgrounds. The trigger rate must be below 30 kHz, the maximum acquisition frequency of DAQ, and the trigger must provide time information with a precision below 10 ns to exploit the potential of the Belle II sub-detectors. To cope with the high background and to the several physics scenarios the trigger system must be robust and flexible.

The Belle II trigger is subdivided in two main stages: a hardware trigger or *Level 1 trigger* (L1) and a software trigger or *High Level Trigger* (HLT). The first one removes most of the background events with the use of raw information from the faster sub-detectors, the second trigger refine the selection with a more exhaustive analysis.

#### 2.4.8.1 Level 1 Trigger

The L1 trigger analyses the signals from detectors with a FPGA structure that provides a very fast response and allows a configurable system (programmable devices). The L1 trigger uses various signal from sub-detectors at low resolution and implements multiple trigger lines. The main sub-detectors that contribute to L1 trigger are CDC and ECL, although TOP and KLM are also used. A *Global Reconstruction Logic* combines the information of various sub-detector to obtain the total trigger logic.

The CDC providse 3D tracking information and in particular the $z$-position of the main vertex of the events that is a strong discriminant for background events (with vertices far away from the IP). The FPGA takes about 1 $\mu$s to reconstruct the $z$ position of the vertex. The ECL provides a trigger signal based on the total energy released in the calorimeter and a signal based on the number of the isolated shower that use the 3D clustering techniques. With this methods the L1 trigger is able to reject most of Bahbha and $2\gamma$ QED backgrounds without a significant degradation of the efficiency for low multiplicity events. The L1 trigger is also able to identify Bhabha and $2\gamma$ events to monitor the luminosity of SuperKEKB.

The latency of the trigger is about 5 $\mu$s and the output trigger rate is 30 kHz as required, with an efficiency over 99.9% for $B\overline{B}$ and continuum events.

#### 2.4.8.2 High Level Trigger

The HLT uses 6000 commercial CPU core to reduce the 30 kHz event rate from L1 trigger to a storable rate of 10 kHz.

The HLT is subdivided in a *Level 3 trigger* (L3) and in a *Physics trigger*. The L3 trigger uses CDC and ECL information only. It performs some selection based on track multiplicity, vertex position and total energy deposit. These selections reduce the event rate to 15 kHz, than the Physics trigger is applied. This trigger uses the full information of all sub-detectors except the PXD and performs the *fast reconstruction*: the events are completely reconstructed with the same software used in the offline analysis (except the PXD information). Then the trigger applies physics requirements to the reconstructed events and reduces to 10 kHz the event rate.

After the HLT reduction, the tracks reconstructed with the SVD and the CDC in the fast reconstruction are extrapolated to the PXD layers and the ROI are evaluated, then the pixels inside the defined regions are read. In conclusion, the full events are built combining the fast reconstruction events with the PXD data before the definitive storage.

Before the fast reconstruction and the ROI extrapolation, the full PXD data are stored inside a pipeline, which receives in input the PXD data from the events selected by L1 Trigger. The length of this pipeline define the latency constraint of HLT (see section 3.2.2 to have more details about the ROI and pipeline system). The current latency of the HLT is some seconds, under the pipeline constraints.

## 2.5 SuperKEKB and Belle II current status

The accelerator upgrades are still in progress and SuperKEKB is going to be completed by the end of 2017. In regard to Belle II, most of subdetectors have been installed: during 2016 the TOP and the CDC have been installed and the ECL readout electronics has been upgraded, while in 2017 the electronics of KLM, the ARICH and the ECL forward endcap have been installed. In early 2017 the detector was transported to the final in-beam position with the *roll-in*, and during the summer it has been commissioned in a global cosmic ray run. In autumn 2017 a complete sector of VXD (i.e one ladder for each layer of SVD and PXD, the detector used for the testbeam described in chapter 6) is going to be installed in the vertex detector volume. In this configuration, the Phase II commissioning will start with the first SuperKEKB collisions in early 2018. Meanwhile the VXD ladders will be assembled and the entire VXD will be installed and commissioned in summer 2018. At the end of 2018 Phase III and the first run with the complete Belle II detector will start. The SuperKEKB luminosity in 2019 start to ramp up from $10^{34}$ cm$^{-2}$s$^{-1}$ to reach the full luminosity around 2022, as in figure 2.9 is shown.

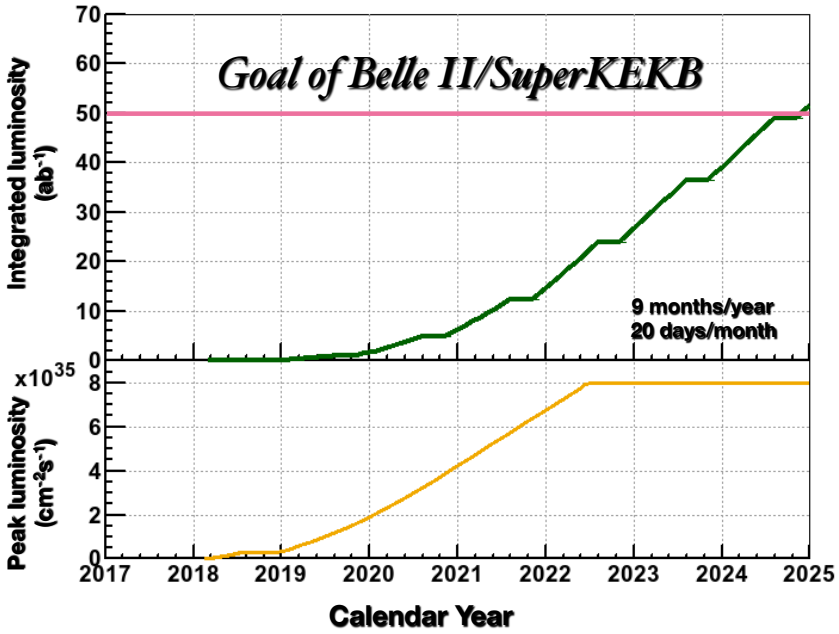



*Figure 2.9: The projected integrated and peak luminosities at SuperKeKB assuming 9 months operation per year, for the first seven operational years. The reduced luminosity in 2018 is due to Phase II operations.*

# Chapter 3

# The Belle II Vertex Detector

The Vertex Detector (VXD) is the innermost sub-detector of Belle II, designed to perform online and offline tracking of charged particles and with a central role in producing the vertex resolution required to perform the Belle II physics program. Four layers of double sided-silicon strip detectors (SVD) and two layers of pixel detectors based on DEPFET technology (PXD) are integrated to realize a 6-layer tracking device. The operating principle of two sub-detectors is different and therefore is described separately in the next section. The rest of the chapter is devoted by a detailed description of the mechanical structure of the detector and to the assembly procedures of the SVD, to which I have participated directly. A global 3D view of the VXD is given in figure 3.1.

The Belle II - Pisa group is involved in SVD detector construction and in particular in the assembly and electrical test of *forward* and *backward* modules of Layer 4, Layer 5 and Layer 6. These modules are assembled and fully qualified in Pisa and then shipped to the ladder production sites: TIFR[1] for Layer 4, HEPHY[2] for Layer 5, IPMU[3] for Layer 6. The R&D

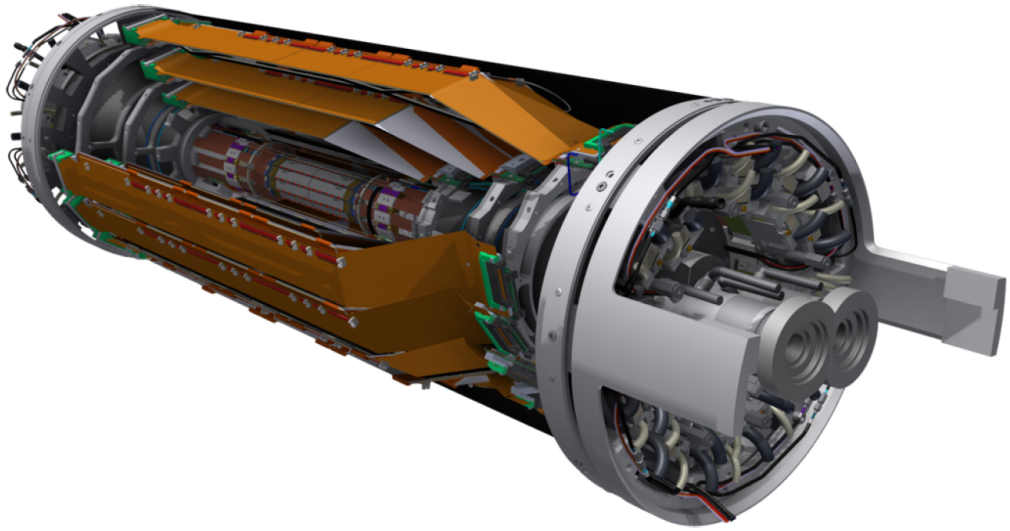


*Figure 3.1: 3D rendering of Belle II VXD in place on the beam pipe. Some ladders are removed to allow to see the internal structure: from the external to the internal part it is possible to see the SVD Layer 6,5,4 with slanted forward modules, then the straight SVD layer 3, and in the central part the PXD layers (only Layer 2 is visible). Around the detector the mechanical structure of the VXD is shown (from [34]).*

[1]Tata Institute of Fundamental Research, Mumbai, India
[2]Institute of High Energy Physics Austrian Academy of Science, Vienna, Austria
[3]Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo, Kashiwa, Japan

activities started in Pisa in 2013 and the actual SVD production started in 2015 and ended in 2017, in parallel to ladder production at assembly sites. In late 2017 and 2018 the final ladder mounting is going to be performed at KEK to assembly the entire VXD.

The modules construction is performed in the clean rooms of the High Technology Laboratories of INFN - Sezione di Pisa and, during the period of this thesis, I participated to the module assembly activities and, in particular performing electrical tests required to determine the module quality and performance.

## 3.1 Operating principle of DEPFET Pixel Detector

The DEPFET (DEPleted Field Effect Transistor) technology was invented in 1987 [38] and several R&D studies have been done in recent years to engineer this concept for X-ray imagers [39] and high energy physics detectors [40]. However Belle II is the first experiment that actually uses a DEPFET-based detector [34].

The DEPFET basic idea is to integrate $p$-channel MOSFET onto a silicon detector substrate. The substrate is fully depleted by a sufficiently high bias voltage. A $p^+$ contact is realized on the back side and a $n^+$ implant, called *internal gate*, is located 1 $\mu$m under the MOSFET channel, as shown in figure 3.2. An incident charged particle produces electron-hole pairs within the depleted bulk and, while the holes drift to the $p^+$ contact, the electrons are accumulated in the internal gate that represents a potential energy minimum for them. When the MOSFET is switched on with inverse polarization, the charge on the internal gate (i.e. the number of electrons) modulates the current $I_{DS}$ in the channel. Therefore it is possible to reveal a particle crossing from the value of the $I_{DS}$ current, which is expected to be about 400 pA per electron.

The readout of the signal is not destructive and can be repeated multiple times, therefore a $n^+$ *clear* contact is present next to the MOSFET gate to remove the signal charge and the thermal electrons from the internal gate. When the clear contact is set to a positive voltage a punch-through channel to the internal gate is created, resetting the internal gate

The amplification of the signal charge, realized with the modulation of the $I_{DS}$ current is just above the position of generation, so that charge transfer before the signal amplification and possible losses are avoided. Another relevant feature of DEPFET sensor is the small capacitance of the internal gate that allows a very low noise performance even at room temperature.

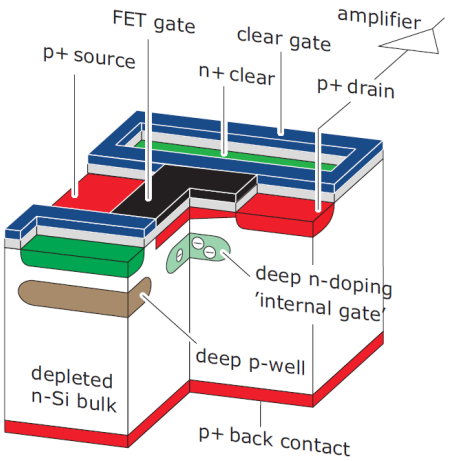


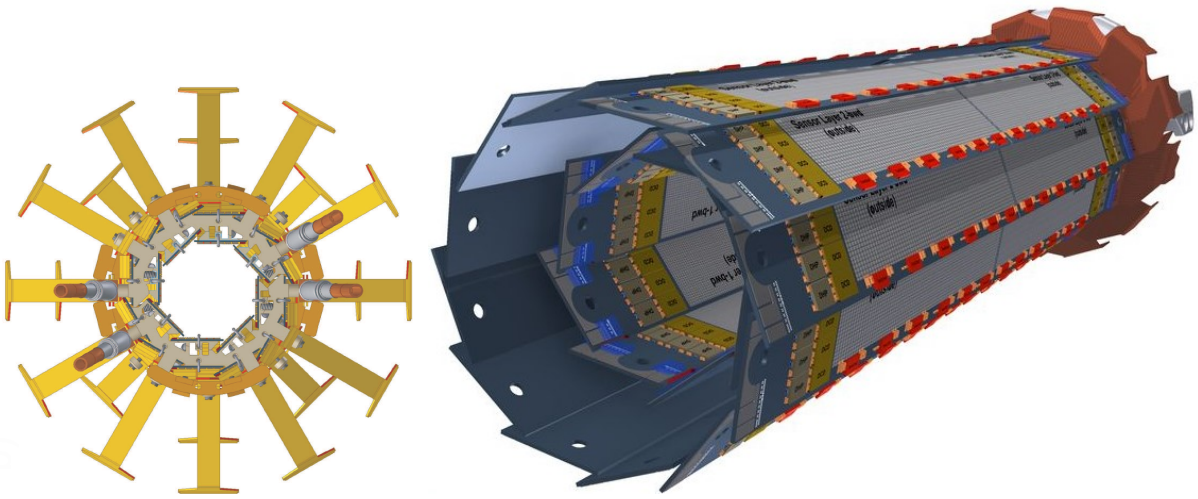

*Figure 3.2: Cross section of a DEPFET pixel, with all sub-structures highlighted (from [34]).*

## 3.2 Pixel Detector (PXD)

To cope with the increased luminosity of SuperKEKB and to have a tracking layer down to 1 cm from IP has been necessary to add pixeled sensor layers, because silicon strip sensors could not sustain, in terms of occupancy, the hit rate expected at this distance from the IP. Instead strip detector layers results the best choice at higher radius. The pixel hybrid sensors used in other experiments, like the LHC ones, cannot be exploited in the Belle II environment because the relatively large amount of material of these sensors would cause excessive multiple scattering for the low energy particles of Belle II (down to 50 MeV/$c$). In these detectors, the readout electronics is bump-bonded to the active sensor, resulting in an overlap of the sensor, electronics and cooling structures inside the acceptance of the experiment.

Therefore the PXD of Belle II has been developed with the use of the monolithic DEPFET technology that allows obtaining a very thin pixel sensor with the readout electronics and the cooling structures placed outside the acceptance region, minimizing the material budget in the region closer to the IP. Although the DEPFET sensors are much less performant than the hybrid pixels in terms of radiation hardness and readout speed, they are well suited for the Belle II environment, in which requirements are less severe than at LHC.

The PXD consists of two layers at 14 and 22 mm from the IP. The innermost layer, Layer 1, is made of 8 ladders, while Layer 2 is made of 12 ladders. Each ladder of both layers is made of two modules, called *half-ladders*. The two half-ladders are symmetric with respect the short edge, they are glued together using a ceramic joint and a self-supporting structure provides the stiffness to the ladder. The angular acceptance is $\theta \in [17°, 150°]$ and the geometry is designed to realize about 1 mm of overlap between adjacent ladders to improve the hermeticity of the detector. In each half-ladder a hole is present to align the module with a clearance of 0.2 mm during ladder mount. More precise alignment is performed after installation with a track based procedure. In figure 3.3 a rendering of the layout is shown.



(a) *Front view* (b) *3D view*

*Figure 3.3: Rendering of PXD layers with support structures and readout electronics. The yellow cable on the front view represents the Kapton Flexes, omitted in the 3D view; in the front view the cooling pipes are shown too (from [34]).*

### 3.2.1 PXD module layout

The DEPFET pixel sensor [41] is a monolithic structure consisting of a matrix of $250 \times 768$ rectangular pixels that realize the sensitive area. In the non-central regions, since the particles hit the detector with higher angles and the charge is distributed on several contiguous pixels, larger pixels are used. The pixel size for Layer 1 is $(50 \div 60) \times 50$ $\mu m^2$, instead for Layer 2 is $(70 \div 85) \times 50$ $\mu m^2$. In conclusion the sensitive area of a single module of Layer 1 is $44.8 \times 12.5$ $mm^2$, instead for Layer 2 is $61.4 \times 12.5$ $mm^2$.

For both layers the thickness of the sensor is 75 $\mu$m but for structural reason the inert silicon support frame is $250 \div 420$ $\mu$m thick. With this configuration the material budget of each layer is $0.2\% X_0$. With these sensors the expected position resolution is about $10 \div 40$ $\mu$m [42]. In table 3.1 the sensor properties for both layers are summarized.

| | Layer 1 | Layer 2 |
|---|---|---|
| Number of ladders | 8 | 12 |
| Distance from IP [mm] | 14 | 22 |
| Sensor thickness [$\mu$m] | 75 | 75 |
| Number of pixels | $250 \times 768$ | $250 \times 768$ |
| Pixel size [$\mu m^2$] | $(50 \div 60) \times 50$ | $(70 \div 85) \times 50$ |
| Sensitive area [$mm^2$] | $44.8 \times 12.5$ | $61.4 \times 12.5$ |
| Material budget [$X_0$] | 0.2% | 0.2% |

*Table 3.1: Module parameters of each layer of PXD (from [34]).*

The readout electronics of each PXD module consists of three ASICs: the *Drain Current Digitizers* (DCD), the *Digital Handling Processor*(DHP) and the *Switchers*. The Switchers are located along the side of the sensor and they have to switch on a pixel row to send the current signal to DCDs and to activate the clean operation to the DEPFETs. Each Switcher has 32 channels (32 gate, 32 clear), and each channel is connected to four rows of pixels to increase the readout speed and use only 6 Switcher per module. The DCDs are loaded out of the acceptance region, and they digitize the currents from the pixels row. The DHPs are set next to DCDs and control it: they apply the zero-suppression i.e. only the pixels with a signal over a certain threshold are processed reducing the data rate from 82 Gbit/s to 5 Gbit/s. The data from each half-ladder are transferred via a flexible Kapton cable to the *Data Handling Hybrid* (DHH), a FPGA for further data processing. The Kapton flex provides the power supply to the entire module. In figure 3.4 the various components of the module are shown.

With this configuration a flushed dry air cooling ($-5$ °C at 1 m/s) is sufficient for the sensor area (i.e to the DEPFETs and Switchers), with no additional material. Instead, a $CO_2$ evaporative cooling ($-20$ °C) is used for the readout electronics region (DCD, DHP), with a copper support structure used as heat sink.

In SuperKEKB the charge losses due to several processes (Touscheck scattering, beam-gas interaction, Bhabha scattering...) requires a continuous replenishing of beam bunches. Therefore each 20 ms two bunches 100 ns apart are injected in each beam. These are called "noisy bunches", and needed 4 ms to cool down. In this period the noisy bunches cause an unsustainable background occupancy in PXD layers. Because of the relatively long integration time of 20 $\mu$s of PXD it is not possible simply to interrupt the readout during noisy bunch crossing, in fact the result would be an "effective" dead time of 20%. Therefore the idea is to make the PXD blind while the noisy bunches cross the detector applying with the Switchers a clear pulse on all the pixel of the sensors, keeping the internal gate charge unchanged. This is called *gated-mode operation*, and it is applied in the 4 cooling ms and it discards about 300 ns of data taking only.

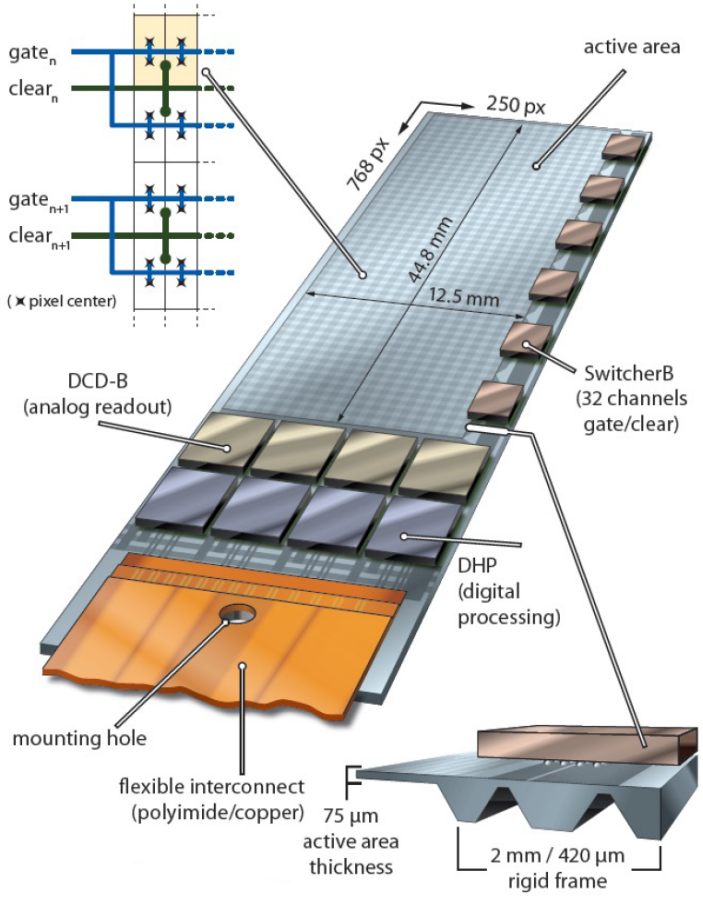



*Figure 3.4: Scheme of a PXD module, with captions for various components. The diagram on the top-left that shows the 4-rows connected to the same switcher channel (from [34]).*

### 3.2.2 PXD Data Reduction and Region of Interest

The readout time of all PXD sensor is 20 $\mu$s with an expected occupancy of 3% and the output rate of the detector is about 20 GB/s, despite the zero-suppression applied at DHP level. In fact most of PXD information collected in this way are background hits. Since it is impossible to determine signal hits from PXD information only, and this amount of data exceeds the limits of global readout and storage, a data reduction system is applied to PXD data [43].

The structure of the system is schematized in figure 3.5. *DHH Controller*s (DHHC) merge the 40 DHH data from each half-ladder in 8 sub-events with the information of 5 DHHs each (2 from Layer 1 and 3 from Layer 2). Then the sub-events are sent to the *ONline SElection Nodes* (ONSEN) system, a FPGA-based system that collects the PXD data and actually applies the data reduction.

Meanwhile the informations from all the sub-detector except the PXD are collected and a first event builder creates a sub-event without PXD information. These sub-events are used in HLT processing to perform trigger selection and extrapolate the tracks from CDC and SVD hits into the Region Of Interest (ROI) on PXD layers. The HLT, depending on the event topology, has a latency up to 5 seconds to evaluate software trigger and ROI evaluation.

To evaluate the ROI another FPGA-based system called *DATa CONcentrator* (DATCON) is also provided: with the use of SVD data only a simplified online reconstruction is performed, based on sector-neighbour finder and fast Hough transformation (see section 4.1 for details on

these methods). The DATCON extrapolates the ROI on PXD layers with a maximum latency of 10 $\mu$s.

The extrapolated ROI from HLT and DATCON are passed to the ONSEN that merges these information in the most conservative way applying a logical OR between HLT and DATCON ROI. The ONSEN also receives general HLT trigger signal, thus if a single event is rejected from the trigger, it will be also discarded in the ONSEN, reducing by a factor 3 the amount of data. After the trigger selection the ROI information is used to read the resulting pixels for each event, reducing of another factor 10 the PXD data per event. After this second selection the data rate remains below 1 GB/s and it can be used to build the complete event and store it.

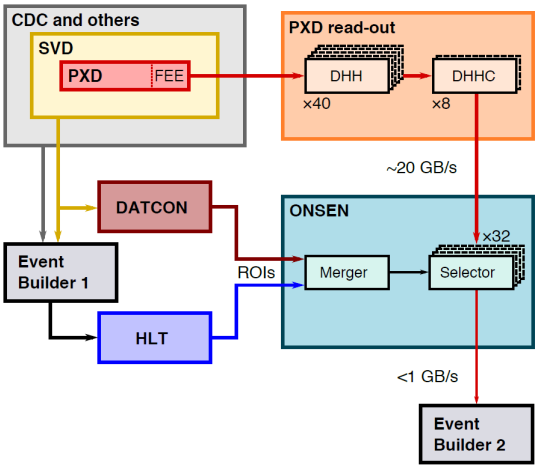



*Figure 3.5: Simplified data reduction scheme of PXD readout, with only the relevant subsystem shown (from [34]).*

## 3.3 Operating principle of a Double Sided Strip Detector

The SVD primary task is to reconstruct efficiently and with high resolution the $B$ vertices and the low momentum tracks. It must be highly segmented and it must have a reduced readout latency to cope with the high rate and to suppress the background hits. On the other hand, because of the typical low momentum of the interesting particles, the material budget must be as low as possible. The Belle II solution to these issues has been the double-sided silicon strip detectors (DSSD).

A charged particle that crosses a silicon layer loses energy for ionization and creates electron-hole ($e$-$h$) pairs are produced. The magnitude of the phenomenon depends on the mean energy released on the particle $\frac{dE}{dx}$, on the energy required to produce the $e$-$h$ pair in the silicon $E_{e\text{-}h} \simeq 3.63$ eV, and of course on the thickness of the crossed layer. Considering a typical sensor thickness of $d_{\text{Si}} = 300$ $\mu$m and a particle at minimum ionization point (MIP) in silicon results $\frac{dE}{dx}\big|_{MIP} \simeq 388$ eV/$\mu$m, therefore:

$$N_{e\text{-}h} = \frac{dE/dx|_{MIP} \cdot d_{\text{Si}}}{E_{e\text{-}h}} \simeq 3.2 \cdot 10^4. \tag{3.1}$$

For any macroscopic sensor dimension, this number is very small compared to the number of intrinsic carriers at room temperature, whose density in silicon at 300 K is $n_i \simeq 1.45 \cdot 10^{10}$ cm$^{-3}$. In fact considering the volume under a single strip (about 100$\mu$m×10cm×300$\mu$m) the resulting number of intrinsic carrier $N_i \simeq 4.3 \cdot 10^7$. Therefore a MIP particle would not be detectable, because the intrinsic noise is 3 order of magnitude larger than the induced signal.

In order to overcome this issue a reverse-biased p-n junction is realized: in Belle II the bulk of the silicon layer is $n$ doped and on one side of the silicon layer a highly doped $p$ implant is realized. The resulting depletion region is increased with the application of a bias voltage $V_{\text{bias}}$ that produce an inverse polarization on the p-n junction. Inside the depleted region the intrinsic carriers are removed thus the $e$-$h$ pairs produced by charged particles are detectable, and thanks to the electric field produced in the junction, they can drift to the edge of the depleted region, where the charge collection electrodes are located.

The DSSD sensor is produced segmenting the $p^+$ implant in strips, realizing on the other side of the silicon layer a $n^+$ implant segmented in strips orthogonal to the $p^+$ ones. Then on top of each strip (on both sides) an aluminium strip electrode capacitively coupled with the implant strip is applied. With this configuration, because of inverse polarization, the $e^-$'s drift to the $n^+$ strips while the $h^+$'s drift to the $p^+$ strips and they induce a signal on the nearest electrodes, from which is readout.

The capacitive coupling is realized typically with an oxide layer between the metallic strip and the doped silicon one. Another expedient is the realization of a $p$ doped implants between the $n^+$ strips (called $p$-stop), to avoid electrical conduction between adjacent strips.

The DSSD layout and a scheme of its operating principle is shown in figure 3.6.

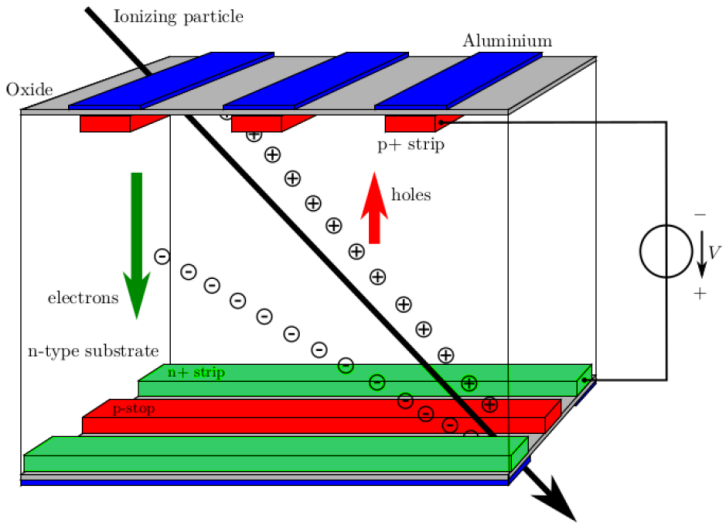



*Figure 3.6: A cross section of DSSD with the operating principle illustrated. The $V_{bias}$ depletes the n bulk thus the e-h pairs are free to drift to the $p^+$ and $n^+$ implants inducing a signal on aluminium strips (from [34]).*

The use of orthogonal strips with a spacing of tens of $\mu$m allows to obtain precise 2-dimensional information on the position of the particle crossing, with a resolution of $d_{\text{pitch}}/\sqrt{12}$, which is known as *digital resolution*. However usually the signal is not induced on a single strip but it is distributed on several adjacent strips, therefore the position of the crossing particle is evaluated from the average between various strip position $x_i$ weighted on signal amplitude $S_i$:

$$\overline{x} = \frac{\sum_i x_i S_i}{\sum_i S_i}.$$

Anyway with the use of this clustering procedure the signal to noise ratio of the detector contribute to the spatial resolution.

Despite the excellent performance that the DSSD detectors provide, their lower cost and readout simplicity with respect to pixel detectors, they suffer of an issue that complicates the pattern

recognition steps. In fact if multiple particles cross the same sensor at the same time, multiple strips in different sensor positions are activated on both sides of the sensor: for $N$ crossing particles $N^2$ strips combination are possible to evaluate the 2D position of each particle i.e. there are $N^2 - N$ *ghost hits* on the sensor. The readout time of a typical DSSD is tens of ns, therefore it is not possible to use the time information to discriminate from the various combination in the same sensor. In conclusion, the ghost hits are limited keeping the occupancy low and then they are removed with the use of multiple layers information inside the track reconstruction. Another possibility being explored is to use very precise measurement of the released energy: for the same particle it should be the exactly the same for both $n$-strip and $p$-strip, and this could allow matching correctly the strip pairs, although it is not clear whether the energy resolution will be sufficient.

## 3.4 Silicon Vertex Detector (SVD)

The SVD of Belle II must satisfy several requirements to perform efficiently its tasks, the main ones are the following:

- an inner radius of 38 mm and an outer radius of 140, to fit the volume between PXD and CDC, and an acceptance covering of $\theta \in [17°, 150°]$.
- efficient track reconstruction, with possibility of extrapolation from CDC to PXD to reconstruct $B$, $D$ and $\tau$ vertices. Efficient stand-alone tracking for low $p_\perp$ tracks is required too (i.e. very low material budget).
- efficient reconstruction $K^0_S$ that decay outside PXD volume (several relevant physics channels involve these decays)
- occupancy lower than 10% to cope with the required vertex efficiency and to be able to sustain the high beam background and trigger rate. Therefore particular geometry and fast readout electronics are required too.

The chosen layout is made of four layers of DSSD. The sensors are equipped with APV25 chips for the readout of the signals, which are compatible with required timing, radiation hardness and trigger tasks. More details about the DSSD and the APV25 is given in section 3.4.1.

Three kinds of DSSD sensors are used in the SVD: two rectangular (*large* rectangular and *small* rectangular from now) and one trapezoidal (*wedge* sensor from now). The layers[4] from 3 to 6 are set at 39 mm, 80 mm, 104 mm and 135 mm from IP, and they are composed respectively of 7, 10, 12, 16 ladders. Each ladder of Layer 3 contains 2 small rectangular sensors. Instead each ladder of Layer 4, 5 and 6 contains one slanted wedge sensor in the forward region and respectively 2,3 and 4 large rectangular sensors. Table 3.2 summarizes the layout parameters

The ladders are glued into aluminium end mounts, and they are screwed on support rings on both ends of SVD, out of its acceptance. These rings are glued onto two carbon fibre cones, mounted to the end flanges. The flanges are connected with a carbon 1 mm thick fibre shell which encloses the whole VXD. In figure 3.7 a cross section of SVD is shown.

For each layer the adjacent ladders present an overlap region (about $8 \div 10\%$ of the area, depending on the layer). The overlap between layers serves to increase the hermeticity of the detector and to facilitate alignment, although it slightly increases the material budget. In fact the short

[4] the layer numeration continue from PXD to avoid ambiguities, thus the SVD Layer 3 is the innermost and Layer 6 the outermost.

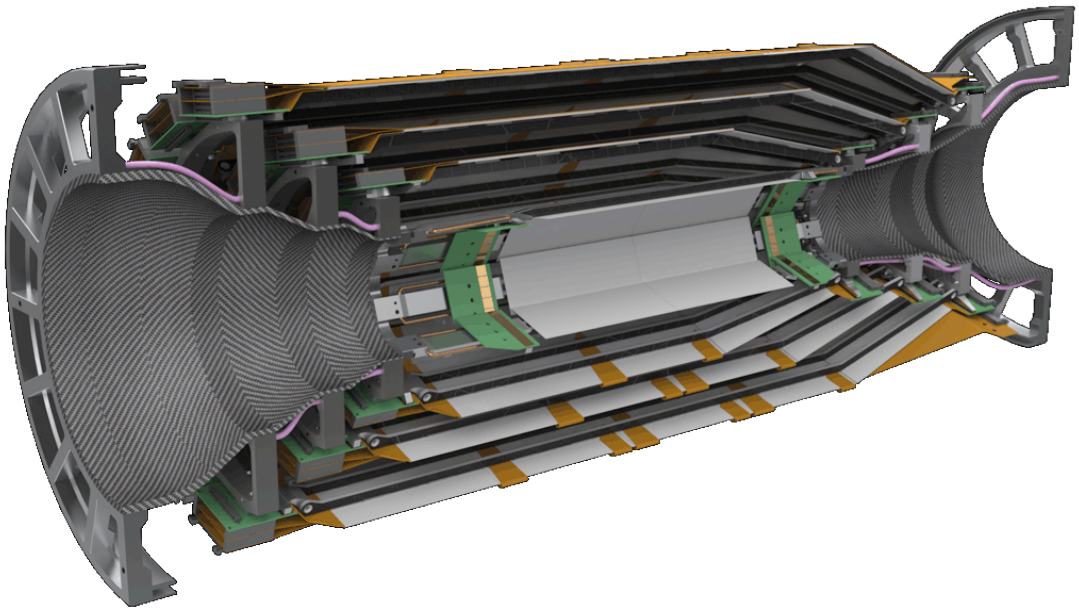


*Figure 3.7: 3D rendering of SVD cross section, with the complete layout of pitch adaptor and readout electronics. Mechanical support and carbon fibre cones are shown too (from [34]).*.

| | Layer 3 | Layer 4 | Layer 5 | Layer 6 |
|---|---|---|---|---|
| Distance from IP [mm] | 39 | 80 | 104 | 135 |
| Number of ladder | 7 | 10 | 12 | 16 |
| Number of sensor per ladder | 2 | 3 | 4 | 5 |
| Kind of sensors | 2 small | wedge + 2 large | wedge + 3 large | wedge + 4 large |
| Slant angle | 0 | 11.9° | 17.2° | 21.1° |
| Material budget [$X_0$] | 0.6% | 0.6% | 0.6% | 0.6% |
| Occupancy | 6.7% | 2.7% | 1.3% | 0.9% |

*Table 3.2: General SVD layout properties (from [34]).*

arm between hits in different sensors of overlap region reduces the uncertainty due to Coulomb multiple scattering in the $\Upsilon(4S)$ tracks used for the alignment of the detector.

The lack of symmetry of the detector on $r-\phi$ plane for the ladder per layer number, due to sensor dimensions, allows a not trivial degree of freedom in layer relative rotation. Has been chosen to have all the overlap aligned in $y$ direction (upward) to facilitate the assembly procedures.

All the sensors have the long strips parallel to the beam axis ($z$ direction) on $p$-side, in the inward sensor face. Instead they have the short strips along the $r-\phi$ direction on the $n$-side, in the outward sensor face. This is almost true also for slanted sensors, but of course the $p$-side is not parallel to the beam axis anymore. This orientation choice is due to minimize the spread of carrier during the drift due to the magnetic field interaction (Lorentz angle different for hole and electrons because of different mobility).

The slanted forward sensors allow increasing the angular acceptance of the SVD with a reduced material budget in the forward region. On the other hand, this particular geometry represents a considerable challenge for ladder mount and its stability and complicate the track finding procedures. However, the BaBar SVT already employed slanted sensors and demonstrated the possibility to realize efficiently the detector and the dedicated tracking software.

### 3.4.1 SVD sensors and readout electronics

The SVD sensors are DSSD with 768 strips on $p$-side and 512 strips on $n-$side, except for Layer 3 sensors (i.e. the small rectangular sensors) which have 768 strips on both sides. The pitch of the strips (i.e the space between consecutive strips) is different for various kind of sensors and sides, and varies from 50 $\mu$m and 240 $\mu$m. The large rectangular and the wedge sensors are read from 6 chips on $p-$side and from 4 chips on $n-$side, instead the small rectangular sensors are read from 6 chips on both sides. All the sensors have a nominal thickness of 300 $\mu$m. The complete geometrical information of the sensors are present in table 3.3.

| | Large rectangular | Small rectangular | Wedge |
|---|---|---|---|
| Area [mm$^2$] | $122.8 \times 57.6$ | $122.8 \times 38.4$ | $122.8 \times 38.4 \div 57.6$ |
| $p$-pitch [$\mu$m] | 75 | 50 | $50 \div 75$ |
| $n$-pitch [$\mu$m] | 240 | 160 | 240 |
| Number of $p$-strips | 768 | 768 | 768 |
| Number of $n$-strips | 512 | 768 | 512 |
| Number of APV25 ($p$-side) | 6 | 6 | 6 |
| Number of APV25 ($n$-side) | 4 | 6 | 4 |

*Table 3.3: Comparison between various kind of sensor parameters (from [34]).*

In the SVD sensors, the DSSD technology is engineered with the following features. The bias voltage is provided by a bias ring that surrounds the active region. It is connected to the implants through an array of bias resistances integrated on the sensors so that each strip implant is connected to a separate $R_{\text{bias}}$ of about 10 M$\Omega$. The capacitive coupling between the aluminium electrodes and the $p^+$ or $n^+$ implants is realized with a $SiO_2$ layer, while the aluminium electrodes are connected to AC pads for readout electronics connections. Additional DC pads directly connected to the implants are present, one per strip, to perform electrical test and characterization of the sensor. Between the strips of the sensors addition floating strips i.e. strips not connected to readout electrodes are introduced to increase the position resolution of the detector. These floating strips transfer the collected charge by a capacitive coupling to the readout strips, opening the possibility to perform an interpolation to improve the spatial resolution [44].

Between the bias ring and the sensor edge, a structure called *guard ring* is realized: several concentric structures that protect the sensor from breakdown effects i.e. from electron avalanche due to edge effects on the electric field. The guard ring makes the electric field reduce in a controlled way toward the edge of the sensor, preventing breakdown effects.

The APV25 chips [45] were originally developed for CMS silicon detector, but their optimal features are also excellent for the SVD readout. The shaping time is about 50 ns, and each chip has 128 channels with low noise preamplifiers. It has an internal analog pipeline compatible with Belle II trigger DAQ scheme and the chips are radiation hard up to an integrated dose higher than 30 MRad, sufficient for Belle II purposes.

### 3.4.2 SVD Ladders

The ladders layout is different for each layer, but the mechanical support structure is conceptually the same: the sensors are glued between a light-weight 1 mm thick Ariex foam layer to ensure electrical and thermal isolation, and two carbon fibre ribs that provide the support structure. The ribs are glued onto aluminium end mounts that allow mounting on the global support structure.

The Layer 3 ladder consists of two small rectangular sensors with readout chips set at both ends of the ladder on a *hybrid sandwich*, out of acceptance region. The chips are connected to the

sensor by a Kapton single layer flexible pitch adapter.

The ladders of Layer 4, 5 and 6 are composed of three sub-assemblies each, called *forward*, *backward* and *origami* subassembly. The backward and the forward subassembly are composed of a single sensor, respectively a large rectangular or a wedge sensor, and of a hybrid readout sandwich. The hybrid sandwiches are, like in Layer 3 case, out of SVD acceptance region and they are connected to the sensor by a single layer Kapton flexible pitch adapter.

The origami subassembly is composed of 1, 2 or 3 DSSDs respectively for Layer 4, 5 and 6. The readout of these sensors is realized with the *Origami chip-on-sensor concept* [46]. The idea is to use a single 3-layer flexible circuit attached on the top of the sensor to power the APV25 and connect their signal and control lines to the outside, while flexible single layer circuits are used to bring the signals from both detector sides to the APV25. There are three different Origami circuit design, electrically equivalent, different only in the length due to the different position: *-Z* (backward), *CE* (central), *+Z* (forward). The ladders of Layer 6 have all three Origami circuits, Layer 5 ladders have -Z and CE ones, and Layer 4 ladders only the -Z one. Figure 3.8 shows the DSSD and Origami circuits configuration for all the layers.

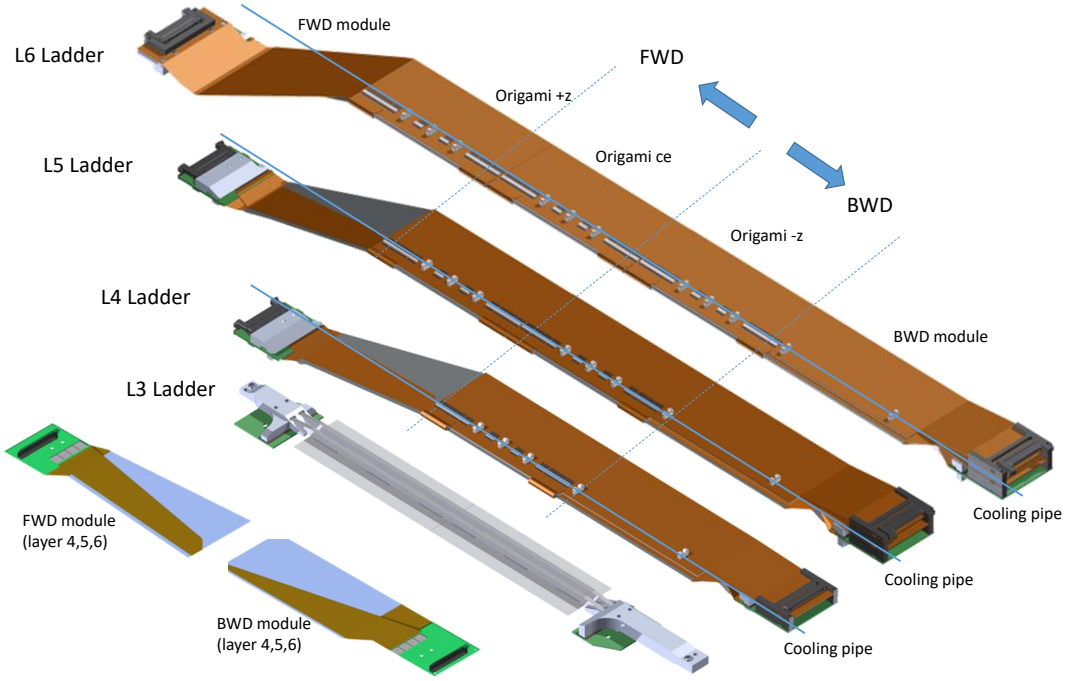
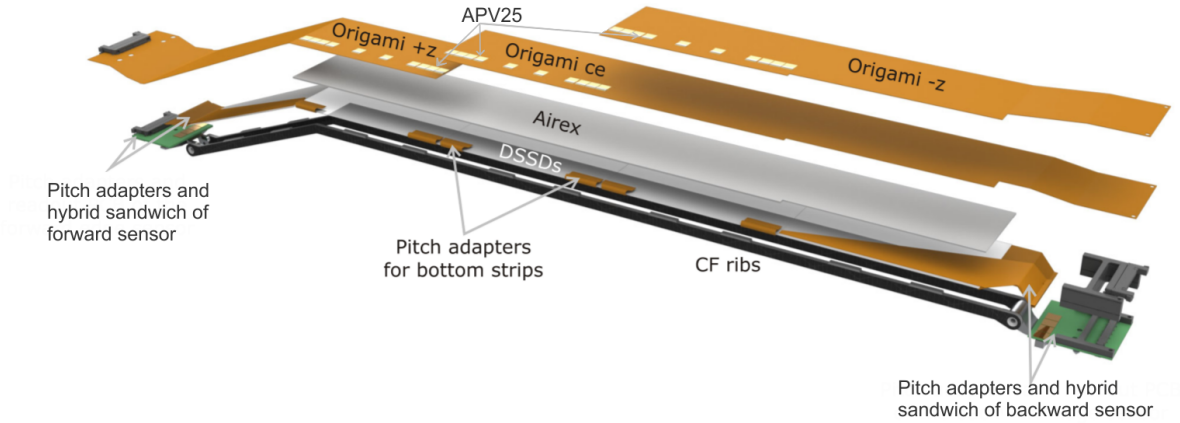


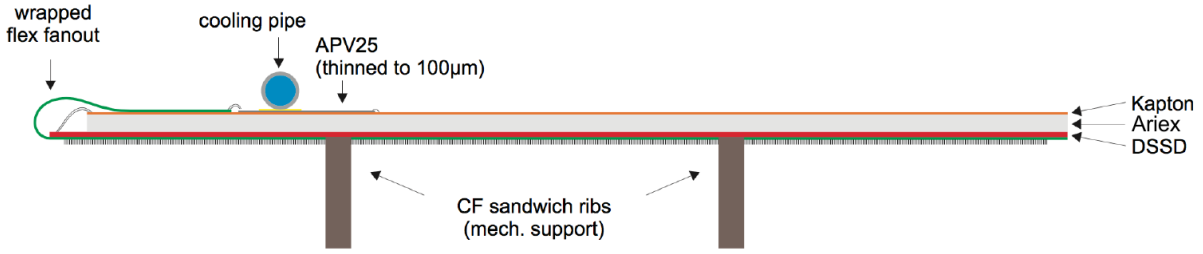

*Figure 3.8: SVD ladders layout for the various layers. The Origami and sensors subdivision and the cooling pipe is higlighted. The forward and backward modules are also shown separately (from Belle II SVD group studies).*

Each design has the same structure: the APV25s are attached on the top of the Origami flexes with a conductive adhesive and then the strips on the *n*-side (outward) are directly connected to the chips by wire bonds and a Kapton flexible pitch adapter. The strips on the *p*-side (inward) are connected to the APV25s by a flexible single layer pitch adapter bent around the edge of the sensor. Then another flexible pitch adapter moves outside the acceptance region the APV25 output signals. With this design all the chips are located on the same side on the sensor, therefore they can be aligned and efficiently cooled with a single cooling pipe for each ladder (which uses the same evaporative $CO_2$ system of PXD ladders). Furthermore, the Origami design, due to the short distance between chips and sensor, ensures low noise and due the thin design of the chips (100 $\mu$m thick) keeps the material budget of each ladder below 0.6% $X_0$.

