

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Alžběta Nečadová

DEPFET – polovodičové detektory pro japonský experiment Belle II

Ústav částicové a jaderné fyziky

Vedoucí bakalářské práce: RNDr. Peter Kodyš, CSc.

Školitel specialista: RNDr. Peter Kvasnička

Studijní program: Fyzika

Studijní obor: Obecná fyzika

Praha 2013

Tímto bych chtěla poděkovat vedoucímu své práce RNDr. Peteru Kodyšovi, CSc. a také svému konzultantovi RNDr. Peteru Kvasničkovi, za jejich vstřícný přístup, rady a připomínky po celou dobu vypracovávání této práce.

Dále bych chtěla velmi poděkovat svému manželovi Markovi a dceři Naděnce za jejich podporu v psaní této práce.

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 24. května 2013

Podpis autorky

Název práce: DEPFET – polovodičové detektory pro japonský experiment Belle II

Autor: Alžběta Nečadová

Katedra: Ústav částicové a jaderné fyziky

Vedoucí bakalářské práce: RNDr. Peter Kodyš, CSc., Ústav částicové a jaderné fyziky

Abstrakt: Tato práce se zabývá přesností křemíkových detektorů, konkrétně pixelového detektoru DEPFET. Z nasimulovaných dat se určovala předpověď pro chyby měření polohy částice v detektoru a hledalo se, jak se tyto chyby liší pro hodnoty různých parametrů, jak detektoru (např. velikost pixelu), tak částice (např. její energie). Jsou zde uvedeny základy principu křemíkových detektorů, jejich použití a přesnosti, experimenty Belle a Belle II v laboratoři KEK (Národní laboratoř pro fyziku vysokých energií v Japonsku) a dále statistické metody jako ANOVA, regresní stromy, rozdělení podle pravděpodobností (využití entropie). Nakonec je v práci uvedeno vlastní zpracování nasimulovaných dat pomocí regresních stromů a výsledné předpovědi chyb.

Klíčová slova: - Belle II, křemíkové detektory, DEPFET, regresní stromy

Title: DEPFET – Semiconductor Detectors on Japan Project Belle II

Author: Alžběta Nečadová

Department: Institute of Particle and Nuclear Physics

Supervisor: RNDr. Peter Kodyš, CSc., Institute of Particle and Nuclear Physics

Abstract: This work deals with accuracy of silicon detectors, especially pixel detector DEPFET. Prediction of particle's in-detector position measurement errors was determined from simulated data and dependence of these errors on several parameters was analysed—some of the parameters relating to the detector (e.g. pixel size) and another relating to the particle (e.g. its energy). Basics about principle of silicon detectors, their applications and accuracy, experiment Belle and Belle II in KEK (The High Energy Accelerator Research Organization) and also statistic method as ANOVA, regression trees, division by probability (using entropy) are summarized here. Finally, original processing of simulated data (using regression trees) and resulting error predictions are presented.

Keywords: - Belle II, silicon detectors, DEPFET, regression trees

Obsah

| | |
|---|-----------|
| Úvod | 3 |
| 1 Polovodičové detektory | 7 |
| 1.1 Stripové detektory | 7 |
| 1.2 Pixelové detektory | 8 |
| 1.2.1 DEPFET [1] | 9 |
| 1.3 Přesnost měření křemíkovými detektory | 9 |
| 1.3.1 Center of gravity (COG) | 9 |
| 1.3.2 Eta korekce | 10 |
| 1.3.3 Analog head-tail (AHT) | 11 |
| 2 Regresní stromy | 13 |
| 2.1 Dělení pomocí analýzy rozptylu | 13 |
| 3 Experimentální část | 17 |
| 3.1 Popis práce, postup | 17 |
| 3.2 Simulace | 17 |
| 3.2.1 Podmínky simulace použitých dat | 17 |
| 3.3 Třídění dat – regresní stromy | 19 |
| 3.3.1 Regresní stromy s použitím míry divergence | 19 |
| 3.3.2 Regresní stromy s využitím transformace dat | 19 |
| 4 Zpracování dat | 23 |
| 5 Diskuse | 27 |
| Závěr | 29 |
| Seznam použité literatury | 31 |
| Příloha 1 – Dodatky k teorii | 33 |
| Příloha 2 – Funkce pro program R | 35 |
| Příloha 3 – Výstupní tabulky pro vytvořené regresní stromy | 39 |

Úvod

Experiment Belle II v laboratoři KEK

Fyzikální motivace experimentu Belle II [1]

Standardní model (SM) částic a interakcí je jednou z nejlépe ověřených fyzikálních teorií mikrosvěta a elementárních částic (v mezích přesnosti současných experimentů). Navzdory jeho úspěchům při popisu základních sil a interakcí v přírodě zůstává několik zásadních otázek v rámci SM nezodpovězeno.

Není jasné, proč by měly být pouze tři generace elementárních fermionů. Mnoho bosonů a fermionů bylo určeno až experimentálně. Původ většiny těchto částic je vyjasněný v rámci SM spontánním narušením elektroslabé interakce. V kosmologickém měřítku je vážný neobjasněný problém s narušením symetrie hmoty a antihmoty ve vesmíru. Zatímco narušení CP symetrie (symetrie, při které se průběh fyzikálních zákonitostí zachovává při změně částice/antičástice a při změně parity) je nezbytnou podmínkou pro vývoj vesmíru, ve kterém dominuje hmota, pozorovaná CP narušení uvnitř kvarků jsou příliš malá pro vysvětlení dominance hmoty oproti malému množství antihmoty ve vesmíru. Takže zde musí existovat ještě neobjevený zdroj CP asymetrie odpovědný za dominanci hmoty ve vesmíru.

Narušení CP symetrie lze měřit např. pomocí rozpadů B mezonů. Nejprve rekonstruujeme dva rozpady $B \rightarrow \phi K_s$ a změříme časový rozdíl Δt mezi nimi. Určíme vůni jednoho z nich. Pak vyhodnotíme asymetrii ze získaného rozdělení časových rozdílů Δt .

Fyzika na Belle

Jedním z experimentů pro fyziku při vysokých energiích jsou právě experimenty Belle a Belle II v laboratoři KEK.