In figure 3.9 the detailed layout of Origami circuits is shown.




(a) *3D exploded view*




(b) *side view (r-$\phi$ plane)*

*Figure 3.9: An example of Origami chip-on-sensor concept in Layer 6 implementation. The relevant components are highlighted (from [34]).*

The ladders are assembled starting from the three subassemblies, realized and electrically tested previously (see section 3.5 for details about forward and backward subassembly), then the subassemblies are aligned and glued onto the ribs and finally the cooling pipe is attached on the ladder. After the assembly, the ladders are electrically tested. Then the assembly of the SVD, called *ladder mount*, will be performed: the ladders of the four layers will be mounted together on the support structures to realize *half-SVD*. The two halves will be realized separately to allow the insertion of PXD on the cooling pipe, and they will have separate cooling pipes. After the PXD mount, the two half-SVD will be installed around it and the entire system will be cooled and electrically tested. It is foreseen to perform a full commissioning a calibration of the system with cosmic rays before is installed inside Belle II.

## 3.5 SVD forward and backward subassembly: Pisa group activity

The Belle II - Pisa group has the responsibility of the assembly and the electrical test of the forward (FW) and backward (BW) subassemblies i.e. the modules at the edges of Layer 4, Layer 5 and Layer 6 of the SVD. The subassembly production started in 2015 and was officially closed and finished in September 2017 including the procedure of spare modules in case of damage or malfunction of the existing ones.

The components of the forward and backward subassembly are the DSSD sensor, the hybrid boards and the pitch adapter:

**Sensor.** The DSSD sensor with large rectangular (BW) or wedge (FW) design. They are fabricated at HPK (Hamamatsu Photoncs) and Micron Semiconductor Ltd. and arrive at the Pisa Laboratoy already fully qualified and tested.

**Hybrid.** The board where 4 APV25 ($n$-side) or 6 APV25 ($p$-side) are allocated with the connector to external readout. The chips are connected to the board through 25 $\mu$m Al-Si micro-bonds, realized with ultrasonic bonding. The single-side hybrids arrive at Pisa Laboratories already assembled, bonded e and electrically tested.

**Pitch Adapter (PA).** The Kapton flex that allows the connection between the sensor pads and the chips. Four PA designs are used, two for FW subassemblies ($p$ and $n$-side) and two BW subassemblies ($p$ and $n$-side). The PAs arrive at Pisa Laboratories already built as a single component and electrically tested.

### 3.5.1 Modules assembly procedure

The purpose of the assembly procedure, or *gluing*, is to realize the FW or BW subassemblies gluing together the components described previously. All the components are very delicate and adequate measures must be taken to guarantee proper handling throughout the procedures.

In a preliminary phase, two single-side hybrids, one $n$-side and one $p$-side, are electrically tested[5] and then glued together. The resulting component is called a *hybrid sandwich* (HS), and is retested electrically before the starting of the subassembly gluing procedure.

In the module assembly procedure, the sensor and the HS are glued with the PA, to allow a connection between the sensor strips and the APV25s channels, which will be subsequently realized with the microbonding technique. The connection must be realized on both sides of the HS, therefore one PA must be glued on each side of the sensor. The procedure is conceptually identical for FW and BW subassemblies, although the geometries are different.

The first step of the procedure is microscope optical inspection of each component to identify possible scratches or defects and to check the quality of the hybrid micro-bonding. Then the HS and the sensor are positioned on the gluing jig (figure 3.10(a)), a metallic structure dedicated to the positioning of the components. During the gluing procedures, four jigs (FB and BW, $n$ and $p$-side) are used. They are equipped with micrometer screws to set the 3-dimensional position of the component and with a vacuum system to keep fixed the components position. Then the jig position is aligned with the reference frame of the Coordinate Measurement Machine (CMM), an optical machine which provides a resolution of 5 $\mu$m. Then the components are aligned to each other using reference crosses.

After the alignment of the sensor and the HS, the PA is positioned by the use of pins to a dedicated chuck with a vacuum system to fix the PA position (figure 3.10(b)). Then the PA chuck is positioned on the top of the gluing jig and the three components are aligned each other with the use of the CMM and the micrometers screws. The maximum allowed distance between PA and chip pads is 100 $\mu$m, instead the one between PA and sensor pads is 150 $\mu$m, to allow safe wire bonds between the PA and the pads. A delicate step is the planarity alignment of the components because it directly affects the gluing quality, therefore the tolerance is lowered to 30 $\mu$m.

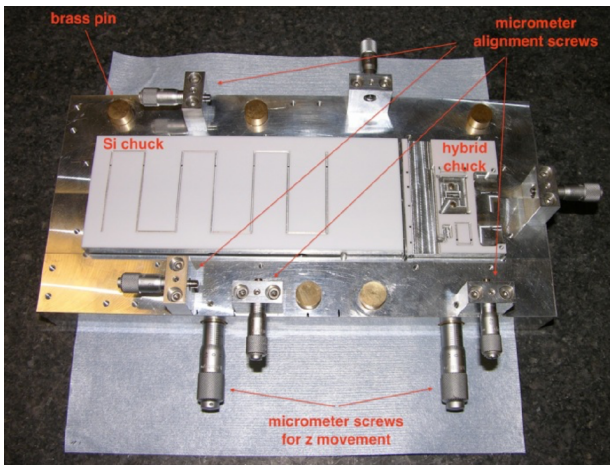
In the next step the PA chuck is moved under the robot dedicated to the glue deposition. For this step it is fundamental that environmental parameters, like temperature, pressure and humidity, are well controlled, because they affect the glue properties like viscosity. The clean room where

[5] all the electrical tests are described in section 3.5.2.

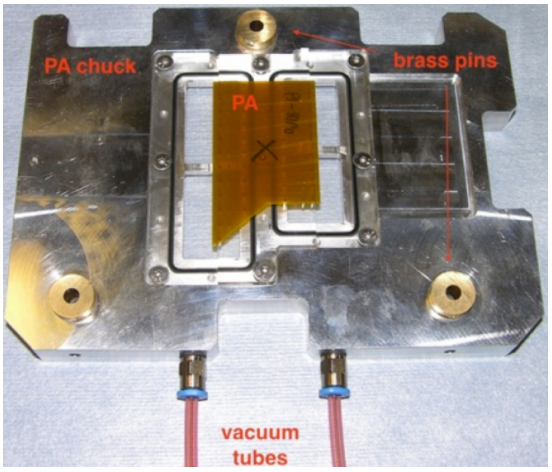
the entire procedure is performed guarantees a stable environment for the gluing deposition and the connected operations. A syringe moved by the robot is filled with the glue, and its ejection is managed with a compressed air system. The PA chuck is aligned with the robot frame (by the use of pins and micrometer screws) and then the glue is deposited by the robot on the edge of the PA (figure 3.10(c)).

After the glue deposition, the PA chuck is superimposed again on the gluing jig (figure 3.10(d)). The alignment system allows to keep and check the previous aligned position. Then the HS is moved up until the glue touches the chips pads. The CMM allows to check that the glue overflows the entire edge of the PA and the pads without covering any sensitive areas (figure 3.11). Then the operation is repeated for the sensor. The glue has a long cure time, about 15 hours, to allow to perform the gluing operation without relevant modification of glue properties during the procedure.

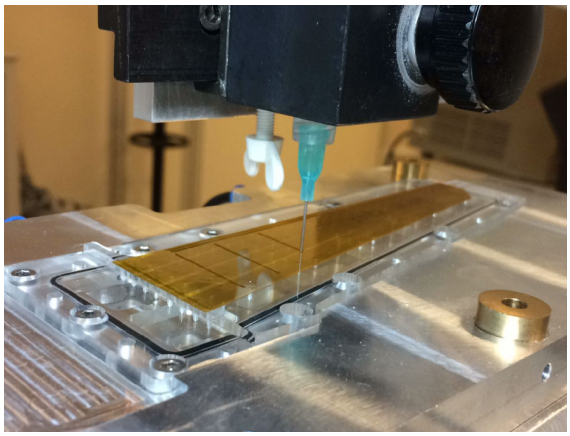
When the glue has cured the module is flipped to perform the gluing of the PA of the other side, with the same procedure described above. After the cure of this second glue deposition the subassembly is optical inspected to check the quality of the assembly. Then the wire bonding between the pads and the PA is performed by an automatic machine and the module can be



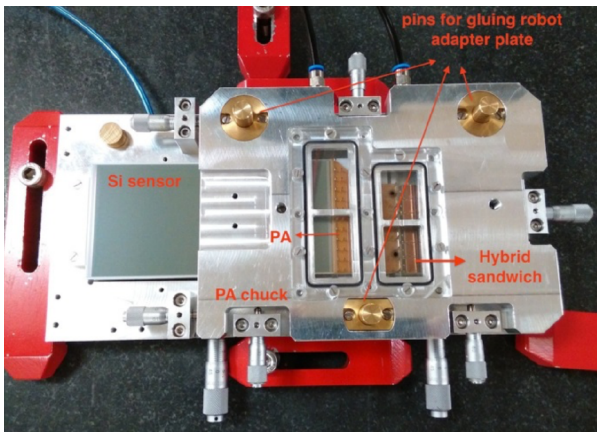

(a) *empty gluing jig of BW modules*




(b) *p-side PA on the dedicated chuck*



(c) *n−side PA and the robot arm with the glue syringe during glue deposition*




(d) *BW module on the dedicated jig covered by the PA chuck in the gluing step*

*Figure 3.10: Some aspects of the gluing procedure (from [47]).*

electrically tested. In figure 3.12 the result of the module assembly is shown.

The main challenges during this assembly procedure are maintaining the alignment and the mechanical precision throughout the procedure and ensuring that the glue is properly distributed under the bonding pads without overflowing onto the pads. In addition, the components must be manipulated and assembled avoiding damage and contaminations.

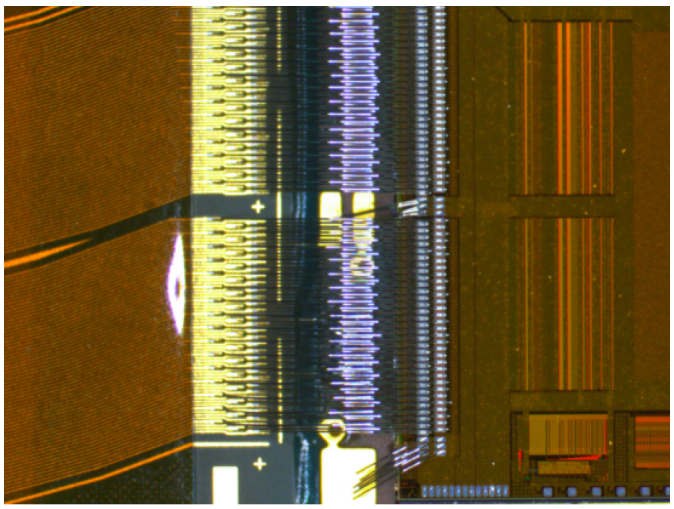


*Figure 3.11: picture of the micro-bond between PA and the chip pads. The glue is the blue halo under the wires (from [47]).*

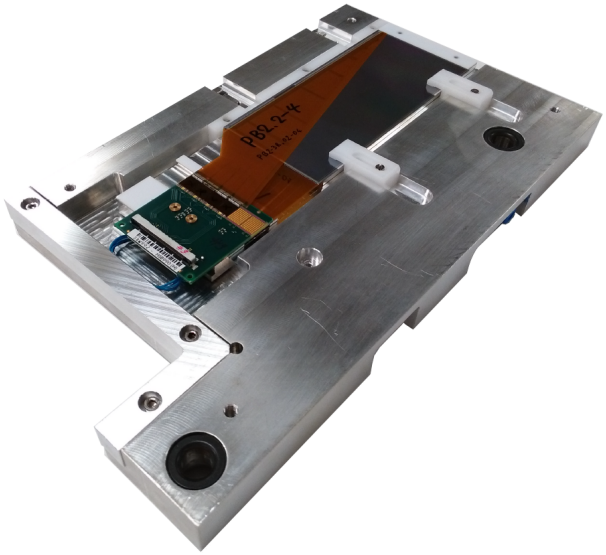


*Figure 3.12: BW subassembly at the end of the procedure, ready for shipment (from [47]).*

### 3.5.2 Electrical test procedures

The electrical active components must be tested after their manufacture to check the correct functionality and to identify possible defects. In general, these defects are not visible from a mechanical test or optical survey. The assembly procedures may produce modules with reduced electric functionality and they may cause additional defects on the original components too.

Therefore after each procedure that may cause malfunction or defects the electric test are repeated, to be able to exactly identify which procedure or step caused the defects and why, and if possible to fix them. On the other hand, the test procedure itself introduce some risk of damaging the components and therefore a good compromise must be found in quality assurance protocol. With this idea, the single-side hybrids are electrically tested at their arrival at Pisa, while the electrical test of the silicon sensors and of the PA are avoided because for the sensors the procedure is more complex and risky, the PAs are unlikely to be damaged during the shipment. The HSs are then tested before the gluing, and in the end the FW and BW subassemblies are tested before the shipment. If some modules need additional procedures (for instance a chip substitution or a defect fixing), the electrical tests are always repeated. This approach is time-consuming, and the same electrical tests are repeated on the same modules several times, but it necessary to guarantee the quality of the final product and ultimately the SVD efficiency. My personal contribution to the Pisa SVD Production has been focused on these electrical test, which I performed in the regular Belle II - Pisa group shifts.

The electrical tests are performed in clean rooms, to work in a controlled environment in terms of temperature and humidity, at a *test-stand* equipped with the following instrumentation:

- a *black-box* which guarantees the electrical and optical isolation of the device under test. A sensor is integrated into the box to measure the temperature of the test-environment during the procedure.
- a data acquisition system (APVDAQ), based on VME hardware and LabView software.
- high voltage power supplies, that provide bias voltage up to 200 V to the subassembly.
- low voltage power supplies used to the acquisition board and to generate the needed power(-5 V to +5 V) required by the APV25.
- a board that provides an additional potential difference $V_{\text{sep}}$ between the chip and the bias ring, from -5 V to +5 V.
- a pulsed infrared laser system with a robot arm for its movements.

The equivalent circuit of the DSSD with the power supply and APV25 readout scheme is shown in figure 3.13. The bias voltage is $V_{\text{bias}} = 100$ V, applied symmetrically on two sides with respect to ground through the Bias Rings and $R_{\text{bias}}$. The APV25s are biased respectively 1.25 V and 2.5 V with respect to the sensor side, to provide the correct operating voltage. The potential difference through the coupling capacitor $C_{AC}$ is as small as possible to limit the stress of the delicate oxide layer which can easily break causing defects, called pinholes. When pinholes are presents an undesiderable DC current flows into the APV25 and an additional voltage, $V_{\text{sep}}$ is provided to the APV25 to tune the reference of chips and adjust the potential difference through $C_{AC}$.

However, the complete procedure is summarized with the following steps, where technical details are omitted:

- **Passive measurements.** Measurement of passive components (resistance and capacitance) of the hybrid sandwich.
- ***I-V* characteristic.** Measurement of the leakage current of the sensor applying a bias voltage that sweeps from 0 to 100 V (FW) or 200 V (BW). This difference is due to the nominal limits given from the two companies which produced the sensors.

– **Pedestal run.** Measurement of the signal mean value (pedestal) and the RMS value (noise) of each APV25 channel (i.e. strip) of the sensor with $V_{\text{bias}} = 100$ V applied. No signal is injected to the channels.

– **Internal calibration run.** Measurement of the gain and noise response of the AVP25 channels with the injection of a simulated signal through the calibration capacitors integrated on the APV25. With the variation of the injection time, the whole shape of the output signal is obtained.

– **$V_{\text{sep}}$ scan.** Measurement of the gain and noise response of the APV25 channels with a simulated signal injection and the application of a $V_{\text{sep}}$ voltage across to the chips i.e. across the $C_{AC}$. The $V_{\text{sep}}$ sweeps between -5 V and +5 V, while $V_{\text{bias}} = 100$ V is kept constant. Channels without defects show a quite constant behaviour with $V_{\text{bias}}$ variation, instead defects have a not trivial trend.

– **Laser scan.** Measurement of the gain and noise response of the APV25 channels with the irradiation of the strips with a pulsed infrared laser beam ($\lambda = 1020$ nm). The laser spot is sufficiently small to illuminate only a few strips at time. Two scans are performed, one in the $z$ direction and one in the $r - \phi$ direction, to cover all the strips of the sensor. A hit is counted when the signal to noise ratio is larger than 6. The choice of the laser wavelength has been done to simulate the energy release of a MIP particle crossing (with different $\lambda$ the silicon results opaque or too transparent).

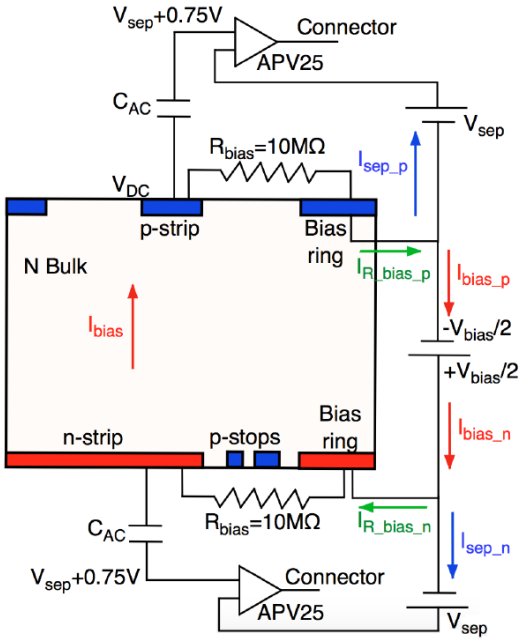



*Figure 3.13: Scheme of subassembly circuit (from [47]).*

The most common defects that can be identified from the tests along with their electrical features are described in the following [47]. The defects are automatically found and classified with a dedicated software that uses the combined result of the tests although often human judgement and detailed observation of the test progress and final results are required to fully understand the effects. The defects involve less than 1% of the total number of channels/strips of the produced modules, and this represents a very high quality for a large production of silicon detectors.

**Open.** A broken strip or bond makes the readout of the channel not possible. If the defect consists in a broken bond it is possible, in some cases, to perform a fix with a manual re-bonding. An open channel is counter-intuitively identified by a high noise and low gain, with adjacent strip also noisy. The reason is the internal common mode subtraction integrated in APV25, which subtracts the average value of the sensor from the open channel output, but since this channel does not follow the average behaviour because is open, it appears as noisy (see figure 3.14(a)).

**Short.** An electrical connection between two adjacent strips or bonds. The effects is a charge sharing between the two strips, with the information of the strip is still available. The short is identified by a reduced gain ($\sim 1/2$ of the expected value) on the two involved channels, because the injected signal is shared between the two strips, reducing the gain. The increased load makes the involved channels also noisy. Depending on the accessibility of the short it is sometimes possible to separate it. (see figure 3.14(b)).

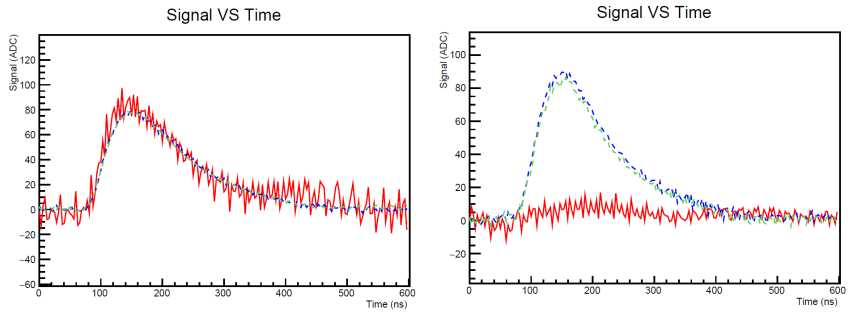
**Pinhole.** A breakdown of the oxide layer between the aluminium electrode and the strip implant, typically on $p$-side. The result is a flow of current from the implant, through the defects, toward the chip (instead of $R_{\text{bias}}$), if a potential difference between the strip and the APV25 is present. This effect produces a saturation of the amplifier and makes it unusable, reducing strongly the gain. Therefore a pinhole is identified by a low gain, and from a clear gain variation in function of $V_{\text{sep}}$: it is peaked at a precise value, typically $V_{\text{sep}} \simeq -0.75$ V, with a maximum similar to the natural expected value, and a lower gain for the other $V_{\text{sep}}$ values. The gain is lower because of the current flow through the pinhole, but if the voltage drop crossing $C_{AC}$ vanishes, the pinhole current disappears and the gain of the strip is restored. In fact 0.75 V is the natural voltage drop across $C_{AC}$ due to the input voltage on the APV25, thus this feature allows to cure the pinhole setting the $V_{\text{sep}} = -0.75$ V for all the pinhole-defected sensor (see figure 3.14(c)).

**Deep pinhole.** A breakdown of the oxide layer between the aluminium electrode and the bulk. The result is a flow of current directly from the bulk toward the chips, affecting the whole sensor and not only a single strip like the simple pinhole. The current produced on $n$-side strip corresponding to the damaged $p$-side strip increases the potential drop on $R_{\text{bias}}$, since $V_{\text{n-strip}} = R_{\text{bias}} I_{\text{n-strip}}$, and it can reach the punch-through threshold (the *punch-through effect* is the charge transfer through a completely depleted region caused by a strong polarization voltage). In this case a large current is established between the strip and the bias ring or the adjacent strips. A deep pinhole is identified by a peculiar gain behaviour in the $V_{\text{sep}}$ scan: the gain results very low below $V_{\text{sep}} \simeq 2$ V, instead the natural expected gain is restored above this value. This can be explained considering that the $SiO_2$ interface on $p$-side is few volts over the voltage of the $p$-implants, due to the application of $V_{\text{bias}}$. Thus the current flows through the defect only if the chip input is below this value, otherwise the current stops. Therefore the deep pinholes are not easily curable and since they can cause a degradation of the performance of the whole detector, they are disconnected (see figure 3.14(d)).

**Double pinhole.** A breakdown of the oxide layer between the aluminium electrode, the bulk and the edge of the nearest strip implant. The APV25 results directly connected both the strip and the bulk. The double pinhole can be identified by a gain behaviour in $V_{\text{sep}}$ scan that results the combination of common and deep pinhole i.e. a low gain except for $V_{\text{sep}} \simeq V_{\text{pinhole}}$, specific of the single defect. As the deep pinhole, also the double pinhole strips are disconnected (see figure 3.14(e)).

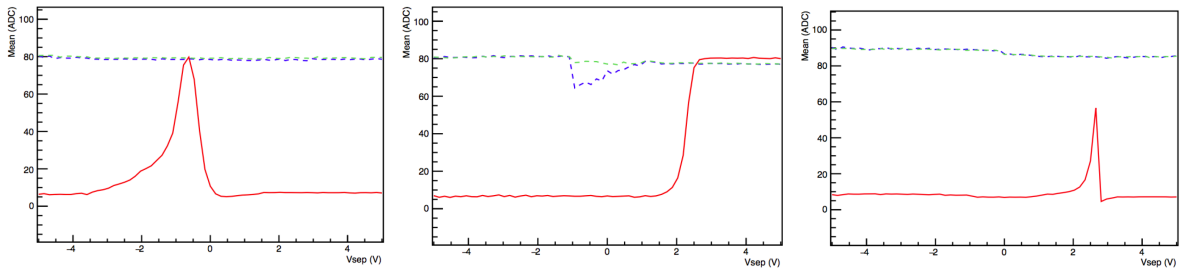
**Particle response.** A lower number of hits in the laser scan test on the strip due to a sensor defect or an electrical defect in connection to the chips.

**Noisy.** A channel or strip shows a noise level higher than expectation. Most of noisy strips are due to another defect (open, short, pinhole), however but some strips do not show the feature of these defects but result still noisy.




(a) *Shape of the signal of an open channel (red) and the shape of the adjacent strips (blue, green).*

(b) *Shape of the signal of a channel in short (red) and the shape of the adjacent strips (blue, green).*




(c) *Gain during the $V_{sep}$ scan for a common pinhole (red) and the adjacent strips (blue, green).*

(d) *Gain during the $V_{sep}$ scan for a deep pinhole (red) and the adjacent strips (blue, green).*

(e) *Gain during the $V_{sep}$ scan for a double pinhole (red) and the adjacent strips (blue, green).*

*Figure 3.14: Some relevant features that allow identifying the defects in the electrical tests*

### 3.5.3 Peel-off issues

During the forward and backward subassembly production several issues were faced, and the solutions developed in these cases are outside of the standard assembly and testing procedure. One perfect example of issues, which developed during the period of this thesis, is the peel-off effect.

After the ladder assembly, in some ladders of Layers 4 and 6, during an optical inspection it has been noticed that the PA glued on the sensor peeled off, separating from the silicon surface. It happened only at the forward edge of forward modules, and affects the inward PA ($\phi$ side), because of the mechanical tension on the PA due to the angle between the hybrid sandwich and the wedge sensor. The peel-off of the PA can propagate to the whole sensor disconnecting completely the PA and making the module unusable. In figure 3.15 it is possible to see a FW module already assembled on the ladder: the peel-off is present on the corner of the PA.

A reinforcement procedure for the FW subassemblies not already assembled on a ladder has been developed, after their electrical tests, to prevent the peel-off. A conceptually similar procedure has been applied to the already assembled ladders to stop the peel-off. These procedures are called respectively *Type 1 and Type 2 reinforcement.*

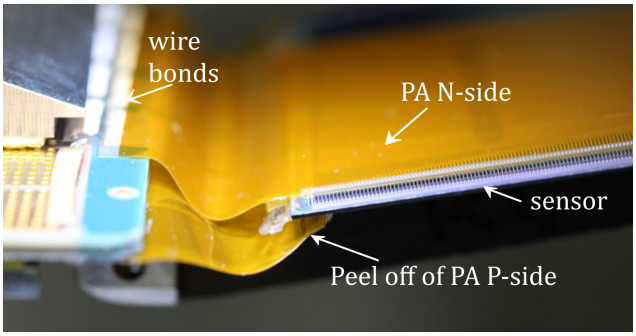



*Figure 3.15: Picture of the peel-off zone in one of the already mounted ladder (from Belle II SVD group studies).*

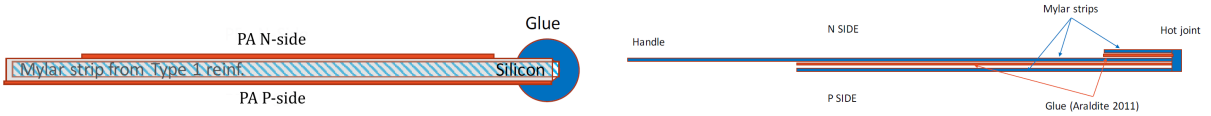
The Type 1 reinforcement consists in the application of a Mylar strip (250 $\mu$ thick, 2.5 mm wide) between the $n$-side PA and the $p$-side PA. The strip is covered by glue, therefore it creates a rigid connection between the two PAs increasing the strength of the gluing. Most of the thickness is represented by the glue and the strip is used to facilitate the application procedure only. Both the glue and the Mylar are electrically inert and do not add relevant material budget. The Type 2 reinforcement procedures are identical, but it uses a thicker and wider Mylar strip.

The entire production of FW modules has been reinforced during 2017 by the SVD group, but after some months has been noticed a peel-off on the reinforced modules. Mechanical stress studies revealed that the reinforcement is not sufficient to keep the PA in the correct position and in particular seems that the adhesive properties of the glue on the silicon are not sufficient to avoid the peel-off.

Therefore two addition reinforcement procedures have been developed to add to each FW module, based on a different concept. They are called respectively *Glue brace* and *Type 3 reinforcement*. The first consists in the application of a glue brace at one side of the Mylar strip of the Type 1 reinforcement, that envelops both the PA keeping them in position from externally. The Type 3 reinforcement consists of a Mylar strip with flaps to mechanically hold the PA of both sides onto the sensor. Between the main strip and the flaps is deposited the glue, but the main reinforcement is due to the Mylar flaps rigidity. In figure 3.16 is shown a schematic representation of Type 2 and Type 3 reinforcement.

These procedures required an R&D work during the SVD production, slowing down the SVD assembly, but this issue would not have been predictable because it depends on the particular adhesive property of the glue on the silicon in the precise position of the SVD layers, a too detailed case to have the information to simulate it before the SVD production. However, after the application of the three reinforcements, the peel-off effect doesn't appear anymore.

All these procedures are very delicate, and they require to work close to fragile parts of the modules (micro bonds, sensor) thus there is a risk to damage them. Therefore, after each reinforcement, the modules are electrically retested.




(a) *Glue brace* (b) *Type 3 reinforcement: Mylar strip with two flaps*

*Figure 3.16: Scheme of Glue brace and Type 3 reinforcements. In left figure the Type 1 or 2 reinforcement is present too, instead the right figure represents only the additional Mylar strip (from Belle II SVD group studies).*

# Chapter 4

# Belle II VXD Track Finder

The high-performance detectors described in the previous chapter must be coupled with a dedicated software to fully exploit their excellent characteristics. The purpose of this software is to select interesting events (*trigger*), connect the signals inside subdetectors to build the particle tracks and vertices (*pattern recognition* and *vertexing*) and then evaluate the track and particle physical properties (*track fitting* and *particle identification*). The set of all these steps is usually called *event reconstruction*. From the set of low level signals that come from the whole detector, the reconstruction software provides to the analyst a compact and accurate description of the physical events.

This chapter focuses on the pattern recognition step: first the general methods of track finding are quickly described and the typical B-Factory software request. Then the features of the Belle II Track finder for *online* and *offline* reconstruction are exposed in detail. The tracking software of Belle II is currently under development and the implementation of some features, from a technical point of view, has not been realized yet. One of the main developments, realized from the Belle II tracking group, that has been done during the course of this thesis is a technical redesign of the entire Track Finder: in the last part of the chapter this redesign is described.

Preliminary it is needed to introduce two concepts widely used in tracking that are necessary to describe the track finding procedures:

**Hit.** A single signal on a generic detector that is able to identify spatially the passage of the particle is called *hit*. Additional information can be provided in the single hit (like timing).

**Cluster.** If multiple signals from spatially close hits are believed to be originated from the same particle they are clustered in a single hit by a specific algorithm that evaluates position and properties of the new object called *cluster*.

## 4.1 General methods of track finding

In high energy physics the increasing performance of colliders and detectors and the rare-events searches make much more complex tracking tasks in the course of time. Due to the high luminosity and the hermeticity of the detectors the interesting signals are immersed in a huge physical background. The required precision from the physics analysis needs an increased accuracy in the reconstruction of the physical parameters. The events under analysis have extreme properties: they span from very low energy events, with few hits inside the detector and pattern distortions from material effects, to extreme high energy, with high track multiplicity and consequent combinatorial background. In addition, the high event rate often requires a time-constrained event

reconstruction. To overcome this huge task several techniques of reconstruction have been developed; some of these techniques will be reported in following sections, for an exhaustive report of the various method see [48], [49] or [50]. Before the description of these methods it is important to recall the main steps of event reconstruction:

**Trigger.** The preliminary step is the raw data reduction to select only the events that are most probably interesting and discard background events. The most significant performance parameters of triggers are the processing time and the rejection rate. The most frequently used triggers are divided in a first level hardware trigger and a High Level Trigger (HLT), a software level that performs a more refined selection (the Belle II one is described in section 2.4.8). The HLT may contain some of the other steps of event reconstruction (like pattern recognition, fitting or particle identification), usually simplified to make it faster. In that case it is called *online reconstruction*, counter the *offline reconstruction* done out of trigger procedure.

**Pattern recognition.** In this step the hits from tracking detectors are combined, using properly track models, to obtain the various separated list that forms a particle track candidate inside the detectors. This is one of the most delicate steps because needs to perform a large combinatorial analysis to be able to match properly the hits and do not produce fake (artificial) tracks from random combination. These fake tracks are one of the sources of combinatorial background. Anyway, the selection of candidates must be quite inclusive because a discarded candidate at this level, especially in case of online reconstruction, can't be recovered.

**Track Fitting.** After the selection of the list of hits that belong to a single particle, the track candidate is fitted to determine the track parameters and then the physical parameters of the particle. This step includes also the valuation of the quality of the estimation performed during the fit because it must discriminate between fake candidate and true tracks.

**Particle Identification (PID).** In this step, sometimes performed in parallel with track fitting, is determined the kind of particle that produces the hit used in the track. Usually, different mass hypotheses are tried and tested (this step is strongly connected with fit quality check).

**Vertexing.** As the track finding use hits as input the vertex finding and vertex fitting use tracks produced in the pattern recognition as starting observables. The vertex information may give strong constraint on the physical parameter, like the momentum of the track. Moreover, the reconstruction of the vertex allows to reconstruct neutral or short lived particles from charged decay products with a common vertex.

In the following sections are described of the most important pattern recognition methods and some track fitting tools connected with pattern recognition.[1]

### 4.1.1 Conformal Mapping

This is a global method, or rather it uses simultaneously all hits of the detector. It can be used in detectors with a homogeneous magnetic field, where the primary charged particles describe circles passing through the origin (IP) on the transverse plane:

$$(x_i - a)^2 + (y_i - b)^2 = a^2 + b^2 \equiv R^2,$$

where $(x_i, y_2)$ are the Cartesian coordinates on the transverse plane, with $i = 1, 2, \ldots N$ is the hit index and $\sqrt{a^2 + b^2}$ the distance from the origin. The idea is to apply a conformal transformation

[1] the trigger and PID methods are not described because are quite independent from pattern recognition.

to the coordinate of the hits to convert the circles into straight lines:

$$\begin{cases} u_i = \frac{x_i}{x_i^2+y_i^2} \\ v_i = \frac{y_i}{x_i^2+y_i^2} \end{cases} \Rightarrow 2bv_i + 2au_i = 1. \tag{4.1}$$

If the track multiplicity in the events is sufficiency small and the momentum sufficiency high, a histogram of $\phi$ (where $\phi$ is the azimuthal angle) filled with the hit distribution in $(u, v)$ coordinate shows peaks corresponding to tracks. The height of the peak corresponds to the number of hits of the track, in fact in this condition each track describes a straight line from the interaction point.

The strength of this method is the speed: the processing time increases linearly with the number of hits. On the other side this method has two important limitations. The $z$ position of the hits (along the beam axis) is ignored, and so high multiplicity events make the $\phi$ analysis and hit assignment ambiguous. Secondly for low momentum tracks, because of the small $R$, the straight lines do not pass through the origin and the $\phi$-histograms results unusable. Moreover, the low momentum tracks are strongly affected by material effects that move the tracks from the perfect circular shape, and further reduce the efficiency of the method.

### 4.1.2 Hough and Legendre transform

To overcome the drawback of the previous method the idea has been extended with the Hough transform, another global approach. In the first step the conformal mapping is used, and then from the $(u, v)$ coordinate (the so-called *image space*) another transformation is applied to move into to the *parameter space*, where the two variables are the lines parameters. From equation 4.1:

$$v_i = cu_i + d \quad \text{where } c = -\frac{a}{b},\ d = \frac{1}{2b} \quad \Rightarrow \quad d = -u_i c + v_i.$$

Now the two variables can be discretized $c \equiv c_j$, $d \equiv d_j$ to search peaks in 2D histogram $c \times d$ filled with all the event hits. These peaks represent hits with the same track parameters, assigned to the same track candidate. This can be qualitatively seen in a plane $c \times d$ without using discretization: each hit describes a straight line, and the intersections are correspondent to hits with same $c, d$ parameters thus if several hits have the same interception on this plane they can be associated to the same track.

A variant of this algorithm is the Legendre transform, frequently used to solve pattern recognition problem in drift tubes detectors. In this case the additional drift time information is used, transforming the drift circles (in $(x, y)$ coordinate) first in circles with the conformal mapping (in $(u, v)$ coordinate) and then into sine curves (in $(c, d)$ coordinate) with the Legendre transform. The tracks are found from the intersections or peaks between these sine curves (see [51] for details).

The main drawback of this algorithm is the processing time: it grows as the square of the number of hits in 2D case or as the cube in 3D case.

### 4.1.3 Track Fitting and Kalman Filter

The track fitting procedure has to estimate the track parameters and evaluate the quality of the track candidate to select or discard it and its evaluation requires the introduction of statistical tools to the procedure. The track fitting methods are composed of some building blocks:

**Track Parametrization:** the optimal choice of the geometrical parameters which describe the track, based on a chosen coordinate system, based on the geometry of the detector and of the physical events. The set of the track parameters define a *state vector* of the track at each time.

**Track model:** the model that describes the propagation of the state vector from a surface of the detector $i$ to a different surface $j$: $\mathbf{v}_j = \mathbf{f}_{j|i}(\mathbf{v}_i)$, where $\mathbf{v}$ is the state vector and $f$ the track propagator.

**Error propagation:** the method of propagation of covariance matrix of the state vector from a layer to another layer.

**Material Effects:** the method used to take into account of the material effects that can modify the propagation of the track (multiple scattering, electron bremmstrahlung).

**Measurement model:** the model $\mathbf{g}$ that describes the measure of the physical quantities $\mathbf{v}_i$: $\mathbf{m}_i = \mathbf{g}_i(\mathbf{v}_i)$, where $\mathbf{m}$ is the measure.

The most used track fitting method belongs to linear least-squares approach with global techniques (detailed description omitted here, see [48]). On the other hand, a local strategy is the Kalman filter: a recursive formulation of the least-square method that "follows" the track closely hit by hit. This method alternates between prediction and update steps: in the first one the state vector is propagated to the next layer, in the second one the measurement of the reached layer is used to update (and correct) the information of the propagated state vector. The same is done for the covariance matrix. At the end of the procedure a *smoother* is applied along the track, equivalent to a pair of Kalman filters that proceed in opposite directions.

An interesting method developed from Kalman filter that includes both pattern recognition and track fitting is the *Combinatorial Kalman Filter* (CKF), an approach which is on the edge between classical and *adaptive methods*. In general adaptive methods are a class of tracking approaches which make several hypotheses in competition during the application of the algorithm. The algorithm starts from an initial state and then evolves with the use of data (the observation: hits for track finding, track for vertexing) several times during iterations. Weights are assigned to each observation depending on hypothesis and to the position of the other observables.

The idea is to realize multiple Kalman filters from a track seed in the detector and explore al the possible combinations of hits in various layers evaluating the track quality with the Kalman filter techniques step by step (in term of $\chi^2$ of the track candidate up to the reached layer). This method realizes a tree of possible tracks that exploit the multiple hypothesis approach typical of adaptive methods. If the quality of a single branch of the tree falls below a definite value the relative candidate is removed, and in the end the best quality candidate is kept.

### 4.1.4 Cellular automaton

The idea of the cellular automaton is to define a cell (a hit or a segment of hits), a cell state and a simple update procedure of the state depending of the cell neighbour state. This realizes an iterative process that produces a tree graph on the cells connections and their weights (states). Then a selection with a proper state threshold is done to obtain track candidate. The cellular automaton is a track finding method hybrid between local and global method: it uses the local connection of hits layer by layer, but does not start from a single seed but proceed in parallel from all the cells of the detector.

## 4.2 Track finding at B-Factories

The B-Factory environment, in term of event properties and physical purposes, imposes strong requirements to the Track Finder. The beam-induced background cross sections of an $e^+e^-$ high luminosity accelerator are much larger than the cross section $e^+e^- \to b\bar{b}$, which is 1 nb at resonant $\Upsilon(4S)$ energy (for instance the radiative Bhabha cross section $\sigma(10.58\ \text{GeV}) \simeq 300$ nb). Thus the Track Finder must be able to reconstruct online the event tracks to discriminate between background and $\Upsilon(4S)$ events, because this task cannot be completely devoted to L1 hardware trigger. In addition in the triggered events the Track Finder performances must be robust in presence of background hits mixed with signal-tracks hits. The trigger task and time constraint from high bunch crossing of the accelerator impose a modular structure to the Track Finder, that gives the possibility to enable a fast reconstruction procedure with the use of partial information from the detector without the loss of high precision tracking.

The track finding efficiency must be high in the whole momentum range, reconstructing *all* the tracks of the $\Upsilon(4S)$ events, to exploit the potential of the high luminosity, the clean environment of the leptonic collision and the knowledge of the initial $\Upsilon(4S)$ state. From detector side the design of a hermetic detector very close to interaction point gives to the software development the possibility to obtain this result, but on the other hand it produces not trivial issues to the Track Finder, like high occupancy and so large background and *ghosts*, asymmetric tracks distribution or tracking in very low momentum range.

A not negligible part of the particles produced in $\Upsilon(4S)$ decays has momentum below 200 MeV/$c$. This raises two main tracking issues: the tracks of the particles do not cross most of the detector and the material effects may produce critical effects on the track geometry. For instance consider the Belle II design: the inner radius of Belle II is $R_m^{CDC} = 16$ cm and it means, from

$$R\ [\text{m}] = \frac{p_\perp\ [\text{GeV/c}]}{0.3B\ [\text{T}]},$$

that the minimum transverse momentum to produce a hit in the drift chamber is $p_\perp(2R_m^{CDC}) \simeq 144$ MeV/c (and 3 hits is the minimum information required to be able to reconstruct a helix). The purpose of Belle II tracking is to be able to reconstruct particles down to 50 MeV/c, thus it is clear that it is necessary to introduce the silicon vertex detector in the online tracking procedure[2] and be able to have a robust tracking with PXD and SVD only for offline reconstruction too. In figure 4.1 is shown the transverse momentum spectrum of various particle type.

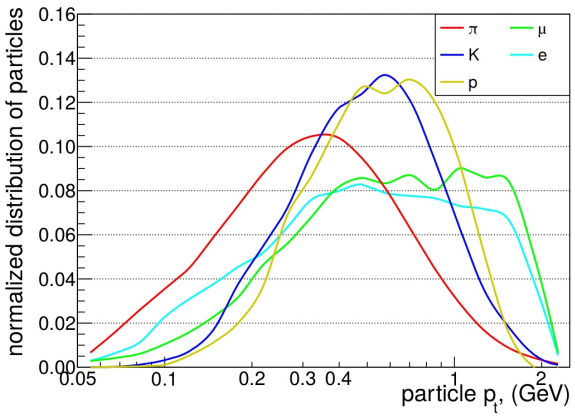
Regarding the material effects, that at low momentum produce not negligible and affects strongly the behaviour of particles, they must be treated with a flexible tracking software. It must be able to identify and not discard also not perfect helicoidal tracks and on the other hand to reject cases with critical interaction with material has produced not fittable and usable tracks. This issue is going to be discussed in detail in chapter 5.

Another important feature required in a B-Factory Track Finder is the vertex resolution that strongly exploits the resolution of the detector and has a central impact on the precision of most of the measurements of the experiment.

It is important to remark that the purpose of low momentum track reconstruction is not only to increase the sample collected by the experiment and to close the kinematic of $\Upsilon(4S)$ event, but also to be able to reconstruct particular vertices. One of the most relevant channel with this features is:

$$D^{*\pm} \to D^0\pi^\pm \tag{4.2}$$

[2] in Belle design the CDC inner radius was about 80 mm, but the pixel detector was not present, reducing the vertex precision




| particle type | fraction |
|---|---|
| $\pi^{\pm}$ | 72.8% |
| $K^{\pm}$ | 14.9% |
| $e^{\pm}$ | 5.8% |
| $\mu^{\pm}$ | 4.7% |
| $p^{\pm}$ | 1.8% |

*Figure 4.1: Normalized transverse momentum distribution from $\Upsilon(4S)$ decay, subdivided in various particle types, simulated in Belle II environment. The average fractions of each kind is reported at side (from Belle II tracking group studies).*

that is is the main decay channel of $D^{*\pm}$ with $BR(D^{*\pm} \to D^0\pi^{\pm}) \simeq 67.7\%$ (and $D^*$ that is one important decay channel of $B^{0,\pm}$). The produced $\pi^0$, necessary to reconstruct the vertex of $D^*$, is a very slow particle, usually called "soft pion", with a momentum of few hundreds MeV/$c$, as it possible to see in figure 4.2. The interest of this class of decays is particularly relevant for the reconstruction of $B \to D^*\tau\nu_\tau$ decays, as is discussed in detail in the section 1.5, where physical interest some features of the reconstruction are described. For instance, one of the main sources of background comes from the decays $B \to D^{**}\tau\nu_\nu$ that where $D^{**}$ decay mimics the $D^*$ one. Accurate vertex measurement with the use of soft $\pi$ information can help to distinguish the background from the signal.

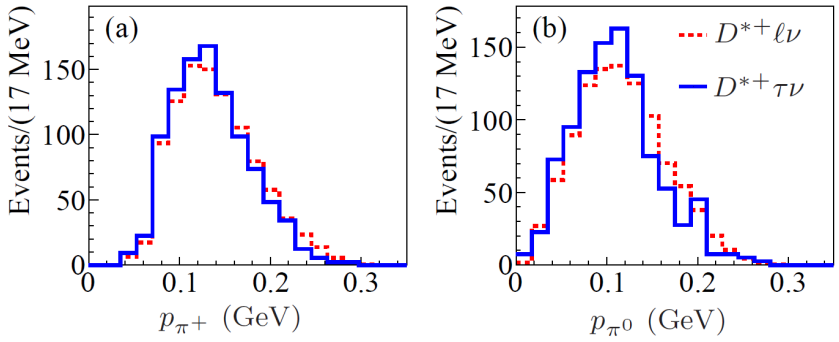



*Figure 4.2: Pion momentum in laboratory frame from $B \to D^{*+}\ell\nu_\ell$ and $B \to D^{*+}\tau\nu_\tau$ decays in $D^0\pi^+$ and $D^+\pi^0$ channels. Histograms are normalized to 1000 entries. (distributions from BaBar measurement [18].)*

### 4.2.1 Comparison with tracking hadron colliders

In current hadron colliders, with the particular reference to LHC [52], the requirements for the Track Finder are quite different. In case of hadronic collider, the $pp$ cross section and the consequent event rate is orders of magnitude higher than the $e^+e^- \to \Upsilon(4S)$ one, despite the lower bunch crossing rate. Besides the single events contain multiple vertices, and so higher *pile-up*, and the complexity in term of number of tracks per vertex is very much higher. These features require a powerful trigger system that includes the Track Finder and vertexing. Some

projects for future hadronic colliders with higher trigger rate, like *High Luminosity LHC* [53], foresee the possibility to include tracking at Level 1 hardware trigger too.

Anyway, the tracking in hadronic colliders does not require to reconstruct the whole event because the initial parton-parton state is in any case unknown. Therefore the trigger, without tracking requirement, may apply strong cuts that select interesting events. These cuts include hard energy or momentum requirement, that discards all the events below some global threshold (with the use of calorimeter information, for example). These choices reflect the different physical purposes of experiments at hadronic colliders at energy frontier where the interesting momentum scale is completely different (GeV to TeV scale). This different energy scale has fundamental implication in tracking methods: much boosted and thus displaced vertices, cleaner tracks, completely negligible multiple scattering effects.

The LHCb experiment at LHC has heavy flavour physical goals in a hadronic environment. This experiment, despite the unknown initial state and the very much larger boost and momentum scale, has some vertex requirement in term or resolution very much similar to B-Factories ones because of the study of the same physics channels. The studied transverse momentum spectrum is also similar to the B-Factory one for some physics channels, despite the much higher longitudinal momentum.

### 4.2.2 Tracking at low momentum at Belle II

Before the description of the Belle II track finder it is important to recall one of the dominant physical effect at low momentum, the *multiple Coulomb scattering*: the electromagnetic interactions between the incident particle and the nuclei of the detector material. In these interactions is relevant the variation of the direction of momentum $\Delta\hat{p}$, in fact the impact between the incident particle and the nucleus is mainly elastic with incident particle $\Delta E \simeq 0$, but it can strongly change the trajectory of the particle. If a material with thickness $X$, with radiation length $X_0$ and atomic number $Z$, and a particle with momentum $p$ and speed $\beta c$ are considered, it is useful to define the *mean square deflection angle*

$$\theta_0 \equiv \sqrt{\langle\theta^2\rangle} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\theta_i^2}$$

where $N$ are the number of nuclei hitten by the particle along the track and $\theta_i$ the angle between the initial momentum and the final momentum directions of a single impact $i$ on a plane containing the initial momentum of the particle. The mean square deflection angle can also be expressed as [4]:

$$\theta_0 = Z\frac{13.6\text{ MeV}}{\beta cp}\sqrt{\frac{X}{X_0}}\Big(1 + 0.038\ln(\frac{X}{X_0})\Big). \tag{4.3}$$

In particular, with respect to the coordinate system of Belle II and considering a layer of detector material with thickness $z$ parallel to the beam pipe (a SVD sensor, for instance) and a particle that impact the layer at angle $\theta$, the "effective thickness" to use in formula 4.3 $X = z/\sin\theta$. Considering that for a single layer of SVD $\frac{X}{X_0} \simeq 0.6\%$ and so $1 + 0.004\ln\frac{X}{X_0} \gtrsim 1$, a lower limit of multiple scattering angle in SVD material is:

$$\theta_0 \simeq Z\frac{13.6\text{ MeV}}{\beta cp}\frac{\sqrt{z}}{\sqrt{\sin\theta}}. \tag{4.4}$$

Because of this deflection the multiple scattering affects the precision in the extrapolation of tracks during the pattern recognition. In particular the spread on the angular track parameters,

both on transverse and longitudinal plane, will be of order of $\theta_0$, instead the spread on linear track parameters, like the impact parameters, will be of the order of $\theta_0 l$, where $l$ is the lever arm of the used hit.

With respect to high momentum tracking, when multiple scattering effects became relevant, the transverse momentum resolution has a different trend. In fact at high momentum the main source of uncertainty of momentum of a tracked particle is the resolution on the sagitta $s$. The sagitta of an arc of circle is the distance between the centre of the arc the subtended chord, and if $L$ is the length of the chord

$$s \simeq \frac{L^2}{8R}, \quad \text{or as function of } p_\perp, \; s = \frac{0.3}{8}\frac{L^2 R}{p_\perp},$$

and so the resolution

$$\frac{\sigma_\perp}{p_\perp}\Big|_{\text{high } p} \simeq \frac{\sigma(s)}{s} \sim \frac{L^2}{8R^2}\sigma(R) \propto p_\perp.$$

Instead at lower momentum the sagitta is measured with higher precision because the tracks are curlier, but the multiple scattering affects the uncertainty on radius [54]:

$$\frac{\sigma_\perp}{p_\perp}|_{\text{low } p} = \frac{\sigma(R)}{R} \simeq \frac{\sigma(\phi)}{\phi} \propto \theta_0 \cdot \frac{R}{L} \simeq \text{const}, \tag{4.5}$$

and so the improvement of $\sigma(p_\perp)/p_\perp$ saturates to a constant value. In all the studies which involve multiple scattering is not relevant the transverse momentum only, but the magnitude of the momentum of the particle, as it is shown in formula 4.5. In fact the result in 4.5 is completely true only for $\sigma(p)/p$.

In this momentum range several other energy loss effects are also relevant, like ionization, strong interaction for hadrons or bremsstrahlung for electrons. All these effects contribute to a degradation of the momentum resolution at low momentum. In addition, in previous discussion has been neglected the granularity of the detector, and in particular the layer-structure of Belle II VXD: if momentum is too low the outer layers are not reached from the particle, so the resolution on momentum deteriorates with a discrete step as a function of momentum.

As an example of the described behaviour in figure 4.3 the transverse momentum resolution of Belle II is reported, with the combined use of all the tracking detectors (CDC, VXD).

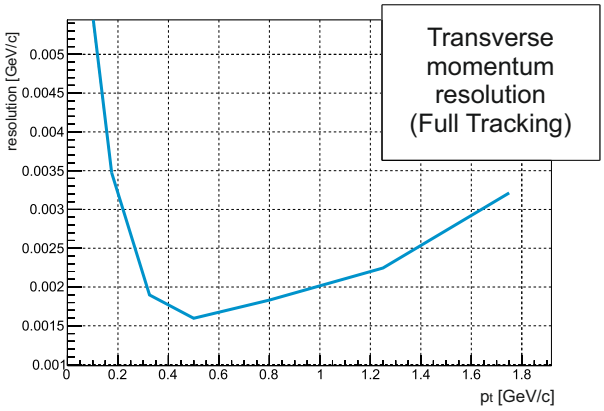



*Figure 4.3: Transverse momentum resolution at Belle II, with the combined tracking devices, from Belle II validation plots.*

## 4.3 VXD Track Finder

The VXD Track Finder (VXDTF) is the tracking software of Belle II that uses the information of silicon detectors only. The VXDTF was developed mainly by J. Lettenbichler in his master and PhD years and detailed described in relative thesis [55] (master thesis), [56] (PhD thesis). The tracking software of Belle II is still in development and, in particular, a technical redesign of VXDTF, called VXDTF2, has been worked out by the tracking group during the course of this thesis [57]. In this section, a summary of the methods used in the VXDTF is given to introduce the environment where my personal contribution to the tracking software (presented in chapter 5) has been developed , and a description of the most recent update on VXDTF2 is given in section 4.4.

The VXDTF has a double task inside the reconstruction software of Belle II: online tracking on the HLT and offline complete reconstruction.

**HLT - Fast Reconstruction.** In HLT the VXDTF performs a 4-layer tracking, with the use of SVD only. The aim is to reconstruct the track candidates used to find ROI on PXD layers to perform the data reduction needed to read the pixel information at HLT level (see section 2.4.8.2 for HLT and 3.2.2 for ROI description). To perform this task the VXDTF has to work with strong time constraints and a limited number of CPU.

**Offline - Full Reconstruction.** In the offline reconstruction the VXDTF performs the full 6-layer tracking with the use of the whole VXD information. This is the tracking used in Belle II analysis software. The tracking software is exactly the same to the fast reconstruction one, but is configured to perform tracking with two additional PXD layers.

Despite its role in the HLT the VXD is not the only tracking sub detector of Belle II, but also the CDC has a central role for track finding and in particular the HLT uses primary the information from Drift Chamber. In section 2.4.8.2 some details are given.

### 4.3.1 Requirement and overview

The VXDTF must be able to perform tracking with the following environmental parameters:

- four layers of double-sided silicon strip detectors (plus two 2 layers of DEPFET pixel detector in full reconstruction only), with the particular geometry described in chapter 3. Particularly relevant features are the overlap between sensors to improve the hermeticity of the detector, and the slanted sensor in forward region to increase the acceptance reducing material budget.

- homogeneous magnetic field $\mathbf{B} = 1.5$ T, parallel to beam axis. All the tracking sub detector (so the detector not included in VXDTF like CDC) are inside the solenoid.

- detectors defects, and so missing hits, and ghost hits (features of double-sided detectors).

- mean occupancies that spans from 6.7% on the innermost SVD layer to a mean occupancy of 0.9% the outermost layer.

- $\Upsilon(4S)$ events, with a center-of-mass boost of $\beta\gamma = 0.28$, a charged track multiplicity per event $\simeq 11$ and a momentum spectrum with the maximum at $\simeq 400$ MeV/$c$ (in laboratory frame).

- not negligible material effects for low momentum particles. Relevant in particular are multiple scattering and $\delta$-ray emission.

- bunch crossing rate of 200 MHz, and an $\Upsilon(4S)$ event rate of 0.9 kHz and total physics rate of 30 kHz. So high beam background and order of 200 kHz of Bhabha and $e^+e^- \to \gamma\gamma$ background. From simulation 75% of the hits inside the tracker come from background and only 25% come from physical tracks.

To be able to perform efficiently its tasks the VXDTF must satisfy stringent requirement:

- the Track Finder must be able to reconstruct tracks with momentum down to $p_\perp = 25$ MeV/$c$. This limit is set because $p_\perp = 25$ MeV/$c$ (up to $p \simeq 80$ MeV/$c$) is the minimum to reach layer 5 in the forward region and so obtain the minimum 3-hit SVD-only track. In backward region this physical limit is $p_\perp = 40$ MeV/$c$ (up to $p \simeq 75$ MeV/$c$) and $p = p_\perp = 40$ MeV/$c$ at $\theta = \pi/2$.

- the Track Finder must be able to reconstruct tracks from $\theta = 17°$ (forward) to $\theta = 150°$ (backward). Therefore inefficiency due to the geometrical acceptance is about 10%.

- the tracking efficiency must be higher than 99% over 300 MeV/$c$ of transverse momentum, and it can undergo excessive degradation down to $50 \div 25$ MeV/$c$ (over $50 \div 60$%).

- the fake rate must be limited below 10%, so the VXDTF must have a strong tool to remove ghost tracks.

- low momentum tracking must take care of multiple scattering effects and include not completely helicoidal tracks.

- the impact parameter resolution good enough to obtain a resolution on $z$-displacement of $B$ veritces $\Delta z(B\overline{B})$ of $\sigma_{\Delta z} \simeq 40 \div 60$ $\mu$m.

- with the SuperKEKB $\Upsilon(4S)$ production rate the fast reconstruction must spend 300 ms second to build the event, extrapolate the ROI on PXD and read it and fit the resulting track. This time is estimated considering a farm of $10^4$ cores dedicated to the HLT operations.

The procedures and algorithms described in the following sections have been implemented inside the software framework `basf2` (Belle Analysis Framework 2), and its naming convention, with proper explanation, is used in this chapter and in the next one.

The input of the VXDTF are the so-called `SpacePoint`, a common hit format for PXD and SVD. The VXDTF approaches the pattern recognition, by applying filters of increasing complexity to simple and then more complex structures: start connecting hits in consecutive layer to build a Segment, then check for adjacent Segments and so on to build the track candidate, applying each step some quality cuts. The cuts values of each step are collected from some lookup tables, called *Sector Map*, created from simulated events information. Then the tracks are collected using a Cellular Automaton, and a quality indicator is assigned to each Track Candidate. As the last step, the overlapped tracks (the track candidate that share hits) are removed by the use of a Hopfield network with the use of the quality indicator. In the end, the output of VXDTF is a list of tracks candidates.

### 4.3.2 Sector Map and Training Sample

The idea of the Sector Map is to manage the pattern recognition with a series of lookup table collectively called Sector Map, that contains all the informations that characterizes the physical track inside the detector. This set is created offline, with the simulation of a large quantity of events and the collection of the Monte Carlo true information of the tracks inside the VXD. The combinations of hits on the various sensors of VXD obtained with this simulation represent the expected patterns of data-hits that will be collected online. Therefore by storing the geometrical information about these simulated tracks inside a lookup table it is possible to compare the data-hit geometrical information with the simulated hit information picking up the latter from the table. Thus with a comparison between the data-hits information and the stored ones inside the Sector Map it is possible to select or discard the combinations. This idea may recall the associative memory strategy, but it is basically different: the Sector Map is not a pattern bank, because it stores geometrical informations about the tracks in term of statical distributions, not the pattern of each simulated track. Secondly, the procedure is completely software and the comparison with pattern bank is made sequentially and not in parallel like in associative memory. Anyway, associative memory solution has been designed for Level 1 tracking-trigger with very high trigger rate not sustainable with CPU tracking, not the case of B-factories. Some other differences are going to be shown in the following paragraphs.

One of the great advantages of this approach is that the tracking software learns from the Training Sample the geometry of the detector. Of course, the VXDTF is optimized for a 4/6-layer detector with the particular VXD geometry, but the VXDTF learns from the distribution of the tracks in the Training Sample how to reconstruct the real tracks, so it can also work with different detector geometry. This is not only a hypothetical advantage: first of all because the VXD lacks trivial geometrical discrete symmetry, secondly because the VXDTF can be used directly in different context like a testbeam geometry (see chapter 6) or to reconstruct a completely different events (like cosmic rays), with the use of a proper trained Sector Map only.

To increase granularity, and so the precision of the geometrical informations, the sensors of VXD are virtually divided in sub-sensors, called *Sectors*. This reduces the possibility to find more than one hit in the same logical position inside the Sector Map, and so the impact of ghost hits. This subdivision allows to take into account separately of overlapped regions of sensors, where the topology of the track patterns is more complex. By default in VXDTF the subdivision of sensor is simple: each real sensor is subdivided in rectangular (BW, ± Z, CE sensors) or trapezoidal (FW sensors) $4 \times 4$ virtual sectors with the same dimension along the $r - \phi$ and $z$ directions, but the software allows to select different subdivision for each sensor of VXD, with different and more complex internal geometries too, but this feature was not optimized.

A Sector with a simulated hit present on it is defined *MC-active*, to differentiate with *active* Sector, that means that a data-hit is present in this Sector. This naming convention is going to be used for all the VXDTF logical elements. The first information that is stored inside the Sector Map is the set of the MC-active Sectors:

$$S^{MC} = \{s_i^{MC}\}, \quad \text{with } s_i^{MC} = \mathcal{S}(p_\alpha^{MC}),$$

where $s_i$ is the Sector with sector index $i$, $p_\alpha^{MC} \in P^{MC} = \{p_\alpha^{MC}\}$ are the simulated Space Points with Space Point index $\alpha$, $\mathcal{S} : P \to S$ is the function which for each Space Point associates the relative Sector.

For each Sector are defined *Friends* Sector the inner ones that are compatible with a track that crosses them consecutively. The network of all the Friends inside the detector build a graph with no loop and so a tree. This feature of the Sector Map allows a fast selection of the tracks candidates using the Cellular Automaton (see 4.3.4). In figure 4.4 an example of Friend Sectors

is graphically shown. The network is completed with the creation of an additional virtual Sector placed at IP with proper Sector dimension, and it is used as a virtual hit for each track produced (the most probable production vertex for the track is inside the IP region). This virtual Sector is very useful for online reconstruction without PXD layers, in case of tracks with a missing layer or for curly tracks with low momentum that do not cross all the SVD layers. Therefore the second information stored inside the Sector Map is the set of friendship relation between MC-active Sectors:

$$F^{MC} = \{f_{ij}^{MC}\} \quad \text{where } f_{ij}^{MC} = f(s_i^{MC} \to s_j^{MC}),$$

where $s_i^{MC} \to s_j^{MC}$ indicates the relation between the sector $i$ and $j$. During the training phase for each pair of Space Points $p_\alpha^{MC}$, $p_\beta^{MC}$ connected by a friendship relation are evaluated five geometrical variables:

- squared 2D distance of the two Space Points in $x - y$ plane,
- squared 3D distance of the two Space Points,
- 1D distance of the two Space Points in $z$ direction,
- angle between of the $z$ direction and the direction defined by the two Space Points.
- angle between of the vector defined by the two Space Points on the transverse plane

Then for each friendship $f_{ij}$ relation are filled these five distribution with the use of all the space points pairs associated to them $p_\alpha^{MC} = \mathcal{S}^{-1}(s_i^{MC})$, $p_\beta^{MC} = \mathcal{S}^{-1}(s_j^{MC})$ (where $S^{-1} : S \to P$ is the inverse image function of $\mathcal{S}$). Finally, for each $f_{ij}$, for each of the five variables variable two threshold values are defined, removing the tails of the distributions, typically at 0.1% and 99.9% quantile of the distributions. This values of quantiles has been set by hand in all the distributions and are the same for all most of connections. The Sector Map strategy in principle has the possibility to make specific cuts-quantiles for different connections but this feature, that needs a specific study, has not been implemented in VXDTF. However, some new features of VXDTF2, like the studies presented in chapter 5, stars to be implemented in this direction. Then this sets of threshold values $V(\mathcal{F}_{ij}^{(2)})^{min}$ and $V(\mathcal{F}_{ij}^{(2)})^{max}$, called 2-hit filters thresholds, are stored inside the Sector Map. The values of these thresholds change for each friendship relation of the Sector Map i.e. for each Sector Pair. It realizes a dependence between the geometrical position of the Sectors in VXD space and the features of the geometrical distribution of the simulated track. This feature, and so specialized selection criteria for different regions, is essential for an experiment like Belle II with so particular geometrical choices as a boosted center-of-mass or three slanted layers of SVD.

Another object is defined to apply more complex filter: a *Segment* is a pair of Space Points with their Sectors connected by a friendship relation which has passed the 2-hit filters. The set of the MC-active Segments is:

$$D^{MC} = \{d_{\alpha,\beta}^{MC}\} \quad \text{with } d_{\alpha,\beta}^{MC} = (p_\alpha^{MC}, p_\beta^{MC})$$

This assignment is injective: for each pair of Space Point that overcome the requirement is defined a different Segment. For each Segment are defined *Neighbour* Segments the ones that share the inner Space Point with it (so the "consecutive Segments"). The relations between MC-active Neighbour as relation between two friendship relation are stored in the Sector Map:

$$N^{MC} = \{n_{ijk}^{MC}\} \quad \text{where } n_{ijk}^{MC} = n(f_{ij}^{MC} \to f_{jk}^{MC}).$$

For each neighbour relation a 3-hit Filter is defined, and so the same procedure of threshold evaluation of the 2-hit filters for several geometrical an kinematical variables also is applied.

It is also defined a *Triplet* as three Space Points with their Segments connected by a neighbour relation that has passed the 3-hit filters. The set of MC-active Triplets is stored inside the Sector Map:

$$T^{MC} = \{t^{MC}_{\alpha\beta\gamma}\} \quad \text{with } t^{MC}_{\alpha,\beta,\gamma} = (p^{MC}_{\alpha}, p^{MC}_{\beta}, p^{MC}_{\gamma}).$$

It is stored another set of threshold values, from the 4-hit filters, but it is not present a 4-hit class of relations because before its application the tracks are collected by the Cellular Automaton (detailed described in the workflow of VXDTF, in section 4.3.3). The description of the various filters and a more detailed description of their application is also described in section 4.3.3.

In conclusion, inside the Sector Map are stored:

- $S^{MC}$, the set of activated Sectors during the training
- $F^{MC}$, the set of friendship relation between Sectors during the training
- $N^{MC}$, the set of neighbour relation between friends during the training
- $V(\mathcal{F}^{(2)}_{ij})^{min}$ and $V(\mathcal{F}^{(2)}_{ij})^{max}$, the set of threshold values of 2-hit filters
- $V(\mathcal{F}^{(3)}_{ijk})^{min}$ and $V(\mathcal{F}^{(3)}_{ijk})^{max}$, the set of threshold values of 3-hit filters
- $V(\mathcal{F}^{(4)}_{ijkl})^{min}$ and $V(\mathcal{F}^{(4)}_{ijkl})^{max}$, the set of threshold values of 4-hit filters

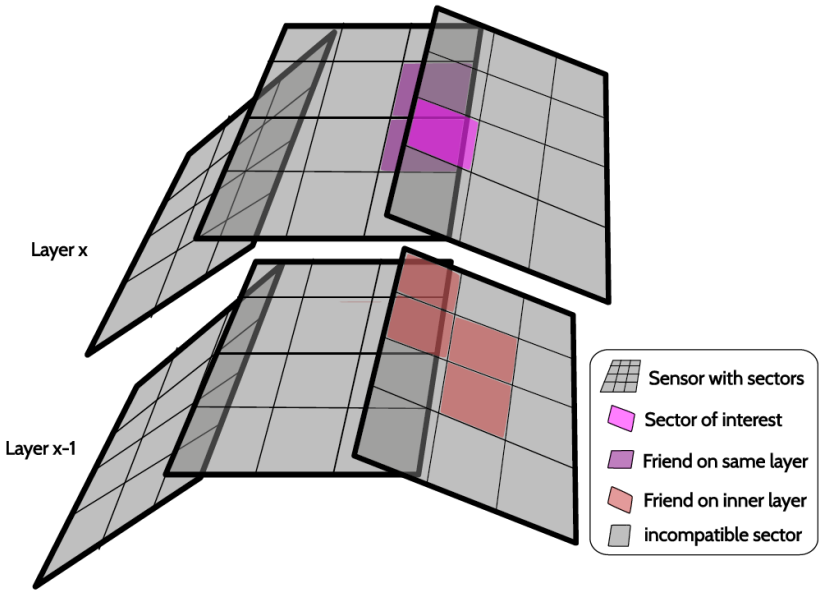



*Figure 4.4: Scretch of some sensor of two layer of SVD with Sector subdivision over them. An example of friendship relations is shown in case a hit is present on the "Sector of interest" (from [56]).*

#### 4.3.2.1 We find only what we look for

Before continuing to describe the structure and the workflow of VXDTF it is necessary to remark a concept. The use of simulated tracks to define the real track features that can be reconstructed

can be seen as a strong limit of the power of the Track Finder, and so on the possibility of the experiment to discover New Physics. In fact it may seem that strange to look for New Physics using "as filter" a Monte Carlo sample of data, and so perfectly known physics. Actually this is partially true, in the sense that *"we find only what we look for"* is the only strategy that it is possible to have, and in fact *all* the experiments do that: to reconstruct tracks it is necessary to give a definition of tracks, and so make some hypotheses on the interaction of the particle with the matter, then make some assumption on the interesting physics to build trigger, apply some cuts to build the events, and then reach the physical quantity of interest. Therefore a lot of physical assumption must be done before producing an interesting plot and make analysis, because it is impossible to look event by event what is happening in the collision making all possible speculation of New Physics. And even if would be possible to produce and look to an emulsion of every single event, we will be limited from the interaction of the produced particles with the photographic plate. In VXDTF case this interaction is the one coded inside the simulation software `Geant4`.

Nobody is expecting to see a $B$ meson decay in a single uncharged lepton only. Somebody is expecting that maybe $B$-meson decays that involve $\tau$ happens a bit more frequently that Standard Model expectation, maybe for some contributions from unknown physical channels.

#### 4.3.2.2 Training sample

The Sector Map is built analysing the distribution of the track-patterns inside the VXD, to be able to determine all the compatibility between Sectors and evaluate the filters threshold values. The sample of events that is used to train the Sector Map has a central role in its efficiency because it determines "how well the Sector Map knows" the events characteristic inside Belle II VXD.

In VXDTF this task is achieved by using a large quantity of $\Upsilon(4S)$ events, obtaining typical 11-tracks events with a momentum spectrum between 50 MeV/$c$ and 3 GeV/$c$. With this choice, for example, cosmic rays events should not recognized from the Sector Map because of the very different topology. Another important class of excluded events are the $K_s^0$: the charged daughter particles of this meson are not coming from virtual IP Sector because of the displaced $K_s$ vertex which is far from the interaction point

Finally, there is a class of tracks that can be treated inside the VXDTF framework but is not already implemented: the *loopers*, the tracks with sufficiently low momentum that loop inside VXD or first reach the CDC and then come back inside the VXD volume. This class of track, for their outward-to-inward connection, cannot be treated automatically like the other friendship connection, and in VXDTF implementation are discarded.

Another possibility that training offers is to build multiple Sector Maps for different classes of events, and then run on each event the filters of the various Sector Maps, removing from the combinatorial problem the hits associated to the tracks found after the application of each maps-filters in order to increase the speed of the procedure. Actually in VXDTF three Sector Maps trained in three different momentum ranges have been used: first the filters from high-momentum Sector Map are applied, then the mid-momentum filters and then the low momentum one (because, of course, it is easier to find the straight patterns in the detector).

### 4.3.3 Filters

The filters are a list of threshold values which define accepted ranges of some geometrical parameters of the patterns inside the VXD. These filters do not act on the hits but on hits relations to

reduce the combinatorial problem, not to reduce the data. The filters have various levels of complexity (single hit, two, three and four-hits filters) applied sequentially to each event processed by VXDTF.

The filters ranges, and so the minimum and maximum threshold values of the cuts are evaluated during the training of the Sector Map, filling the distribution of interest as described in section 4.3.2. In the following are described the various filters and their application. The Sector Map training is done both in fast or in full reconstruction in a preliminary phase and does not during the tracking step. In the workflow of VXDTF the filter application is the first step of the procedure.

**Single hit filter.** The first step is to assign to each data-hit (each Space Point) $p_\alpha$ of the event the relative Sector $s_i = \mathcal{S}(p_\alpha)$. Then for each active Sector $s_i$ the relative friendship relations $f_{ij}$ are picked up from Sector Map for each $j$, and is checked if there is at least one Friend Sector in the event. If the hit $p_\alpha$ is inside a MC-active Sector and if has at least one Friend the hit pass the filter and it is accepted. This filter is very fast, the time consumption linearly increases with the number of hits, but the rejected hits are almost zero: with a reasonable Sector Map all the Sectors of the VXD are MC-active (otherwise it means that there are useless regions of VXD!) and so for each Sector at least one Friend Sector always exists. Anyway, the result of this filter is the set of friendships $F$.

**2-hit filters.** For each friendship relation $f_{ij}$ of $F$ a 2-hit filter is defined. The filter is a function that from two Space Point returns a boolean that decides if keep the friendship relation induced by the two Space Point or to discard it.

$$\mathcal{F}^{(2)}_{ij}(p_\alpha, p_\beta) : P \to \text{Boolean} \quad \text{with } p_\alpha = \mathcal{S}^{-1}(s_i), p_\alpha = \mathcal{S}^{-1}(s_j),$$

Each filter to make the decision comparing the threshold values $V(\mathcal{F}^{(2)}_{ij})^{min}$, $V(\mathcal{F}^{(2)}_{ij})^{max}$ stored in the Sector Map with the relative variables evaluated from the two detected Space Points $p_\alpha$, $p_\beta$. Actually almost 50% of rejected are filtered from 3D distance requirement. The relations that pass this 2-hit filter are used to build the Segment $d_{\alpha,\beta}$ with the relative Space Points. Then for each Segment, are picked up from the Sector Map the possible Neighbours and so the set $N$ is built. An illustrative example of the geometrical region defined by the 2-hit filter is shown in figure 4.5.

**3-hit filters.** For each neighbour relation $n_{ijk}$ found in $N$ it is defined a 3-hit filter, a function that takes in input three Space Point and decides if to keep the neighbour relation induced by the three Space Point or to discard it:

$$\mathcal{F}^{(3)}_{ijk}(p_\alpha, p_\beta, p_\gamma) : P \to \text{Boolean} \quad \text{with } p_\alpha, p_\beta, p_\gamma \text{ collected from } \mathcal{F}^{-1}(f_{ij}), \mathcal{F}^{-1}(f_{jk})$$

where $\mathcal{F}^{-1} : F \to P^2$ is the function that from each $n_{ij}$ return the relative Space Points. The variables used in the filtering are the angular information (various plane, 3D), a rough estimation of the $p_\perp$ and rough estimation of helix parameters. Actually, to index the triplet filter only the informations of the two outer Sectors of the relation $n_{ijk}$ are used because the in the Sector Map it is not possible to store the information of three Sector relations. The relations that pass this 3-hit filter are used to build the Triplets $t_{\alpha,\beta,\gamma}$ with the relative Space Points.

**4-hit filters.** The four-hit filters are not applied immediately after the 3-hit filters, but the tracks that pass the 3-hit filter are collected by the Cellular Automaton, then the four-hit filters are applied only on four-hit tracks. In fact the VXDTF accepts tracks that have also a missing hit, and so only 3 hits are left to build the track candidate. The four-hit parameters are rough

estimation of $p_\perp$ and an evaluation of the shape of the track. Also in this case only the two outermost Sectors of the four hits are used to index the threshold values because of the limitation of the Sector Map. This is the last filter in order of complexity because all the procedure has been thought to be applied in HLT, and so without the two additional PXD layers information.

**High-Occupancy filter:** this filter can be applied after 2-hit and 3-hit filters. It uses as addition hit the virtual IP, and so it evaluates the 2-hit and 3-hit parameters between the virtual IP and the outermost hit of the Friends/Neighbours. Of course, this introduces a bias, because the production of the tracks is not for sure the IP but the use of this additional filter can be useful to reject a large number of track candidates in case of high occupancy events.

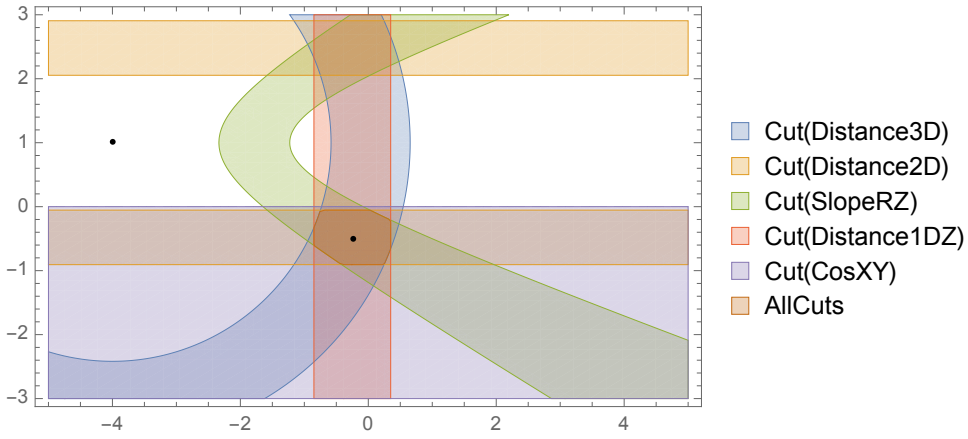



*Figure 4.5: Example of the geometrical regions defined on a single sector by the 2-hit filter. The two black dots represent the hit on the sector of interest and the hit on the friend sector, not shown (the lengths on exis are in arbitrary units).*

### 4.3.4 Collecting Track Candidates

After the application of the filters based on Sector Map algorithm, the remaining Segments must be collected building the complete path inside the SVD: it is necessary to join properly the Segments in order to find the most probably chains to produce the track candidates. The chosen method inside VXDTF is the Cellular Automaton.

#### 4.3.4.1 Cellular Automaton

The Cellular Automaton implemented in the VXDTF uses as Cell the Segments and then a simple procedure of updating and selection is defined. A segment $d_{\alpha,\beta}$ is defined *neighbour* of a segment $d_{\beta,\gamma}$, if a neighbour relationship between them is present inside the Sector Map, so if the second segment is composed of hits in the same or inner layer of the first one and they share one hit. To each Cell is assigned a initial state $s(0) \equiv 0$, then the procedure follow 2 steps in loop:

1. **Check Step:** for each Cell if there is at least one neighbour with the same state. If yes the cell is marked as "to be updated"
2. **Update Step:** for each Cell marked as "to be updated" the state is incremented by one: $s(t+1) = s(t) + 1$

The two steps are repeated until the states of all the Cells become stable. In the final situation there are outer Segments with higher states and inner Segments with lower states: the track candidates length depends on the maximum states reached in the chain. The Track Candidate collection is done with the following procedure: is performed a loop over all Cells and if a Cell has a state $s > s_{thr}$, where $s_{thr} = 2$ is a threshold value (to be able to collect chain of three or longer chains), the algorithm starts to collect the cell and the next inner neighbour with a state $s - 1$. The loop continues until there is no more inner neighbour with $s - 1$ state. If at some steps more than one neighbour fulfils the requirement both are collected defining two different Track Candidates, but after collection, the four-hit filter may cut some branches.

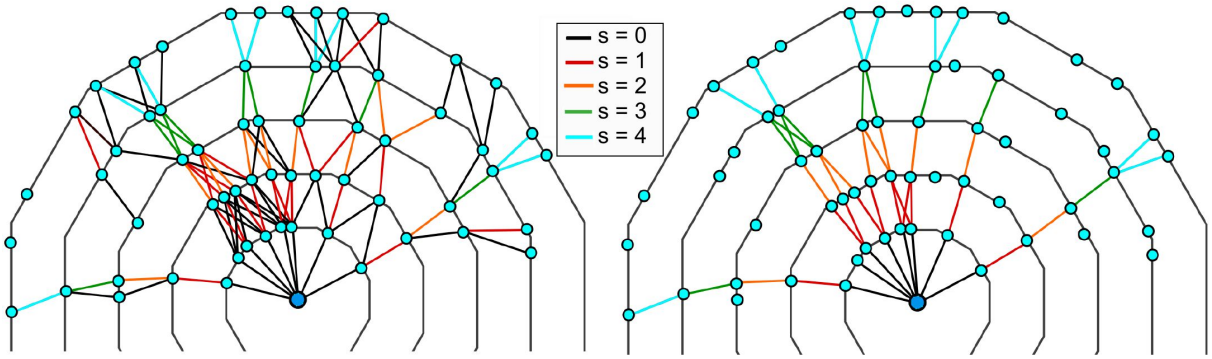



*Figure 4.6: The Cellular Automaton state assignment (on the left, after the loop procedure, when stable state has been reached) and the collection of best paths (on the right, using the procedure with the starting $s_{rhr}$. Notice that not all the consecutive Segments are neighbour, but the filters requirement removes some combination, and so the state assignment on the left is realized. In addition in the final situation there are some branched tracks: they can be removed by four-hit filters or in the overlap removing step. The detector is not the VXD, but a 5-layer detector used only to show the behaviour of the Cellular Automaton (from Belle II tracking group studies).*

### 4.3.5 Overlap Remover

Despite the geometrical constraints imposed by multiple filters there is a not negligible probability that two track candidates collected by the Cellular Automaton share one or more hit. Physically one of these tracks is a fake, and so it is necessary to define a selection criterion to keep only the true of the *overlapped* tracks. Inside the VXDTF are defined two methods to remove the overlapping tracks: the simpler methods is based on a quality index of the tracks, and a more sophisticated procedure combines the quality index with a Hopfield Neural Network. The choice between the two methods can be done during the configuration of the reconstruction.

#### 4.3.5.1 Greedy algorithm

To apply this overlap remover it is first necessary to assign a *Quality Indicator* (QI) to each track, a single real number between 0 (bad track candidate) and 1 (good track candidate) that can sort by quality the track candidates. There are a lot of possibilities to define this indicator: the simplest is the length of the track candidate, or in absence of magnetic field the p-value of a straight line fit. In presence of magnetic field, a circle fit on transverse plane can be used instead of the linear fit. A helix fit is not used inside the VXDTF because of some instability due to very bad track candidates issues (very low QI), but it is used to evaluate the momentum seed of the track candidate (this momentum seed is used to define the ROI on PXD layers). Another possibility is to use a Kalman Filter to obtain a QI, but in VXDTF it is not used because a

suitable Belle II Kalman Filter was not developed and the default one is too slow for HLT time constraints. In conclusion, the circle fitter is the best method provided by VXDTF to evaluate the QI of each track.

The Greedy algorithm sorts all the overlapping track candidates by descending QI, then for each track checks all the overlapped tracks and remove them from the list (they have a QI lower than the first one), then descend the list and repeat the procedure. The algorithm is very fast but the efficiency is low and so it is not used in most of the cases (in fact by default the Hopfield Network is used).

#### 4.3.5.2 Hopfield Neural Network

Another more refined overlap remover in the VXDTF implements a Hopfield Neural Network (HNN). In general in this adaptive method the state is defined by the neurons $s_i$ which evolve with the dynamics:

$$s_i = \mathrm{sgn}\Big(\sum_j T_{ij} s_j\Big)$$

where $T_{ij}$ is the weight of the connection between $i$-$j$ neurons and sgn the sign function[3]. The purpose of the algorithm is to minimize the energy function:

$$E = -\frac{1}{2}\sum_{i,j} T_{ij} s_i s_j$$

varying the configuration of neurons. The strategy is iterative, and the purpose is to reach the minimum of an energy surface. To avoid the issue to stop in a local minimum it is usually introduced a temperature function $T$, that smooths the energy surface to a simpler one with a clear global minimum. This temperature parameter is decreased each iteration to show up the real energy surface only when the iteration process is near convergence; this strategy is called *annealing*.

The purpose of the HNN is to obtaining the optimal subset of compatible neurons, from a set of compatible and not compatible neurons. In the network the neurons represent the track candidates, instead compatible means not overlapped and incompatible means overlapped; overlapped tracks are nodes (neurons) connected by an edge.

The compatibility between neurons is mapped with a weight matrix $n \times n$, where $n$ is the number of neurons of the network. The element $w_{ij}$, that describe the weight between the $i^{th}$ and $j^{th}$ neuron and its value is:

$$w_{ij} = \begin{cases} -1 & \text{if } i \text{ and } j \text{ are not compatible} \\ \frac{(1-\omega)}{n} & \text{if } i \text{ and } j \text{ are compatible} \end{cases} \tag{4.6}$$

This weights steer the behaviour of the HNN and enter in the state evolution of the network, introduced later on. With the weight proposed in 4.6 it is guaranteed that the absolute value of a weight for two incompatible neurons is higher than the weight for two compatible neurons. The tuning parameter $\omega \in [0, 1]$ is used to favour subset with the highest number of compatible neurons ($\omega \to 0$) or sets with the best sum of QI ($\omega \to 1$). By default in VXDTF is set $\omega = 0.5$, a trade-off between the two limit cases. The procedure to minimize the energy of the network is

[3] $\mathrm{sgn}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$ .

the following: the initial condition of each neuron is a random state picked up from a uniform distribution in the interval $[0, 1]$. The state $s_i$ of each neuron is:

$$s_i = \frac{1}{2}\big(1 + \tanh(a_i/T_k)\big),$$

where the $a_i$ is the input from other neurons plus a bias QI-dependent:

$$a_i = \sum_{j=1}^{n} w_{ij} s_j + \omega_i(QI). \tag{4.7}$$

$T_k$ is the temperature parameter used for the annealing of the HNN. This parameter starts from $T_0 = 3.1$ and the evolution over the iteration $k$ is:

$$T_{k+1} = \frac{1}{2} T_k + T_{\min},$$

where $T_{min} = 0.1$.

The update of the neurons is asynchronous to avoid oscillations of the whole network between two global states [58]. The iteration of the HNN continues until no neurons change the state value by more than 0.01, then all the neurons with a state higher than 0.75 are accepted in the subset of non-overlapped tracks candidates, instead the other neurons (tracks candidates) are rejected.

The performance of the HNN presented above is higher than the greedy algorithm, therefore has been used as a standard method of overlap removing in the VXDTF.

### 4.3.6 Multiple passes

It is possible to run the VXDTF multiple times over the set of hits of an event, using different Sector Map trained and optimized for different classes of tracks. In the standard reconstruction procedure the VXDTF uses three momentum ranges. The idea is to reduce the combinatorial issue by finding the easiest tracks before, then their hits are reserved to make simple the pattern recognition of the following steps. Each pass of VXDTF sorts the found track by their QI, and then a parameter defines the percentage of hits that must be reserved, and so excluded from following passes. The standard value of this percentage is 60%.

The performances of the VXDTF are presented in section 4.5 after the description of VXDTF2 to allow a comparison between the two track finders.

## 4.4 Track Finder redesign: VXDTF2

The VXDTF is a software of pattern recognition fully implemented, but the performances are not completely satisfactory. Because of that, in 2014, the Belle II tracking software started a complete refactoring of VXDTF, called VXDTF2. The Belle II-Pisa tracking group has a primary role inside this redesign programme, and finally in 2017 the VXDTF2 reached a stable level, with superior performances than VXDTF and with more refined features. It was tested both in simulations and in real data-taking during the Testbeam of February-2017. Despite these promising results the VXDTF2 is still a work in progress, and is going to be improved until end of 2017 before Belle II - Phase 2 is going to start.

During my thesis period, I contributed both in software redesign, as described in 5, and in testbeam VXTF2 data analysis, as described in chapter 6.

### 4.4.1 Motivations and requirement

The Belle II Track Finder in VXDTF implementation needed a redesign for several motivations. In the first place the tracking performances, in term of efficiency of fake tracks rate, are below the expectation. For instance, the final tracking efficiency with VXDTF is lower than the BaBar one at $p_\perp > 100$ MeV/$c$. Conceptually this is not a critic motivation to write another version of the track finding software, because there would be time to improve the performances before the start of data-taking of Belle II. Unfortunately, the VXDTF has been written in a monolithic way: a single module of more than $4.4 \cdot 10^4$ lines of code mainly developed from a single person. The software is very hard to maintain for the developer himself because of the reached complexity level, and the flexibility is limited. In this framework to add new features or to modify the existing ones is a hard work, and is some case improvement the software without redesign the entire code are not possible.

The Belle II tracking group made the decision to develop a new version of the track finding software with a different approach, because of these features of VXDTF. The idea is to port the working features of VXDTF in a new software, more flexible and with a modular structure. During the redesign, the bugs, the limits and the missing parts of VXDTF have been completed.

In the end, the main developer of VXDTF left the Belle II collaboration, making an optimal maintenance of VXDTF code impossible, so that the development of VXDTF2 became one of the main goals of Belle II tracking group.

The first requirement of VXDTF2 is to reach the performances of the first version, providing a modular design and then to implement some missing features of the latter. Among the most relevant ones are:

- store inside the Sector Map the threshold values correctly indexed for three four-Sector combinations.
- optimize the multi-passes performances in term of the number of passes and Sector Map type.
- optimize the number and the geometry of the Sectors on each sensor, also in term of the position of the sensor in the VXD.
- develop a studied training of the Sector Map to optimize the performances (kind of events, track properties, size of Training Sample...)
- study and optimize the quantiles of filter parameters that define the thresholds.
- develop more refined quality estimators and overlap remover tools, like the helix fit or Kalman Filter.

### 4.4.2 Implementation of VXDTF2

The monolithic approach is abandoned and a multi-module approach is used, as is shown in figure 4.7. The idea behind the structure is the same of VXDTF, but a general improvement of this modular approach is the possibility to switch between various modules (and so between various tracking solution at different steps) at configuration level, with big freedom in the choice of configuration parameters, and so enable the possibility to change methods between various physical scenarios to optimize the performances.

First of all the informations from the detector, PXD and SVD, are converted in a detector-independent hit format, called `SpacePoint`. Then the `Segment Network Module` applies to the

hits the $n$-hit filters from the Sector Map and then stores the directed graphs of Space Points, Sectors or Segments. After the application of the filters the Space Points are collected by the Cellular Automaton to form the first preliminary Track Candidates (called `SpacePointTrackCandidates`, or SPTC). Then the overlap remover is applied to remove fake tracks in two steps: first an SPTC Network is created, where the overlapping SPTC are mapped and QI informations are collected, secondly the actual overlap remover (the Hopfield Network) is applied to produce the final Track Candidates. Before the conclusion, the repetition of all the procedure is allowed, if multi-passes Sector Maps are provided.

Some of the main differences from VXDTF implementation are listed and described with more details in following paragraphs, without follow the complete reconstruction workflow, conceptually identical to the first version of VXDTF.