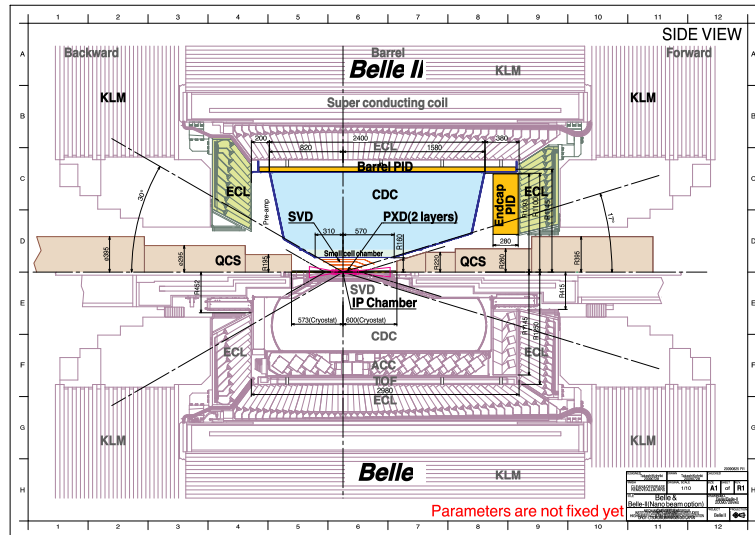
Hlavní vědecké objevy na experimentu Belle, na které naváže připravovaný experiment Belle II:

- 2002: Je pozorováno časově závislé CP narušení (TCPV) v systému neutrálních B mesonů a uskutečněno měření pro jeden úhel z unitárního trojúhelníku $\sin 2\phi_1$. Pro určení tohoto parametru pak byla vyžadována přesnější měření.
- 2003: Byl podán důkaz TCPV v rozpadu $B^0 \rightarrow \tau^+\tau^-$ a byla změřena hodnota dalšího úhlu ϕ_2 . Měření úhlu byla poté také zpřesňována.
- Ve stejném roce bylo na Belle pozorováno TCPV v penguin dominovaných rozpadech $B^0 \rightarrow \phi K_S, K^+K^-K^S, \eta'K^S$. Pozorovaný parametr $\sin 2\phi_1$ se mírně odlišoval od hodnoty naměřené v tree dominovaných rozpadech $B^0 \rightarrow J/\psi K^0$. Tento rozpor byl významně zmenšen s větším množstvím dat, ale potvrzení nebo vyvrácení jeho existence naráží na nedostatečný soubor dat. Větší citlivost experimentu Belle II bude dovolovat rozlišit přítomnost těchto efektů v penguin modech rozpadů.

- Dále byla objevena nová rezonance $X(3872)$, produkovaná v B rozpadech podobným způsobem jako charmonia, ale vykazuje vlastnosti, že klasifikovat je jako obvyklý mezon lze jen obtížně. Bylo pozorováno mnoho vlastností a rozpadů $X(3872)$ a také velké množství dalších zajímavých nových částic s podobnými vlastnostmi. Studium těchto stavů na Belle II bude snad možné objevit další, což pomůže našemu chápání kvantové chromodynamiky.
- 2004: Data umožnila první pozorování časově přímého CP narušení (DCPV) v rozpadech B na $\pi^+\pi^-$ a $K^+\pi^-$.
- Ve stejném roce bylo vykonáno první měření posledního úhlu ϕ_3 s použitím nové metody: časově nezávislého měření Dalitzova rozdělení v rozpadu $B \rightarrow D^{(*)}K$. Toto by mělo být prokázáno citlivějšími metodami a využíváno na Belle II.
- 2005: První měření TCPV v rozpadu $B^0 \rightarrow K^{*0}\gamma$, kde se b kvark rozpadá na s kvark a foton.
- 2006: Byl poprvé pozorován čistě leptonový rozpad $B^+ \rightarrow \tau^+\theta_\tau$.
- 2007: Belle a BaBar poprvé zaregistrovaly jev míchání v systému neutrálních mezonů D^0 .
- 2008: Byly změřeny části větvení a energetické spektrum fotonů pro rozpad $B \rightarrow X_s\gamma$ s energetickým prahem 1,7 GeV. Tento výsledek odpovídá SM a přináší velké omezení pro nabitý Higgsův boson. Tvar spektra pomáhá určit hmotnost b kvarku.
- Ve stejném roce měření DCPV v rozpadu $B^+ \rightarrow K^+\pi^0$ ukázalo, že je odlišné od rozpadu $B^0 \rightarrow K^+\pi^-$, v rozporu s předpovědí SM. V kombinaci s dalším měřením $B \rightarrow K\pi$ a s větší sadou dat z Belle II bude nezávislou cestou otestována platnost SM.
- 2009: Vzácný penguin rozpad $B \rightarrow K^*l^+l^-$ byl změřen se zlepšenou přesností.

Fyzika na Belle II

Několik příkladů rozpadů, které se budou měřit v experimentech na Belle II: $b \rightarrow s\bar{s}s$, $b \rightarrow s\gamma$, $B \rightarrow \tau\theta$, $B \rightarrow K\tau$. Toto je pouze pár příkladů z širokého fyzikálního programu na Belle II, který přináší hledání nové fyziky v moderní B továrně se zcela novou úrovní citlivosti a významně přispívá k dnešní experimentální snaze. Skutečná hodnota Super B továrny je v její schopnosti vykonávat měření ve všech oblastech fyziky vůně, rozpadů mezonu $B^{0(\pm)}$ a $B_s^{(*)}$, fyzice půvabu, fyzice τ leptonu, spektroskopii a čistých elektroslabých měřeních. Velký počet plánovaných měření bude přehodnocovat parametry SM a případně jeho rozšíření.



Obrázek 1: Předpokládaná geometrie Belle II. Převzato z [1]

Belle II

Pixelový detektor tvoří nejvnitřnější část vertexového detektoru. Následující změny oproti Belle zachovávají srovnatelné nebo zlepšují vlastnosti Belle II:

- křemíkový stripový detektor je doplněn dvouvrstvým křemíkových pixelovým detektorem DEPFET,
- křemíkový stripový detektor je umístěn za pixelovým detektorem a zasahuje do většího poloměru než u Belle,
- změny ve vyčítání z křemíkového stripového detektoru, v dráhovém detektoru, v elektronice elektromagnetického kalorimetru,
- vertexové rozlišení je zlepšeno výborným prostorovým rozlišením nejvnitřnějšího dvouvrstvého pixelového detektoru,
- účinnost rekonstrukce rozpadů při srážkách dvou nabitých pionů v křemíkovém stripovém detektoru je lepší, protože dráhový detektor zabírá větší objem,
- nová zařízení pro identifikaci částic velmi dobře rozlišují piony a kaony,
- nová elektronika elektromagnetického kalorimetru značně zmenšuje šum.