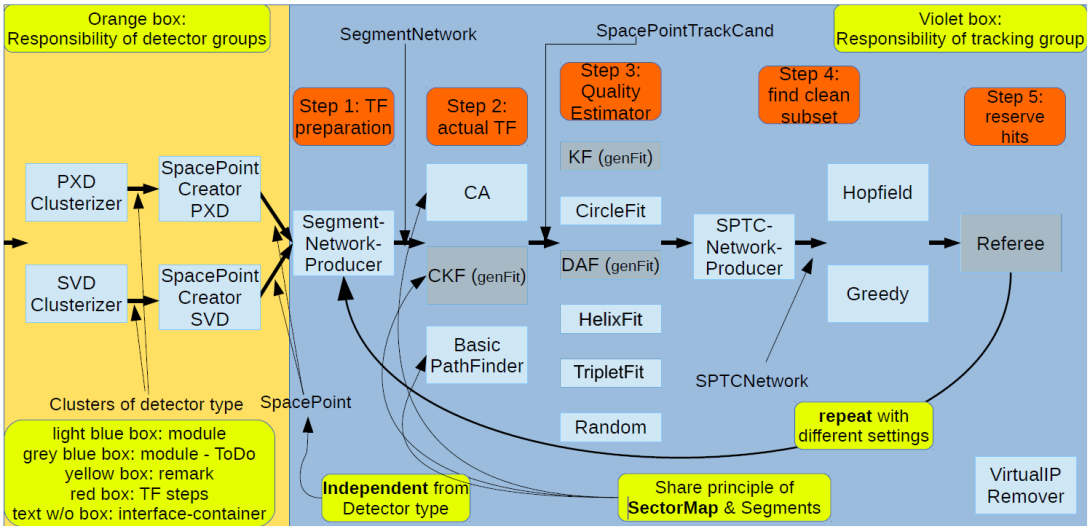



*Figure 4.7: The structure of VXDTF2; the modularity is shown with the incomplete modules are shown too (marked in gray). Input data preparation (out of VXDTF2, in orange box) is shown too, instead the Sector Map training steps are not shown (from [56]).*

#### 4.4.2.1 Segment Network Module

This module represents one of the most relevant upgrade of VXDTF2. It is able to manage the Space Point and apply filters of different complexity level and then produce a versatile container to store the connection between Space Points, Sectors of Segments. In this way, for example, it is possible to use specialized filters for three and four-hit filters correctly indexed with three or four hit Sectors relations, increasing the efficiency of the cuts (anyway by default the four-hit filter is disabled because it results inefficient in VXDTF2). The container is filled with directed graphs that can be used from different algorithm of pattern recognition (different, for example, from default Cellular Automaton).

#### 4.4.2.2 Training Sample and quantiles studies

A new selection of the Sector Map Training sample was introduced. The detail of this new approach is given in chapter 5. Previously the quantiles 0.1% and 99.9% of various geometrical distributions that set the threshold of the $n$-hit filters are removed, and the thresholds are selected as the minimum and the maximum values of the variable distributions. The reason for this choice is detailed explained in section 5.1.1.

#### 4.4.2.3 Combinatorial Kalman Filter

The Combinatorial Kalman Filter method, presented in section 4.1.3 can be used in Belle II track finding as an alternative method to VXDTF. The can CKF use as seed the CDC track and then extrapolate with the Kalman Filter update step the track inside the VXD layers.

The implementation combined with VXDTF2 framework is still under construction and not completely usable. The current development of CKF shows that it can be used as an alternative Track Finder and it can be applied in series with VXDTF2. Two different ways and under study: apply CKF before the VXDTF2, to reduce the combinatorial issue and increase the efficiency or apply CKF after the VXDTF2 to reduce the fake rates. At the current level of development, it is not possible to describe the CKF and its application with more details.

#### 4.4.2.4 Quality Indicators

Various new quality indicators are developed in VXDTF2: the p-value of helix fit (complete 5-parameter fit to the complete 3D helix) and the p-value of triple-fit (fit to the Triplet) and the output of a neural network with use as input the properties of the track (number of hits, cluster information, momentum estimation...).

#### 4.4.2.5 SPTC Network Module

The SPTC Network Module has the task to collect the preliminary Track Candidates and remove the overlapped tracks. It is faster than the VXDTF overlap remover, it allows different definitions of *overlapping tracks*:

- two SPTC that share a single cluster (clusters are the data-hit of each side of SVD sensor, two clusters form a Space Point. Clusters and Space Points coincide for PXD.)
- two SPTC that share a Space Point.
- two SPTC that share a Segment.

Study on the best definition to optimize the performance must still be done. The Hopfield Neural Network continues to be the default choice for the overlap remover method.

#### 4.4.2.6 High Occupancy cut

The VXDTF uses a *high occupancy cut*, that automatically discards events with more than $5 \cdot 10^3$ overlapping Track Candidates or more than $5.5 \cdot 10^3$ Segments. This cut is necessary to make VXDTF's overlap remover work (otherwise it crashes), but it introduces a bias in the event reconstruction. In fact these "high occupacy events" may occur for several reasons (particular topology, more loopers particles, more than average beam-induced background). If real tracks are present, they are discarded too. The VXDTF2 develops an overlap remover that is able to analyse these events too, even if with an increase of fake rate.

Unfortunately, the removal of the high occupancy cut brings to light loops inside the Sector Maps: two Sector connected in both directions, an unexpected feature of the Sector Map, that breaks the "tree" hypothesis needed from the Cellular Automaton to collect tracks (that may fall in an infinite loop state). The Sector Map has stored these connections anyway because no cuts are applied on the Training Sample, but the removal of the high occupancy cut make this connection realized during the track reconstruction. In figure 4.8 it is represented an example diagram of the Sector Map connection, that shows a loop. Each node is a Sector and each arrow

represents the oriented directions, from the last hit to the first hit (Cellular Automaton collection direction).

The presence of loops inside the Sector Map makes the VXDTF processing time increase, and some specific studies are needed to remove the loops. For example, the Training Sample selection could in principle reduce or remove the presence of looping connections, as discussed in 5.5.

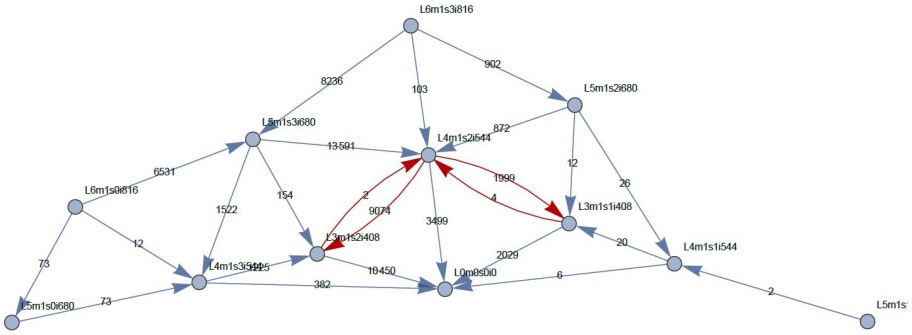



*Figure 4.8: Sector connection diagram of a simplified geometry detector (only a single $\phi$-Sector of VXD) realized with 1000 $\Upsilon(4S)$ events on the VXDTF2 default Sector Map. It is possible to notice a loop between two sensors (in red). The format of the sector name is Lllmaasbbicc, where ll is the layer ID, aa is the module number, bb is the sensor number and cc is the sector number. The numbers on the edge correspond to the number of tracks in the Training Sample connecting the two sectors (from Belle II tracking group studies).*

## 4.5 Performance comparison

One of the primary goals of VXDTF2 is to ensure the performance of the first version, and this result was reached in the first months of 2017. In this section, these latest results are presented (Sector Maps, track finder modules, standard validation scripts updated to August 2017). All these results are evaluated on Monte Carlo simulated samples. Instead the performance of the Track Finders on real data, with the application of fast reconstruction inside the HLT of Belle II, is presented in chapter 6 where testbeam data has been used.

There are three parameters that are important to evaluate the performance of the Track Finder:

**Tracking efficiency:** the number of correctly reconstructed tracks over the number of Monte Carlo tracks that are present inside the detector (the geometrical acceptance is factored out). A *correctly reconstructed* track is defined as a track found by the Track Finder having more than 2/3 of hits matched with a single Monte Carlo Particle. Therefore: $\varepsilon_{\text{tracking}} \equiv \frac{N_{\text{reco }>2/3}}{N_{\text{tot}}}$. The efficiency can be evaluated before fitting (pattern recognition or finding efficiency) or after fitting (total tracking efficiency) step. In this case, the additional requirement is that the fitting procedure converges.

**Fake rate:** the number of fake track over the number tracks that are reconstructed by the track finder. A track is defined *fake track* if it is found by the Track Finder and less than 2/3 of its hits are matched with a single Monte Carlo particle. Therefore: $\mathcal{R}_{\text{fake}} \equiv \frac{N_{\text{fake}}}{N_{\text{reco}}}$, usually expressed as a probability or a percentage. This value can be evaluated before or after the fit. In the latter case are counted both in numerator and denominator only the tracks in which the fitting procedure converges, and typically $\mathcal{R}_{\text{fake}}$ is reduced (the fit improves the quality of the track sample). Notice that $\mathcal{R}_{\text{fake}}$ and $\varepsilon_{\text{tracking}}$ are not complementary, because $N_{\text{tot}} = N_{\text{reco}} + N_{\text{miss}} - N_{\text{fake}} = N_{\text{reco }>2/3} + N_{\text{reco }<2/3} + N_{\text{miss}} - N_{\text{fake}} = N_{\text{reco }>2/3} + N_{\text{miss}}$, therefore the two parameters contain different information.

**Time consumption:** the CPU processing time needed by the Track Finder module only to reconstruct a single event. It is useful to distinguish between time needed for pattern recognition and time needed for track fitting step.

The comparison between two track finder results biased from some difference in configuration. Unfortunately, it is not possible to remove this bias because the configuration of VXDTF is not accessible because of this monolithic structure, and some features of VXDFT2 are not implemented yet. First the VXDTF uses three passes reconstruction with three different Sector Maps trained in different momentum ranges (0.35 MeV/$c < p_\perp <$ 100 MeV/$c$, 100 MeV/$c < p_\perp <$ 400 MeV/$c$, 400 MeV/$c < p_\perp$). Instead for VXDTF2 the multiple passes are not yet implemented, so the track finding is done in a single pass with a single Sector Map. In particular, the Training Sample of VXDTF2 Sector Map is composed by 0.9 million of $\Upsilon(4S)$ events. In addition, the VXDTF uses an *high occupancy cut*, instead in VXDTF2 this cut has been removed as described in the previous section, but it is not completely compensated with a proper Sector Map study. The specialized training of VXDTF and the high occupancy cut reduces the combinatorial background and improve its ability to reject false combinations with respect of a single pass tracking, therefore fake rate and the time consumption are expected higher in VXDFT2 than VXDTF.

The performances are evaluated testing the track finders on $10^4$ $\Upsilon(4S)$ events and in the following

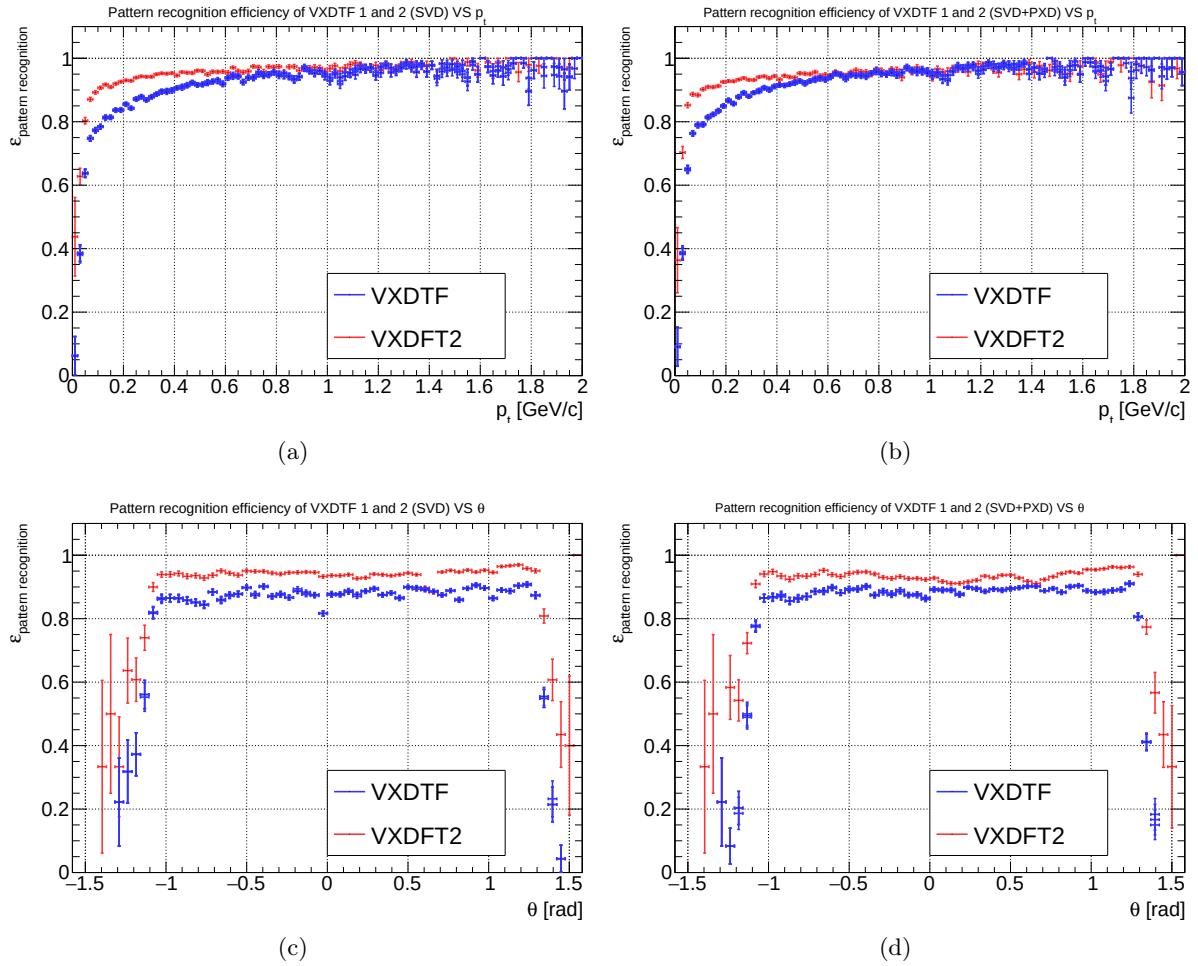



*Figure 4.9: Comparison between pattern recognition efficiency of VXDTF and VXDTF2 with SVD only (4.9(a), 4.9(c)) or SVD+PXD (4.9(b), 4.9(d)) tracking in $p_\perp$ and $\theta$ trends.*

plots a redefinition of $\theta$ has been used: $\theta \in [-\pi/2, \pi/2]$, is the angle between the momentum and the transverse plane, where $\theta = \pi/2$ in the forward direction. This choice has been to keep the coherence with the Belle II standard validation plots.

In figure 4.9(a) is shown the $p_\perp$ behaviour of the pattern recognition efficiency of the two Track Finders, with the geometrical acceptance factored out: the VXDTF2 has a quite higher efficiency in particular at low transverse momentum (below 400 MeV/$c$) and the use of PXD (figure 4.9(b)) increases the efficiency of both track finders, with a larger impact on the VXDTF, because of the more optimized training. Looking at the pattern recognition in bins of $\theta$ the VXDTF2 is globally more efficient. Both track finders show degradation at larger angles due to the geometry of the detector: the VXD acceptance is $\theta \in [-1.05, 1.27]$, therefore the tracks out of this range are secondary particles, for which the efficiency is expected reduced. The integrated efficiency values are (Full Reconstruction):

$$\varepsilon^{VXDTF}_{\text{patt.rec.}} = (93.5 \pm 0.2)\%$$
$$\varepsilon^{VXDTF2}_{\text{patt.rec.}} = (95.5 \pm 0.1)\%$$

The fake rate distribution as a function of $p_\perp$ is shown in figure 4.10(a) and 4.10(b), plotted only below 0.5 GeV/$c$ to better see the low momentum shape (over 0.4 GeV/$c$ the fake rate is

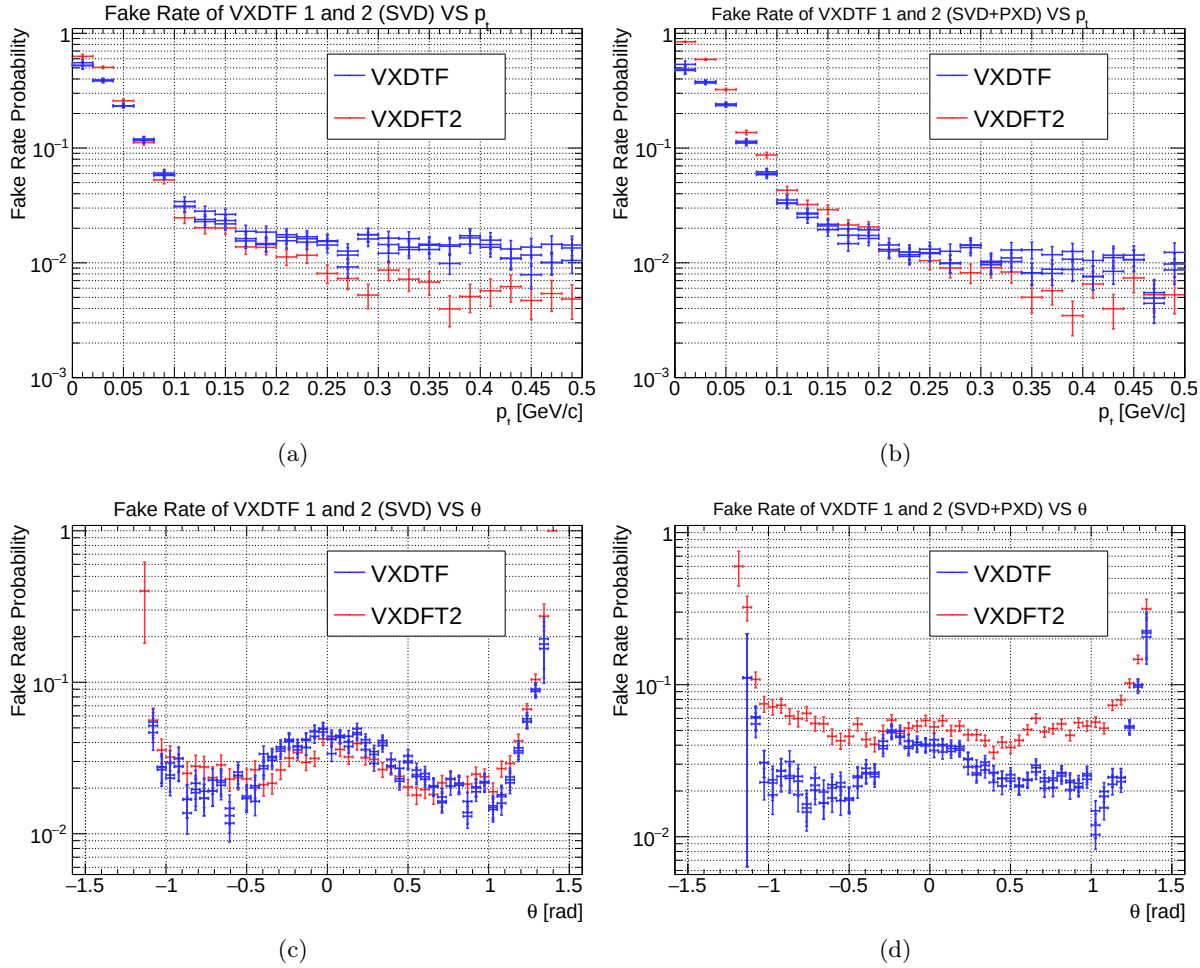



*Figure 4.10: Comparison between fake rate of VXDTF and VXDTf2 with SVD only (4.10(a), 4.10(c)) or SVD+PXD (4.10(b), 4.10(d)) tracking in $p_\perp$ and $\theta$ trends.*

quite constant). The fake rate is evaluated before the fit and $p_\perp$ is the momentum seed given to the fit algorithm. For both Track Finders the mean fake rate is below 2%, except in the low momentum region: from 100 MeV/$c$ to 20 MeV/$c$ the fake rate arise from 10% to 40% $\div$ 60%. The inclusion of PXD degrades the performances as expected for more hits tracks. This effect is particularly relevant for low momentum tracks of VXDTF2. The $\theta$ analysis of fake rate shows a small increase around $\theta = 0$, where are present more tracks, and degradation also at very large $\theta$, due to the geometry. In this case too the VXDTF has quite better performances than VXDTF2, in particular with the inclusion of PXD.

In the end, it can be useful to analyse the behaviour of the Track Finder when the track fitting step is applied. In figure 4.11 the distributions of fitting-only efficiency are plotted, defined as the number of correctly reconstructed tracks in which the fitting procedures converges over the number of correctly reconstructed tracks. In this way the inefficiency due to the fitting step only, that affects the overall tracking efficiency, is isolated. From figure 4.11(a) it is possible to see that the $\varepsilon_{fit}$ is very close to 1 except below 70 MeV/$c$, and anyway it remains over 80% down to 25 MeV/$c$. The inclusion of PXD (figure 4.11(b)) improves the performances, and for both the distributions the VXDTF2 has quite better performances than the VXDTF. In $\theta$ distribution (figure 4.11(c) and 4.11(d)) the improvement of VXDTF2 is more evident.

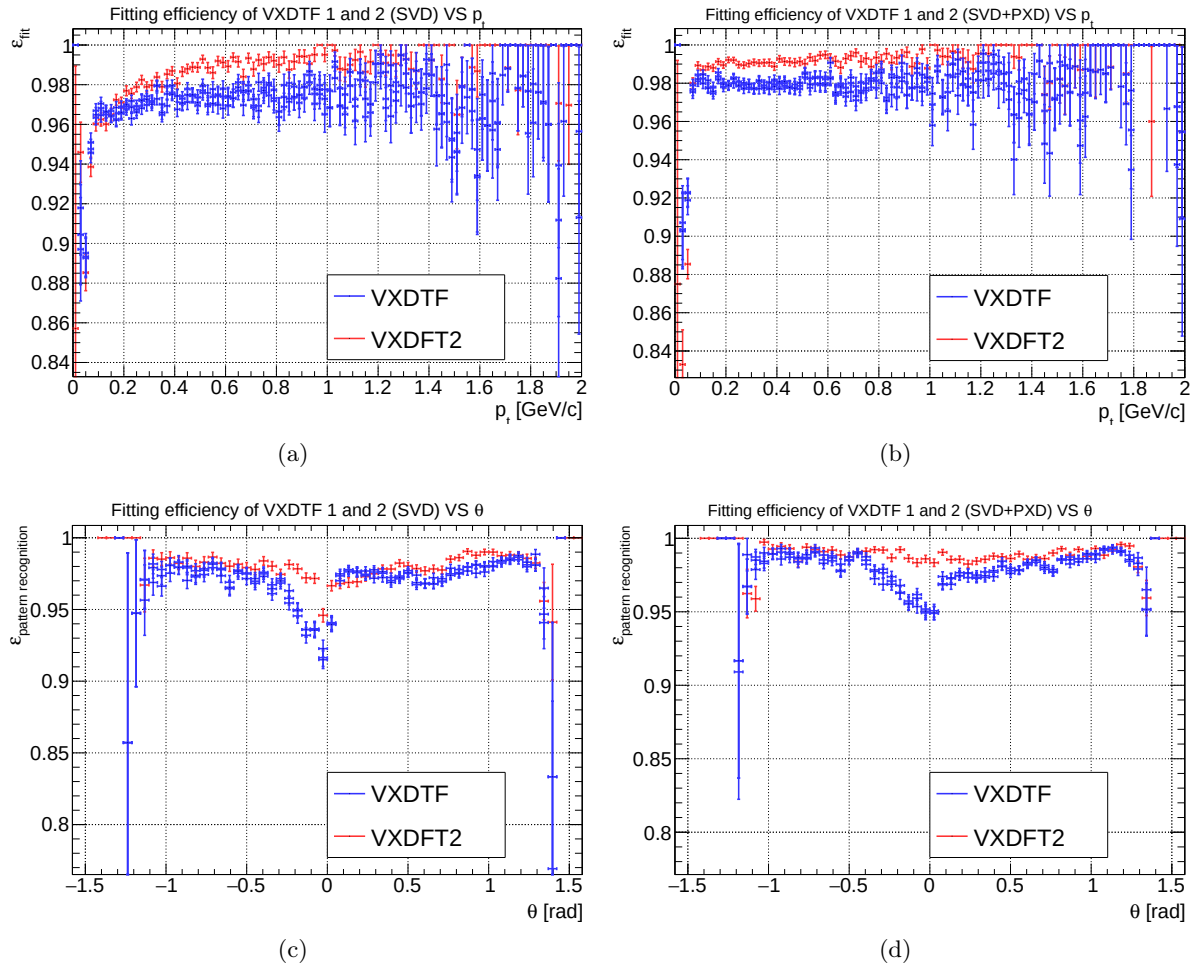



*Figure 4.11: Comparison between fitting efficiency of VXDTF and VXDTf2 with SVD only (4.11(a), 4.11(c)) or SVD+PXD (4.11(b), 4.11(d)) tracking in $p_\perp$ and $\theta$ trends.*

The total tracking efficiency as a function of $\theta$ and $p_\perp$, evaluated as the product of the pattern recognition efficiency and the fitting efficiency, is reported in the appendix A, and it is coherent with the expectations from previous results.

Finally the mean time consumption of the single module of VXDTF is

$$t_{VXDTF} \simeq 3.6 \text{ ms/event (SVD only)}$$
$$t_{VXDTF} \simeq 9.3 \text{ ms/event (SVD+PXD).}$$

Instead for VXDTF2 the sum of the time consumption of various modules is

$$t_{VXDTF2} \simeq 10.6 \text{ ms/event (SVD only)} \tag{4.8}$$
$$t_{VXDTF2} \simeq 32.3 \text{ ms/event (SVD+PXD).} \tag{4.9}$$

The VXDTF2 is slower by a factor 3 with respect to the first version. In both cases, the time required for the fitting procedure, external to the Track Finder, is not considered. The main modules which contribute to this time consumption are the overlap remover and the Segment Network module. The cellular automaton module, to a lesser extent, uses also a relevant fraction of the time.

These times are evaluated as a mean time per event from the total CPU time of the validation, but the fluctuations of the single event time are up to $100 \div 200\%$, both for VXDTF and VXDTF2.

One of the critical points that must be improved is the reduction the fake rate before the application of the overlap remover, to reduce the computation burden of these modules.

The time required for the fitting procedures only is:

$$t_{VXDTF} \simeq 229 \text{ ms/event (SVD only)}$$
$$t_{VXDTF} \simeq 340 \text{ ms/event (SVD+PXD)}$$
$$t_{VXDTF2} \simeq 270 \text{ ms/event (SVD only)}$$
$$t_{VXDTF2} \simeq 425 \text{ ms/event (SVD+PXD).}$$

These values make completely negligible the pattern recognition time consumption, and they are on the upper limit of the time constrains of the HLT ($\sim 300$ ms). Therefore it is fundamental, on one hand optimize the fitting procedures, and on the other hand improve the pattern recognition in term of fake rate, so that provide the fitting module tracks as clear as possible.

### 4.5.1 Summary of tracking performances

The performances of the VXDTF2 are satisfactory: the efficiency is reasonable high in whole momentum range and the efficiency of VXDTF, the primary goal, has been reached and passed. The fake rate is low enough, but the VXDTF has still better performance, thus there is room of improvement in particular for low momentum tracks. The reduction of the fake rate can improve strongly the time consumption too. In fact the time consumption is close to the upper limit of HLT time requirement, in particular because of the fitting step. The high performances of pattern recognition are uphold from a very high fitting efficiency: almost all the tracks pass the fitting steps revealing the goodness of the pattern recognition.

# Chapter 5

# Definition of the Training Sample

The Belle II group of Pisa is strongly involved in tracking development, and during the late 2016 and 2017 one of the main tasks of the group was the redesign of VXD standalone track finder software (i.e. the development of the VXDTF2). One of the open points since the design of the first version of the track finder is how to define the sample needed to train the Sector Map in order to optimize it for various physics scenarios.

In this chapter the results of some studies on the optimization of the training procedures are presented. The issues of the Training Sample definition are described together suitable strategy to solve them. In this context a set of selection variables is introduced together with the procedure developed for the Training Sample selection. In the final part of the chapter the results of some test on Monte Carlo dataset are presented to show the performances of the VXDTF2 with the selected Training Sample.

## 5.1 Sector Map status

Conceptually the Sector Map represents the set of connections between Sectors together with the geometrical feature that a charged particle tracks has inside the Belle II inner tracker. The Training Sample is the set of simulated tracks used to teach to the software of pattern recognition which Sector connections are expected and so which connections must be present inside the Sector Map.

It would seem at first glance that the best choice for the Training Sample is to make it as big as possible in order to reproduce all the possible physical connections inside the inner tracker, taking care also of all very rare events. In this case, the plan could be to produce a huge amount of simulated data and then to train the Sector Map with this sample. This was the idea during the development of the first version of VXDTF, and the original plan was to use a Training Sample of hundreds millions of $\Upsilon(4S)$ events.

This approach has various drawbacks. First a large Training Sample allows the Sector Map to collect information from rare events, which make the geometrical requirements less stringent. Therefore a larger number of Track Candidates is produced, and so a larger number of fake tracks. Secondly the Sector Map complexity is increased, because of the inclusion of rare tracks that have not easily predictable patterns inside the detector. A complex Sector Map makes the Track Finder slower, and in principle it should not overcome the HLT requirement in the Fast Reconstruction. The rare tracks included in the Training Sample in this approach also allow the collection of Track Candidates that are actually not useful for tracking, because the tracking software will remove them away at a certain point of track reconstruction, for example at during

fitting procedure, when the tighter constraints are applied. In conclusion, the inclusion of these tracks inside the Training Sample, and so inside the Sector Map, does not largely increase the efficiency of the track finder, but might reduce the performance of the VXDTF2.

### 5.1.1 VXDTF version 1 strategy

In the first version of VXDTF the removal of these rare events from the analysis is devoted to the filters: the definition of the thresholds by the use of quantiles of parameters distribution (so the removal of the tails) has been done with this approach.

This solution is inefficient and unsatisfactory. First of all, because these events have been included in the Training Sample and so in the Sector Map and only in the application of the latter (in the real data reconstruction) the algorithm tries to exclude them, with a waste of disk space and processing time. Secondly this solution has a not well-controlled efficiency: the correlation among various filters has not been studied and it is not easy to foresee the overall tracking efficiency because these filters act on hits, segments or triplets. Moreover, a study on the definition of best quantiles has not been done, and so these filter selection has a not negligible probability to remove also useful tracks, introducing a bias in the reconstruction.

In view of these difficulties during the redesign of VXDTF it has been decided to select the Training Sample and the quantile-thresholds has been removed to define the threshold simply as maximum and minimum of the selection variables distributions, removing the bias.

### 5.1.2 A strategy to improve the performances in VXDTF2

One of the possible sources of patterns hard to fit is the multiple scattering effect (more details in paragraph 4.2.2). This interaction produces on low momentum particles large deviations from the ideal helical trajectory, with not easily predictable pattern inside the tracker. If the effect is too large the interested track should be removed from Training Sample because it may become one of the sources of time consumption of the VXDTF2 which does not increase the tracking efficiency, as discussed before.

The idea is to select a subset of $\Upsilon(4S)$ events, defining an algorithm to tag tracks with "*catastrophic* multiple scattering interaction" and to remove these tracks from the Training Sample. The purpose is to reduce the complexity of the Sector Map, with a consequent reduction of time consumption, without affecting the overall tracking efficiency.

The general strategy is composed of three parts. In the first part (section 5.2) the best variables that are able to distinguish the catastrophic interactions are identified and their properties are studied. In the second part (section 5.3) it is developed a software to analyse a large sample of simulated data to define automatically a list of selection criteria from the study of the tracks that are present in the sample. In this phase it is actually useful to use a huge amount of simulated data in order to exploit the properties of the variables chosen in the first phase. In the end of the analysis this list of selection criteria is stored permanently to be used for reconstruction also on a different sample of data. In the third part (section 5.4) of the strategy, these selection criteria are applied in a specific point of the Training Sample definition, as a filter in track selection. The starting Training Sample which must be filtered can be composed of the same simulated data that was used to define the list of selection criteria or can be another sample of simulated data, for example the default one. In fact the purpose of the definition of these selection criteria is to be general and independent from the specific events used in the selection criteria definition.

One of the main difference between this method of rare-events removal and the quantiles defined in the first version of VXDTF is that with this procedure is required a precise efficiency to

the selection criteria that define the Training Sample, so the percentage of the tracks that are discarded from the Training Sample and so from the reconstruction is exactly known. This is the guideline in the development of the procedure of the selection (in section 5.3.4.1 this required efficiency is described in detail).

It is necessary to remark that all the procedures that involve the Sector Map uses Monte Carlo simulated data only, both now and when real data will be available: the Training Sample is composed by simulated data, the procedure of selection need to use Monte Carlo Truth information and also the procedure of training. As explained in paragraph 4.3.2.1 this is not a limit of the Track Finder.

The final result is a selected Training Sample and a Sector Map trained with this particular Training Sample. This new Sector Map can be compared with another Sector Map trained with the default Training Sample running the Full or Fast Reconstruction on the same simulated events and study the performances in two cases.

## 5.2 Selection variables

The first step is to define an algorithm to select the tracks with catastrophic multiple scattering interaction identifying which variables could be useful to distinguish these tracks from the more common tracks. Anyway it is necessary to introduce the object "track" inside the Belle II tracking and analysis software, `basf2` (belle analysis framework 2) before understand the best way to select catastrophic tracks.

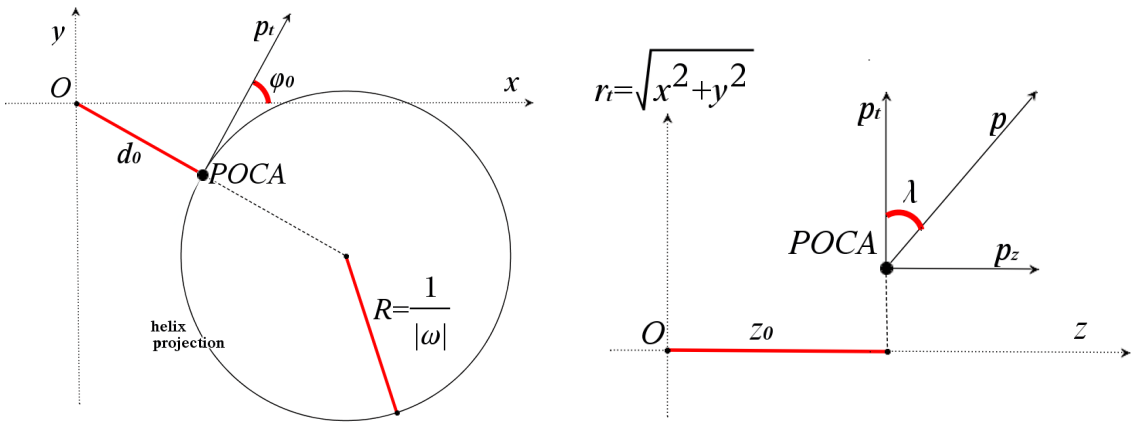
### 5.2.1 Helix parameterization

The concept of track inside the tracking software changes during the chain of reconstruction: the track is built starting from some hits inside the detector, and following the reconstruction it acquires step by step more "identity", and so physical properties, as a single object, the `RecoTrack`. Anyway, for the use inside the selection criteria definition procedure a low level definition of track is needed: a track is a list of hits inside the VXD detector connected using the Monte Carlo Truth information (inside `basf2` this is called "Monte Carlo Track Finder").

Ideally a track, as a charged particle inside a constant magnetic field, describes a helix. To describe this helix have been chosen the following 5 parameters with respect to the POCA, the *Point Of Closest Approach* between the helix and the $z$ axis:

- $\omega \in [-\infty, \infty]$, is the inverse of the radius of curvature, signed with the charge of the particle
- $d_0 \in [-\infty, \infty]$, is the signed distance of the POCA with respect to the $z$ axis (the *impact parameter* of the track), where the sign is the same of the $z$ component of angular momentum at POCA with respect to the origin.
- $\varphi_0 \in [-\pi, \pi]$ is the angle defined by the $p_\perp$ at the POCA and the $x$ axis.
- $z_0 \in [-\infty, \infty]$ is the $z$ coordinate of the POCA
- $\tan\lambda \in [-\infty, \infty]$ is the tangent of the angle defined by the momentum at the POCA and the $x$-$y$ plane (the *dip angle* of the track).

In figure 5.1 these are shown in a graphic representation of the track projected on the transverse and the longitudinal plane.




(a) *Variables on transverse plane* (b) *Variables on longitudinal plane*

*Figure 5.1: Geometric representation of track variables on the transverse and the longitudidal planes. Momentum vector of the track at* POCA*,* POCA *and the track projection on the transverse plane are shown.*

With the use of the Monte Carlo truth information it is possible to analyze the track properties completely at hit level. A single hit contains the position $\mathbf{x}$ and the momentum $\mathbf{p}$ of the particle at that point (local momentum and position, from now on). From these informations it is possible to evaluate all the track parameters in that point using the following formulas, where $Q$ is the charge of the particle, $B_z$ the $z$-component of the magnetic field, and $p_i, x_i$ the components of momentum and position.

$$\chi \equiv \arctan\left[\frac{\mathrm{sgn}(B_z Q)\left(\frac{|\mathbf{p}_\perp|^2}{B_z Q} + p_y x_x - P_x X_y\right)}{\mathrm{sgn}(B_z Q)\left(-p_x x_x - p_y x_y\right)}\right] \tag{5.1a}$$

$$\omega = \frac{B_z Q}{|\mathbf{p}_\perp|} \tag{5.1b}$$

$$d_0 = \mathrm{sgn}(B_z Q)\left[\sqrt{\left(\frac{p_y}{B_z Q} + x_x\right)^2 + \left(x_y - \frac{p_x}{B_z Q}\right)^2} - \frac{|\mathbf{p}_\perp|}{B_z Q}\right] \tag{5.1c}$$

$$\varphi_0 = \arctan\left(\frac{p_x}{p_y}\right) - \chi \tag{5.1d}$$

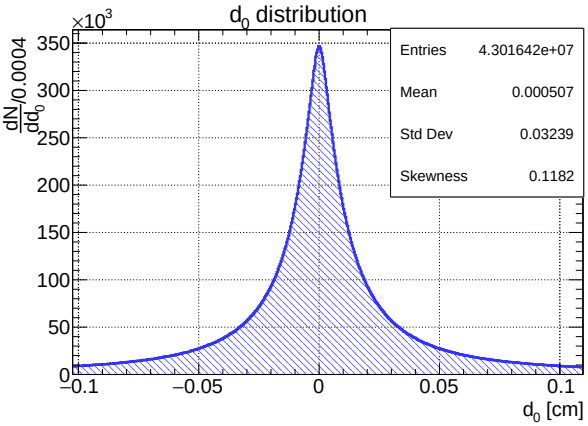
$$z_0 = \frac{p_z \chi}{B_z Q} + X_z \tag{5.1e}$$

$$\tan\lambda = \frac{p_z}{|\mathbf{p}_\perp|}. \tag{5.1f}$$

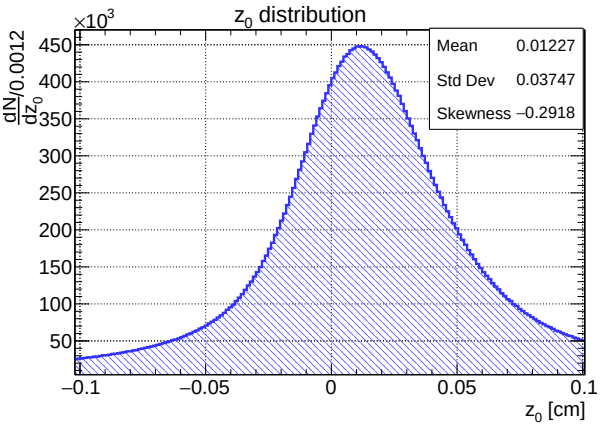
In conclusion, it is possible to evaluate the track parameters for each hit inside the VXD and so to study the evolution of the simulated track parameters inside the detector.

In the figure 5.2 are shown the distributions of the five track parameters. These distributions are realized from a simulated sample of $9 \cdot 10^5$ events of $\Upsilon(4S)$ and each distribution has been filled with the track parameter evaluated using formulas 5.1 for each hit inside SVD.

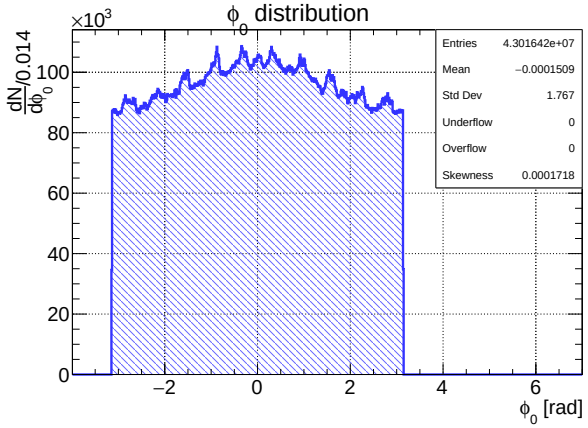
The two impact parameters, $d_0$ and $z_0$, are strongly peaked in 0, as expected: these tracks come mainly from the IP ($d_0 = 0$, $z_0 = 0$ by definition) or from its close proximity. The $z_0$ distribution shows a quite large asymmetry and the peak result about 100 $\mu$m far from the IP, and this is consistent with the boost of the center-of-mass. Both $d_0$ and $z_0$ distributions show long tails for positive and negative values. These tails and some interesting features visible in logarithmic scale only are described in details in section 5.3.2.



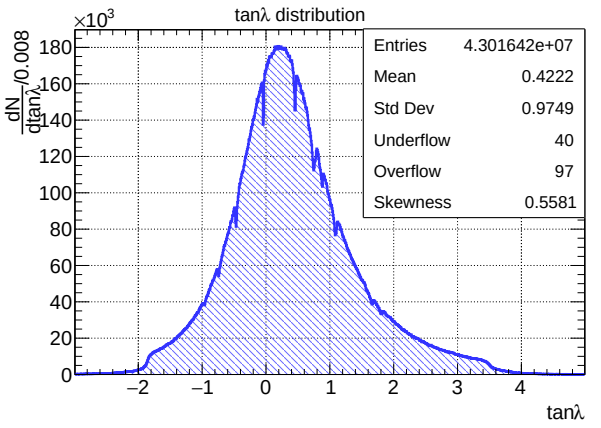

(a) $d_0$ *distribution around peak region.*



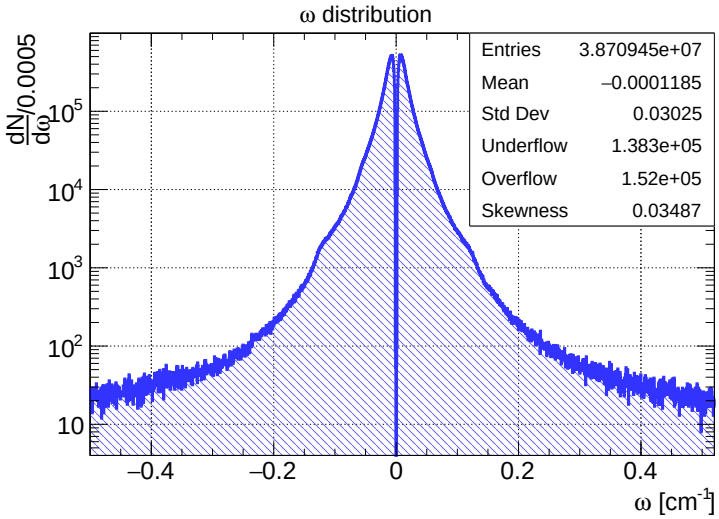

(b) $z_0$ *distribution, around peak region. The asymmetry and the displaced peak are visible.*




(c) $\varphi_0$ *distribution, the small peaks are due to overlap regions.*




(d) $\tan\lambda$ *distribution. The dips due to the secondary particle-hits are clearly visible.*




(e) $\omega$ *distribution, in logarithmic scale, to appreciate the long tails from low momentum particles that do not reach all SVD layers.*

*Figure 5.2: Track parameters distributions from SVD MC-Truth hits of* $9 \cdot 10^5$ $\Upsilon(4S)$ *events.*

The $\phi_0$ distribution is quite flat in all the range: the events $\Upsilon(4S)$ events are symmetric on transverse plane and the detector (i.e. the SVD material) is almost symmetric too. More events around $\phi = 0$ with respect to larger angles are visible, and then some small peaks on the top of the $\phi_0$ distribution. This is consistent with the geometry of the detector: as already said each entry of these histograms is a single SVD hit, thus in region with thicker material more hits are expected: the peaks correspond to the overlap regions of SVD, but the lack of symmetry in the ladder numbers of various SVD layers (7, 10, 12, 16) does not allow to identify clearly the ladder boundaries. The global behaviour with the maximum at $\phi = 0$ is instead due to the transverse boost of the $\Upsilon(4S)$.

The $\tan\lambda$ is strongly asymmetric with the maximum value around $\tan\lambda \simeq 0.42$, consistently with the geometry of the detector and the boosted center-of-mass. Various dips in $\tan\lambda$ shape are visible. These dips are due to lack of hermeticity of the detector in the small space between sensors accessible for secondary particles born inside the detector volume (in fact with the requirement of primary particles only they disappear, where primary means particles that came from IP region and not from decay of long-lived particles or from interaction between the latter and detector material). For these decays the geometry of the detector is not optimized and in some directions the acceptance is lower. The sum of primary and secondary SVD hits distributions shows these dips.

The $\omega_0$ distribution is not peaked in zero because $p_\perp = \frac{qB_z}{|\omega|}$, while the sign of $\omega$ is the charge of the particle, thus the distribution of $\omega$ is a direct access to transverse momentum distribution of the tracks. The $\omega$ of primary particles in confined in region $|\omega| < 0.513$ cm$^{-1}$, the minimum radius to leave hits in the inner layer of SVD. Anyway, larger value tails are present (not shown in the figure), due to secondary particles. A structure at $|\omega| \simeq 0.15$ cm$^{-1}$, or $R = 6.7$ cm, is visible: this is half of the radius of layer 6 of VXD, so the minimum radius (momentum) for a particle to leave at least one hit on each layer. Therefore for a larger value of $|\omega|$ the efficiency (in term of geometrical acceptance) decreases.