Význam odhadu přesnosti detektorů

Pro samotná měření potřebujeme velmi přesné detektory. Neméně důležité ale je vědět, s jakou chybou naše detektory měří. Vertexovými detektory určujeme polohu vrcholů z rozpadových diagramů a potřebujeme vědět, s jakou chybou byly tyto polohy určeny. Přesnost detektoru se může měnit pro různé jeho části, stejně tak pro různé parametry prolétávajících částic (např. jejich energie).

Cíle práce

Prolétá-li částice detektorem DEPFET, zanechá v místě průletu signál, který změříme. Naměříme tedy polohu částice. Otázkou je, s jakou velkou chybou tuto polohu určíme a hlavně, zda se velikost chyby určení polohy mění v závislosti na parametrech detektoru (např. velikost pixelu) a na dalších parametrech, které jsme schopni měřit (např. velikost naměřených clusterů, energie částice). Cílem práce bylo zkoumat, které parametry jsou z tohoto hlediska důležité a které nehrají roli. Poté určit chyby pro jednotlivé kategorie parametrů, u kterých se tyto chyby liší, a zda lze na základě znalosti těchto parametrů zlepšit odhad nepřesnosti polohy.

1. Polovodičové detektory

Pro studium polovodičových detektorů jsem nejvíce využívala text [2]. Polovodičové detektory jsou velmi využívány, především pro vysoký počet e-h párů vzniklých ionizací (až 10x více oproti plynovým detektorům). To je způsobeno jednak vysokou hustotou použitého materiálu, která způsobuje značné ionizační ztráty záření, a navíc malá šířka zakázaného pásu vede k malé energii potřebné na vytvoření e-h páru. Získáme tedy vysoký počet e-h párů. Je tedy možné konstruovat tenké detektory s velkým signálem. Tyto detektory mají většinou přesnost na mikrometry. Polovodičové detektory mají další výhody: mechanická pevnost, malá mrtvá doba, možnost dotování, možná integrace s elektronikou. Nejčastěji používané materiály jsou germanium (má vysokou absorpční schopnost pro fotony, používá se tedy jako RTG detektor nebo detektor IR záření, musí být chlazen, pro spektrometrické účely se používá driftovaný litium), křemík (využívá se k registraci těžkých částic, často užívaný driftovaný litium) a GaAs (vysoká pohyblivost, tedy se užívá pro aplikace s vysokou rychlostí, materiál spíše pro elektroniku).

Náboj vytvořený ionizujícím zářením zůstává lokalizován v malé oblasti, takže polovodičové detektory mají dobré prostorové rozlišení. Pro lepší výsledky se používají diskrétní detektory. Toto rozdělení detektoru na menší plošky je obvykle řešeno proužky (= stripový detektor) nebo obdélníčky (= pixelový detektor).

Dále nás zajímá, jaký je šum systému. Hlavním zdrojem šumu jsou fluktuace počtu a rychlosti nositelů náboje. Šum má tři hlavní složky:

- tepelný šum – lze jej snižovat ochlazováním,
- objemový šum – tvoří jej fluktuace závěrného proudu detektorem,
- šum na odporech – jeho spektrální hustota je nepřímo úměrná frekvenci.

Většinou se měří náboj, a proto se amplituda šumu převádí na veličinu ENC, tzv. ekvivalentní šumový náboj (toto je taková reprezentace šumu, která vstupuje přímo do měřené veličiny). Obecně platí, že šum s napětím klesá a navíc s napětím a s tloušťkou vyprázdňené vrstvy roste signál (tj. velikost sebraného náboje). Poměr S/N roste s napětím až do úplného vyprázdnění detektoru.

1.1 Stripové detektory

Elektrody zde tvoří proužky o šířce desítky až stovky mikronů a délce několik centimetrů. Hlavní omezení je fakt, že každý strip potřebuje svou vlastní vyčítací elektroniku (to limituje počet a vzdálenost jednotlivých stripů, vzdálenost stripů je limitována hlavně vzájemnou kapacitou – dlouhé, blízké stripy budou silně sdílet své signály).

Pro vyčítání můžeme použít několika způsobů. Nejjednodušší je předávat pouze logickou hodnotu ano/ne (binární vyčítání). Účinnost závisí na nastaveném prahu (ten se obvykle nastavuje na trojnásobek šumu). Přesnost měření polohy je pak dána jako $\sigma_x = \frac{p}{12}$, kde p je rozteč stripů. Při analogovém vyčítání

získáváme z detektoru informaci i o amplitudě signálu na jednotlivých stripech, čímž lze dosáhnout lepšího rozlišení. Lze rekonstruovat tvar oblaku náboje a pak spočítat jeho těžiště. Současně získáváme i informaci o celkové předané energii.

Lze použít i uspořádání, kde je jen část stripů připojena k vyčítací elektronice. Náboj, který je sebraný nepřipojenými stripy, se převádí ke čteným stripům. Mezi stripy lze nanést odporovou vrstvu pro dělení náboje. Tyto odpory ale generují šum, a tedy snižují rozlišení. Druhou možností dělení náboje na nečtených stripech je jeho rozdělení v poměru sériově zapojených kondenzátorů. Toto se dobře technicky realizuje, ovšem působí zde ještě kapacita mezi stripy a zemí a mezi stripy ob jedno navzájem. Mezistropy je potřeba udržovat na stejném potenciálu jako vyčítané stripy.

U detektorů se uvádí poměr signál/šum u jednoho kanálu. Šum je úměrný kapacitě a zpětnému proudu.

Ze získaných dat lze určit polohu několika způsoby:

- poloha je určena kanálem s nejvyšší amplitudou, chyba určení polohy je stejná jako u digitálního vyčítání $\sigma_x = \frac{p}{12}$,
- poloha je určena těžištěm signálu ve všech zasažených stripech (viz odstavec 2.3),
- používá se i η korekce, lineární nebo nelineární model. U lineárního η algoritmu (těžišťová metoda pouze pro 2 stripy) se poloha určí jako $x = x_L + p\eta$, kde $\eta = \frac{s_R}{s_L + s_R}$ a $s_{R(L)}$ je signál na pravém (levém) stripu. Toto předpokládá lineární rozdělení náboje na oba sousední stripy, což však většinou není splněno (náboj se ztrácí při přesunu - nositelé rekombinují nebo se zachytávají v materiálu, část náboje se přesouvá kondenzátory). Nejpresnější metodou určení polohy měřeného bodu je tedy nelineární η algoritmus $x = x_L + pf(\eta)$, kde $f(\eta)$ je taková funkce, aby výsledné rozložení bylo rovnoměrné v x .