### 5.2.2 Track parameters variation along the track

Multiple scattering is a stochastic process and for Training Sample selection purposes in necessary to identify some rare multiple scattering events. For these reasons it is not possible to evaluate the global effect of multiple scattering after the passage of the particle inside the detector, but is necessary to study the evolution of the track inside the material to understand where the single catastrophic punctual process appears.

The parametrization of helix, function of local momentum and position only, is very useful to follow how the parameters change along the track. For a perfect helix these parameters must be constant along the trajectory. Instead for a real track, because of material effects, these parameters change. Multiple scattering contributes to this variation, and a strong modification of track parameter at a certain point of the track could be the signature of a catastrophic process.

The variables that have been chosen for the analysis are $\Delta X$, where $X$ is a generic one of the track parameters (i.e. $X = \{\omega, d_0, \phi_0, z_0, \tan\lambda\}$), and

$$\Delta X_n \equiv X_n - X_{n+1} \tag{5.2}$$

where $n$ means the layer where the parameter is evaluated from the $\mathbf{x}$ and $\mathbf{p}$ of the hit and the entry point of the layer. So, the chosen variables represent the variation of track parameters during the $n$th layer crossing.

This definition can be extended for a generic pair of layers, not strictly consecutive. This extension can be useful for tracks with a missing hit on a given layer or to be able to consider the

variation of parameters between a certain layer and another sensor of the same layer (because of overlap between sensors there are regions where a particle can cross twice the same layer). Furthermore it is useful to add a virtual hit at interaction point, called "Layer 0", to evaluate the difference of parameter in beampipe crossing.

The distributions of $\Delta X$ are expected to be quite symmetric around 0 also subdividing the charge distributions. In fact these parameters describe geometrical features of the track that multiple scattering increases or decreases with the same probability. An exception is $\Delta\omega$: this parameter, as $|\omega| = \frac{1}{R}$, is strongly connected with energy loss, in fact from formula 5.1b, $|\omega|$ can increase only in track propagation. Anyway, the multiple scattering can produce anomalous modification or radius of curvature and so $\mathbf{p}_\perp$ because of the change of direction, and so there are cases of $|\omega|$ decreases. Because of this asymmetry is going to be necessary a particular study to be able to include properly $\Delta\omega$ in the analysis.

Multiple scattering effects, from formula 4.4, has an explicit dependence on momentum $\mathbf{p}$ and angle of incidence $\theta$ of the particle. So it is necessary to study the variation of $\Delta X$ distribution in function of $\theta$ and $\mathbf{p}$, and not only in function of the chosen layers pair. In conclusion, the useful distributions to define the catastrophic multiple scattering interactions are:

$$\Delta X(\mathbf{p}, \theta, \text{first layer}, \text{second layer}).$$

From this point the notation will be: $\Delta X(\mathbf{p}, \theta)_{\text{first,second}}$.

## 5.3 Selection criteria definition and evaluation

After the identification of the set of variables that can measure the multiple scattering effects on the tracks it is possible to define the limits that is needed to put on these variables in order to select properly the Training Sample for the Sector Map. The naif idea could be is to look at distributions of $\Delta X(\mathbf{p}, \theta)_{n,m}$ and to put by hand reasonable cuts. This strategy is not practical because the number of distributions is necessary to analyse is large (consider 13 layer pairs, taking into account of the 6-layer VXD, the missing layer connections, and for instance 3 bins in $\theta$ and 10 bins in $\mathbf{p}$: over 300 distributions). These selection criteria obviously depend on the sample of simulated data used to look at distributions of $\Delta X$ therefore if they are set by hand they are not easily reproducible if the Sector Map must be trained on different Training Sample.

The better choice is to develop an automatic analysis of the $\Delta X(\mathbf{p}, \theta)_{n,m}$ distributions from a generic simulated data input integrated with a procedure of selection criteria definitions. A sample of 0.9 million of generic event of $\Upsilon(4S)$ decays is used as input.

### 5.3.1 Hit selection

The first step in the track selection procedure is the creation of an object that collects all the informations about the true trajectory of the particle, useful to evaluate $\Delta X$. This object is a `C++` vector of `hitXP` objects. The `hitXP` represents an interface between the reconstructed tracks of the Monte Carlo Track Finder and the procedure of selection criteria evaluation. The same interface is going to be used in the application of the selection criteria in the selection of the Training Sample. This vector actually represents a redefinition of reconstructed track inside the selection criteria definition algorithm, and from now on this is the meaning of *track*, or *hit* of the track.

The `hitXP` objects contains the position and the momentum of the hit in local and global coordinates, the sensor informations (sensor, ladder and layer number), the track parameters evaluated

as in 5.1 formulas, the information about the IP (momentum, track parameters at IP), the charge of the particle and its particle ID.

The hits on VXD are scanned along the track, then it is selected only the first hit, at the entrance of sensitive volume, on each layer and eventually the second hit if it is present on a different sensor from the first one. This selection is needed because in the sensor crossing are produced several hits, but the first one gives sufficient informations. The second hit is needed to take into account of overlap regions: if track cross an overlapped sensor create a hit on two sensors of the same layers, but the different sensor requirement allows to select only these patterns. The vector of `hitXP` is created with the selected hits. This vector can contain from 1 to 12 hits, ordered ascending in time, so consecutive elements in the vector corresponds to consecutive hit in the track. In figure 5.3 a sector of the SVD is represented with a crossing track and `hitXP` hits are shown to a better comprehension of the hit selection.

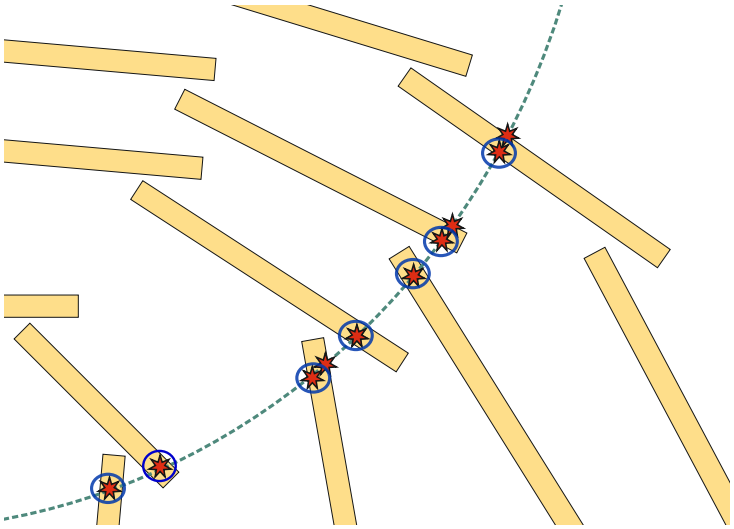


*Figure 5.3: A sector of the four layers SVD (yellow rectangles) with a charged track that crosses all the detector (green line). The red stars represent the hit inside the detector, the blue circles the selected hits inside this step of the selection criteria definition. Multiple hits selected are for overlap regions so that in this example from a 10-hits reconstructed track is built a vector of 7 `hitXP` elements.*

### 5.3.2 Global requirements

In section 5.2.1 has been noticed that the $z_0$ and the $d_0$ distributions have long tails. In figure 5.4 and 5.5 it is possible to see the distribution in logarithmic scale in a wide range. Focusing on the blue distribution of the figure 5.4, the major feature are the 8 clear peaks at radii corresponding to the layer.

The hit information is collected from the cluster list which form the Monte Carlo Track Finder reconstructed track. If a primary particle interacts with the material of the detector, for each interaction which produce more than one particle in the final state, the simulation software `Geant4` changes the unique ID of the particle after the interaction, despite in some cases (nuclear interactions, $\delta$-ray emission) the primary particle result mainly unchanged after the interaction. Unfortunately, an unhappy feature of `basf2` software merges in the same cluster hits from primary particle and its decay products, if the latter release hits close to the production point (i.e. inside the cluster of the mother), appearing as hits of the primary particle. Keeping in mind this feature, using MC-Truth informations it is possible to identify two main classes of hits which produce the 8 peaks in the $d_0$ distribution:

– last hits of primary particles ($K^{\pm}$, $\pi^{\pm}$, $e^{\pm}$, $\mu^{\pm}$, $p$) that interact with the material. This is the "last hit" in term of `Geant4` particle assignment: of course these particles are not stopped inside the material, but they simply interact and change their particle ID. The possible wrong assignment of the hits from the secondary particles, may produce a totally wrong evaluation of the track parameter, because the very low **p** of the secondary make the helix radius so small that the POCA results the production vertex itself and so the $d_0$ and $z_0$.

– first hit of secondary particles originated from the interaction of primaries with the material. As in the other class of hits the low momentum artificially produces $d_0 \simeq r_{\text{hit}} \simeq R_{\text{layer}}$ and $z_0 \simeq z_{\text{hit}}$ because of the very low momentum of these particles.

The red distribution of figure 5.4 contains just these two class of SVD hits for layer 4 only: all the last hits of particles that have their last hit on layer 4 and all the first hits of particles that has their first hit on layer 4. These hits reproduce quite well the peaking structure. In the blue distribution is possible to note a small peak at $d_0 \simeq 1$ cm too: these are the particles that interact with the beampipe material.

Because of the geometry of the detector the $z_0$ distribution does not show any evident structure at particular values (the sensors are parallel to the $z$-axis). Anyway, from figure 5.5 it is possible to note that the statistic decreases suddenly at $z_0 \simeq -25$ cm and $z_0 \simeq -35$ cm. These are the geometrical boundaries of the outer layer of SVD, so the distribution is coherent with the interpretation of $d_0$ tails and peaking structure.

These hits with very large $d_0$ and $z_0$ bias every study on $\Delta X$ distributions. These hits populate the non-Gaussian tails of the $\Delta X$ distributions and the Training Sample selection goal is to reject a prescribed small fraction of tracks laying away from the Gaussian core (i.e. the rare

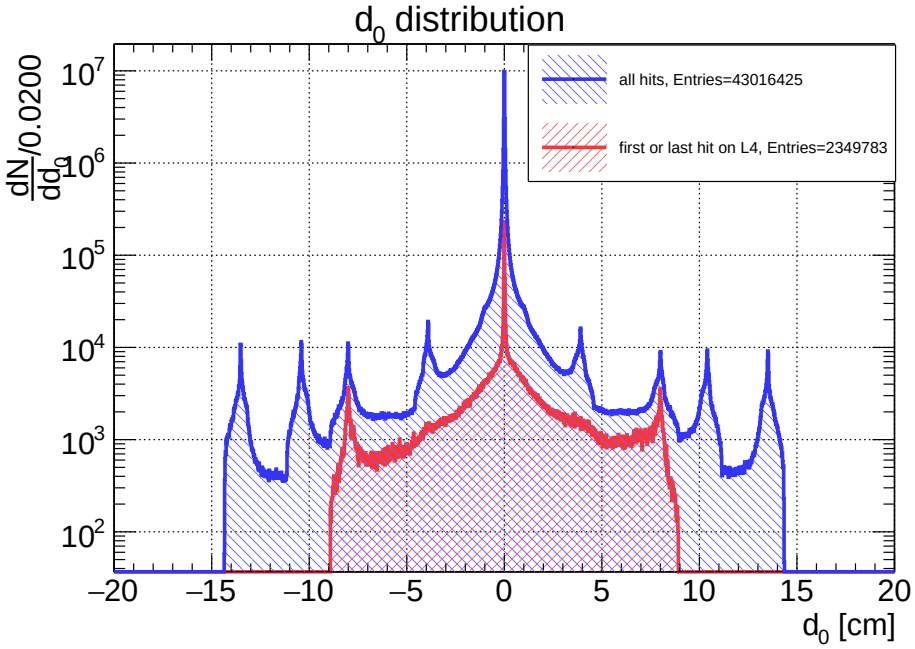



*Figure 5.4: Distribution of $d_0$ in logarithmic vertical scale. The blue distribution contains all the SVD hits from $9 \cdot 10^5$ $\Upsilon(4S)$ events. The red one contains only the hit from the tracks with the first or the last hit on layer 4. The peaking structure at layers radii are evident in the blue distributions, instead the red distributions reproduce the peak at layer 4 only, revealing the source of the $d_0$ peaks.*

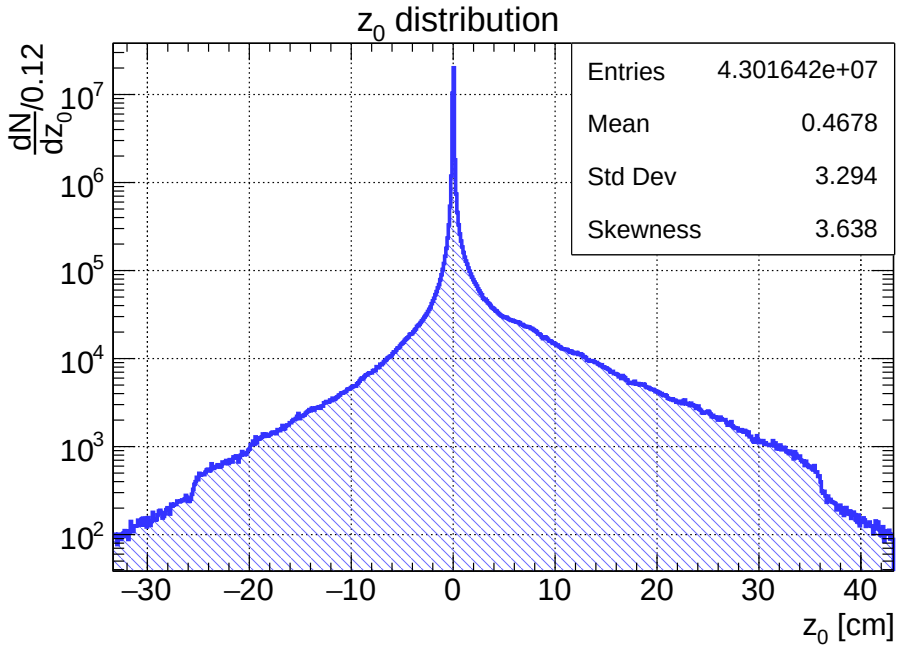



*Figure 5.5: $z_0$ distribution, in logarithmic scale, of all the SVD hits from $9 \cdot 10^5$ $\Upsilon(4S)$ events. It is possible to notice that the main part of the distribution is between the physical limit of the detector and the asymmetry, due to the geometry and to the boost of the center-of-mass.*

events), hence the inclusion of these hits artificially enlarge the acceptance range of $\Delta X$. In other words, the inclusion of these hits leads to an overestimate of the multiple scattering effect. If these tracks are removed a priori from the $\Delta X$ analysis the tails results very much lower and it is possible to evaluate properly the multiple scattering catastrophic interaction.

In order to exclude these particles from the analysis some *global requirements* must be satisfied by the tracks for the rest of the analysis, before the study in function of layer's pairs, of momentum and of $\theta$ angle. In case a track doesn't pass these requirements is automatically excluded from the analysis and its hits are not included in $\Delta X$ distributions used to define the specific cuts. These requirements are applied also in the procedure of the selection of the Training Sample.

The global requirements are:

- a track must have one hit on at lest three different layers of SVD (counting the overlap region as a different layer)
- for each hit in the track the local $|d_0| < 1$ cm,
- for each hit in the track the local $|z_0| < 1$ cm.

The first requirement is necessary to exclude tracks with two or less hit: they can't be fitted at all. The second and the third requirement reject the particles stopped inside the materials of the detector, in term of the discussed `Geant4` and `basf2` feature. 1 centimetre has been chosen to exclude the interaction with the beampipe. Another parameter that can be affected from these class of particles is $\omega$, but another requirement on it is not needed because the $d_0$ and $z_0$ requirement remove automatically the $\omega$'s long tails, due to the the correlation between these variables.

### 5.3.3 Specific selection criteria analysis

After the application of the global requirement on the tracks, a rank-5 tensor[1] of 1-dimensional histograms has been created to contain the histograms on which the *specific selection criteria* will be defined. The first index of the tensor is the track parameter, so it runs from 0 to 4. The second and the third index are the layer pairs, so both they run 0 to 6, where layer 0 is the IP. The fourth index is the momentum, divided in 20 is from 25 MeV/$c$ to 2 GeV/$c$. The lower limit is the acceptance boundary of VXDTF2. The upper limit imposed because over 2 GeV/$c$ the statistic becomes too low to to perform efficiently the analysis (and anyhow the multiple scattering effect is negligible at these momenta). The fifth index is the polar angle $\theta$, divided in 3 bins from 17 degrees to 150 degrees, the acceptance boundary of the SVD. Have been only used 3 bins because of the high statistic needed from this multidimensional analysis. In this way the total sample of simulated data is divided in $7 \times 7 \times 20 \times 3 = 2940$ histograms). Anyhow the $\theta$ behaviour is not critical in the specific selection criteria definitions: at 17 degrees, the worst case, from equation 4.4, $\frac{1}{\sqrt{\sin(\theta)}} \simeq 1.8$.

For each track, for each pair of consecutive hits, $\Delta X(\mathbf{p}, \theta)_{n,m}$ is evaluated for each parameter, and the appropriate 1D histogram is filled. To reduce the computational time of the procedure and the dimension of the file where the selection criteria distribution are stored, not all the layers pair distribution has been filled: denoted with $n$ =layer 1, and $m$ =layer 2, are filled only the combination with:

- $m \geq n$, so layer 2 external with respect to layer 1 (a track that is going away from the IP) or layer 2 equal to layer 1 (overlap)
- $m - n \leq 2$, so no more than one layer is missing between consecutive hits

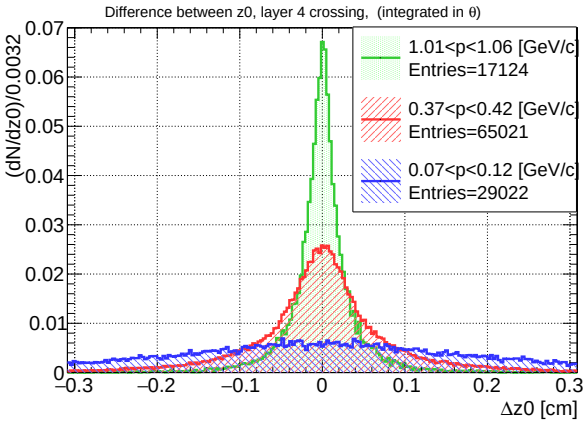
The various $\Delta X(\mathbf{p}, \theta)_{n,m}$ distributions share some typical behaviour that is relevant to remark. The same track parameter in the same layer crossing, in the same range in $\theta$, has a smaller width at higher momentum as expected. This is true for each parameter and for each layer crossing, an example in figures 5.6(a) is shown. This is consistent with our interpretation of multiple scattering interaction as main responsible of $\Delta X$ spread.

The same track parameter at the same layer crossing, in the same momentum range has quite the same spread for various $\theta$ intervals. This feature is quite expected because of $1/\sqrt{\sin(\theta)}$ angular trend of multiple scattering, quite constant for $\theta \simeq \frac{\pi}{2}$. In 5.6(b) this trend is shown.

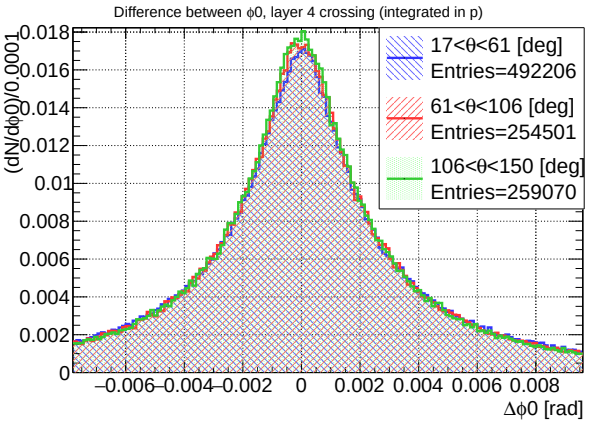
Looking at a given track parameter at a given $\mathbf{p}$ and $\theta$ range for different layer crossing, if $X = d_0$ or $z_0$, it is possible to notice that outer layer crossing has a larger width. For $X = \omega$ the behaviour is the opposite: outer layer crossing has a smaller width. Instead for $X = \phi_0$ and $\tan\lambda$ the distributions are quite similar for all the layer crossing. All these behaviours are expected: the width of material is quite the same for all the layer crossing, so the expected multiple scattering effect is roughly the same. However the same variation in local position has different effects on the track parameters in function of distance from interaction point: for angular variables ($\phi_0$ and $\tan\lambda$) the distance is not relevant, for $d_0$ and $z_0$ the arm and so the distance increase with the distance, for $\omega$ the effect is opposite because $|\omega| = \frac{1}{R}$. For instance, in figures 5.7(a) and 5.7(c) the behaviour of $\Delta d_0$ and $\Delta \tan\lambda$ is shown. The behaviour of the latter is coherent with the expectation, and the modulation is various layer-crossing is due to different material width. In fact for layer 4 and 5 the shape is almost the same, the layer 3 has a quite larger width due to the ladder material to keep it in position. The PXD layers instead are thinner than the SVD one, and the beam-pipe crossing has a larger one. Note that the effect

[1] tensor in term of a 5 indices object only

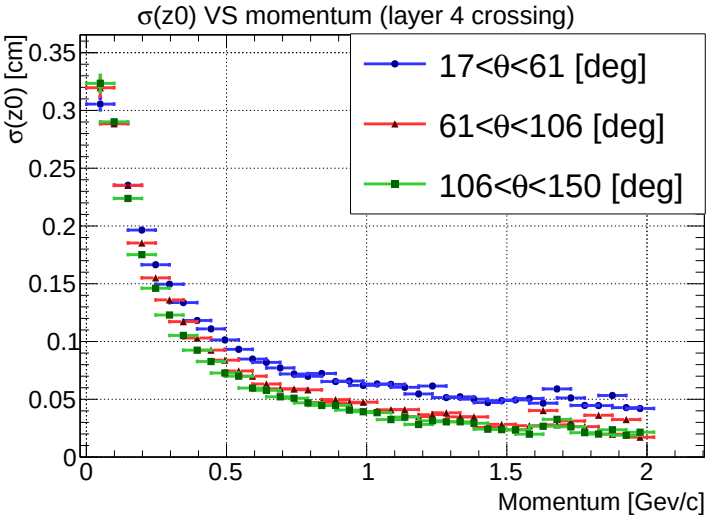
of the thickness of the beam pipe ($\sim 1$ mm) also dominates over the opposite effect short lever arm (figure 5.7(a)).




(a) *Distribution of $\Delta z_0(\mathbf{p},\theta)_{3,4}$ integrated in all $\theta$ bins, for three momentum bins: the spread increase with the reduction of $|\mathbf{p}|$.*




(b) *Distribution of $\Delta\phi_0(\mathbf{p},\theta)_{3,4}$ integrated in all $\mathbf{p}$ bins, for three $\theta$ bins: the spread is quite constant, except a small increasing for forward region*

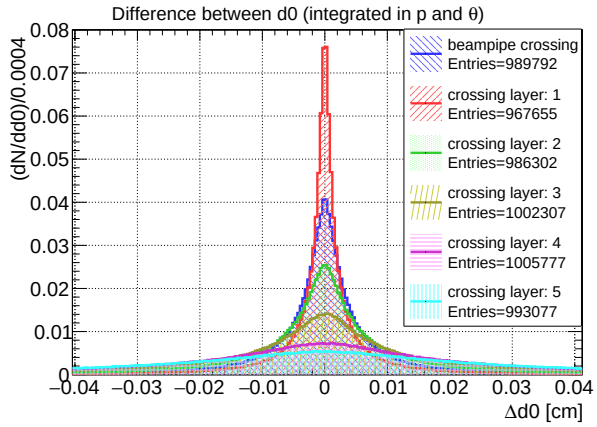



(c) *Momentum trend of $\sigma(\Delta z_0)_{3,4}$ distribution for three $\theta$ bins, for layer 4 crossing.*

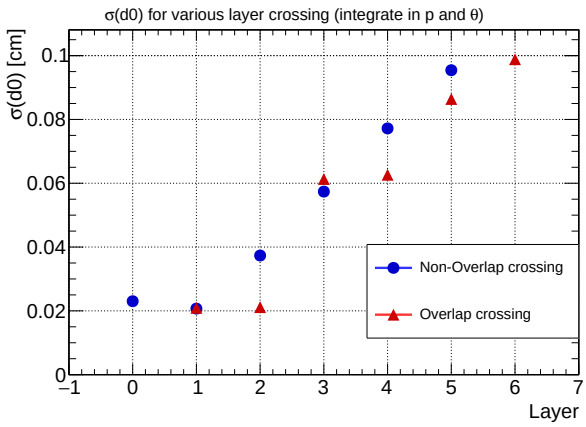
*Figure 5.6: Some features of $\Delta X(\mathbf{p},\theta)_{n,m}$ distributions in momentum and angular analysis.*

### 5.3.4 Selection criteria definition

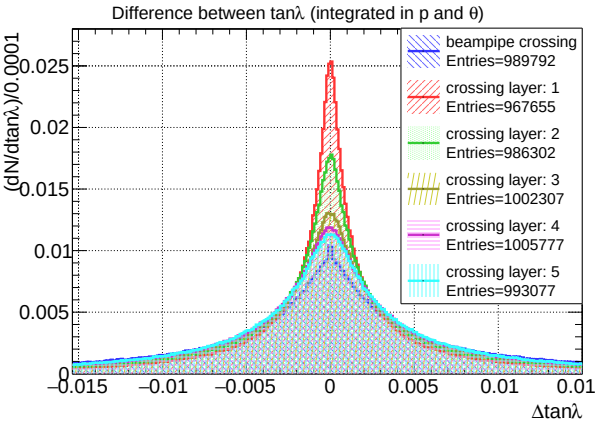
Once the multi-dimensional set of histograms has been filled they are analyzed to define the allowed ranges of $\Delta X$. Two different algorithms have been developed. The first one define the ranges by selecting a fixed fraction of the distributions. The second method fits the $\Delta X$ distributions and the ranges are defined from the fit parameters. The first method has been chosen in the definitive procedure because it is more stable, is easier to cross-check the results and because the efficiency is better under control. The following sections are dedicated to the first method, instead the second method is briefly described to show the possibility of a different approach as a final short section (5.3.4.4).



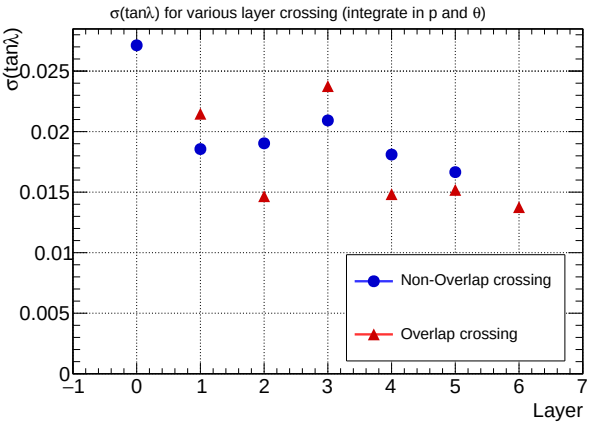

(a) *Distribution of $\Delta z_0(\mathbf{p}, \theta)$ for various layer crossing of VXD (excluding overlap), integrated in all $\theta$ and $\mathbf{p}$ bins: the spread increase for outer layers*




(b) *Trend of $\sigma(\Delta d_0)$ for various layer crossing: the spread increase for outer layers. Layer-crossing in overlap region are plotted separately.*




(c) *Distribution of $\Delta \tan\lambda(\mathbf{p}, \theta)$ for various layer crossing of VXD (excluding overlap), integrated in all $\theta$ and $\mathbf{p}$ bins: the spread is constant, weighted with the material of various layers*




(d) *Trend of $\sigma(\Delta \tan\lambda)$ for various layer crossing: the spread doesn't have an evident behaviour, but is consistent with material thickness only. Layer-crossing in overlap regions are plotted separately.*

*Figure 5.7: Some features of $\Delta X(\mathbf{p}, \theta)_{n,m}$ distributions analysing differences between various layer-crossing*

#### 5.3.4.1 *Cut-function*

The purpose is to develop a definition of the ranges of $\Delta X$ to reject a small fraction of tracks that represents very rare patterns. A reasonable choice is to define the allowed range so that the impact of the selection criteria fixed, and a known fraction of $\Delta X$ distributions in rejected. So, an efficiency curve in function of momentum it is imposed by hand: in table 5.1 the imposed efficiency points are reported. These points are fitted with a reasonable function that is able to reproduce well this dataset (actually has been used two functions, one over 40 MeV/$c$, another below this value). The final function, called from now on *cut-function*, defined from the fit is:

$$\varepsilon_{\text{req}}(p) = -\frac{7.5 \cdot 10^{-7}}{p^{3.88}} + 1 \qquad 0.041 < p < 1 \text{ GeV/c} \tag{5.3}$$

$$\varepsilon_{\text{req}}(p) = 6.3p + 0.57 \qquad 0.02 < p < 0.041 \text{ GeV/c} \tag{5.4}$$

With the imposed cut-function, in the of the momentum range over 200 MeV/$c$ the inefficiency

of the single cut is lower than 1 per mil. With this value a rough upper bound of the total inefficiency of the selection criteria is feasible: in the worst case on every track 60 selection criteria are applied (5 parameters and 12 layer crossing), so the overall inefficiency should be less than 6%. This is an overestimation of the inefficiency: is rare that a track crosses 6 overlap regions and moreover the selection variable (i.e. track parameters) are not completely uncorrelated. On the other hand if the selection criteria affect mostly particles with momentum below 200 MeV/$c$ the inefficiency for each selection criterion rises.

| $\mathbf{p}$ [GeV/$c$] | $\varepsilon$ | $N_\sigma$ |
|---|---|---|
| 0.2 | 0.999 | 3.3 |
| 0.1 | 0.995 | 2.8 |
| 0.07 | 0.98 | 2.3 |
| 0.05 | 0.9 | 1.6 |
| 0.04 | 0.84 | 1.4 |
| 0.03 | 0.75 | 1.1 |
| 0.02 | 0.7 | 1.0 |

*Table 5.1: Required efficiency to define the cut-function. If the $\Delta X$ distributions are treated as Gaussian with standard deviation $\sigma = 1$ and mean $\mu = 0$, a selection with efficiency $\varepsilon$ obtained by defining the acceptable ranges as the interval $[-N_\sigma \cdot \sigma, N_\sigma \cdot \sigma]$.*

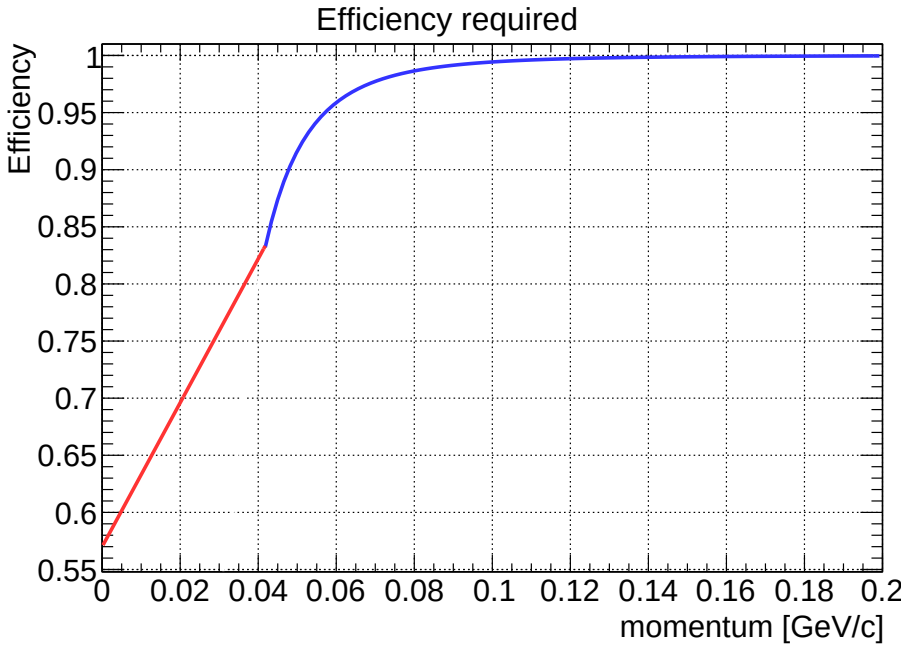



*Figure 5.8: Result of cut-function from the fit of required efficiency points. The blue fit function is $\varepsilon = -\frac{A}{p^B} + C$ (fit over 41 MeV/$c$) and the red one $\varepsilon = Dp + E$ (fit below 41 MeV/$c$). Two fits has been performed separately.*

#### 5.3.4.2 Algorithm of definition

The allowed ranges of $\Delta X(\mathbf{p}, \theta)_{n,m}$ are defined once the cut-function $\varepsilon_{req}(\mathbf{p})$ has been defined. For each distribution two threshold values are defined: a maximum value $\mathcal{C}_M$ and a minimum

value $\mathcal{C}_m$. These two thresholds are defined as the values of $\Delta X$ for which:

$$\int_{-\infty}^{\mathcal{C}_m} \mathscr{P}(\Delta X(\mathbf{p},\theta)_{n,m}) d\Delta X = \frac{1-\varepsilon_{\text{req}}}{2} \tag{5.5}$$

$$\int_{\mathcal{C}_M}^{+\infty} \mathscr{P}(\Delta X(\mathbf{p},\theta)_{n,m}) d\Delta X = \frac{1-\varepsilon_{\text{req}}}{2}, \tag{5.6}$$

where $\mathscr{P}(\Delta X(\mathbf{p},\theta)_{n,m})$ is the pdf of $\Delta X$ obtained from the histogram. The two values have not been imposed symmetric with respect of the central bin of the distribution, in fact $\mathcal{C}_M$ is evaluated starting from overflow bin and then include bins of the right tail of the distribution until the integral of the tail reach $(1-\varepsilon_{\text{req}})/2$. The same procedure has been applied for $\mathcal{C}_m$ starting from underflow bin.

The $\Delta\omega$ analysis the threshold values $\mathcal{C}_M$ and $\mathcal{C}_m$ are defined on $|\Delta\omega|$ distribution because the sign of $\Delta\omega$ is due to the charge only, and the hypothesis of vanishing most probable value is not true anymore. Instead the "rare events" are contained in the tails $\Delta\omega \to 0^{\pm}$ and $\Delta\omega \to \pm\infty$, therefore the tail of $|\Delta\omega|$ only contains these events.

With the use of described procedure of selection criteria definition, in case the distribution of $\Delta X$ is strongly asymmetric with respect to 0, the two values turn to be asymmetric, because all the histograms have been created with the same symmetric limit with respect to 0. This is the case, for instance, of $\Delta\omega$ distributions. The choice to make the histogram limits symmetric is the only reasonable one: a priori, before the filling of these histograms is not possible to foresee the distribution real limits. For this reason reasonable limits from the distribution of various parameters (i.e. $X$ distributions) has been chosen. These set of limits, one for each parameter, are relevant in the accepted range definition, in fact from the procedure exposed before, the limits on the histogram represent a upper limit on the value of $|\mathcal{C}_M|$ and $|\mathcal{C}_m|$.

Various studies have been performed in order avoid "saturation effects" on the limits of threshold values: if the limits of the histograms have chosen too close to 0 may happen that the underflow and the overflow bins contain more than $(1-\varepsilon_{\text{req}})/2$. In this situation the values $\mathcal{C}_M$ and $\mathcal{C}_m$ "saturate" systematically at the limits of the histograms. On the other side, the limits of the histograms cannot be set too wide because it would require a too fine binning in term of the dimensions of the histograms (in this algorithm about $1.5 \cdot 10^4$ histograms of 5000 bins each have been created). The limits that have been chosen are:

- $\pm 1$ cm for $\Delta d_0$ and $\Delta z_0$: by definition of the global requirements
- $\pm 0.4$ cm$^{-1}$ for $\Delta\omega$: it contains all the $\omega$ distributions after global requirements applications, equivalent to a radius of 2.5 cm (PXD only radius is removed away from global requirements)
- $\pm 0.3$ for $\Delta\phi_0$ and $\Delta \tan\lambda$: it contains most of the $\varphi_0$ and $\tan\lambda$ distributions.

To validate these limits an artificial file containing all the selection criteria have been produced, where all the threshold values have been set to the limit-value ($\mathcal{C}_{M,m} = \pm 1$ for $\Delta d_0$ and $\Delta z_0$, $\pm 0.4$ for $\Delta\omega$, $\pm 0.3$ for $\Delta\phi_0$ and $\Delta \tan\lambda$) independently from layers pair, momentum and $\theta$ angle. The result of the selection is reasonable, with the inefficiency of this selection is about 4% i.e. 4% of the tracks are removed from the Training Sample with the application of specific selection criteria.

Another limit of this procedure comes from the limited statistic. In case of low statistic $\Delta X$ distribution, the histograms could actually look like isolated spikes in tails. This will cause strong fluctuations in the evaluation of $\mathcal{C}_M$ and $\mathcal{C}_m$. In addition in case of low statistic the

tails of $\Delta X$ distributions do not reproduce the rare events and so the correct the values of $|\mathcal{C}_m|$ and $|\mathcal{C}_M|$, and they are going to be underestimated to reach the value of the $(1-\varepsilon_{\text{req}})/2$. For these reasons a big sample has been used to define the accepted range, and the binning in $\theta$ and $\mathbf{p}$ has been chosen large. To make the algorithm more stable if one of the analysed $\Delta X$ distributions contains less than 100 entries the two values $\mathcal{C}_m$, $\mathcal{C}_M$ are set automatically to the underflow and the overflow bin centre, i.e. to the limits of the histogram. This cure follows a conservative principle: it is better to include in the Training Sample a not trackable track (as the track with catastrophic multiple scattering interactions) than to exclude it from the Training Sample, because is prevalent the requirement of high efficiency of the VXDTF2.

The threshold values $\mathcal{C}_m$, $\mathcal{C}_M$ evaluated with this procedure are stored inside a pair (for minimum and maximum) of rank-5 tensor (the 5 index are the track parameter, the first and the second layer, the momentum and the $\theta$ angle, as before).

#### 5.3.4.3 Fit

The coarse binning in momentum and $\theta$ angle could cause discontinuities in the selection of the tracks. To obtain a smooth efficiency curve in $\mathbf{p}$ and $\theta$ for the specific selection criteria application the best choice is to be able to apply the selection criteria without binning these variables. To obtain this feature a pair of rank-3 tensor of 2-dimensional histograms is created. The indices of the tensor are the track parameters, the first and the second layer (the pair is for maximum and minimum cuts, of course). The two indices of the histograms are momentum and $\theta$ angle. Then this 2-dimensional histograms are properly filled from 5-dimensional tensor of threshold values defined in previous section.

With an appropriate model for $\Delta X$ distributions in function of $\mathbf{p}$ and $\theta$ it is possible to fit these 2-dimensional histograms and so obtain a continuous parametrization of the threshold values. The model that has been developed come from directly from multiple scattering behaviour in $\mathbf{p}$ and $\theta$, and so equation 4.4. Inside the $\Delta X(\mathbf{p},\theta)_{n,m}$ spread can enter more effect than simple multiple scattering so some generalization has been applied. In conclusion, the fit function that has been used is:

$$\mathcal{C}_{m,M}(p,\theta,n,m) = \frac{A_{m,M}(X,n,m)}{(p)^{B_{m,M}(X,n,m)}\sqrt{\sin\theta}} + C_{m,M}(X,n,m) \tag{5.7}$$

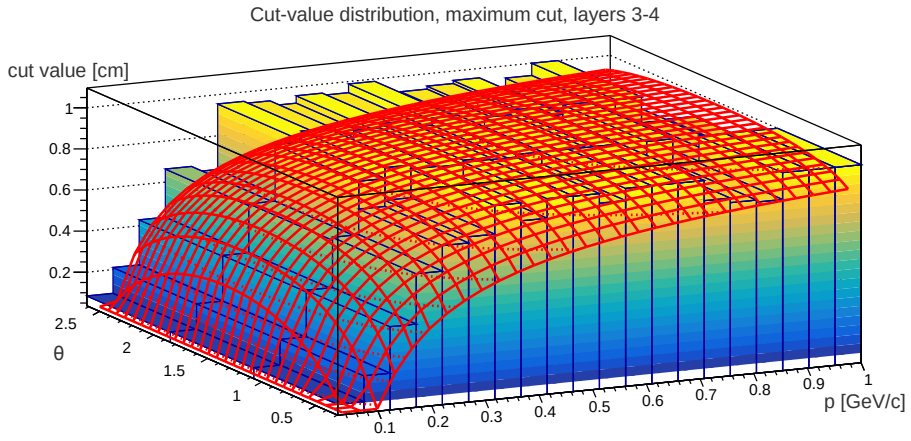
Where $X$ is the parameter, $n$ and $m$ the layers and the 3 free parameters of the fit are $A, B, C$. The initial values of these parameters has been set differently for each fit:

- $A$ equal to the value of the histogram in the central bin in $\theta$ and the first bin in $\mathbf{p}$.
- $B = 1$ for all the fits, as expected for multiple scattering only.
- $C$ equal to the value of the histogram in the central bin in $\theta$ and the last bin in $\mathbf{p}$.

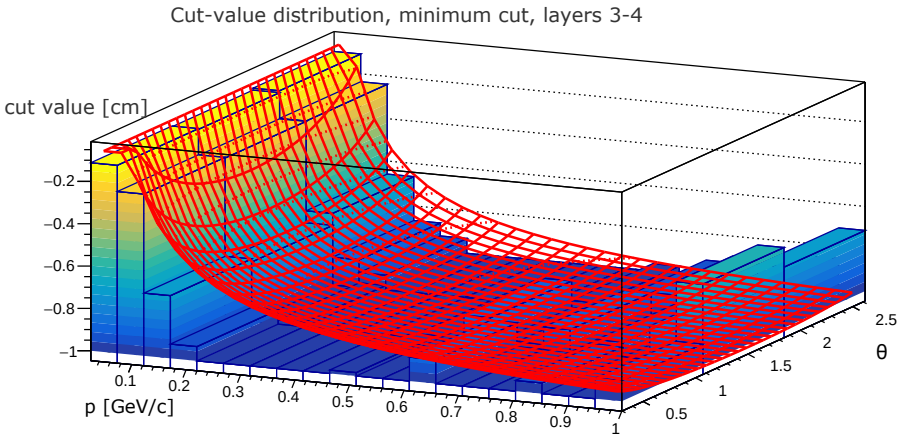
The fit procedure is stable for all the histograms, and the shape of the fit function results qualitatively similar for all the maximum threshold values and for all the minimum threshold values. For instance, in figure 5.9(a) the distribution of $d_0$ distribution of $\mathcal{C}_M$ for layer crossing between layer 3 and 4 is shown. The fit is shown as a red mesh, and it follows properly the shape of the 2-dimensional distribution. Qualitatively the threshold values have lower value at small momentum, than rises reach a plateau at 200-300 MeV/c. This behaviour is the opposite of what is expected from multiple scattering behaviour, but the cut-function saturate to $\varepsilon_{\text{req}} \simeq 1$ over 200 MeV/$c$, instead at low momentum the requirement in rejection is higher and so the threshold

values reach value closer to zero. The behaviour in $\theta$ is as expected from multiple scattering: has a maximum $\theta = \frac{\pi}{2}$ and quite smooth decrease symmetrically to $\pi$ and 0 directions. The cut-function, in fact, has no dependence in $\theta$, so the threshold values follow only their "intrinsic" structure from multiple scattering physics. In figure 5.9(b) the same distribution, but for the minimum threshold value is shown and it has a qualitatively symmetric behaviour with respect to $\mathcal{C}_M$ distribution. In both figure 5.9(a) and 5.9(b) that the maximum values of the histograms are very close to the "saturation limit" of 1 cm. In some distributions, not shown, actually the saturation at high value of $\mathbf{p}$ is reached.

For each function the variables $A, B, C$ are collected and stored in six (three variables for maximum and minimum) 3-dimensional histograms that are stored the final file of the specific selection criteria that have to be used to select the Training Sample (called *NoKickCuts* file, from now on).




(a) $\mathcal{C}_M$ *distribution*




(b) $\mathcal{C}_m$ *distribution*

*Figure 5.9: Distribution of $\mathcal{C}_M$ and $\mathcal{C}_m$ for $d_0$ in layer crossing 3-4, the fit is shown as the red gird. On the two axes there are momentum $|\mathbf{p}|$ and $\theta$ respectively*

#### 5.3.4.4 A different approach: "Double-Gauss-Fit" method

The requirement of the fixed efficiency and so the cut-function $\varepsilon_{\text{req}}$ turns out to be a stable method to obtain a selection with a known impact on the tracking efficiency. The drawback of this method is that is mostly independent on the shape of $\Delta X$ core distribution and evaluate only the density of the tails. The $\Delta X$ distribution contains more information about multiple scattering, energy losses effects and their behaviour in function of $\mathbf{p}$, $\theta$.

A possibility to extract these informations is to modify the approach to the accepted $\Delta X$ range definition: a function to fit the $\Delta X$ distributions has been studied, and then a model to extract the physical information from the fit parameters has been developed. Some studies have been performed to select the best fit function and it been chosen:

$$\mathscr{P}\Delta X(p,\theta)_{n,m} = N_1 e^{-\frac{(\Delta X-\mu_1)^2}{2\sigma_1^2}} + N_2 e^{-\frac{(\Delta X-\mu_2)^2}{2\sigma_2^2}} + C \tag{5.8}$$

That is the sum of 2 Gaussian distribution and constant. The initial value that has been chosen for the free parameters of the fit result fundamental for the convergence of these fits:

- $\mu_1 = \mu_2$, equal to the mean value of the $\Delta X$ distribution
- $\sigma_1 = \sigma_2$, equal to the standard deviation of the $\Delta X$ distribution
- $N_1 = 0.9b_0$ and $N_2 = 0.1b_0$, where $b_0$ is the content of the central bin of the distribution
- $C$ equal to the content of the first bin of the $\Delta X$ distribution (so the height of the tail)

It is possible to interpret the two Gaussian as a core distribution of standard events, and a secondary distribution, the one with wider $\sigma$, as the distribution of the rare catastrophic multiple scattering events. So the threshold values could be defined as:

$$\mathcal{C}_m, \mathcal{C}_M = \mu_1 \pm K\sigma_1$$

The multiplicative factor $K$ depends on the required efficiency of the selection: with a low value of $K$ the selection turn out tighter, but with very high purity (where the impurities are the residual catastrophic multiple scattering tracks). On the other hand, a big value of $K$ increase the efficiency, but reduce the purity.