Sběr náboje má na přesnost vliv ze dvou důvodů, jednak elektronická a pak také detektorová složka (např. separace e a h elektrickým polem, difúze nositelů náboje).

Výhodou je použití oboustranných detektorů, jelikož elektrony a díry jsou sbírány nezávisle na opačných stranách detektoru, tedy je lze sbírat nezávisle (použijí se stripy na obou stranách, kolmo na sebe). Výhodou je získání úplné prostorové informace o poloze prolétávající částice a možnost vyloučení kolizních událostí vzniklých korelací různých signálů ze stejných stripů. Toto řešení má však určité technické komplikace. Problémem je akumulace elektronů mezi N^+ stripy (lze řešit např. přidáním dalšího p stripu mezi N^+ stripy nebo implantace p vrstvy).

1.2 Pixelové detektory

Plocha senzoru je rozdělena na menší části, do matice pravoúhelníků. Jsou-li větších rozměrů (milimetry), jedná se o ploškový detektor, v případě jemnějšího dělení se jedná o pixelový detektor. Sběr náboje je obdobný jako u stripových detektorů. U ploškových detektorů (lze použít k detekci světla místo

fotonásobičů) lze náboj vést pomocí vodivých tras na detektoru. U pixelového detektoru je jednou z možností vyvedení náboje mimo rovinu senzoru, aby mohl být odveden do vyčítací elektroniky. To se řeší tak, že se k senzoru vodivě připojí čip, jehož vstupní plošky zrcadlí rozložení pixelů. Detektory i elektronika jsou vyrobeny z křemíku. Problémem však je, že detektory potřebují málo vodivý křemík a elektronika naopak vodivý křemík. Používají se jak detektory přizpůsobené elektronice, tak naopak (toto je použito u detektorů DEPFET).

1.2.1 DEPFET [1]

DEPFET (DEpleted Field Effect Transistor) je systém sestavený z pixelových detektorů. Detektor DEPFET je velmi přesný, používá se k přesnému určení pozice rozpadu částice. Na výrobu je použit vysoce nevodivý křemík. Lze jej používat v experimentech částicové fyziky pro částice s vysokou energií a mohl by být využit i v experimentech s rentgenovými paprsky a v biomedicině. Hlavní výhoda tohoto detektoru je kombinace detekce a zesílení.

Princip

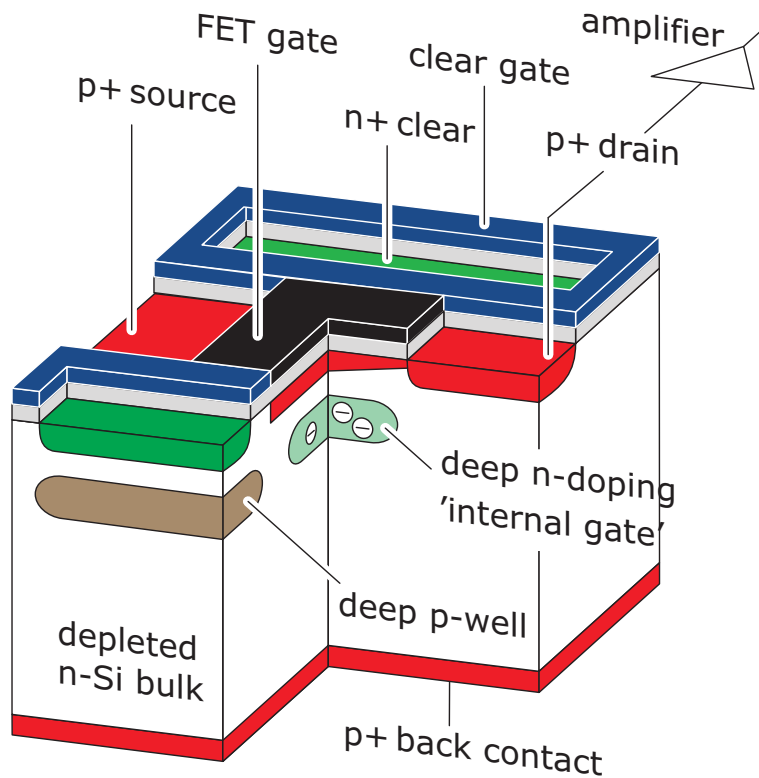
(Viz schéma na obrázku 1.1.) Jednotlivé části detektoru jsou tvořeny n křemíkem, tyto plně vyprázdněné oblasti mají tloušťku 50–450 μm . Vletí-li částice do této oblasti, vytvoří e-h páry podél své trajektorie a elektrické pole uvnitř oblasti oddělí elektrony a díry. Zadní plocha detektoru p+ má velké záporné napětí a přitahuje díry. Naopak elektrony jsou mezitím shromážděny v místě minima potenciálu (to je vytvořeno přidáním n+ materiálu), které se nazývá vnitřní brána. Vnitřní brána je umístěna pod transistorem, u detektorů DEPFET se používají transistory MOSFET (Metal Oxide Semiconductor Field Effect Transistor). Měříme výslednou změnu potenciálu na vnitřní bráně, která ovlivňuje proud v tranzistoru. Nashromážděné elektrony ve vnitřní bráně jsou po vyčtení převedeny pomocí kladného napětí na další tranzistor nazvaný clear gate. Jednotlivé pixely nemusejí být čteny neustále, elektrony zůstávají ve vnitřní bráně, dokud nejsou odstraněny do clear gate. Další výhoda detektoru DEPFET je nízká spotřeba energie, protože jednotlivé pixely jsou zapnuté, jen když jsou čteny.

1.3 Přesnost měření křemíkovými detektory

Křemíkové mikrostripové a pixelové detektory se obecně používají jako přesné vertexové detektory na urychlovačích v experimentech s fyzikou při vysokých energiích. Pro dráhy kolmé k povrchu detektoru vykazuje detektor prostorové rozlišení lepší než 1,8 μm při použití rozteče detektoru 25 μm . Existují různé algoritmy pro určování prostorového rozlišení detektoru, dále popíši nejpoužívanější z nich (podle [3]).