The drawback of this method is the stability: the procedure adds a not trivial fit, and it requires a proper validation. One of the sources of instability are the limits of the histogram and the binning: to obtain a fittable $\Delta X$ distribution a proper binning is required, but it may change from various $p$, $\theta$ and layers pair. This increase the burden of the whole analysis. For these reasons in the end this method is not applied in the final version of the Training Sample selection.

## 5.4 Selection criteria application

To prepare a sample of simulated data to be able to train the Sector Map inside `basf2` exists a standard procedure, in which the selection criteria can be applied to each track as a filter which keeps or discards the track from the Training Sample. In the following section is developed the application of the filter in detail.

### 5.4.1 Selection of hit and global cuts

The first step of the filter follows the same procedure presented in section 5.3.1 to convert the reconstructed track in one simplified internal object (a vector of `hitXP`) built from the hits of the reconstructed track inside the VXD.

Then on this vector the same global requirements discussed in the section 5.3.2 are applied (at least 3 layer of SVD with hit on it, $|d_0| < 1$ cm, $|z_0| < 1$cm). If the track does not fulfil these requirements it is rejected. These global requirements represent the tighter requirement of the filter, and remove about 10% of the tracks from a generic Training Sample of $\Upsilon(4S)$.

### 5.4.2 Selection and applications of specific selection critera

Each track that passes the global requirements is given as input to the specific selection criteria filter. Inside this filter the hits of the track are examined from the virtual IP to the more external ones. Each pair of consecutive hits defines a *segment*.

The segments are the fundamental objects of this filter: each segment is defined by the layer's pair of the two hits, 5 $\Delta X$ values, a momentum $\mathbf{p}$ and a $\theta$ angle. The $\Delta X$ are evaluated from the methods of `hitXP` to access to the local track parameters that use the position and the momentum of the hit. The momentum $\mathbf{p}$ and $\theta$ of the segment are defined as $\mathbf{p}$ and $\theta$ of the more internal hit, because it represents the value of this quantities before the interaction of the material of a certain layer.

For each segment of a track it is given as the effective segment-filter. When a segment is given to this method first of all the proper allowed range is evaluated from the informations that are present in the NoKickCuts file. Inside this file are present 6 values for each layers pair and for each track parameter: $A, B, C$ parameter of the function 5.7 from the fit of maximum and minimum distributions of the threshold values. From the segment, the correct layers pair is selected and then for each track parameter the relative value of $\Delta X$ is evaluated as follow:

$$\Delta X^{\text{thr}}_{m,M} = \frac{A_{m,M}}{p^{B_{m,M}}\sqrt{\sin\theta}} + C_{m,M} \tag{5.9}$$

Where $p$, and $\theta$ are the properties of the selected segment and $m, M$ are referred to the minimum and the maximum values. From this procedure is apparent the advantage to store $A, B, C$ $(X, n, m)$ parameters and do not store $\Delta X(p, \theta)_{n,m}$ only: the value of $\Delta X_{\text{thr}}$ is evaluated with the precise information of $\mathbf{p}$ and $\theta$ of the segment (i.e. of the hit) without the need of bin these variables. For each track parameter of each segment are evaluated $\Delta X^{\text{thr}}_m$ and $\Delta X^{\text{thr}}_M$ and then the biggest in absolute value is chosen, or more precisely:

$$\Delta X_{\text{thr}} = \max\{|\Delta X^M_{\text{thr}}|, |\Delta X^m_{\text{thr}}|\}$$

This step makes the selection symmetrical between $\mathcal{C}_m$ and $\mathcal{C}_M$ with respect to 0 is needed because some strongly asymmetric distributions are present, and the application of these asymmetric threshold values has a too high effect on the total efficiency (this is the case, for example of $\Delta\omega$, because of the strong asymmetry).

After the evaluation of $\Delta X_{\text{thr}}$ the segment-filter checks if $-\Delta X_{\text{thr}} < \Delta X < \Delta X_{\text{thr}}$, for each $\Delta X$ parameter of the analysed segment. If for all the tracks parameter of the segment this is true the segment is flagged as "good segment". This procedure is repeated for all the segment of the track, and if all the segment is flagged as good the track passes the total filter. Instead when one of the segment is not flagged as good the filter rejects all the track, and the track is removed from Training Sample. In a more physical way it means: if there is one or more catastrophic multiple scattering interaction the track must be removed from Training Sample.

## 5.5 Results

The application of the procedure described in the previous section on each track of the Training Sample has direct effect on Training Sample composition and indirect effect on Sector Map structure. Therefore a modification of the Sector Map can have a strong impact on the tracking performances. In this section these three levels of consequences of the selection (Training Sample, Sector Map, track finder) are presented, with a comparative study with VXDTF2 standard performances.

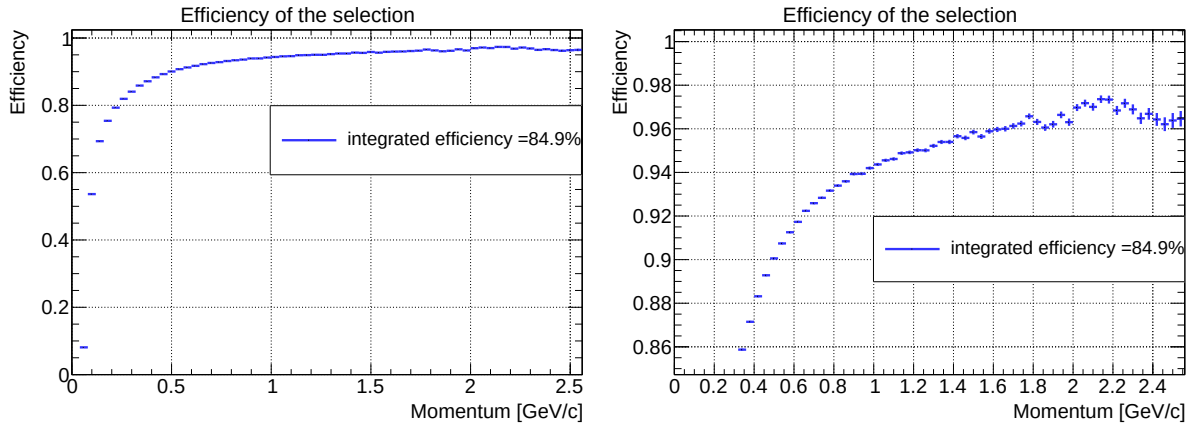
### 5.5.1 Effect on Training Sample

To study the direct effect of the selection procedure a sample of $9 \cdot 10^5$ $\Upsilon(4S)$ events has been used as starting Training Sample on which apply the selection. This initial sample is the one that has been used to train the default Sector Maps of VXDTF2. The sets of specific selection criteria (i.e. the NoKickCuts file) is evaluated on the same events, to remove all possible bias to the selection. The selection procedure applies the complete set of selection criteria i.e. the 6-layer requirements, for the Training Sample that will be used to in the Fast Reconstruction too. In fact if a catastrophic multiple scattering interaction occurs in PXD layers but not in SVD ones, it must be removed from Training Sample anyway (despite the PXD layers are not used in the Fast Reconstruction) because probably will have anomalous geometrical properties.

The first variable of interest that is studied is the total efficiency of the selection procedure. With an original Training Sample composed 0.9 million of events (9.1 millions of tracks), the selection procedure removes 1.4 millions of tracks. It means a total rejection ratio of 15.1%. The application of global requirements selects 90% of the Training Sample thus is consistent with the raw estimation given in section 5.3.4.1 of 6% for specific selection criterion only.

This source of inefficiency is not flat in momentum, and in figure 5.10 the efficiency curve in function of the magnitude of momentum is shown. Global requirements over 2 GeV/$c$ has a quite flat impact on efficiency, and about 3%-4% of the tracks are removed by them. Below 2 GeV/$c$ the efficiency drops quickly as expected from the efficiency cut-function.

A study on the momentum distribution of rejected and selected tracks is interesting to understand




(a) (b)

*Figure 5.10: Efficiency of NoKick Cuts selection in function of momentum. A zoom in the region around $\varepsilon = 1$ is shown too.*

the properties of the catastrophic multiple scattering interaction tracks. From figure 5.11 it is possible to notice that rejected tracks have the peak of momentum distribution at lower value that the selected tracks, as expected for low momentum effect and below 50 MeV/$c$ the rejected tracks becomes are more than the selected tracks. It is not present a visible change of the shape of the momentum distribution at 2 GeV/c, and this shows that the global requirements and specific selection criteria are properly merged. The distribution of rejected tracks decreases faster than the selected track one, as expected from momentum-dependent cuts.

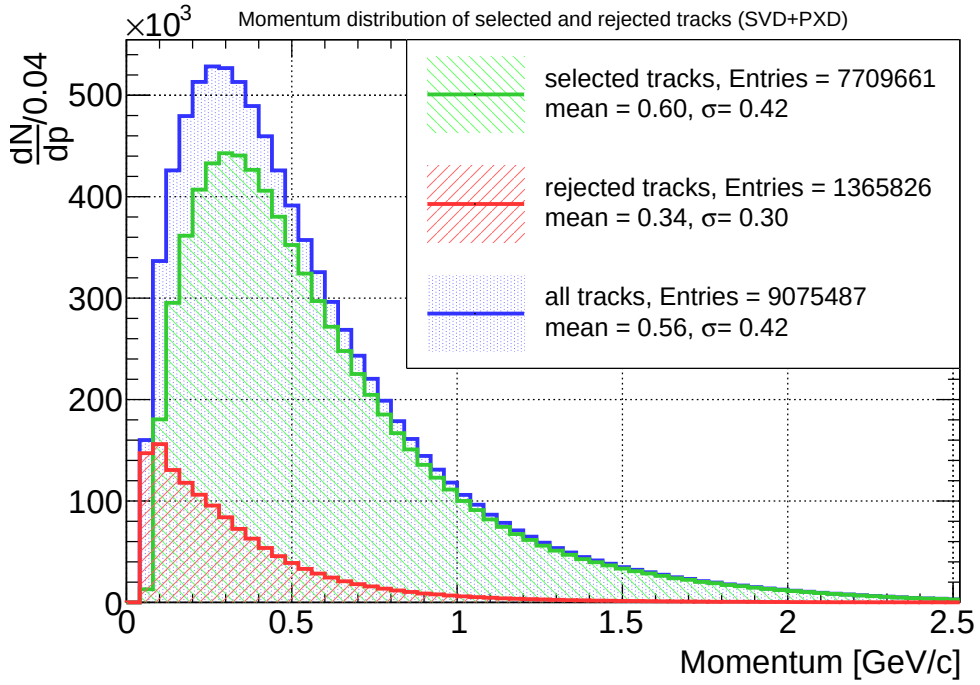



*Figure 5.11: Momentum distribution of the track for the entire sample of simulated events ( $9 \cdot 10^5$ $\Upsilon(4S)$ events), for the selected track and of the rejected tracks by the selection procedure.*

### 5.5.2 Effect on Sector Map

With the selected sample it is possible to train the Sector Map, via the standard procedure. The first apparent result is the size of the produced Sector Map. Without the selection of the Training Sample the size is 12.3 MB (Fast Reconstruction) or 21.0 MB (Full Reconstruction). To be compared with 5.1 MB (Fast Reconstruction) or 8.3 MB (Full Reconstruction) obtained with the selection procedure. It means, for instance in the SVD+PXD[2] configuration, a Sector Map size reduced by 60%, with only 15% of the tracks removed: this is a first evidence of the reduction of the complexity of the Sector Map. To avoid ambiguities the selection procedure developed in the previous chapter is called from now on *NoKick* selection.

To evaluate the effects of the removal of catastrophic multiple scattering events from the Training Sample has been used a third Sector Map in the comparison, which uses a third Training Sample. This Training Sample is composed of $2 \cdot 10^6$ events with 10 muons tracks only, produced with a flat momentum spectrum. These events are therefore composed of clear and easy fittable tracks which avoid strong material effects. The side of this Sector Map is reduced with respect to the

[2]in all the section is used the notation SVD+PXD to indicate the Full Reconstruction case, while the Fast Reconstruction is indicated with SVD only. Sometimes the Fast/Full Reconstruction notation is used to stress the different roles.

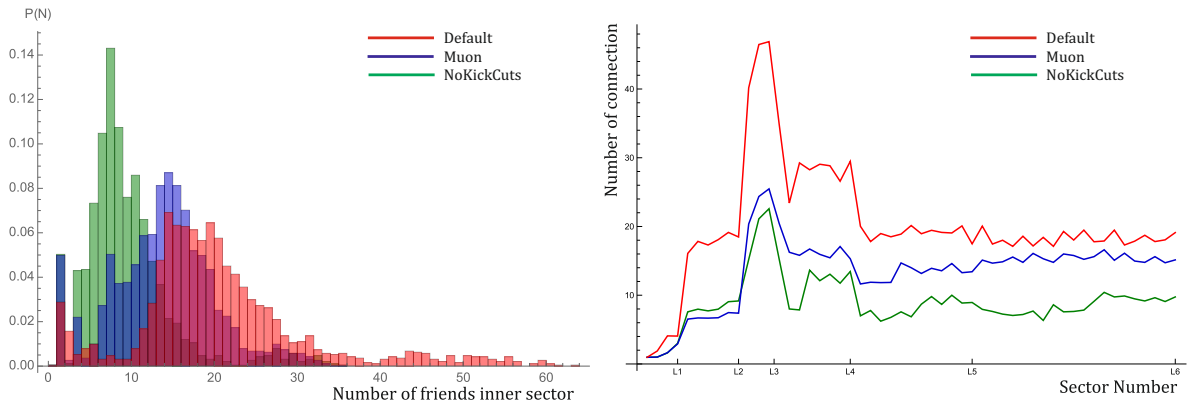
default one, but quite larger than the NoKick one: 13.0 MB (Full Reconstruction) or 6.8 MB (Fast Reconstruction).

Because it is observed the same behaviour of Fast and Full Reconstruction the following studies on Sector Map complexity are reported only in the Full Reconstruction case. The number of friendship connection inside the Sector Maps, of course, correlated with the Sector Map size, is a clear indicator of the Sector Map complexity:

$$\begin{aligned} N_F^{\text{default}} &= 38689, \\ N_F^{\text{NoKick}} &= 17545, \\ N_F^{\mu} &= 25785. \end{aligned}$$

The complexity can be further investigated studying the distribution of the number of internal friendship connection for a given sensor. In figure 5.12(a) the three distributions correspondent to the various Sector Map are shown: the mean number of connection is higher in the default Sector Map with respect to the muons and the NoKick Sector Map, and the long tails of high number of connection is removed. Another point of view is the number of friendship connection of each Sector (figure 5.12(b)): all the Sector Maps shows that the Layer 3 and the Layer 4 are the ones most related to other Sectors. This plot reveals the possibility of increasing the sectors granularity on the most connected layers to reduce the complexity of the Sector Map and so to reduce the time consumption. Instead for all the sectors the NoKicK Sector Map has a reduced number of connections, except in layer 2 where it overcomes the muons Sector Map connections. The number of connections per sector of the muon Sector Map are closer to the NoKickCuts one in the inner layer, instead for outer layer it is more similar to the default Sector Map. This feature suggests that the NoKickCuts selection is able to reproduce the cleanliness of the low momentum muons events, without increase the complexity also for more-hit patterns.

In conclusion, the NoKick Sector Map reduces of a factor 2 the number of connection inside the Sector Map in term of mean value and number per sensor. The complexity is reduced also with respect to the muon Sector Map.




(a) *Distribution of the Number of inner connections for a given sector.*

(b) *Number of connection for each Sector. The Sectors are numbered with a progressive number from 0 (Virtual IP) to outer Layers.*

*Figure 5.12: Sector Map complexity plots, evaluated for the three Training Samples*

### 5.5.3 Effect on tracking performances

The most important result is the impact of the selection of the Training Sample on the tracking performances. To evaluate the performances of VXDTF2 with the selected Training Samples has been used the same selected Training Sample described in the previous section and Sector Map trained with it, and then the performances of the Track Finder are evaluated testing the tracking software on another sample of $10^4$ $\Upsilon(4S)$ events. This second dataset is the same that has been used in chapter 4, so it is possible to directly compare the performances of VXDTF2 with the two different trainings.

The tracking performances are also evaluated for the muons Sector Map on the same testing dataset. The expected performances of the are similar to the selected Training Sample by NoKick Cuts: a lack of efficiency, because the muons do not describe the complete properties of the tracks from $\Upsilon(4S)$ decay, but a reduced fake rate and execution time due to the cleanliness of the $\mu$ tracks. This muons Sector Map is already used by the VXDTF2 to processing large quantity of simulated data for Monte Carlo studies because the VXDTF2 with the default Sector Map sometimes crashes on particular events with very high occupancy. Instead with the muons Sector Map the VXDTF2 is able to process and reconstruct them.

In the following plots it is shown a panel under each distribution where the normalized difference between the variable evaluated with the Selected Sector Map and the others Sector Maps i.e.:

$$\frac{\varepsilon_{\text{No Kick}} - \varepsilon_{\text{other SecMap}}}{\varepsilon_{\text{other SecMap}}} \quad \text{or} \quad \frac{\mathcal{R}^{\text{fake}}_{\text{No Kick}} - \mathcal{R}^{\text{fake}}_{\text{other SecMap}}}{\mathcal{R}^{\text{fake}}_{\text{other SecMap}}},$$

where "other SecMap" can be the muons (in light blue) or the default Sector Map (in orange), $\varepsilon$ is the efficiency and $\mathcal{R}^{\text{fake}}$ the fake rate (evaluated before the fit, like in 4). These panels allow to better look at the discrepancy between the various performances, and the $y$-scale is limited to the values around 0, not distinguishable in the main plot, thus the very low momentum bins or very high angle bins are not present in these panels.

In the following plots a redefinition of $\theta$ has been used: $\theta \in [-\pi/2, \pi/2]$, is the angle between the momentum and the transverse plane, where $\theta = \pi/2$ in the forward direction. This choice has been to keep the coherence with the Belle II standard validation plots.

In figure 5.13 the pattern recognition efficiency of the VXDTF2 is shown using a Sector Map with the default Training Sample, one where the selection has been applied and a third with the Sector Map trained on the muons events (VXDTF2 - default, VXDTF2 - NoKick, VXDTF2 - $\mu$, from now on). The VXDTF2 - NoKick and the VXDTF2 - $\mu$ have almost the same efficiency, both in momentum and angular analysis. Both the Sector Maps produce a degradation in efficiency of up to 10% a low momentum with respect to the default Sector Map, while they reproduce the default efficiency curve over 500 MeV/$c$. The only relevant discrepancy between the VXDTF2 - NoKick and VXDTF2 - $\mu$ is at very large $\theta$, where the muons Sector Map is more efficient (up to 10%), but this is compatible with the tighter selection criteria at higher angle due to the increased material budget.

In figure 5.14 the fake rate for VXDTF2 - default and VXDTF2 - NoKickCuts are compared. The VXDTF2 - NoKick and produces reduced fake rate in all the momentum spectrum with respect to the default Sector Map. The muons Sector Map produces a quite similar result in SVD only case, but with the insertion of the PXD the NoKicKCuts Sector Map has a fake rate reduced of 25% at $\theta \simeq 0$ and of 50% $|\theta| > 1$ also with respect to the muons Sector Maps.

In figure 5.15 is shown the efficiency of fitting step only. In this case, both VXDTF2 - NoKick and VXDTF2 - $\mu$ increase the efficiency with respect to the default Sector Map in all the momentum

range, but the muons map obtains better results in SVD only case instead the NoKicK one in the SVD+PXD one (of about 0.5%).

The original idea of the "NoKickCuts" was to remove tracks that have suffered catastrophic multiple scattering interaction and they became not fittable because of the consequent anomalous trajectory. Therefore the purpose of the selection is to remove tracks that otherwise discarded from the fitting step, thus an improved fitting efficiency is a fundamental result to validate this hypothesis: the reduction of the pattern recognition efficiency can be partially compensated below $p_\perp \simeq 300$ MeV keeping the total tracking efficiency, evaluated as the product of tracking and fitting efficiencies, almost constant. On the other hand, thanks to the selection the fake rate is also reduced. The total tracking efficiency plots is reported in the appendix A, and coherently with the expectations from previous results, it shows a small degradation at low momentum for VXDTF2 - $\mu$ and VXDTF2 - NoKick.

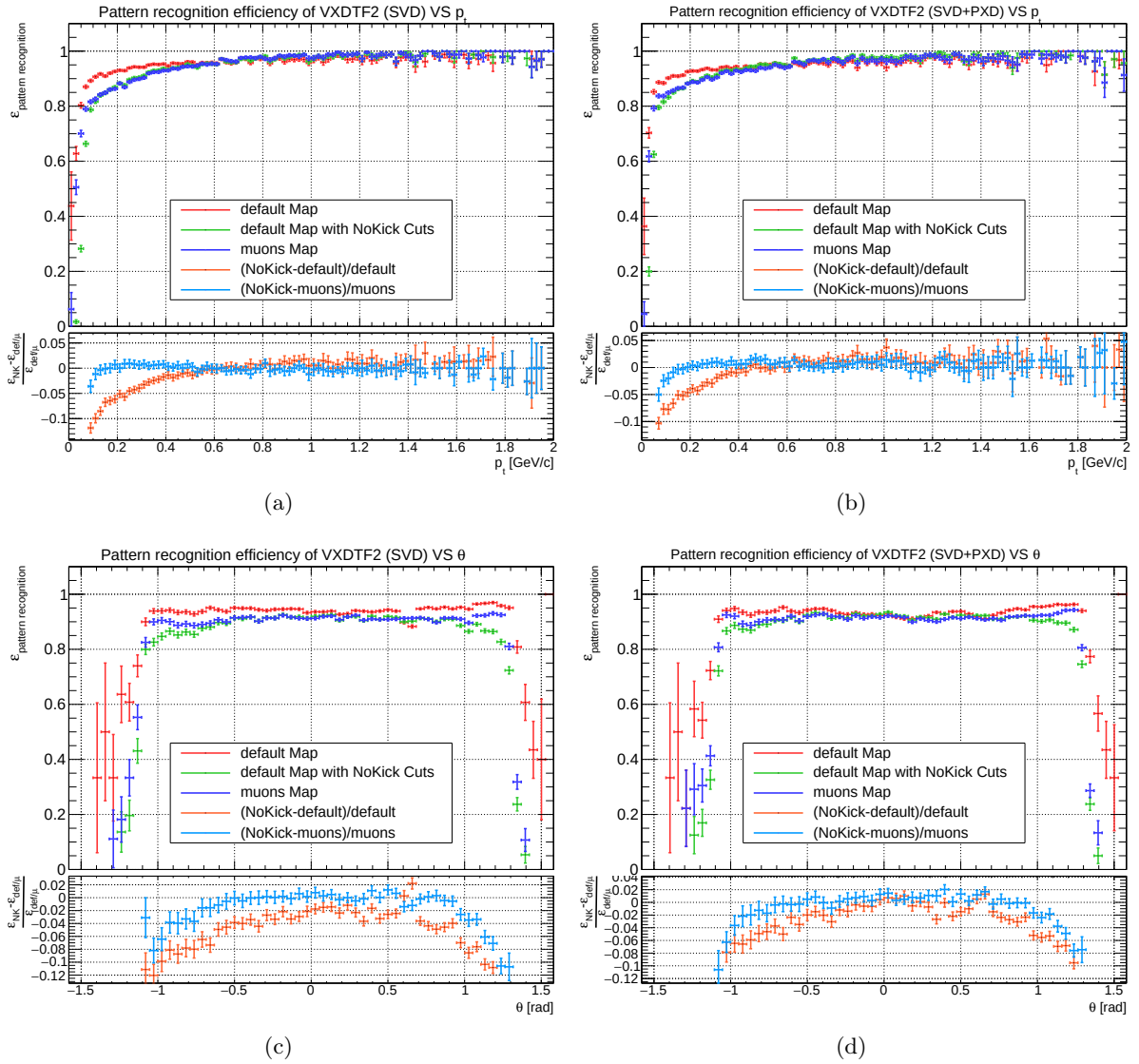



*Figure 5.13: Comparison between pattern recognition efficiency of VXDT2 with Sector Map trained with default Training Sample and the one with the selection applied, in case of SVD only (5.13(a), 5.13(c)) or SVD+PXD (5.13(b), 5.13(d)) tracking in $p_\perp$ and $\theta$ trends.*

The last relevant parameter for tracking performance is the time consumption. For VXDTF2 - NoKick the total time consumption by all the modules is:

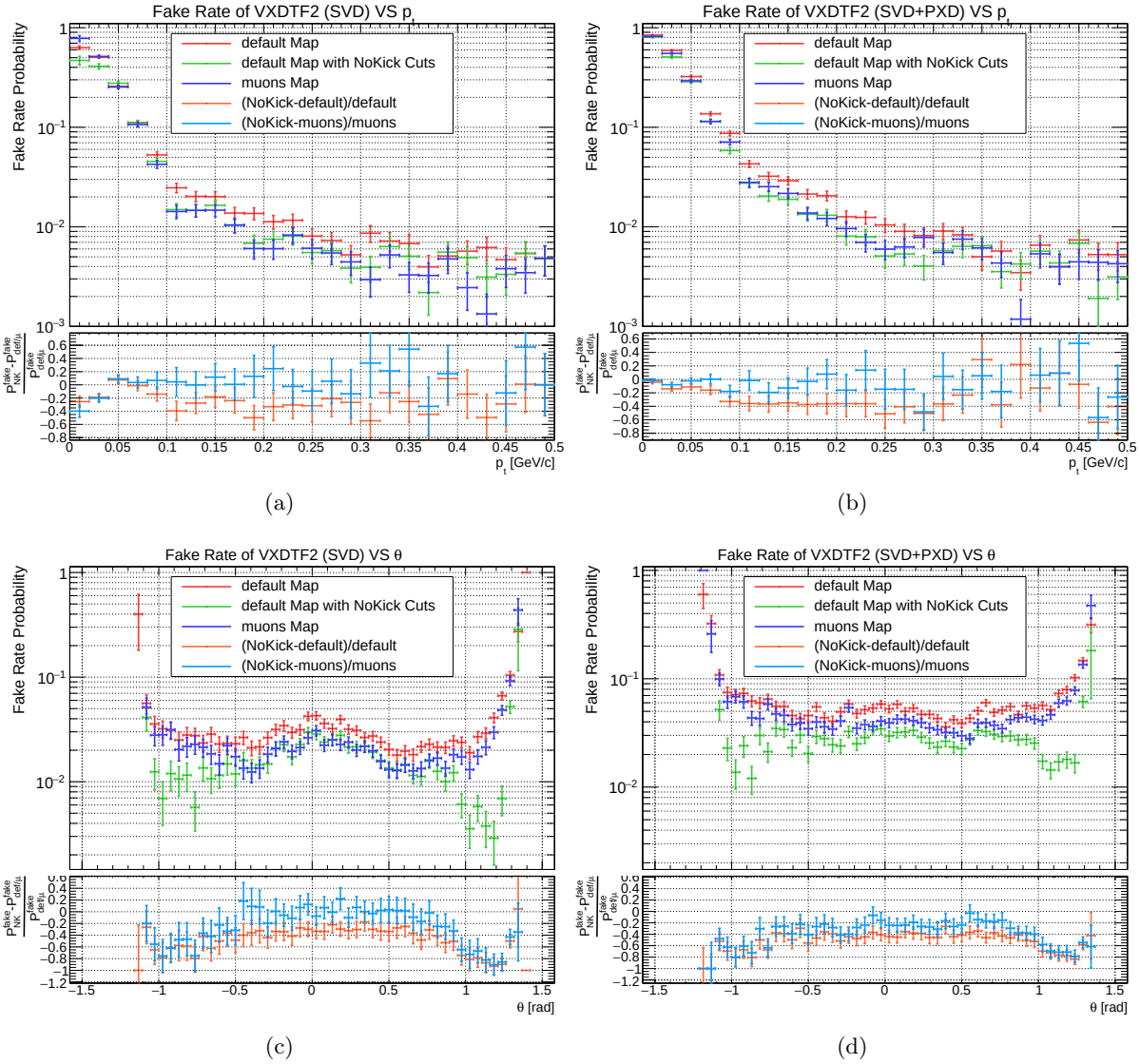
$$t_{VXDTF2} \simeq 4.53 \text{ ms/event (SVD only)}$$
$$t_{VXDTF2} \simeq 9.58 \text{ ms/event (SVD+PXD)},$$

to be compared with the values for VXDTF2 - default reported in section 4.5: it means a gain of a factor two for the SVD case and over a factor three for the SVD+PXD case. The two modules which have the larger gain are the Overlap Remover and the Cellular Automaton, a signature of the reduced number of fake candidates.

It is interesting to report the time consumption for the VXDTF2 - $\mu$ too:

$$t_{VXDTF2} \simeq 3.24 \text{ ms/event (SVD only)}$$
$$t_{VXDTF2} \simeq 12.37 \text{ ms/event (SVD+PXD)}.$$

In this case, the track finder is faster than VXDTF2 - default and also than VXDTF2 - NoKick




(a) (b)

(c) (d)

*Figure 5.14: Comparison between the fake rate of VXDT2 with Sector Map trained with with default Training Sample and the one with the selection applied, in case of SVD only (5.14(a), 5.14(c)) or SVD+PXD (5.14(b), 5.14(d)) tracking in $p_\perp$ and $\theta$ trends.*

in SVD case, instead it is slower than the latter in SVD+PXD case. This is coherent with the other performances results: slightly higher than VXDTF2 - NoKick in Fast Reconstruction but not in Full Reconstruction.

The time required for the fitting procedures only with the NoKick and muons Sector Map is:

$$t_{VXDTF2-\mu} \simeq 267 \text{ ms/event (SVD only)}$$
$$t_{VXDTF2-\mu} \simeq 361 \text{ ms/event (SVD+PXD)}$$
$$t_{VXDTF2-NoKick} \simeq 251 \text{ ms/event (SVD only)}$$
$$t_{VXDTF2-NoKick} \simeq 396 \text{ ms/event (SVD+PXD).}$$

The values reveal that unfortunately for both NoKick and muons Sector Map the times consumption of the fitting step is reduced with respect to the VXDTF2 - default, but the improvement is marginally (below 10%) both for Fast and Full Reconstruction.

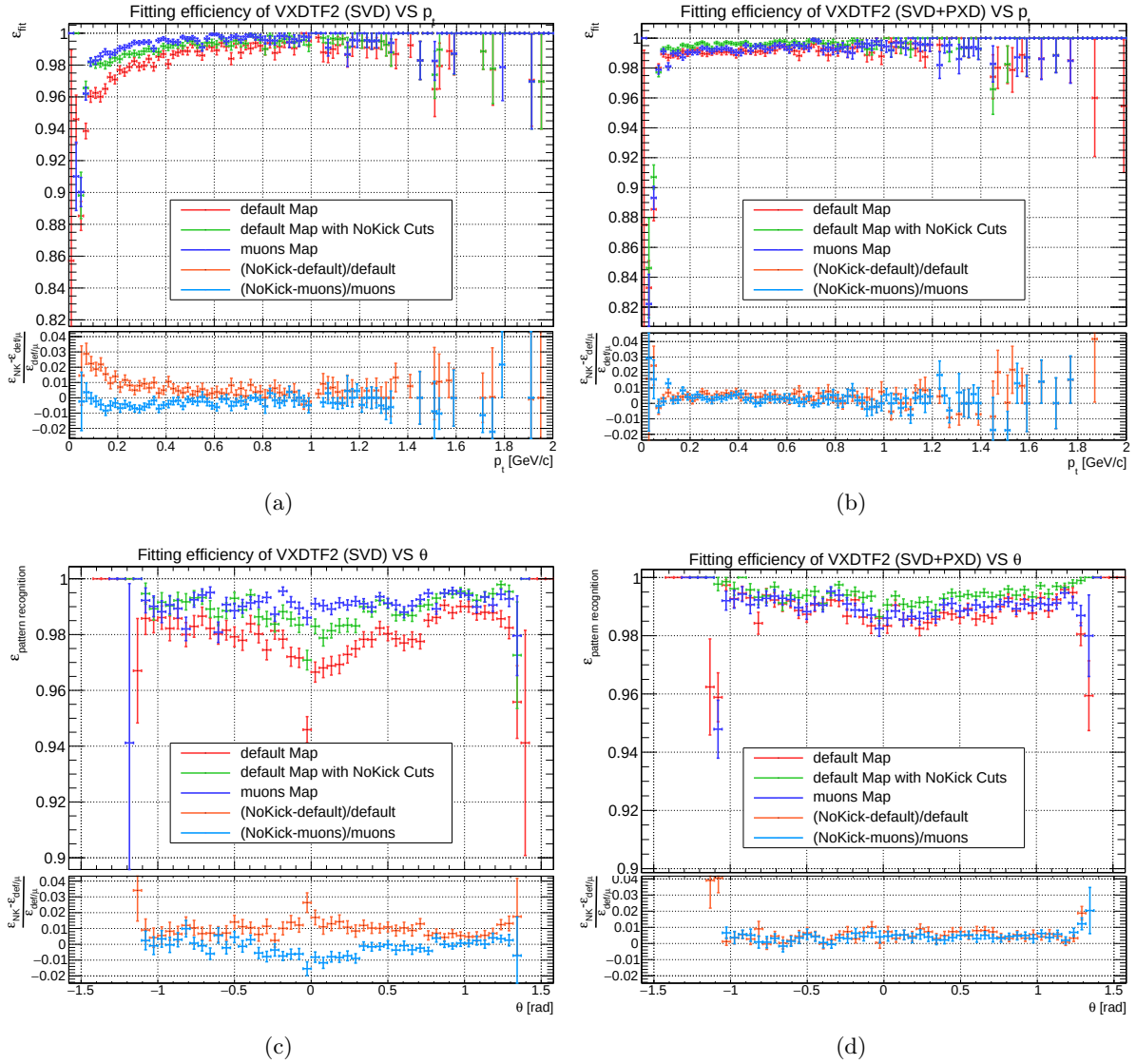



*Figure 5.15: Comparison between fitting efficiency of VXDT2 with Sector Map trained with default Training Sample and the one with the selection applied, in case of SVD only (5.15(a), 5.15(c)) or SVD+PXD (5.15(b), 5.15(d)) tracking in $p_\perp$ and $\theta$ trends.*

### 5.5.4 Tracking Performance in presence of background

In the Belle II environment a not negligible beam-induced background is expected, as described in section 2.3.2. The background hits inside the VXD affect the performances of the Track Finder by increasing the combinatorial burden. The selection of the Training Sample, removing "dirty" tracks, may improve the background hits rejection. The expected background is overimposed to the simulated $\Upsilon(4S)$ sample which has been used to testing the Track Finder in the last section, then the VXDTF2 - default, VXDTF2 - NoKick and VXDTF2 - $\mu$ are tested on the new sample. The SVD background rate in this simulation is higher by a factor 6 because the APV25 features are not yet fully exploited by the VXDTF2 to reject out-of-time hits.

The pattern recognition efficiencies obtained with the three Sector Maps are compared in figure 5.16. They exhibit similar efficiency in the whole momentum range with a slight advantage for

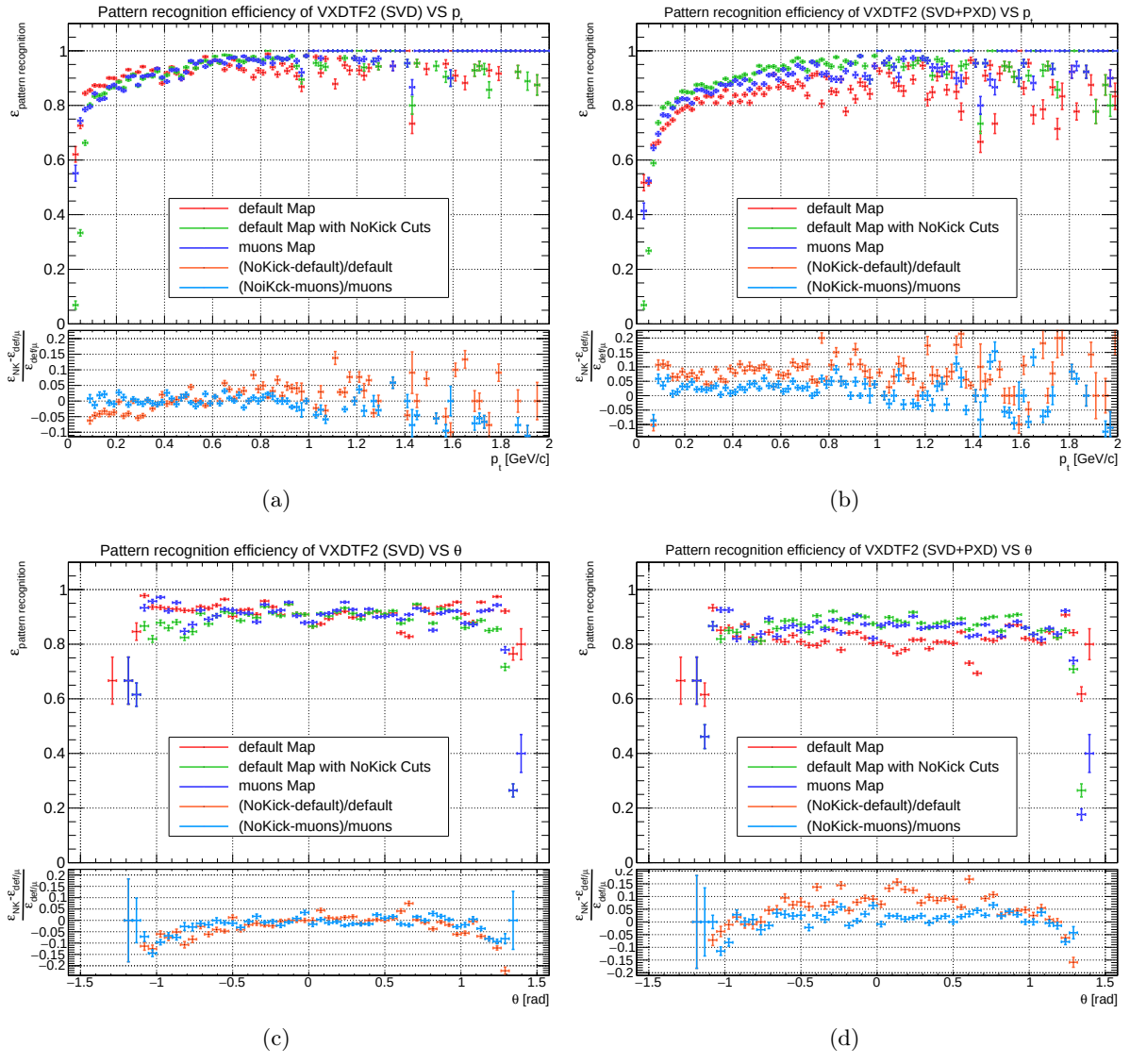



*Figure 5.16: Comparison between pattern recognition efficiency of VXDT2 with Sector Map trained with default Training Sample and the one with the selection applied, in case of SVD only (5.16(a), 5.16(c)) or SVD+PXD (5.16(b), 5.16(d)) tracking in $p_\perp$ and $\theta$ trends. The nominal mixed background is added to the $\Upsilon(4S)$ simulated sample.*

the muons Sector Map at high $p_\perp$ (above 1 GeV/$c$). The insertion of PXD reduces the efficiency of the VXDTF2 - default and, to a lesser extent, also the VXDTF2 - $\mu$ one. This is due to the large occupancy of these events which degrade the performance Cellular Automaton. The NoKicK Sector Map avoids this degradation and so it exhibits an efficiency $0.5\% \div 1\%$ higher of the others Sector Maps.

The fake rate increase by about a factor 3 with the addition of the nominal background for al the Sector Maps, despite the fake rate of the VXDTF2 - Nokick and of the VXDTF2 - $\mu$ are halved with respect the default one. In particular with the insertion of the PXD, which increase the fake overall fake rate, the NoKick Sector Map exhibits the better results.

The addition of the nominal background degrades also the fitting efficiency in all the momentum ranges as is shown in figure 5.18, despite the better performances of the muons and the NoKick Sector Maps holds with respect to the default Sector Map. The total tracking efficiency is

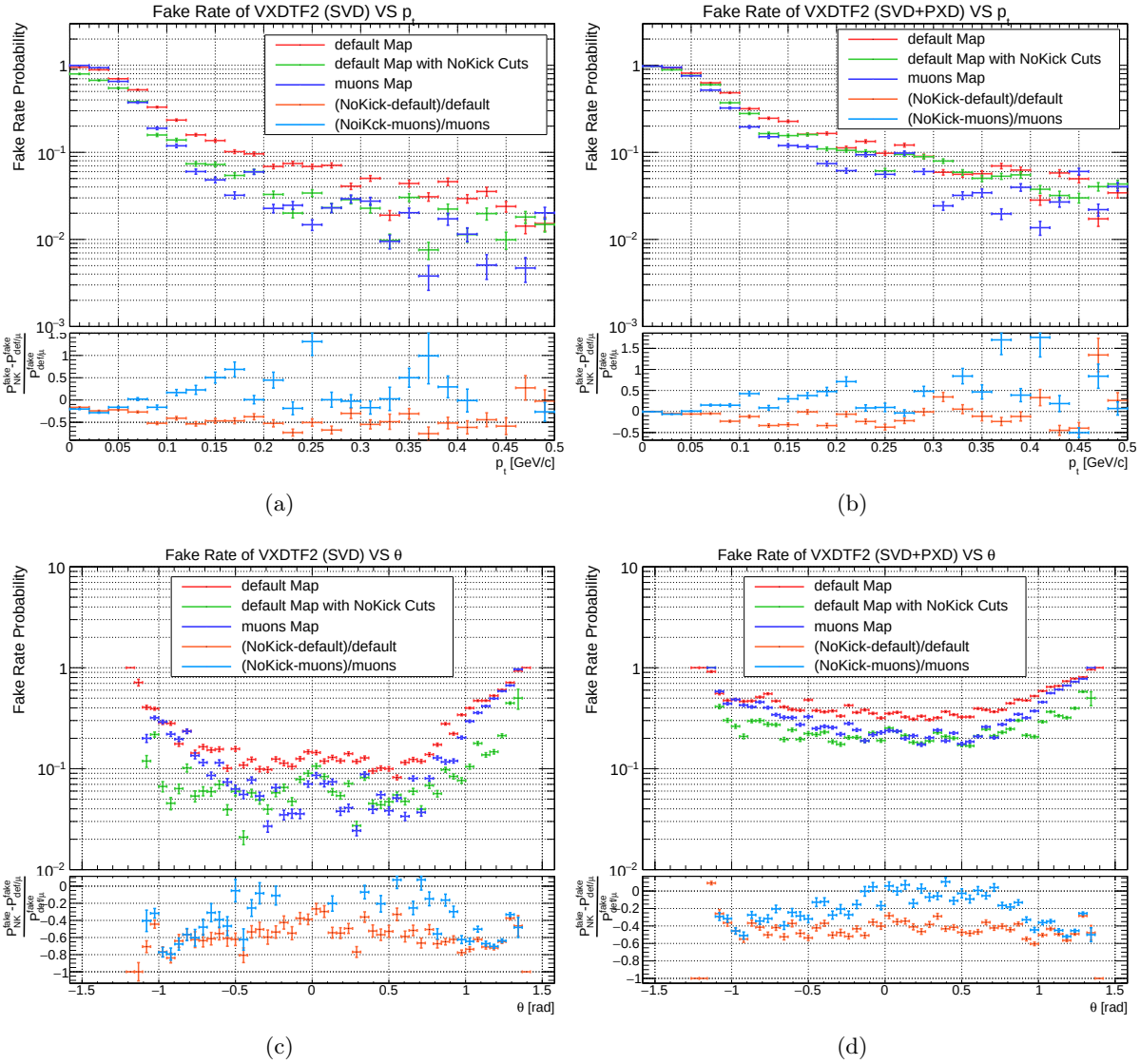



*Figure 5.17: Comparison between the fake rate of VXDT2 with Sector Map trained with with default Training Sample and the one with the selection applied, in case of SVD only (5.17(a), 5.17(c)) or SVD+PXD (5.17(b), 5.17(d)) tracking in $p_\perp$ and $\theta$ trends. The nominal mixed background is added to the $\Upsilon(4S)$ simulated sample.*

reported in the appendix A, and it is coherent with the expectations from previous results.