1.3.1 Center of gravity (COG)

Poloha je dána váženým průměrem $x_{\text{COG}} = \frac{\sum s_i x_i}{\sum s_i}$, kde x_i je poloha i-tého stripu (pixelu) obsaženého v clusteru a s_i je signál tohoto stripu (pixelu). Chyba určení



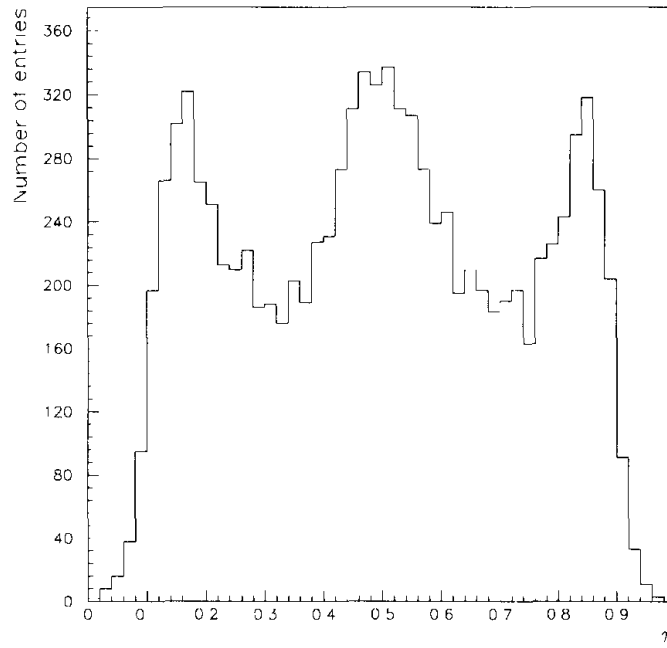
Obrázek 1.1: Schéma jednoho pixelu detektoru DEPFET. Převzato z [1].

této polohy je $\sigma_x = aP \frac{ENC}{s}$, kde a je koeficient závisící na modelu šumu systému a na počtu zasažených stripů, P je velikost stripu, s je signál a ENC je equivalent noise charge (nábojový ekvivalent šumu), způsobený hlavně elektronikou. Toto používáme pouze pro clustery malých velikostí.

1.3.2 Eta korekce

Proměnná η je definována jako $\eta = \frac{s_R}{s_R + s_L}$, kde $s_{R(L)}$ je velikost signálu na pravém (levém) stripu v naměřeném clusteru.

Běžně je v experimentálních výsledcích pík v η rozdělení. To znamená, že rozdělení náboje není blízké lineárnímu, protože jinak by η rozdělení bylo ploché. Tato pozorovaná nelinearita může být plně vysvětlena šířkou difuzního oblaku. Pro dráhy kolmé na senzor rozhoduje o tvaru clusteru difúze, pro dráhy pod velkým úhlem Landauovy fluktuační převažují nad difúzí. Šířka nábojového oblaku závisí na parametrech detektoru jako jeho tloušťka, odpor a použité napětí. Pokud částice narazí do detektoru přímo mezi dvěma stripy, rozdělení náboje je rovnoměrné (zhruba lineární) – platí pro kolmý průlet. Naopak prochází-li částice detektorem na hranici jednoho stripu, většina náboje se odvede na tento strip. V této oblasti η nabývá hodnot blízkých 0 nebo 1. Šířka oblasti, kde je rozdělení náboje rovnoměrné, závisí pouze na difuzi, šířka druhé oblasti je větší, čím větší je rozteč stripu (pixelu). Difúzi lze vysvětlit hlavní znak η rozdělení, jeho nelinearitu, ale nevysvětluje všechny jeho charakteristiky. Uvažujeme-li také šum, píky se rozšiřují a jejich šířka je určena velikostí šumu. Šířka píků je přibližně dána $\sigma_{pk} = \frac{ENC}{s}$, kde s je signál. Tato závislost je



Obrázek 1.2: Simulované rozdělení parametru η pro stripový detektor s jedním mezilehlým stripem, pitch = 50 μm , $S/N = 20$. Převzato z [3].

potvrzena jak experimentálními výsledky, tak simulacemi (v. obrázek 1.2).

Prostorové rozlišení, které můžeme očekávat, pokud používáme lineární model, je správné za předpokladu absence šumu.

1.3.3 Analog head-tail (AHT)

Pro šikmé dráhy je prvním přiblížením výpočet polohy průletu částice pomocí polohy začátku (head) x_h a polohy konce (tail) x_t clusteru $x_{DHT} = \frac{x_h + x_t}{2}$. Energetické ztráty jsou zhruba úměrné dráze částice v křemíku, podle AHT pak můžeme definovat jako $x_{AHT} = x_{DHT} + P \frac{s_h - s_t}{s_\theta}$, kde $s_h(s_t)$ jsou signály prvního a posledního stripu (pixelu) v clusteru, odpovídajícího dráze částice v detektoru. s_θ je průměrný signál zbylých (tj. vnitřních) stripů nebo pixelů. Velikost ztrát se řídí Landauovým rozdělením, s malou pravděpodobností nastávají i velké energetické ztráty. Abychom redukovali vliv tohoto rozdělení, můžeme definici vylepšit jako $x_{AHT} = x_{DHT} + P \frac{\min(s_h, s_\theta) - \min(s_t, s_\theta)}{2s_\theta}$. Výsledky simulací ukazují, že tato korekce je lepší než předchozí definice. Chyba určení polohy z měření je dána $\sigma x = \frac{N_t - N_h}{2s_0} t \sin \theta$, kde hodnoty N_t, N_h vyjadřují odchylku naměřené hodnoty energie od nejpravděpodobnější hodnoty. Tyto odchylky jsou způsobeny jednak Landauovým rozdělením depozice náboje, jednak šumem v elektronice.

2. Regresní stromy

Regresní stromy jsou typem nelineárního prediktivního modelu (základní popis v [4]).