The time consumption of three track finders in presence of the nominal background are summarized, separately for pattern recognition and fitting steps, in table 5.2. In SVD only case the time consumption of the pattern recognition for VXDTF2 - NoKick and VXDTF2 - $\mu$ is similar, instead a factor 4 is present with respect to the VXDTF2 - default, therefore as expected from the efficiency performances, the VXDTF2 - $\mu$ is less affected by the background presence. With the PXD insertion instead the VXDTF2 - $\mu$ time consumption increases over a factor 100, while it increases about a factor 3 both for VXDTF2 - NoKick and VXDTF2 - default. Therefore seems that the muons Sector Map pays the reduced the fake rate with a much larger CPU time. However this large time consumption can be a result of the large fluctuations ($100 \div 200\%$) which affects all these times. On the other hand the time required for the fitting procedures in quite the same of the case without background. This result, in addition to the small difference between

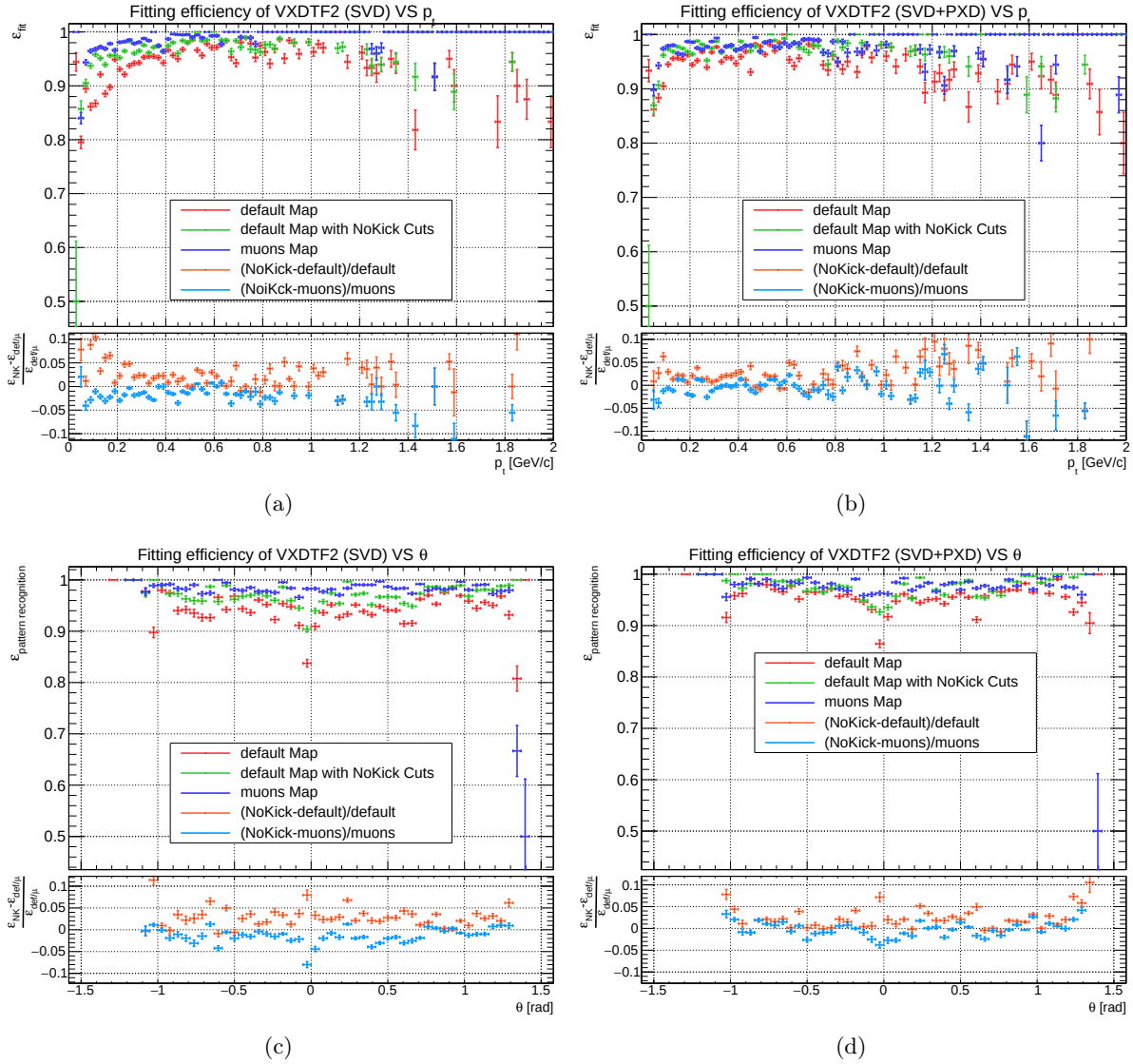



*Figure 5.18: Comparison between fitting efficiency of VXDT2 with Sector Map trained with default Training Sample and the one with the selection applied, in case of SVD only (5.18(a), 5.18(c)) or SVD+PXD (5.18(b), 5.18(d)) tracking in $p_\perp$ and $\theta$ trends. The nominal mixed background is added to the $\Upsilon(4S)$ simulated sample.*

| | Pattern Recognition [ms/ev.] | | Fitting [ms/ev.] | |
|---|---|---|---|---|
| | SVD | SVD+PXD | SVD | SVD+PXD |
| VXDTF2 - default | 48.26 | 125.51 | 280 | 394 |
| VXDTF2 - $\mu$ | 11.47 | 1610.08 | 272 | 378 |
| VXDTF2 - NoKick | 12.84 | 39.56 | 237 | 370 |

*Table 5.2: Time consumption of VXDTF2 trained with Sector Maps trained with different Training Samples.*

the fitting time with the selected Sector Maps and the default one, suggests that most of fitting time is due to an unoptimized fitting algorithm, instead of a low-performance input from pattern recognition.

### 5.5.5 Summary of the Training Sample selection effects

The selection of the Training Sample has an efficiency compatible with the requirement, and the momentum distribution of the selected track is reasonable: the algorithm of selection works as expected. The selection of the Training Sample reduces the complexity of the Sector Map, in term of size and number of connections, but unfortunately the loops are not completely removed from this selection, and their origin must be investigated. The test of the VXDTF2 that use a Sector Map trained with the selected Training Sample reveals a reduced pattern recognition efficiency compatible with the expectation, partially compensated by an increased fitting efficiency. In addition, the fake rate is strongly reduced on the whole momentum range. In conclusion, the selection affects mainly the low momentum a high angle regions (both in terms efficiency and fake rate), as expected for multiple scattering effects. Therefore the selection of the Training Sample seems to be an improvement over the default one. Deeper studies may improve the results by a further optimization of the selection procedure. The most delicate step is the definition of the efficiency cut-function, which should be done automatically in function of the required tracking performance. With the selection of the Training Sample, the time consumption in pattern recognition is reduced up to a factor 3, a very useful improvement for the HLT purposes. On the other hand, the time required for the fitting steps seems to be weekly dependent from the Training Sample, and the VXDTF2 with a Sector Map trained with selected Training Sample gains below 10% of the fitting time.

The alternative muons Training Sample is satisfactory in term of performances and in particular the fitting efficiency and the time consumption result slightly better than the case with the NoKickCuts applied for the Fast Reconstruction. In the Full Reconstruction the results are instead worse. Further studies are needed to understand if the muons Training Sample bias the reconstruction with the rejection of particular classes of interesting tracks, but the studies on this Training Sample are outside of the goals of this thesis.

A test with the nominal beam-induced background injection reveals that the VXDTF2, with the use of the Sector Map trained with the selected Training Sample, obtains a slightly increased efficiency, and the fake rate is also reduced. The time consumption gains a factor 4.

The tracking group foresee in the next future to apply these selection criteria to the $\mu$ Training Sample with the hope that the same benefits in terms of reduced CPU time, Sector Map complexity and the fake rate will be observed also in this case. Further studies will be needed when the excellent time resolution of the SVD, a few of ns, will be exploited to reject out-of-time background hits.

# Chapter 6

# VXD Testbeam

During February 2017 a testbeam took place at DESY (Deutsches Elektronen-Synchrotron, Hamburg, Germany) where the SVD and PXD modules have been tested with the Belle II 6-layer setup, reproducing a sector of the Belle II VXD. Both the SVD and PXD groups have been involved in the testbeam from the hardware side (detectors, DAQ, trigger) as well as the tracking group from the software side (reconstruction, trigger).

I personally participated to the testbeam activities as a member of both the tracking and the SVD groups. During the testbeam I took part in monitoring activities of the track finder and the SVD performances and in particular I implemented the measurement of the hit efficiency for the SVD with the VXDTF2 as a track finder, previously performed in a previous teastbeam with VXDTF. In this chapter, after an introduction about the motivation and the facilities of the testbeam, both the monitoring activities and the hit efficiency measurements are reported.

## 6.1 Testbeam motivations

Testing a full VXD sector on a beam is motivated by the need to know the VXD performances, in term of detector response, data acquisition system, trigger performances and tracking software integration before the detector is installed. A testbeam where the final VXD layer setup has been tested and validated took place in April 2016 [42]. Since then several updates in data acquisition system and in tracking software have been done requiring an additional testbeam. The 2017 testbeam goals can be summarized as follows:

- test the performances of VXD modules in their final setup (4 SVD layers and 2 PXD layers) in term of efficiency of detection, stability, slow control and readout. In particular it is the first testbeam with complete PXD ladders.
- test the full DAQ chain of Belle II and in particular test the HLT and the ROI performance.
- test for the first time the VXDTF2 both for online tracking (HLT) and offline Full Reconstruction.
- collect data to perform offline detector and tracking analysis to optimize the performances in several conditions (with different beam energy, angle of incidence, magnetic field).

The testbeam has achieved the online goals and the results from offline analysis indicate that the expected performances are satisfactory, but further studies are necessary to fully exploit the collected data.

## 6.2 DESY Facilities and setup

DESY offers several testbeam lines realized from the DESY II accelerator, a circular collider that accelerates $e^+$ and $e^-$ up to 7 GeV and is used as the main injector for DORIS and PETRA colliders. From the main beam a bremsstrahlung photon beam is generated by a carbon fibre target and then the photons are converted into $e^+e^-$ pairs by a metal converter target. The $e^+$ and $e^-$ are spread out with a dipole magnet and the final beam is selected by a collimator. The energy of the final $e^+$ or $e^-$ is tunable from 1 to 6 GeV with an energy spread of $\sim 5\%$ and a divergence of $\sim 1$ mrad regulating the magnet current. In figure 6.1 the entire map of the DESY testbeam facility is shown, more details can be found at [59].

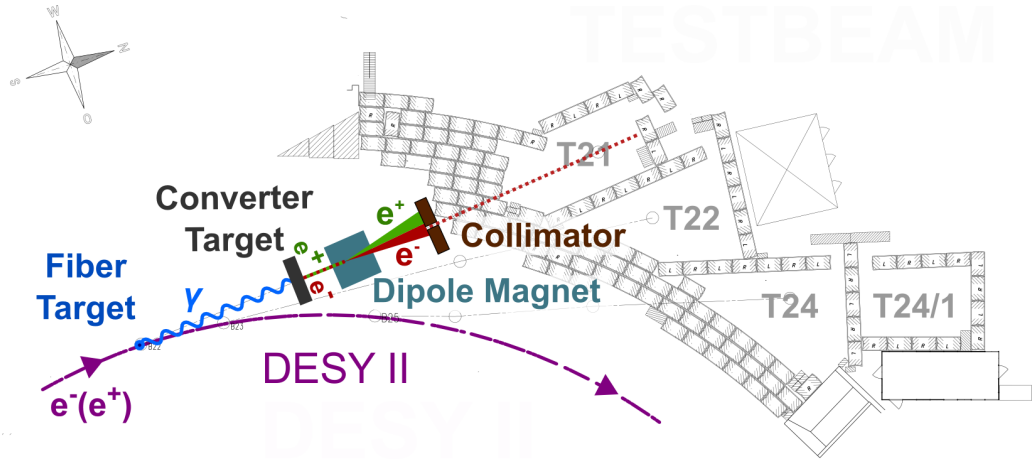



*Figure 6.1: Top view of DESY testbeam facility; one of the three testbeam lines is highlighted.*

### 6.2.1 Belle II testbeam setup

The Belle II VXD Testbeam took place on the testbeam line 24/1, with the setup shown in figure 6.2. The tested detector consists in one sector of the VXD: four complete ladders of SVD and two complete ladders of PXD, one for each layer. The sector of VXD is mounted with the same setup that will be used for the complete VXD, inside a carbon fibre crate, called *cradle*. The ladders are cooled with the same evaporative $CO_2$ system that is used in the entire VXD with an operating temperature of the detector in the range $-15 \div -20$ °C.

The electron or positron beam hits the VXD sector orthogonally to the layers' surfaces defined as the $x$ direction according to the global Belle II coordinate system. The VXD craddle is placed inside a superconducting solenoid which produces a constant magnetic field up to 1 T in the inner region. The direction of the magnetic field is parallel to the virtual Belle II's beam pipe (along $z$, in the global Belle II coordinate system), simulating the Belle II magnetic field.

At the entrance of the testbeam room, after the collimator, a telescope realized with two layers of scintillator fingers, to provide the event trigger event and to define the beam position and direction. It is possible to change the incident beam angle from $-5° < \theta + 25°$, by rotating the entire solenoid (the VXD cradle is rigidly connected to the solenoid). The beam energy is tunable from 1 GeV to 6 GeV, and the magnetic field is tunable from 0.25 T to 1 T. During the testbeam many *runs*[1] have been performed, spanning the energy, the incident angle and the magnetic field in all the accessible ranges.

[1] it is called *run* a single set of data taking with the same environmental parameters.

To collect the VXD data the full readout chain of Belle II as explained in section 3.2.2 is used, including of HLT (of course without CDC information) to evaluate the ROI on PXD layers and read them. The DATCON was not operative when the testbeam was performed.

In the entire testbeam about $320 \cdot 10^6$ events subdivided in 300 runs have been collected. In 170 runs the VXDTF has been used in the HLT, while in the last 130 runs VXDTF2 has been used. This is the first time of the redesigned track finder is used with a real detector. Unfortunately, in February the selection of the Training Sample was not completed yet, thus a default Training Sample has been used. In figure 6.3 the event display of the testbeam is shown. While in the 2016 testbeam a high resolution pixel telescope was installed allowing the measurement of the VXD resolution, for the 1017 testbeam whose focus was shifted to the system integration and operation, the telescope was not available and the resolution measurement has not been reproduced.

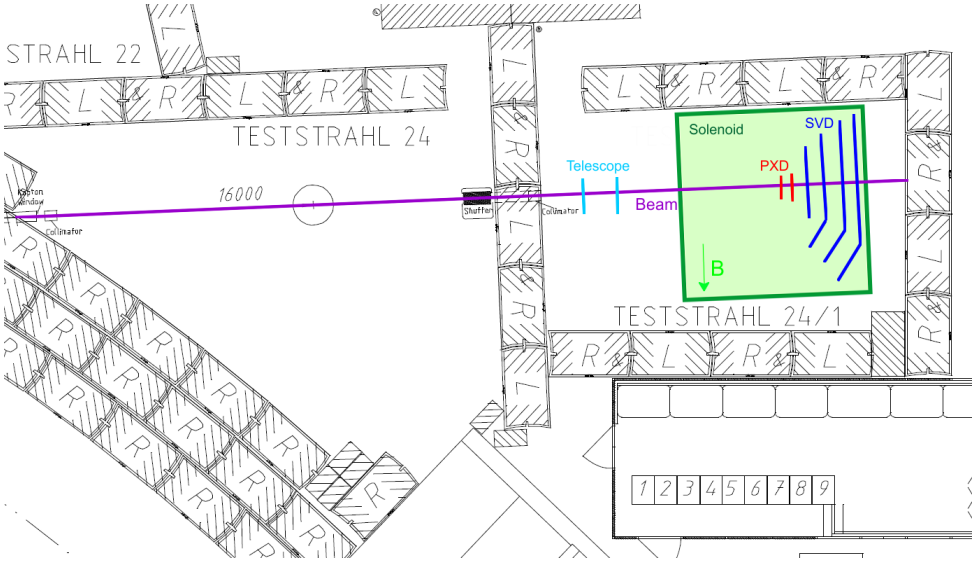



*Figure 6.2: Top view of testbeam area (superconducting solenoid and VXD not in scale to make the detector visible.)*

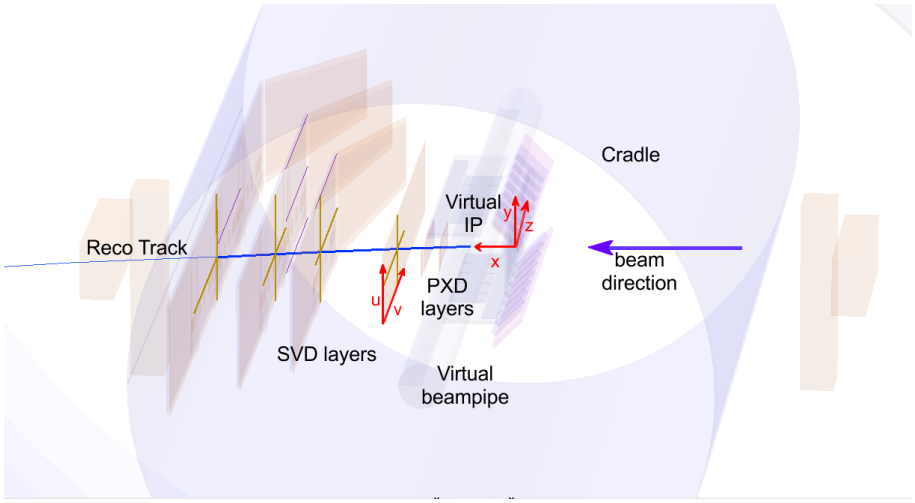



*Figure 6.3: Display event of the testbeam: are shown the SVD and PXD layers, the cylindric craddle, the inert material that mimics the beampipe and some support structures. A track is shown too, with the hit strips/pixel highlighted. Some noisy strips triggered in the events are visible too.*

## 6.3 VXDTF2 validation

During the testbeam data taking it was possible to monitor online the beam parameters and the detector response and to perform several online quick analyses to monitor the general properties of the collected data. These analyses were useful to identify any setup or electrical issue, but it was not possible to access more complex information, like full tracking information or physics quantities, and study them online. For the purposes of the testbeam it was fundamental to understand during the data taking if the track finder and the HLT have the expected performances, to be able to solve software bugs and immediately test them. Therefore after each run the data were quickly analysed offline and the results could be used to improve the performances of the following runs.

This *semi-online* validation gave not trivial results in this teastbeam, because the VXDTF2 has been tested in HLT for the first time, and several software issues appear during the tests. In the following sections, some relevant validations are reported, performed on the run 399. In this run the beam energy is 4 GeV, the incident angle $\theta = \pi/2$ (orthogonal beam) and the magnetic field $\mathbf{B} = 1.0$ T. The VXDTF2 has been used in the HLT to reconstruct the Track Candidates and extrapolate the ROI on the PXD layers. Two sensors are redout per layer both for PXD and SVD (i.e. the complete Layer 1, Layer 2 and Layer 3, -Z and BW sensors for Layer 4, -Z and CE for Layer 5 and Layer 6). Although complete ladders for outer layers is present and electrically operative has not been possible to read them at simultaneous because of a limited number of readout connection. However, the connection in different runs has been switched to test the complete set of sensors. In this run $17 \cdot 10^6$ events have been collected and it has been chosen to show the VXDTF2 performances after the bugs fixing, to give the most updated results.

For each layer it is possible to look to the *hitmap* i.e. the spatial distribution of the hits on the sensors. For instance in figure 6.4(a) and 6.4(b) are reported respectively the projection on $z$ and $y$ axes of the Layer 5, while on figure 6.4(c) the complete hitmap is shown. On the $p$ side of the SVD sensor the strips are along $z$ (or in local coordinate along $v$ direction), instead on $n$ side they are along $y$ (or $u$), therefore the $z$ projection allows to see a scan of the $p$-strips and the $y$ projection a scan of the $n$-strips. The Gaussian-like beamspot is well visible on the -Z sensor, instead on the CE sensor shows mostly noise hits. The narrow peaks visible in the $z$ and $y$ projection are due to noisy strips, corresponding to known defects of $n$ or $p$ strips respectively. These noisy strips appear on the 2D hitmap as lines with quite flat distribution folded with the beamspot shape. For completeness in figure 6.4(d) is shown a hitmap of the firs PXD layer: in this case the beam is not centred in the PXD sensor, and visible defects are not present. The hitmaps are used to locate unexpected defects on the sensors, check the beam alignment, identify possible basic-reconstruction issues (like wrong cluster or hit position mapping) or readout problems.

A second class of validation plots useful to monitor these possible issues are the *correlation plots*. For each pair of layers two 2-dimensional plots are realized, and in each plot the hit distribution of the first layer in one coordinate versus the hit distribution of the second layer in the same coordinate is shown. It is expected to find spatially correlated hits, like the ones that came from the same particle that crosses both layers, mostly along a narrow straight line with slope 1. Widening and deviation from this ideal distribution are due in the first place to the beam spot spread and detector resolution, and secondly because the track curvature caused by the magnetic field. Thus the linear correlation is more evident for consecutive layers and less evident for far layers. Material effects are expected instead to be negligible because of the high energy of the beam. In figure 6.5 some examples of correlation plots both for PXD and SVD layers are shown. In this class of plots is possible to notice horizontal and vertical lines,

due to the noisy strips of the sensors that produce uncorrelated hits. The correlation plots are very useful to identify misaligned sensors. In particular for the relative position of PXD with respect to SVD, a misalignment between the two sensors implies a strong degradation of the ROI finding algorithms. The discussion of alignment procedures is out of the purposes of this thesis. Anyway, the correlation plots provide information on the translation misalignment only and to evaluate rotation misalignment more refined method must be used. The correlation plots can't provide information on rotation misalignment because use the straight line position information only and a rotation transforms straight lines in straight lines. A quick method to see rotation misalignment can be to plot $u \times \mathcal{R}_v$ and $v \times \mathcal{R}_u$, where $\mathcal{R}_{u,v}$ are the residuals between the hit position and the extrapolated track position in $u$ and $v$ coordinate respectively.

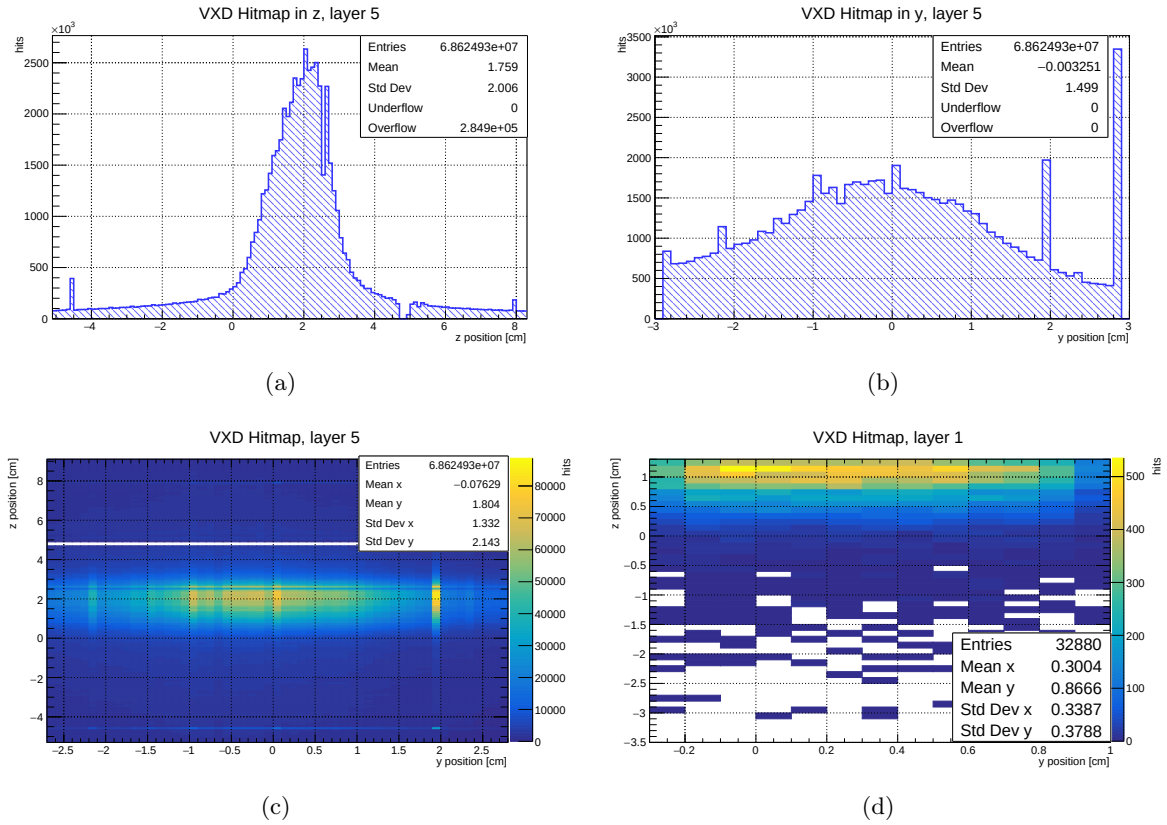
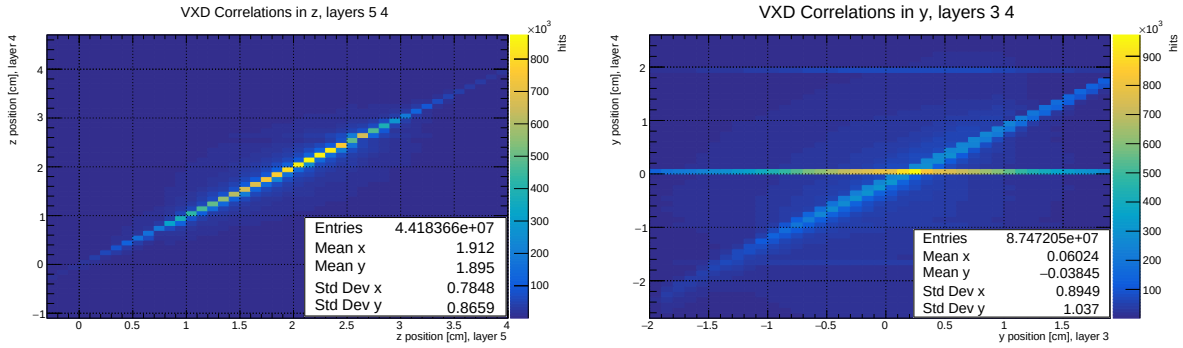



*Figure 6.4: Some examples of hitmaps for SVD and PXD layers. The white dip in figure 6.4(a) and 6.4(c) is the gap between two sensors. In figure 6.4(c) the noisy strip at the end of the sensor has been removed to improve the color resolution.*




(a) *An example of well-correlated layers, with an evident straight line*

(b) *An example of correlated layer where some noisy strips on Layer 4 degrade the result*

*Figure 6.5: Two examples of correlation plots between SVD layers*

Another relevant feature that is needed to check is the capability of HLT to extrapolate ROI on PXD layers. One of the relevant variables are the ROI residuals, defined as:

$$\mathcal{R}_u^{ROI} = y_{\text{ext}} - y_{\text{PXD}},$$
$$\mathcal{R}_v^{ROI} = z_{\text{ext}} - z_{\text{PXD}},$$

where $y_{\text{ext}}$ ($z_{\text{ext}}$) is the extrapolated $y$ ($z$) coordinate, and $y_{\text{PXD}}$ ($z_{\text{PXD}}$) is the $y$ ($z$) hit position on the sensor. In figure 6.6 an example of the ROI residual distribution is shown. A fast cross-check can be done with the comparison between the RMS of the central peak of the ROI residual distribution in $y, z$ direction, independent from the global ROI dimension, and the sum of the intrinsic PXD resolution and the RMS fit position error: $\sigma_{\text{PXD}} \oplus \sigma_{\text{fit}} \simeq \sigma_{\text{ROI}}$.

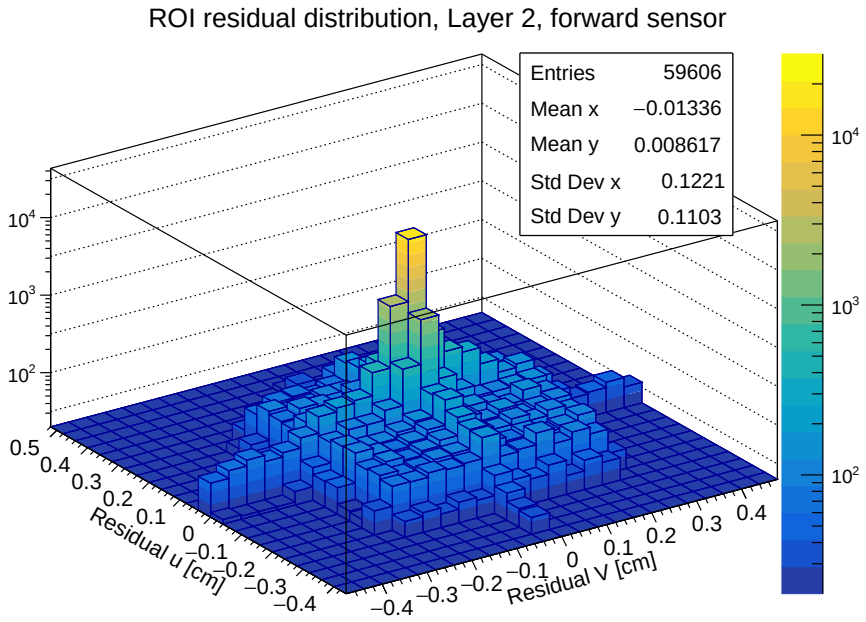
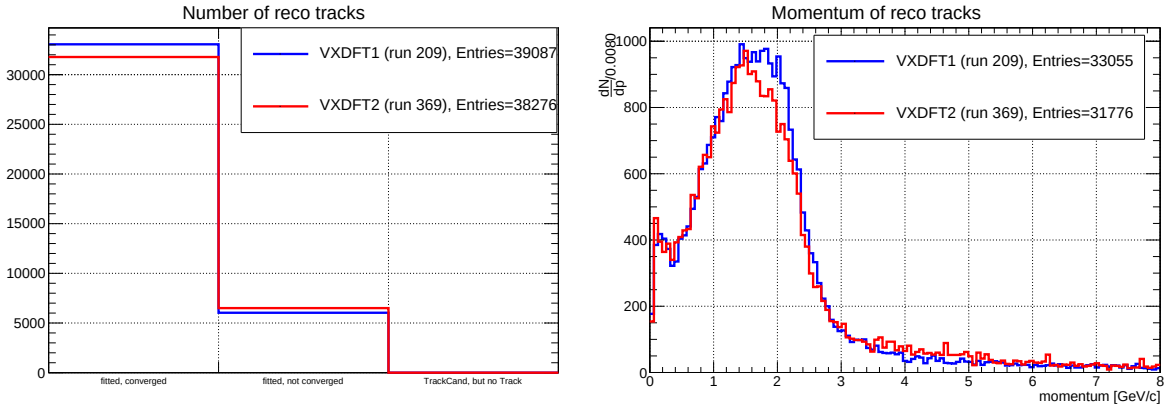



*Figure 6.6: Distribution of ROI residuals in $u \times v$ plane. Logarithmic scale is used to allow to appreciate the rectangular ROI shape under the main gaussian peak. Because of the chosen scale the standard deviation showed is not the one of the ROI peak but of the entire ROI.*

Finally it is interesting to show the comparison between two track finder online performances. This comparison has been done after some initial VXDTF2 runs, where the main bugs have been fixed, to understand if it would have been possible to collect the following data with the VXDTF2 in the HLT. The comparison has been done on online performances and cannot use the same data. Therefore two runs with the similar environmental parameters ($E_e = 2.4$ GeV/$c$ (VXDTF), $E_e = 2.0$ GeV/$c$ (VXDTF2), $\mathbf{B} = 1$ T, $\theta = \pi/2$) and the similar statistic are identified and they are compared. Some behaviour differences are expected because of the different energy. Without the DESY pixel telescope as a clear reference it is not possible to evaluate the tracking efficiency of VXDTF and VXDTF2 and therefore only the relative efficiencies of two the track finders can be shown. In figure 6.7(a) it is shown the number of reconstructed track subdivided between ones in which the fitting procedure converged and the ones in which it did not. A third category, empty for both the samples, are the Track Candidates removed from the overlap remover (the high energy leptonic events are very clear, in term of the number of tracks and topology). The result is similar for two the track finders, but the VXDTF2 found only 98% of the tracks found by VXDTF. To compare the capability of correctly reconstruct the tracks properties it possible

to look at the reconstructed momentum distribution. The result, in figure 6.7(b), is coherent with the efficiency one. Despite this lower efficiency, there is room for improvement optimizing the tracking procedures, therefore has been decided to collect about half of runs with VXDTF2 HLT.




(a) (b)

*Figure 6.7: Comparison between the performances of VXDTF and VXDTF2 in run 209 and 369 runs ($E_e = 2.4$ GeV/c, $\mathbf{B} = 1$ T, $\theta = \pi/2$). The number of entries is different between two plots because the overflow entries. The dips in VXDTF2 momentum distribution at 2 GeV/c could be due to the different nominal energy of the two runs. The large number tracks with non-converged fit can be recovered offline*

## 6.4 Hit efficiency measurement

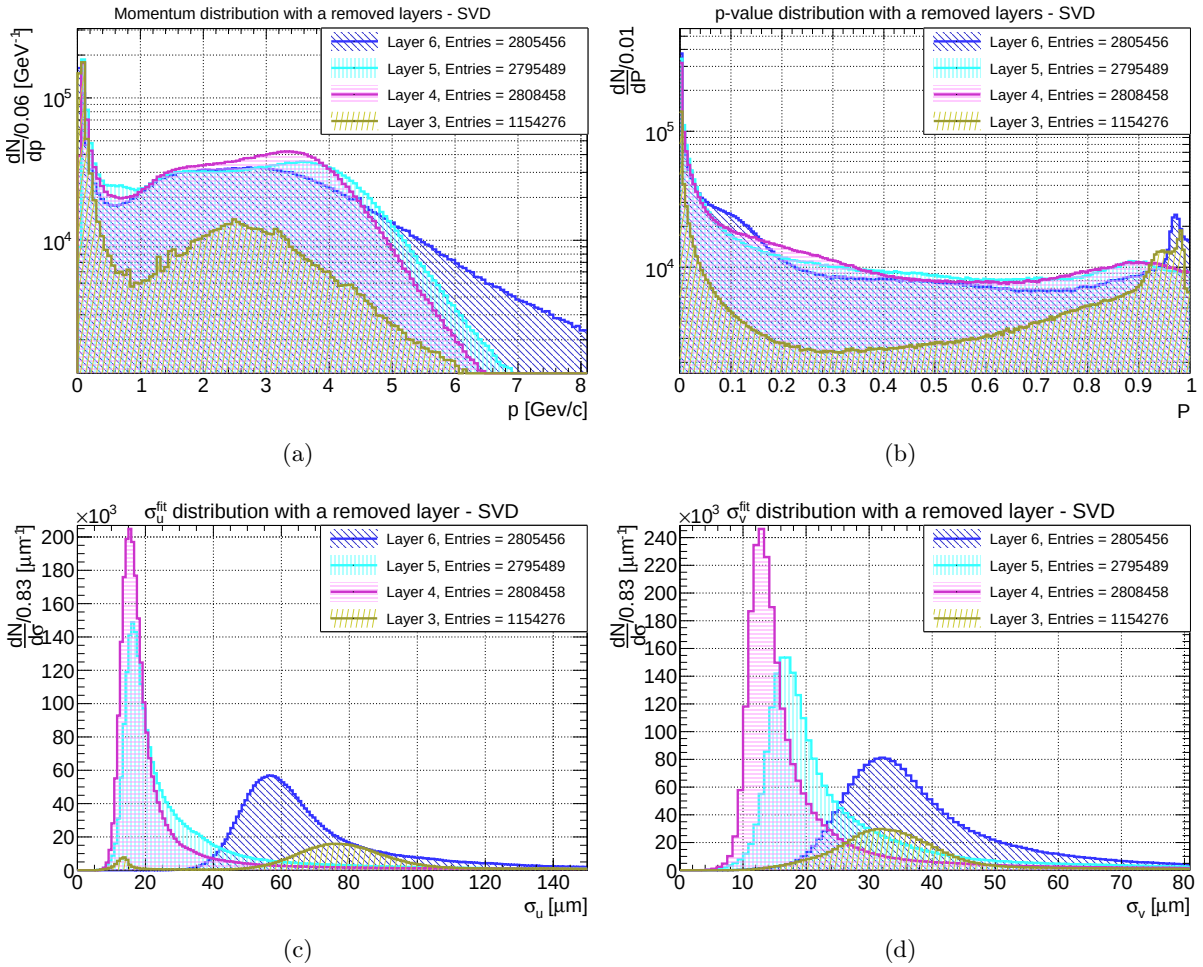
One of the relevant measurement feasible with the described testbeam setup is the hit efficiency of the SVD sensors. Conceptually the hit efficiency is the probability that a hit will be detected if a trackable particle crosses the detector: $\varepsilon_{\text{hit}}^{\text{teo}} = \frac{N_{\text{hits}}}{N_{\text{tracks}}}$. Therefore this property depends both on the tracking software and on the detector performances. The numerator depends on SVD modules only (sensor efficiency, readout efficiency...), can be different for each strip of the SVD and is affected by local defects. The denominator depends on the track finding capability of the tracking software, and should be independent of the local properties of the sensors but only on the general property of the tracks (momentum, angular direction, number of hits in the track...). The same measurement was performed during the Belle II VXD testbeam of 2016, thus the detector properties have been already measured and no differences are expected. Instead the track finding software has been completely redesigned since 2016 with the development of VXDTF2, therefore the measure of hit efficiency can return different results and should be reimplemented and redone with the new track finder. In this section the developed procedure is explained in detail and the result of the measurement is given. For offline reconstruction the VXDTF1 is still used to perform reconstruction, because when the measurement has been implemented it had better tracking efficiency.

The basis of the procedure is the method used in 2016 testbeam [42]. The events are reconstructed and fitted offline inside the `basf2` framework. The analysis is performed layer by layer: for each analysed SVD layer the tracks are reconstructed and fitted using the other three SVD layers and the two PXD layers only, i.e. without the information of the layer under study. To perform the reconstruction, the VXD is aligned with the complete track-based alignment framework of Belle II and specific Sector Maps are trained for the particular testbeam geometry with the different layers used in the tracking. In fact, the track finder learns the geometry of the detector from Sector Map tracks information, thus if a Sector Map trained with different geometry is used to reconstruct tracks, the result will be inefficient. The PXD layers are used to improve the

tracking efficiency, but the hit efficiency analysis is not extended to PXD layers because the detector properties are quite different and the extension is not trivial.

On the selected set of reconstructed tracks some quality cuts based on three estimators are applied: the reconstructed momentum, the $p$-value of the fitted track and the errors on the fitted position. In principle the distribution of these estimators i.e. the quality cuts may change for each layer under study, therefore for each estimator is performed a specific study to define the relative cuts.

**Momentum.** The reconstructed momentum of the tracks shows the sum of two distributions: the main beam-track distribution and a secondary distribution of tracks peaked at $p \simeq 0$. The latter events are very low momentum tracks (secondary particles, back-scattered particles... ), with an incorrectly reconstructed momentum because the Sector Map is not trained for this class of events. The momentum distributions show a long tail for high momentum too, another class of event with miseconstructed momentum. Both low momentum peak and high momentum tail should be removed to obtain a track sample with the expected momentum only, therefore for all the layers under study are set the quality cuts: $p > 0.7$ GeV/$c$, $p < 6$ GeV/$c$. The figure 6.8(a) shows the momentum distributions for each layer under study.




(a) (b)

(c) (d)

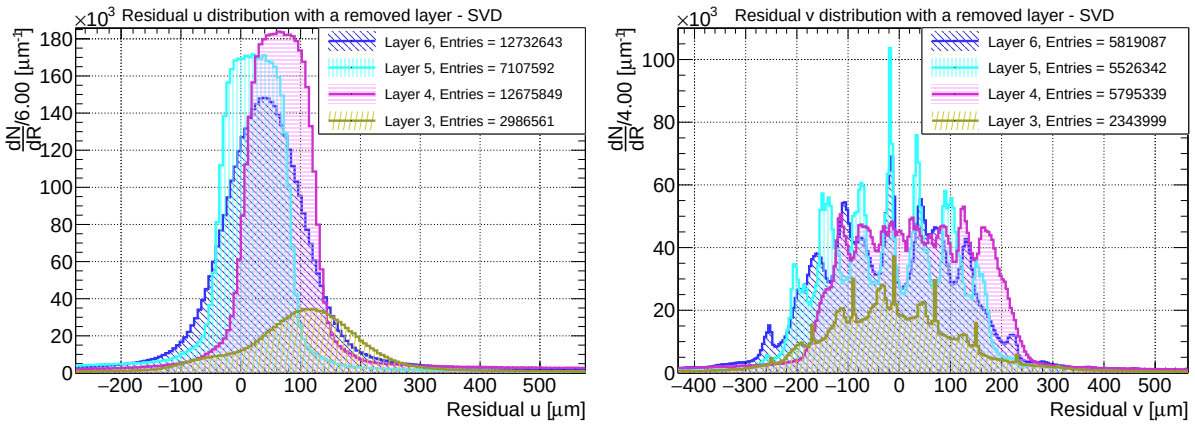
*Figure 6.8: Original distribution of the variables used to define the quality cuts, before the application of any selection.*

**$p$-value.** The $p$-value of the fit has almost the same distribution for the four layers under study: quite flat down to $P \simeq 0.2$ and with a large at $p$-value $P \simeq 0$. The slope of the peak slightly change between the various layer, but a cut $P > 0.2$ is set for all the layers to remove tracks with an unsatisfactory fit. In figure 6.8(b) the various $p-$value distributions are shown. Another peak is visible at $P \simeq 1$, probably due to track with very large errors (i.e. very low $\chi^2$). This class of track is not directly removed with an additional high-$p$-value cut, but the cut on $\sigma_{\text{fit}}$ is used.

**Fit errors.** The distribution of the errors assigned in the fitting procedure to the extrapolated position in the layer under study gives an additional information about the fit quality. The widths of the various distribution are quite different, and in particular the spread arises in extremal (Layer 3 and Layer 6) layers with respect to the inner ones (Layer 4 and Layer 5). This is reasonable: removing one extremal layer the fitting lever arm is reduced, while removing an inner layer it does not change. A layer and coordinate dependent cut $\sigma_{\text{thr}}$ is applied on the fit error $\sigma_{\text{fit}}$, with values spanning from 40 $\mu$m to 120 $\mu$m. The complete list is present in table 6.1. In figure 6.8(c) and 6.8(d) the various distributions are shown.

In addition to these cuts only the events with only one track are selected, retaining clean events. The purpose of these track selection is to obtain an effective tracking efficiency very close to 1 in all the analysed range, because in the efficiency definition the assumption of zero hit losses for inefficient tracking is implied.

Another relevant tracking parameter necessary to evaluate efficiency is the unbiased residual distribution, i.e. the distribution of the difference between the extrapolated position on a certain layer removed from tracking and the actual position of the hit in this layer. This quantity can be evaluated for each strip, considering the tracks that use the hit on the strip only, thus it is possible to understand the spatial limit of the hit assignment to a track during the reconstruction and its dependence from the layer. For each layer $l$ a threshold value $\mathcal{R}_{th}(l)$ has been set that will be used in the hit efficiency evaluation. These values, reported in the table 6.1, are 300 $\mu$ for each layer except 400 $\mu$m for Layer 3, because the layer 3 result strongly misaligned with respect to the rest of the SVD. In figure 6.9 the distributions for the entire layers (i.e. the sum of various hit distributions) are shown.




(a) (b)

*Figure 6.9: Residual distributions in $u$ and $v$ direction for the entire SVD layers. It is possible to notice a residual misaligment, in particular of layer 3. In the $v$ distribution the peaking structures are due to the strip design of the detector: despite the use of the weighted mean of the cluster position for multiple clusters hits, persits a not-continuum position distribution because of the discrete combination of strips that may build a track.*

| Layer | $N_{\text{track}}$/ev. | $p^{\text{th}}$ [GeV/$c$] (min/max) | $P^{\text{th}}$ (min/max) | $\sigma_u^{\text{th}}$ [$\mu$m] | $\sigma_v^{\text{th}}$ [$\mu$m] | $\mathcal{R}_{\text{th}}$ [$\mu$m] (min/max) |
|---|---|---|---|---|---|---|
| 3 | 1 | 0.7 / 6.0 | 0.2 | 100 | 50 | -400 / 400 |
| 4 | 1 | 0.7 / 6.0 | 0.2 | 40 | 40 | -300 / 300 |
| 5 | 1 | 0.7 / 6.0 | 0.2 | 60 | 65 | -300 / 300 |
| 6 | 1 | 0.7 / 6.0 | 0.2 | 120 | 80 | -300 / 300 |

*Table 6.1: Summary of the entire set of cuts applied on the track sample and the residual intervals that define the hit efficiency, for each layer. The large residual force to use a large $R_{th}$ which produces a loss in strip-by-strip resolution i.e. the capability to assign to each strip its missing or found hits.*

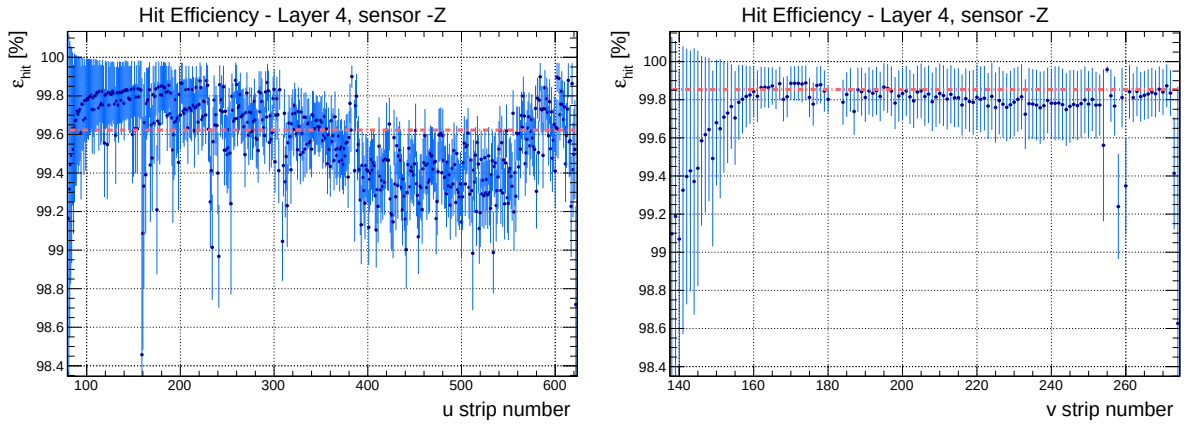
Then, after the application of the quality cuts on the track sample, for each track the strip closest to the interception point on the layer under study is identified and the hits within $\mathcal{R}_{\text{th}}(l)$ to the interception point are counted. Hence the hit inefficiency of each strip is evaluated as:

$$1 - \varepsilon = \frac{N_{\text{track}} - N_{\text{hit}}}{N_{\text{track}}} + B_{\text{norm}}, \tag{6.1}$$

where $N_{\text{hit}}$ is the number of the counted hits within $R_{\text{th}}(l)$ and $N_{\text{track}}$ the number of track whose closest strip is the analysed one. $B_{\text{norm}} \equiv N_{\text{bkg}}/N_{\text{out}}$ is the background probability (from noisy strip, secondary particles), evaluated as the number of hits in the event out of acceptance region defined by $\mathcal{R}_{th}$, normalized to the number of strips out of the acceptance region. This value is typically below $10^{-3}$. The errors on $\varepsilon$ are evaluated as:

$$\sigma^2_{(1-\varepsilon)} = \frac{N_{\text{track}} - N_{\text{hit}}}{N^2_{\text{track}}} + \frac{N_{\text{bkg}}}{N^2_{\text{out}}} + B^2_{\text{norm}}, \tag{6.2}$$

where a Poissonian error is assigned to $N_{\text{bkg}}$ and $\Delta N \equiv N_{\text{track}} - N_{\text{hit}}$ (because is expected $\varepsilon \simeq 1$ $N_{\text{hit}}/N_{\text{track}}$ has not a Poissonian distribution, while $\Delta N$ has it, see appendix C for details). The error on the number of tracks inside and outside of the acceptance region instead is neglected, because $N_{\text{track}}$ and $N_{\text{out}}$ define the studied sample and they are affected by reconstruction inefficiencies only. The last term $B^2_{\text{norm}}$ is added as a systematic error in the background evaluation, used to obtain a conservative result. In addition if for some events in formula 6.1 $\Delta N = 0$ is




(a) (b)

*Figure 6.10: Hit efficiency distributions for layer 4, p-strips (u) on the left and n-strips (v) on the right, the average fitted value is shown too.*

assigned $\Delta N = 0.7$, the upper limit of the mean (90% CL) of a Poisson distribution for zero observed events (i.e. if 0 events are observed at 90% the mean of $\Delta N$ is below 0.7).

In figure 6.10 are shown some examples of the hit efficiency distributions obtained with this procedure. In the efficiency shape are present several dips correspondents to known defects in the sensors. Depending on the type of defect, it may cause an excess of noisy hits in the strip or losses in the detection of real hits, but in both cases the efficiency results reduced. To obtain an average value for each sensor-side the hit efficiency distributions are fitted with a constant shown, while in figure 6.11 the entire set of fitted averages is reported. The values are quite different sensor by sensor because of the different number of defects which affects the average efficiency. The values $\mathcal{R}_{\text{th}}$ have been varied to study a possible bias due to too large residual windows, but the efficiency results are stable in quite large range (the efficiency appears degraded with $R_{\text{th}} < 100$ $\mu$m only, because of misalignment and intrinsic residual distribution of the sensors). the average values are stable in function of From these values is extracted a weighted average between the all the SVD sensors. This value is:

$$\varepsilon_{SVD} \simeq 0.9966,$$

and for all the sensors $\varepsilon_{\text{average}} \geq 0.992$. This value is coherent with the previous testbeam measurement.

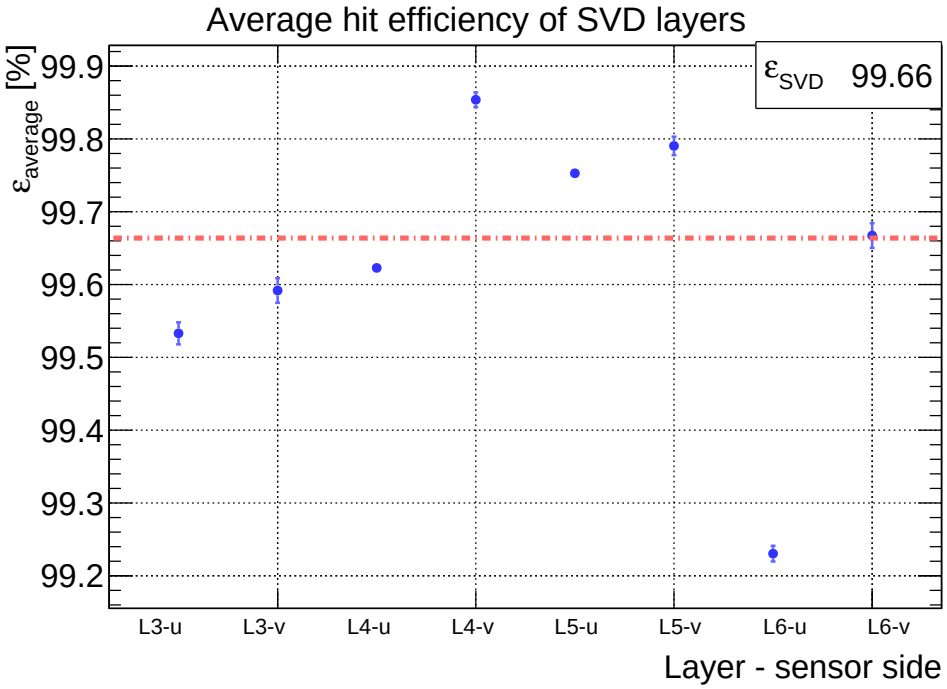



*Figure 6.11: Summary of fitted average values of the hit efficiency for various sensor-sides. The fitted average between various sensors is shown too.*

# Conclusions

The new generation B-Factory Belle II is going to start data taking in 2018, and thanks to the precision that it will achieve on the flavour physics measurement, has the possibility to discover New Physics signatures in $B$, $D$ or $\tau$ sectors. One of the central ingredients to perform competitive precision measurement in these sectors is a high-performance tracking system, both from the hardware and software side. During this thesis, both of these aspects of Belle II tracking have been studied.

From the hardware side, I developed the expertise required to test the Belle II Silicon Vertex Detector (SVD) modules inside the Forward and Backward SVD modules production of Belle II - Pisa group. In the thesis period I participated to the laboratory activities mainly performing the electrical qualification of the SVD modules. The production ended in September 2017 and in the last month of the year the assembly of the entire SVD will be performed at KEK, before the commissioning of the detector in which the Belle II - Pisa group will be involved.

From the software side I joined the Belle II tracking group, currently involved in a refactoring of the Vertex Detector Track Finder, the VXDTF2. I explored the Belle II tracking software and the pattern recognition techniques. One of the central objects of the VXDTF2 is the Training Sample of the Sector Map, the concept on which the Track Finder is based. The main part of my thesis work concerns the development of a robust procedure to define the Training Sample for the VXDTF2, since the previous definition produces unsatisfactory performances in term of low momentum fake rate, time consumption and complexity of the Sector Map. The procedure consists of a selection of the simulated tracks which must be contained in the Training Sample, to remove the classes of tracks which reduce the performance of the VXDTF2. The procedure of the selection automatically learns the sets track-properties useful to identify the interesting tracks from the characteristic of the simulated physical events and the geometry of the detector and store them. Then, during the training of the Sector Map the tracks contained in the Training Sample are filtered using the stored selection criteria values.

The tracking performance tested with the selected Training Sample are satisfactory: with a $3 \div 10\%$ degradation of the pattern recognition efficiency, depending on the momentum, the fake rate is halved, the time consumption reduced by a factor 3 and the fitting efficiency increases. In addition, the Sector Map complexity is reduced. A test on the VXDTF2 with the selected Training Sample on a background enhanced sample reveals that it is possible to gain efficiency from the selection, despite the fake rate is still halved and the CPU time reduced.

The procedure of Training Sample selection is not optimized, and the obtained results show the possibility of obtaining much better results with further studies in the Training Sample selection. One of the main upgrades that could be useful to develop is the automation of the required efficiency of the single selection criteria, currently settled by hand. This efficiency should be optimized on one or more figures of merit based on the tracking performance required.

During the thesis period, I had the possibility to test the SVD detector and the online tracking software on beam, in the DESY testbeam of February 2017. In this context, I performed Track Finder performance studies in the online and offline event reconstruction, which serve to validate the VXDTF2 on real data taking. In addition, I implemented the measurement of the hit efficiency of the SVD, which average value is $\varepsilon_{SVD} \simeq 99.66\%$. However, the collected testbeam data contains much more information both from tracking and Vertex Detetctor side and further studies are needed to exploit them completely. In particular, it would be interesting to evaluate the hit efficiency selecting data collected with a tilted geometry in which the tracks cross the SVD layers with a tilting angle.

# Appendix A

# Total tracking efficiency plots

In this appendix the plots of total tracking efficiency are reported, in bins of $\theta \in [-\pi/2, \pi/2]$ or $p_\perp$. The total tracking efficiency is evaluated as the number of tracks found by the Track Finder having more than 2/3 of hits matched with a single Monte Carlo Particle whose fit converged, over the number of Monte Carlo tracks found inside the detector (the geometrical acceptance is factored out), therefore:

$$\varepsilon_{\text{tracking}} \equiv \frac{N_{\text{reco}>2/3 \text{ \& fit converged}}}{N_{\text{tot}}}.$$

This definition is equivalent to the product of the pattern recognition efficiency and the fitting efficiency.

As following the comparison between VXDTF and VXDTF2 (figure A.1) and the comparison between various Sector Map Training Sample of VXDTF2 (figure A.2) are reported, together with the comparison between various Sector Map Training Sample in presence of nominal mixed background (figure A.3).

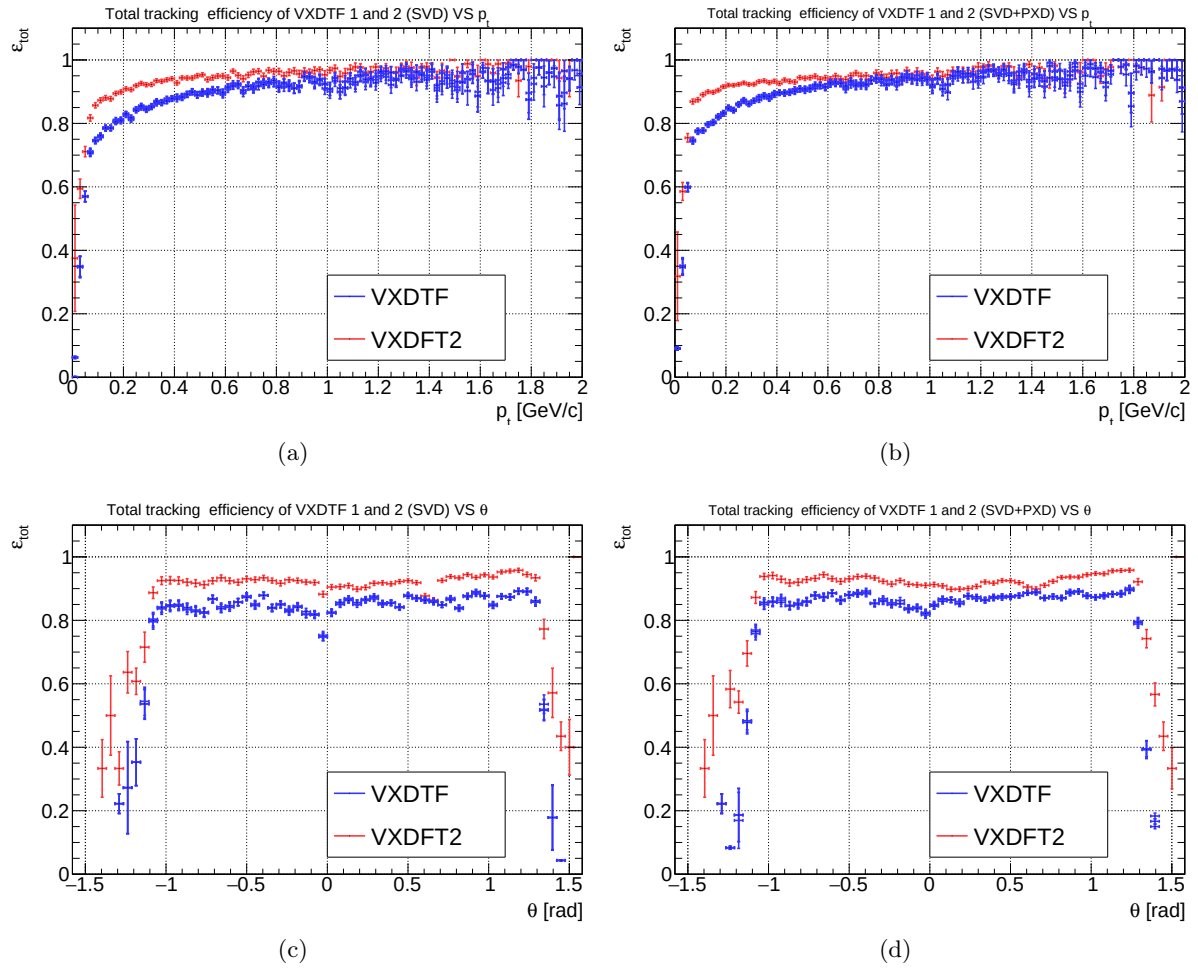



*Figure A.1: Comparison between total tracking efficiency of VXDTF and VXDTf2 with SVD only (4.11(a), 4.11(c)) or SVD+PXD (4.11(b), 4.11(d)) tracking in $p_\perp$ and $\theta$ trends.*

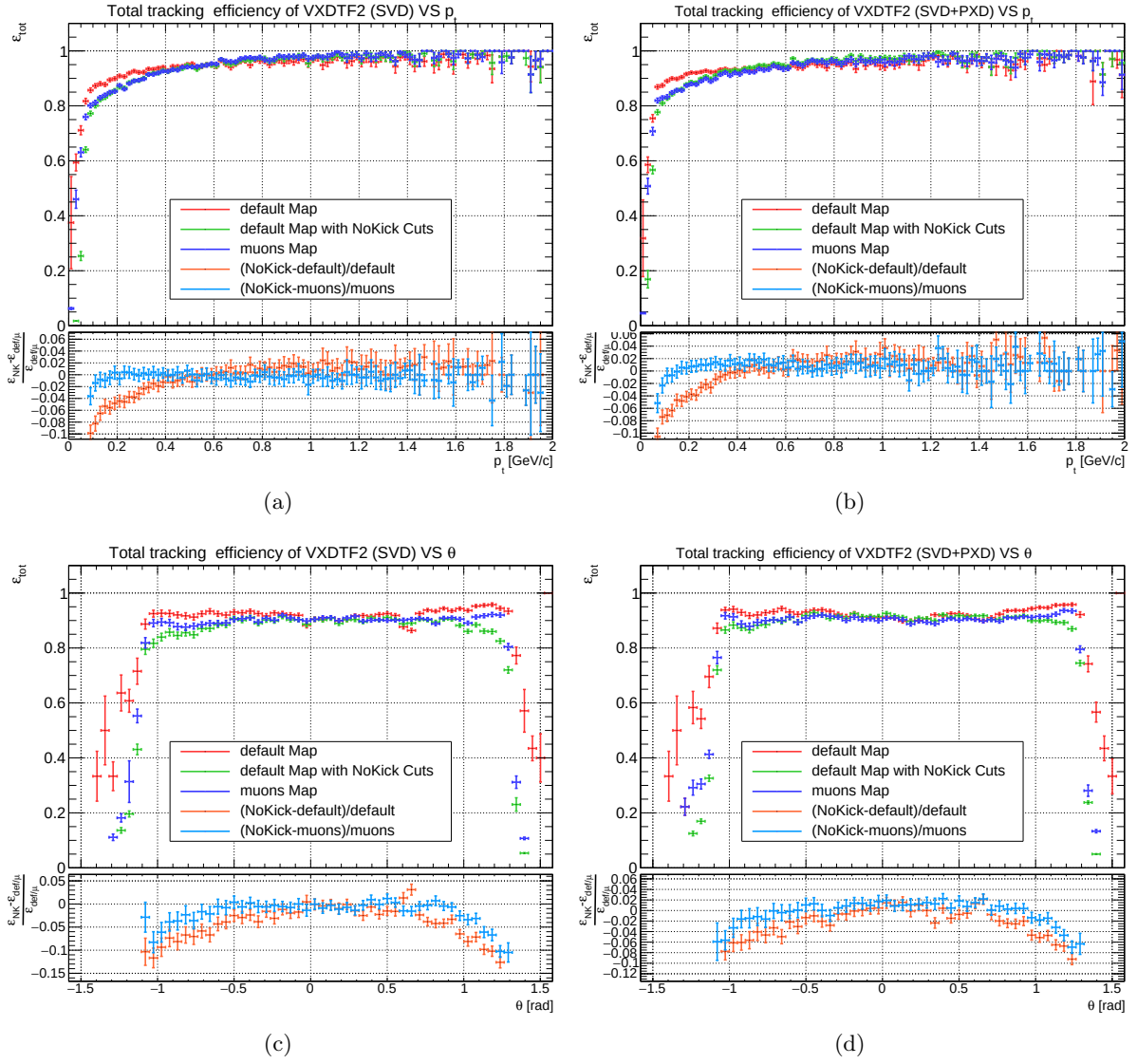
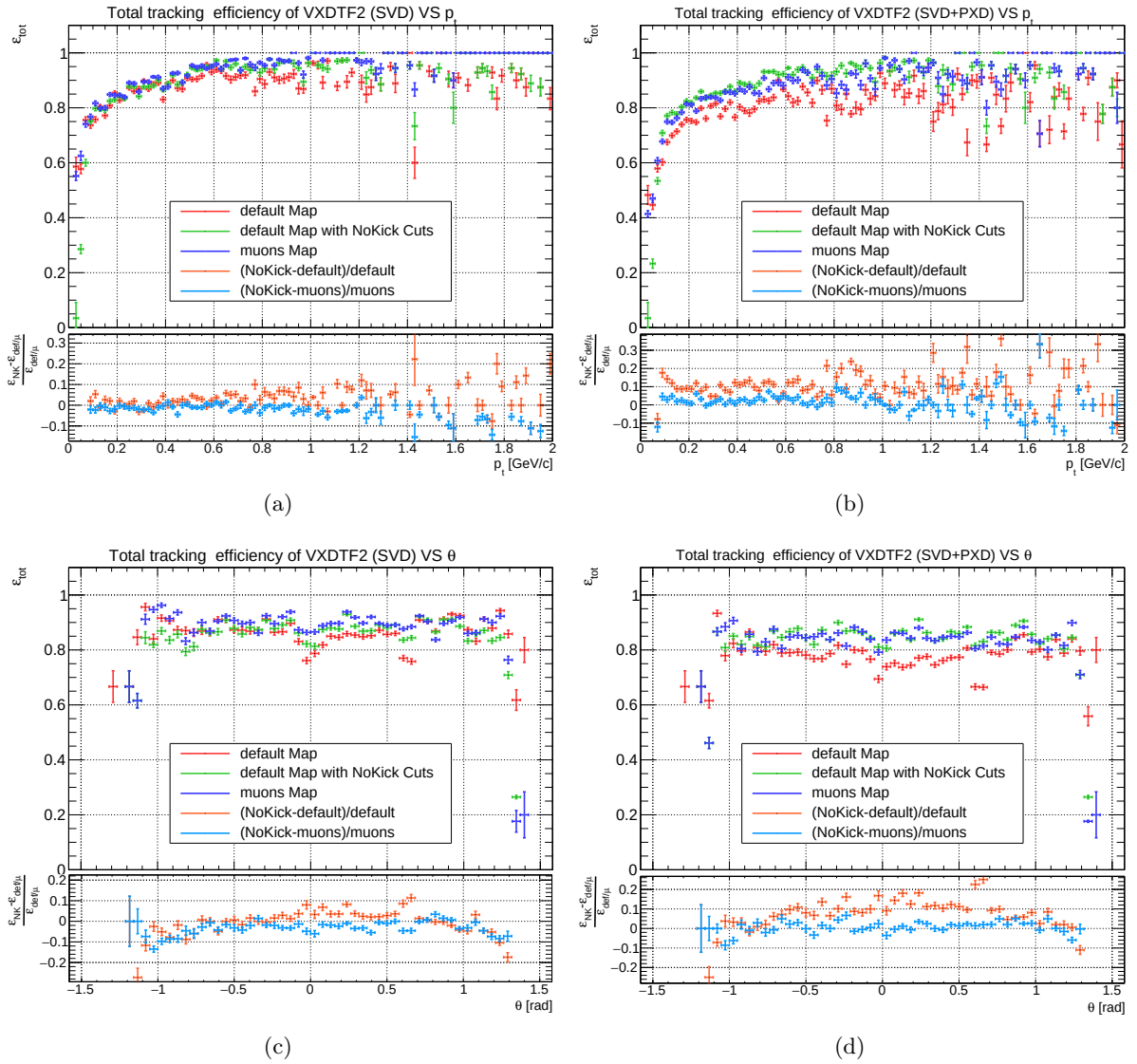



*Figure A.2: Comparison between total tracking efficiency of VXDT2 with Sector Map trained with default Training Sample and the one with the selection applied, in case of SVD only (A.2(a), A.2(c)) or SVD+PXD (A.2(b), A.2(d)) tracking in $p_\perp$ and $\theta$ trends.*




(a) (b)

(c) (d)

*Figure A.3: Comparison between total tracking efficiency of VXDT2 with Sector Map trained with default Training Sample and the one with the selection applied, in case of SVD only (A.3(a), A.3(c)) or SVD+PXD (A.3(b), A.3(d)) tracking in $p_\perp$ and $\theta$ trends. The nominal mixed background is added to the $\Upsilon(4S)$ simulated sample.*

# Appendix B

# Structure of the Training Sample selection inside basf2

The development of the procedure of the selection of the Training Sample required a structured organization of the code inside basf2 software (Belle Analysis Framework 2), to build the proper framework to evaluate and apply the selection criteria inside the Sector Map's training procedure of the VXDTF2 code.

The basf2 framework is used for online and offline data handling. The Belle II code has a modular structure: a typical processing chain consists in a linear arrangement of smaller processing block called *modules*, organized in a *path*. All the operations are done inside the modules, while the path specifies the execution order and the linking of various modules only. The data which have to be processed by the modules are stored in a common DataStore, from which each module reads and writes data. The DataStore contains the StoreArrays i.e. the data-objects used by basf2, and the RelationArrays i.e the relations between these objects. The figure B.1 illustrates the described structure. The path and the modules configuration are managed from python *steering files*, instead the modules and the core framework is written in C++ language with an intensive use of the root libraries[1]. Typically only software developers work on C++ code, and the analysis-level users write python steering file only.

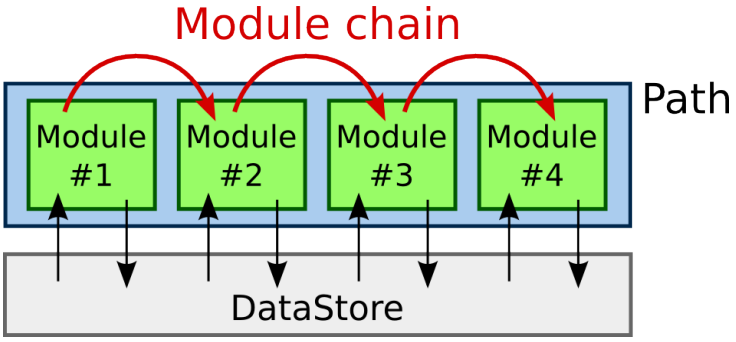



*Figure B.1: Structure of a typical basf2 processing chain.*

Each module is built with the same general structure: for each module are defined 5 standard methods, which if present are executed in a precise order:

– The method initialize() contains general declarations, the StoreArrays and the StoreArrays and needed for the module. The initialize() of all the modules of the path are executed

[1]Data analysis framework developed by CERN composed of several C++ libraries

preliminarily with respect to the rest of the methods which access to `DataStore` information.

- The method `beginRun()` contains all the definition and declaration which are used inside the data processing. The `beginRun()` of each module is executed before the module execution inside the path.

- The method `event()` performs a sequence of operation for each event of the given input file. This is the central part of the module.

- The method `endRun()` contains the operations which must be executed after the entire input file events processing.

- The method `terminate()` delete the objects created in the `initialize()`. The `teminate()` of all the modules of the path are executed at the end of the path, after the conclusion of the `endRun()` of the last module.

The purpose of this appendix is to describe the structure of the code required to select the Training Sample inside the `basf2` framework.

## B.1 NoKickCuts evaluation code

The selection criteria are evaluated in a preliminary phase with respect to the training preparation and the consequent training of the Sector Map. Therefore has been developed an independent steering file devoted to the NoKick Cuts evaluation only. This steering file uses a pre-generated sample of simulated events and reconstructs the events with the Monte Carlo Track Finder which builds the events with the Monte Carlo truth information. On these reconstructed events, a dedicated module called `NoKickCutsEval` evaluates the sets of the selection criteria and store them inside a `root` file, `NoKickCuts.root`.

In the `initialize()` the rank-5 tensor of 1-dimensional histogram is created. Then the module enters in the `event()` method, where each reconstructed track of the event is passed as a parameter to an object `NoKickRTSel`, a class which contains several methods useful to manage the selection criteria. With the use of this class from the reconstructed track is defined a vector of `hitXP` objects (discussed in the section 5.3.1) using the method `hit8TrackBuilder`. Then, on this object, with the method `globalCut`, is checked if the track satisfies the global requirements, otherwise it is discarded from the evaluation procedure. After this selection the histograms of the rank-5 tensor are filled with the `hitXP` information for track of each event of the input file.

The selection criteria are evaluated in the `endRun()` of the module because the complete set of complete histograms is needed for the analysis. First, for each histogram, the selection criteria are evaluated with the procedure of tail-removing described in section 5.3.4.2. Then the threshold values of the selection criteria are stored inside two rank-3 tensors of 2D histogram. These histograms are fitted with the function 5.7 and the values of the three fit parameters `norm, pow, bkg` for each selection criteria is stored inside six 3D histograms as explained in section 5.3.4.3.

In addition, the module allows to activate, by a public parameter, the production of several validation plots, which contain the $\Delta X$ distributions and the fitted selection criteria histograms.

In conclusion, the evaluation of the selection criteria is linear and a single module perform the entire procedure. It uses the `NoKickRTSel` class to manage the `hitXP` class to store properly the selection criteria values.

## B.2 NoKickCuts application code

The application of the selection criteria is integrated into the pre-existing algorithm of Training Sample definition. The first step of this procedure is the reconstruction of the tracks the Monte Carlo Track Finder from the sample of simulated data previously generated. After the reconstruction, these `RecoTracks` are converted in `SpacePointTrackCand` inside a module called `RT2SPTCConverter`. This conversion picks up one by one each track of the sample and applies the conversion. In this point of the procedure the selection criteria can be applied as a filter: if the `RecoTrack` that must be converted passes the filter it is converted, if it fails the filter it is discarded from the Training Sample.

Each analysed `RecoTrack` is passed to a `NoKickRTSel` object as a parameter. The `trackSelector` method applies the selection criteria and return a boolean used to select or discard the track from the `RT2SPTCConverter` event processing i.e. from the Training Sample. The method `trackSelector` first builds from the `recoTrack` a vector of `hitXP` object using `hit8TrackBuilder` method, then apply the global requirements. If the track passes the global requirements the for each pair of consecutive hits the correspondent selection criterion is found and then the method `SegmentSelector` compares the $\Delta X$ of the segment with the threshold values and discards or selects the segment with a flag. In case all the segments of the track are flagged as "selected" the result is passed to `RT2SPTCConverter` module and the track is included in the Training Sample. The procedure of the selection criteria, reading from the `NoKickCuts.root` file produced by `NoKickCutsEval` module and the finding of the selection criteria correspondents to each segment (in term of `X,l1,l2,p,t`), is managed from ad additional `NoKickCuts` class object with dedicated methods.

In conclusion, the procedure uses two classes, `NoKickRTSel` and `NoKickCuts`, to manage the selection criteria applications and the selection criteria reading respectively. It use a third class `hitXP` as a class-object which allows the comparison between the `recoTrack` parameters and the stored threshold values. It is completely integrated inside `RT2SPTCConverter` module processing, and require a `NoKickCuts.root` file where must present the threshold values.

# Appendix C

# Testbeam hit efficiency uncertainty

In this appendix are presented three different methods to obtain the result in equation 6.2, focusing on the main term of the efficiency and neglecting the background contribution. Therefore the definition of efficiency in this appendix is:

$$\varepsilon = \frac{N_{\text{hit}}}{N_{\text{track}}}, \tag{C.1}$$

according to the nomenclature of equation 6.1. The efficiency has a Bernoulli distribution where the two possible values are 1 and 0 with probability $\varepsilon$ and $(1-\varepsilon)$ respectively. The mean and the variance of the efficiency distribution are

$$\begin{aligned} \mu(\varepsilon) &= \varepsilon \\ \sigma^2(\varepsilon) &= \varepsilon(1-\varepsilon). \end{aligned}$$

Therefore if $N_{\text{track}}$ efficiency measurements are repeated, with $N_{\text{track}}$ high, thanks to the Central Limit theorem the expected mean of the efficiency means distributions is $\varepsilon$ and the variance is:

$$\overline{\sigma}^2(\varepsilon) = \frac{\sigma^2(\varepsilon)}{N_{\text{track}}} = \frac{\varepsilon(1-\varepsilon)}{N} = \varepsilon\frac{N_{\text{track}} - N_{\text{hit}}}{N_{\text{track}}^2}. \tag{C.2}$$

If $\varepsilon$ is expected close to 1 the result in equation C.2 can be simplified to

$$\overline{\sigma}^2(\varepsilon) \simeq \frac{N_{\text{track}} - N_{\text{hit}}}{N_{\text{track}}^2}, \tag{C.3}$$

a coherent result with formula 6.2 using the definition C.1.

The result C.3 is true independently from the error on the counts $N_{\text{hit}}$ in the efficiency definition. However, if it is possible to assign a poissionian error to all the counts, the error propagation from the formula C.1 produces the same result of formula C.2. The correlations between the counts must be taken into account, in fact in this framework $N_{\text{track}} = N_{\text{hit}} + N_{\text{missing}}$, and $N_{\text{hit}}$, $N_{\text{missing}}$ can be used as uncorrelated variables, on which is assigned a poissionial error.

The strategy adopted in the chapter 6 is to consider $1-\varepsilon$ distribution. In this case the binomial distribution resulting from the sum of the Bernoulli inefficiency distribution can be approximated to a Poisson distribution because $(1-\varepsilon) \ll 1$ and a poissonian error can be assigned to the $\Delta N = N_{\text{track}} - N_{\text{hit}}$ variable. In this case the error propagation directly brings to equation C.3, because the approximation $\varepsilon \simeq 1$ has been already applied considering $\Delta N$ a poissonian variable.

# Ringraziamenti

Quest'anno in Belle II è stata un'intensa, impegnativa e stimoltante avventura per la quale devo in primo luogo ringraziare Francesco Forti. Un relatore che ha sempre avuto la pazienza di trovare un momento per ascoltare un mio dubbio o valutare una mia proposta, e che fin dai tempi di Laboratorio 3 è stato un punto di riferimento nell'intraprendere nuovi sentieri. Insieme a Francesco ringrazio tutti i membri del gruppo di Belle II di Pisa, dai più esperti (Stefano, Giuliana) ai più giovani (Laura, Alberto, Giulia, Giacomo) o alle nuove reclute (Michael, Daniele) che hanno sempre saputo creare un ambiente accogliente dove passare le proprie giornate, che sia per testare un modulo o per alzare un boccale di birra in un OBM. Un ringraziamento particolare va ad Eugenio e Thomas (perdona l'italiano!), che quest'anno hanno pazientemente tentato di colmare le mie infinite lacune come sviluppatore software e mi hanno accompagnato passo passo sulla via del tracking, sorbendosi ogni mia più sciocca domanda. Grazie a tutti voi per avermi permesso di girare un po' il mondo lavorando su Belle II.

Questa tesi rappresenta anche la fine del mio percorso universitario come studente uniPi, e anche se resterò a Pisa so che molte cose cambieranno. Per questo vorrei ringraziare le persone che hanno passato questi cinque anni, o parte di essi, sui libri con me. In primo luogo Gregorio, un amico e un coinquilino con cui poter sempre trovare un contraddittorio, che si parli di fisica o di politica. Quindi voglio ringraziare Alfredo per la pacata allegria, tutta sua, che riesce sempre a portare. Non posso non ricordare tra un biscotto e una pizza, Luca, Marta e Gioele e le nottate passate con loro davanti a fit che non ne volevano sapere di convergere. Vorrei poi ringraziare Nick e Olmo, per aver affrontato con me i primi passi nel mondo dei pallettari, ed essere sempre pronti e carichi, che sia per discutere di fisica o per andare a bere.

Vorrei pure ringraziare i miei storici amici che hanno condiviso con me anche questi anni, condividendo momenti di studio, come Jacopo, Jimbo, Paolo e Alessandra, oppure solo di divertimento come tutto il gruppo delle *Feste Plebee*. Al suo interno un grazie particolare va al *Cineforum intellettuale e introspettivo* che ha spesso allietato le mie serate con capolavori di oggi e di ieri. Un grossissimo ringraziamento lo voglio fare a Viviana, un'amica che mi è stata molto vicina in questi anni, e ha sempre saputo e voluto ascoltarmi. Un ulteriore ringraziamento va a Isabella, coinquilina che è stata un po' il mio riferimento per imparare a vivere da solo.

Infine voglio ringraziare delle persone lontane che però sono spesso una ventata d'aria fresca rispetto all'ambiente pisano, con proposte nuove e interessanti: grazie Bort, Vacchi e Virginia.

Concludo ringraziano i miei genitori e tutto il supporto che non mi fanno mai mancare, nonostante la mia antipatia e il mio continuo sfuggire alla vita di casa.

Mi scuso con tutte le persone che ho dimenticato in questi ringraziamenti, scritti in un'ora davvero troppo tarda e davvero troppo vicino alla consegna per poter ricordare tutti coloro che hanno contribuito a questi cinque anni di università.

# Bibliography


[1] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the {ATLAS} detector at the {LHC}". In: *Physics Letters B* 716.1 (2012), pp. 1 –29. DOI: http://dx.doi.org/10.1016/j.physletb.2012.08.020. URL: http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[2] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the {CMS} experiment at the {LHC}". In: *Physics Letters B* 716.1 (2012), pp. 30 –61. DOI: http://dx.doi.org/10.1016/j.physletb.2012.08.021. URL: http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[3] A.J. Bevan et al. "The Physics of the B Factories". In: *Eur. Phyis. J.* C74 3026 (2014).

[4] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[5] Peter W. Higgs. "Broken symmetries, massless particles and gauge field". In: *Physics Letters* 12.2 (1964), pp. 132–133.

[6] N. Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[7] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev. D* 2 (7 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285. URL: https://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[8] J. J. Aubert et al. "Experimental Observation of a Heavy Particle $J$". In: *Phys. Rev. Lett.* 33 (23 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1404.

[9] J. E. Augustin et al. "Discovery of a Narrow Resonance in $e^+e^-$ Annihilation". In: *Phys. Rev. Lett.* 33 (23 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1406.

[10] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[11] J. H. Christenson et al. "Evidence for the $2\pi$ Decay of the $K_2^0$ Meson". In: *Phys. Rev. Lett.* 13 (4 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[12] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652. URL: http://dx.doi.org/10.1143/PTP.49.652.

[13] C. Bebek et al. "Evidence for New-Flavor Production at the $\Upsilon(4S)$". In: *Phys. Rev. Lett.* 46 (2 1981), pp. 84–87. DOI: 10.1103/PhysRevLett.46.84. URL: https://link.aps.org/doi/10.1103/PhysRevLett.46.84.

[14] Ashton B. Carter and A. I. Sanda. "CP Nonconservation in Cascade Decays of B Mesons". In: *Phys. Rev. Lett.* 45 (12 1980), pp. 952–954. DOI: 10.1103/PhysRevLett.45.952. URL: https://link.aps.org/doi/10.1103/PhysRevLett.45.952.

[15] H. Albrecht et al. "Observation of B0-B0 mixing". In: *Physics Letters B* 192.1 (1987), pp. 245 –252. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/0370-2693(87)91177-4. URL: http://www.sciencedirect.com/science/article/pii/0370269387911774.

[16] P. Oddone. "Detector Consideration". In: *Proceedings, Conference on Linear Collider B-B Factory Conceptual Design, Los Angeles, California, 26-30 Jan. 1987*. Vol. C870126. 1987, pp. 423–446.

[17] "CKM Fitter". In: *Eur. Phys. J. C* 41 (2005). [hep-ph/0406184] updated results and plots available on website, pp. 1 –131. URL: http://ckmfitter.in2p3.fr.

[18] J. P. Lees et al. "Measurement of an excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays and implications for charged Higgs bosons". In: *Phys. Rev. D* 88.7 (2013), p. 072012. DOI: 10.1103/PhysRevD.88.072012. URL: https://link.aps.org/doi/10.1103/PhysRevD.88.072012.

[19] S. Hirose et al. "Measurement of the $\tau$ Lepton Polarization and $R(D^*)$ in the Decay $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$". In: *Phys. Rev. Lett.* 118.21 (2017), p. 211801. DOI: 10.1103/PhysRevLett.118.211801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.211801.

[20] M. Huschle et al. "Measurement of the branching ratio of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ relative to $\overline{B} \to D^{(*)}\ell^-\overline{\nu}_\ell$ decays with hadronic tagging at Belle". In: *Phys. Rev. D* 92.7 (2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. URL: https://link.aps.org/doi/10.1103/PhysRevD.92.072014.

[21] A. Bozek et al. "Observation of $B^+ \to \overline{D}^{*0}\tau^+\nu_\tau$ and evidence for $B^+ \to \overline{D}^0\tau^+\nu_\tau$ at Belle". In: *Phys. Rev. D* 82.7 (2010), p. 072005. DOI: 10.1103/PhysRevD.82.072005. URL: https://link.aps.org/doi/10.1103/PhysRevD.82.072005.

[22] Y. Sato et al. "Measurement of the branching ratio of $\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau$ relative to $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$ decays with a semileptonic tagging method". In: *Phys. Rev. D* 94.7 (2016), p. 072007. DOI: 10.1103/PhysRevD.94.072007. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.072007.

[23] R. Aaij et al. "Measurement of the Ratio of Branching Fractions $\mathcal{B}(\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$". In: *Phys. Rev. Lett.* 115.11 (2015), p. 111803. DOI: 10.1103/PhysRevLett.115.111803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.111803.

[24] J. G. Körner and G. A. Schuler. "Exclusive semileptonic heavy meson decays including lepton mass effects". In: *Zeitschrift für Physik C Particles and Fields* 46.1 (1990), pp. 93–109. DOI: 10.1007/BF02440838. URL: http://dx.doi.org/10.1007/BF02440838.

[25] Florian U. Bernlochner et al. "Combined analysis of semileptonic $B$ decays to $D$ and $D^*$: $R(D^{(*)})$, $|V_{cb}|$, and new physics". In: *Phys. Rev. D* 95.11 (2017), p. 115008. DOI: 10.1103/PhysRevD.95.115008. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.115008.

[26] M. Tanaka and R. Watanabe. "Tau longitudinal polarization in $\overline{B} \to D\tau\overline{\nu}$ and its role in the search for the charged Higgs boson". In: *Phys. Rev. D* 82.3 (2010), p. 034027. DOI: 10.1103/PhysRevD.82.034027. URL: https://link.aps.org/doi/10.1103/PhysRevD.82.034027.

[27] E. Kou, P. Urquijo, and Belle2 Collaboration B2TiP theory Community. "The Belle II Physics Book". it is going to be published in: Progress of Theoretical and Experimental Physics.

[28] The Heavy Flavor Averaging Group (HFLAV). *Average of R(D) and R(D*) for Moriod EW 2017*. 2017. URL: http://www.slac.stanford.edu/xorg/hflav/semi/moriond17/RDRDs.html.

[29] M. Tanaka and R. Watanabe. "New physics in the weak interaction of $\overline{B} \to D^{(*)}\tau\overline{\nu}$". In: *Phys. Rev. D* 87 (3 2013), p. 034028. DOI: 10.1103/PhysRevD.87.034028. URL: https://link.aps.org/doi/10.1103/PhysRevD.87.034028.

[30] Y. Sakaki et al. "Testing leptoquark models in $\overline{B} \to D^{(*)}\tau\overline{\nu}$". In: *Phys. Rev. D* 88.9 (2013), p. 094012. DOI: 10.1103/PhysRevD.88.094012. URL: https://link.aps.org/doi/10.1103/PhysRevD.88.094012.

[31] Ilja Doršner et al. "Minimally flavored colored scalar in $\overline{B} \to D^{(*)}\tau\overline{\nu}$ and the mass matrices constraints". In: *Journal of High Energy Physics* 2013.11 (2013), p. 84. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2013)084. URL: https://doi.org/10.1007/JHEP11(2013)084.

[32] M. Jung, A. Pich, and P. Tuzón. "Charged-Higgs phenomenology in the aligned two-Higgs-doublet model". In: *Journal of High Energy Physics* 2010.11 (2010), p. 3. DOI: 10.1007/JHEP11(2010)003. URL: http://dx.doi.org/10.1007/JHEP11(2010)003.

[33] Roel Aaij et al. "Measurement of the ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong $\tau$-lepton decays". In: LHCB-PAPER-2017-017. LHCB-PAPER-2017-017 (2017). URL: https://cds.cern.ch/record/2281448.

[34] Belle2 Collaboration. *Belle II Technical Design Report*. arXiv:1011.0352. June 2015.

[35] Y. Ohnishi et al. "Accelerator design at SuperKEKB". In: *Progress of Theoretical and Experimental Physics* 2013.3 (2013), 03A011. DOI: 10.1093/ptep/pts083. URL: http://dx.doi.org/10.1093/ptep/pts083.

[36] M. Bona et al. *SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report*. 2007. arXiv: 0709.0451 [hep-ex].

[37] M. Akatsu et al. "Time-of-propagation Cherenkov counter for particle identification". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 440.1 (2000), pp. 124 –135. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/S0168-9002(99)00819-0. URL: http://www.sciencedirect.com/science/article/pii/S0168900299008190.

[38] J. Kemmer and G. Lutz. "New detector concepts". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 253.3 (1987), pp. 365 –377. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/0168-9002(87)90518-3. URL: http://www.sciencedirect.com/science/article/pii/0168900287905183.

[39] J Treis et al. "Pixel detectors for x-ray imaging spectroscopy in space". In: *Journal of Instrumentation* 4.03 (2009), P03012. URL: http://stacks.iop.org/1748-0221/4/i=03/a=P03012.

[40] J.J. Velthuis et al. "DEPFET, a monolithic active pixel sensor for the ILC". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 579.2 (2007). Proceedings of the 6th Hiroshima Symposium on the Development and Application of Semiconductor Detectors, pp. 685 –689. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/j.nima.2007.05.278. URL: http://www.sciencedirect.com/science/article/pii/S016890020701162X.

[41] F Mueller. "Some aspects of the Pixel Vertex Detector (PXD) at Belle II". In: *Journal of Instrumentation* 9.10 (2014), p. C10007. URL: http://stacks.iop.org/1748-0221/9/i=10/a=C10007.

[42] T. Lueck et al. "Performance studies of the Belle II Silicon Vertex Detector with data taken at the DESY test beam in April 2016". In: *25th International Workshop on vertex detectors*. PoS(Vertex 2016)060. Isola d'Elba, Italy: Proceedings of Science. URL: https://pos.sissa.it/287/060/.

[43] T. Geßler et al. "The ONSEN Data Reduction System for the Belle II Pixel Detector". In: *IEEE Transactions on Nuclear Science* 62.3 (2015), pp. 1149–1154. ISSN: 0018-9499. DOI: 10.1109/TNS.2015.2414713.

[44] M. Valentan. "The silicon Vertex Detector for b-taging at Belle II". PhD thesis. Vienna: Vienna Technical University, 2013. URL: https://inis.iaea.org/search/search.aspx?orig_q=RN:46127226.

[45] M.J. French et al. "Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 466.2 (2001). 4th Int. Symp. on Development and Application of Semiconductor Tracking Detectors, pp. 359 –365. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/S0168-9002(01)00589-7. URL: http://www.sciencedirect.com/science/article/pii/S0168900201005897.

[46] C. Irmler et al. "Construction and test of the first Belle II SVD ladder implementing the origami chip-on-sensor design". In: *Journal of Instrumentation* 11.01 (2016), p. C01087. URL: http://stacks.iop.org/1748-0221/11/i=01/a=C01087.

[47] A. Martini. "Measurement of the time dependent asymmetry in the decay $B^0 \to K^0_S \pi^0 \gamma$ using the Silicon Vertex Detector of the Belle II experiment". Pisa: Pisa University, 2016.

[48] A. Strandlie and R. Fruhwirth. "Track and vertex reconstruction: From classical to adaptive methods". In: *Rev. Mod. Phys.* 82 (2010), pp. 1419–1458. DOI: 10.1103/RevModPhys.82.1419.

[49] R. Fruehwirth et al. *Data Analysis Techniques for High-Energy Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge: Cambridge University Press, 2000.

[50] R. Fruhwirth and A. Strandlie. "Pattern recognition and estimation methods for track and vertex reconstruction". In: *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research*. PoS(ACAT2010)003. Jaipur, India: Proceedings of Science. URL: https://pos.sissa.it/093/003/.

[51] T. Alexopoulos et al. "Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 592.3 (2008), pp. 456 –462. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.04.038. URL: http://www.sciencedirect.com/science/article/pii/S0168900208005780.

[52] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. URL: http://cds.cern.ch/record/782076.

[53] G. Apollinari et al. *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2116337.

[54] M. Valentan, M. Regler, and R. Frühwirth. "Generalization of the Gluckstern formulas II: Multiple scattering and non-zero dip angles". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 606.3 (2009), pp. 728 –742. DOI: http://dx.doi.org/10.1016/j.nima.2009.05.024. URL: http://www.sciencedirect.com/science/article/pii/S0168900209009474.

[55] J. Lettenbichler. "Pattern recognition in the Silicon Vertex Detector of the Belle II experiment". Vienna: Vienna University, 19 May 2012.

[56] J. Lettenbichler and R. Frühwirth. "Real-time Pattern Recognition in the Central Tracking Detector of the Belle II Experiment". Presented on 28 11 2016. PhD thesis. Vienna: Vienna Technical University, 2016.

[57] Bilka, Tadeas et al. "The track finding algorithm of the Belle II vertex detectors". In: *EPJ Web Conf.* 150 (2017), p. 00007. DOI: 10.1051/epjconf/201715000007. URL: https://doi.org/10.1051/epjconf/201715000007.

[58] Y. Shrivastava, S. Dasgupta, and S. M. Reddy. "Guaranteed convergence in a class of Hopfield networks". In: *IEEE Transactions on Neural Networks* 3.6 (1992), pp. 951–961. ISSN: 1045-9227. DOI: 10.1109/72.165596.

[59] Deutsches Elektronen-Synchrotron DESY. *Test Beams at DESY*. 2017. URL: http://particle-physics.desy.de/e252106.