Pokud máme velký soubor dat, která mají mnoho různých vzájemně se ovlivňujících znaků, může být použití globálního modelu, tedy aproximující funkce na celé oblasti hodnot všech parametrů obtížné, a navíc, povede-li se, výsledky mohou být zavádějící. Alternativní přístup stromové regrese spočívá v dělení oblasti parametrů problému na menší části, kde už jsou vzájemné vztahy lépe zvládnutelné. Vzniklé části zase dělíme (tzv. rekurzivní dělení), až nakonec dostaneme malé části, které už lze zvládnout nějakým jednoduchým modelem. Nehledáme tedy nějakou složitou funkci, kterou bychom problém popsali v celém rozsahu hodnot parametrů, ale hledáme oblasti (dělíme tedy podle parametrů, znaků), kde už můžeme problém popsat funkcí konstantní nebo lineární. Globální model tedy obsahuje dvě části, jednak rekurzivní dělení a pak nějaký jednoduchý model pro každou vzniklou část problému.

Stromy reprezentují rekurzivní dělení. Každý z koncových vrcholů (listů) stromu představuje oblast hodnot parametrů, které přísluší jednoduchý model týkající se pouze této části (viz příklad na obrázku 2.1). Pro nalezení předpovědi modelu pro dané hodnoty parametrů začneme v jeho kořeni a ptáme se sekvencí otázek na jednotlivé znaky. Uzly představují otázky a hrany (větve) představují odpovědi. V základní verzi se jedná o otázky týkající se pouze jednoho znaku s možnou odpovědí ano, nebo ne.

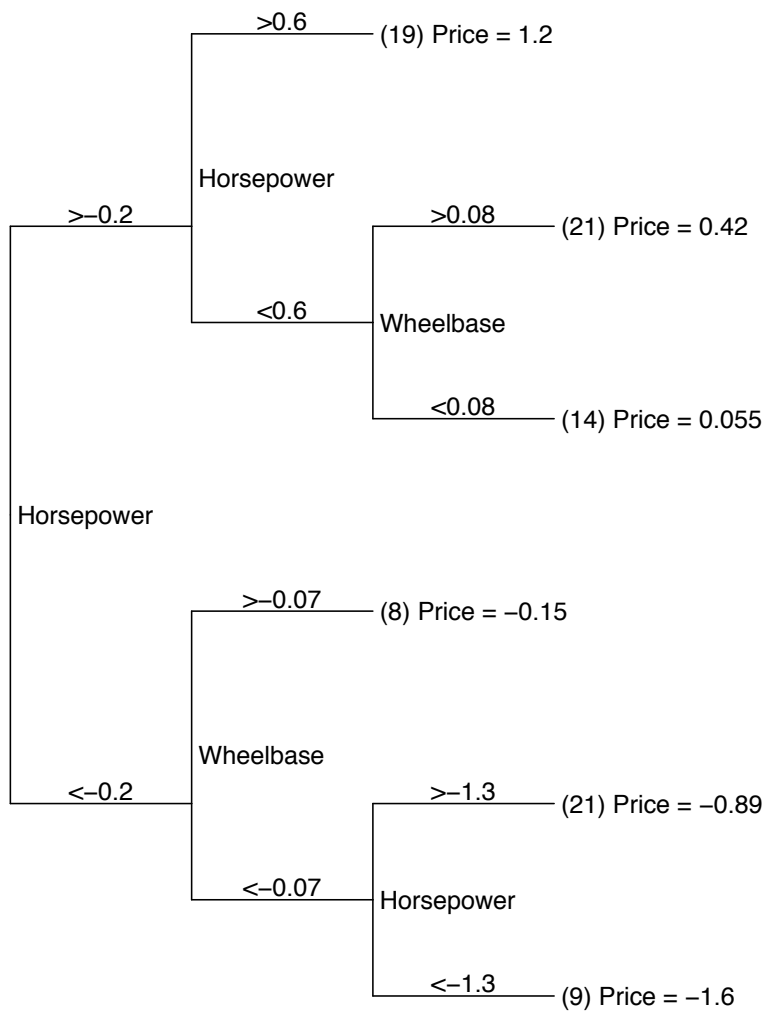
Výhody tohoto přístupu jsou, že vytváření predikce je rychlé (máme jen jednoduché modely pro rozdělené části), je jednoduché zjistit, které proměnné jsou pro předpověď důležité, a neznáme-li hodnoty některých parametrů, můžeme udělat predikci jako střední hodnotu přes příslušné větve stromu.

Jedním z problémů při vytváření stromu je zvolení vhodného dělení. Nevhodné je např. to, aby každý prvek měl ve výsledku svou vlastní část. Typicky je zadaným parametrem, kdy má už strom přestat růst, počet prvků v listu, pro který už se další dělení neprovádí, nebo počet prvků v nově vznikajícím listu, který se musí překročit, aby se toto dělení provedlo.

Nejdůležitější částí vytváření stromů je splitovací (dělící) funkce. Je-li proměnná numerická a má N možných různých hodnot, pak je možných $N - 1$ různých dělení. Pokud máme kategoriální proměnnou, zkouší se všechna možná dělení, tedy $2^{N-1} - 1$ různých dělení.

2.1 Dělení pomocí analýzy rozptylu

ANOVA (neboli analýza rozptylu, [5]) je metoda matematické statistiky, pomocí které lze ověřit, že na hodnotu náhodné veličiny má významný vliv hodnota některého znaku. Jednotlivé znaky nabývají konečného počtu hodnot (pro spojitě znaky mluvíme o lineární regresi) a slouží k rozdělení prvků veličiny do různých skupin. Tato metoda předpokládá, že všechny náhodné veličiny jsou nezávislé a mají normální rozdělení. Předpoklad normálního rozdělení však není zcela nutný, rozdělení náhodné veličiny by však mělo být aspoň blízké



Obrázek 2.1: Příklad regresního stromu. Převzato z [4].

normálnímu rozdělení. Dále se předpokládá stejný rozptyl u všech podskupin. Prvky se pak dělí do skupin podle znaků, v každé skupině se určí střední hodnota a určí se, zda se tyto střední hodnoty významně liší. Hledáme takové dělení do podskupin, aby se jejich střední hodnoty co nejvíce lišily. Pokud jsou střední hodnoty v obou podskupinách stejné (nebo podobné natolik, že jejich rozdíl není významný), dělení podle tohoto znaku se vyhodnotí jako zbytečné. Tuto metodu lze využívat také např. pro tvorbu regresních stromů.

3. Experimentální část

3.1 Popis práce, postup

K určení předpovědi byla využita simulovaná data, takže byly k dispozici jak naměřené hodnoty polohy, tak skutečné polohy prolétávajících částic. Odtud se určovala chyba měření polohy pro jednotlivá měření (u kterých byly zároveň známy, jaké hodnoty měly dané parametry). Jednotlivé parametry jsou zapsány v tabulce 3.1 společně s možnými hodnotami, jednalo-li se o kategoriální proměnnou.

Dále byly k dispozici ještě úhlové parametry, kterými jsem se ale při určování chyb nezabývala.

Pro zpracování dat se využívalo třídění pomocí regresních stromů. V již hotových stromech se pak pro data v jeho jednotlivých listech vypočítaly hodnoty rozptylu, standardní odchylky a systematického posunu (biasu).

3.2 Simulace

Po získání dat (simulace, měření) jsou data zapsána a následně zpracovávána offline softwarem (podrobný popis softwaru v [1]). Ve zpracování se jedná o rekonstrukci drah letu částic, Monte Carlo simulaci a fyzikální analýzu. Offline software je navržen jako prostředí takové, že uživatel je schopen veškeré funkce provádět v tomto prostředí. Lze přidávat i nové části, navržené uživatelem. Typicky si uživatel píše vlastní moduly pro fyzikální analýzy.

V experimentech pro fyziku vysokých energií znamená simulace tři kroky: generování jednotlivých událostí (procesů, generujících částice), simulací dráhy částice v detektoru a digitizace (tj. simulace odezvy detektoru). První krok obsahuje generování případů pro různé fyzikální studie, např. rozpad B mezonu nebo simulaci pozadí. Druhý krok slouží k simulaci interakcí jednotlivých částic při jejich průletu detektorem, popisuje jeho materiály a geometrie a zaznamenává energii částice odezvanou v objemu detektoru. Třetí krok odpovídá jednotlivým citlivým částem detektoru, bere v úvahu detaily fyzikálního procesu, vlastnosti elektroniky a vytváří finální nasimulovaný zásah.

3.2.1 Podmínky simulace použitých dat

Pro práci byl využit soubor dat PXDHitErrorOutputRT_15.txt, v němž jsou odezvy částic létajících v Belle II geometrii v rozmezí $\pm 90^\circ$ od roviny kolmé na přímkou srážky částic (úhel měřený od kolmé roviny na srážku). Částice vyletují z oblasti 10–15 mm mimo střed v ose x a 0 mm v ose y (osy kolmé roviny) a v rozmezí -10 až 10 mm v ose z (ve směru letících částic. Částice létají ve všech směrech homogenně.

Létající částice jsou elektrony, pozitrony, piony a antipiony, všechny homogenně v rozmezí energií 0,1–2 GeV. Magnet byl zapnutý.

Zdroj byl takto zkonstruován proto, aby částice létaly detektory i pod velmi velkými úhly a všemi směry (zvoleno jen pro případ vytváření korekcí).

| parametr | užívaný název | možné hodnoty |
|---|---|---|
| vrstva detektoru velikost naměřeného clusteru ve směru u velikost naměřeného clusteru ve směru v celková velikost naměřeného clusteru typ clusteru podle jeho tvaru celková energie zanechaná částicí energie v pixelu s největším signálem rozteč pixelu ve směru u rozteč pixelu ve směru v | Layer Size.u Size.v Size ClusterShapeType Charge.ADU Seed.ADU Pitch.u Pitch.v | 1, 2 celočíselné hodnoty celočíselné hodnoty celočíselné hodnoty 0, 1, 2, 3, 40, 41, 42, 43, 44, 5, 6, 7, 8, 23 spojitá proměnná spojitá proměnná 60 55, 60, 70, 85 |

Tabulka 3.1: Jednotlivé parametry pro nasimulovaná data.

Tato koncepce bude ještě upravována, protože nezabezpečila dostatečně dobré počty částic pro všechny možné kombinace. Pro účely této práce by však měla být dostatečně dobrá.

3.3 Třídění dat – regresní stromy

Samotné třídění bylo řešeno pomocí regresních stromů v programu R.¹ R je softwarové prostředí s vlastním programovacím jazykem zaměřené na statistiku. V tomto programu byly upravovány datové soubory a pak také vytvářeny samotné regresní stromy pomocí balíku RPART ([8]). Samotná funkce `rpart` má již vytvořené splitovací funkce, v úvahu přicházela metoda ANOVA. Ta však dělí data podle středních hodnot. To se v našem případě nedalo použít, jelikož střední hodnota byla pro podskupiny dat víceméně shodná, podskupiny se lišily velikostí rozptylu a právě tyto rozdíly pro nás byly zajímavé. Našli jsme dva různé přístupy, kterými se toto dalo řešit. Jedná se o transformaci dat, aby již bylo možno používat splitovací funkci s metodou ANOVA nebo využití entropie. Druhý přístup byl elegantnější v tom, že by bylo možné používat přímo naše data a dělit je do podskupin pomocí entropie, ale bylo by potřeba napsat vlastní splitovací funkci (a také inicializační a vyhodnocující funkce, [9]).

3.3.1 Regresní stromy s použitím míry divergence

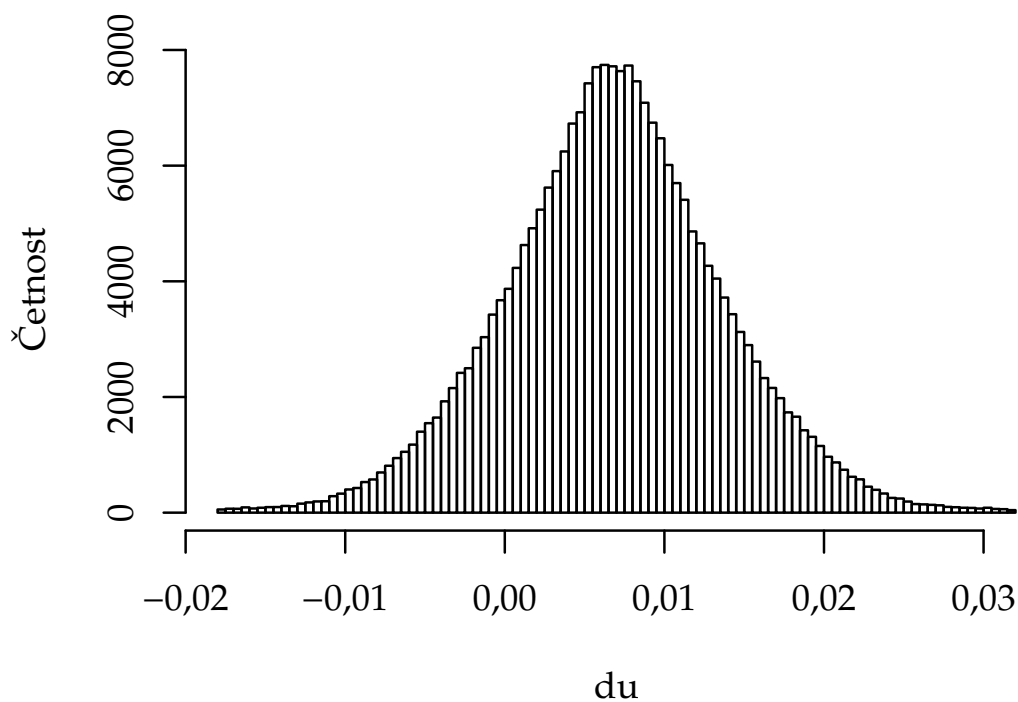
Dělení by bylo prováděno podle rozdělení pravděpodobnosti. Tato rozdělení pravděpodobnosti by byla porovnávána pomocí nějaké míry podobnosti, např. pomocí entropie. Data by byla reprezentována pomocí histogramu. Histogram by byl vytvořen tak, aby v každém binu bylo stejně (nebo alespoň zhruba stejně) prvků. Jednotlivé biny by tedy byly různě široké, okolo střední hodnoty by byly potřeba úzké biny a na okrajích rozdělení zase naopak spíše široké, aby neobsahovaly málo prvků. K tomuto histogramu by se určila entropie jako $H = -\sum \frac{n_i}{n} \ln \frac{n_i}{n}$, kde n je celkový počet prvků a n_i jsou počty prvků v jednotlivých sloupcích histogramu. Po rozdělení dat do dvou podskupin by vznikly dva histogramy, u nichž by byla stejným způsobem určena entropie, a hledalo by se takové rozdělení, aby se smísením těchto částí ztratilo maximum informace. Tedy by se hledalo dělení $D = D_a + D_b$ takové, aby $H_a + H_b - H$ bylo minimální. Což by znamenalo, že se z původního složeného rozdělení podařilo oddělit dvě různá rozdělení (tj. rozdělení s různými velikostmi rozptylu).

3.3.2 Regresní stromy s využitím transformace dat

Pro využití již hotové funkce v programu R používající metodu ANOVA bylo potřeba transformovat data. V této metodě se po rozdělení do dvou podskupin porovnávají střední hodnoty a hledá se takové rozdělení, aby rozdíl středních hodnot byl co největší. Jsou tedy potřeba různé střední hodnoty, nikoliv různé rozptyly.

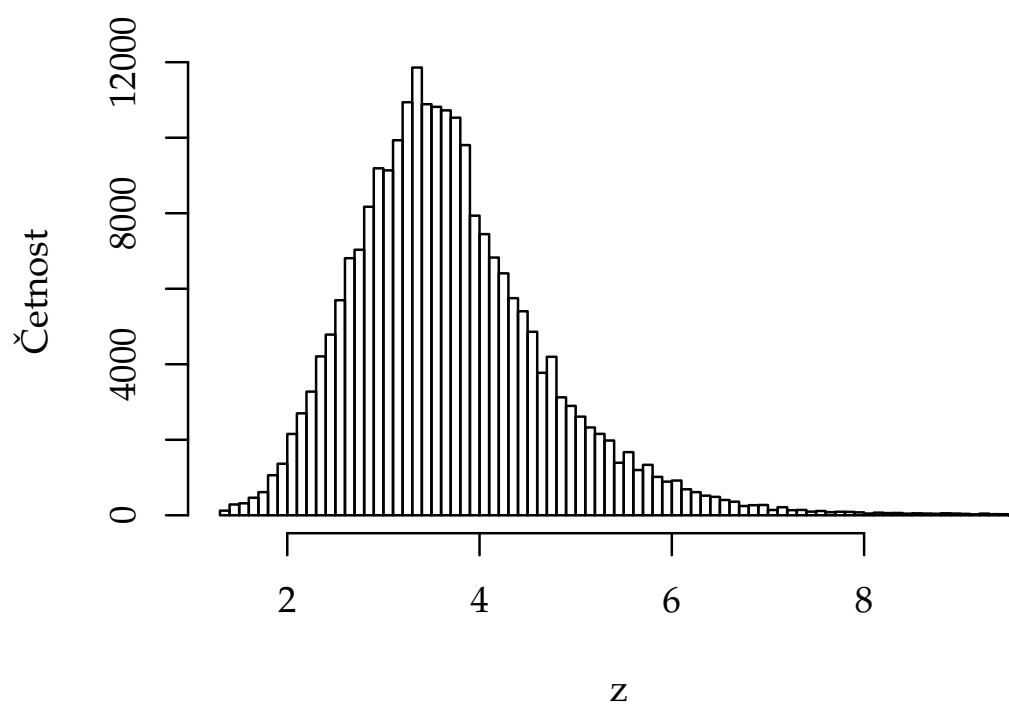
K této transformaci lze využít log-normálního rozdělení, jelikož střední hodnota log-normálního rozdělení závisí na velikosti rozptylu. Byla použita

¹<http://cran.r-project.org>



Obrázek 3.1: Histogram chyby du určení polohy ve směru u .

transformace $z = e^{a+b \cdot x}$, kde x je původní proměnná (tedy chyba určení polohy). Konstanty a , b byly určeny jako $a = 1$, čímž se maximum rozdělení posune dále od nuly a $b = \frac{0,25}{\text{mad}(du)}$, kde $\text{mad}(du)$ je mediánová absolutní odchylka chyby určení polohy a platí $\text{mad}(x) = \text{med}(|x_i - \text{med}(x)|)$, kde med je medián. Tato transformace je znázorněna v histogramech na obrázcích 3.1 a 3.2.



Obrázek 3.2: Histogram pro transformovaná data, byla použita transformace

$$z = e^{1 + \frac{0,25}{\text{mad}(du)} \cdot du}.$$

4. Zpracování dat

Před samotnou tvorbou regresních stromů jsem si musela připravit data. Nejprve jsem vybrala parametry, které byly pro toto dělení relevantní: Size.u, Size.v, Size, ClusterShapeType, Charge.ADU a Seed.ADU. Pitch.v, tedy velikost pixelů ve směru v, jsem nakonec nepoužila jako parametr pro dělení, ale vyhodnocovala jsem data zvlášť pro každou hodnotu rozteče ve směru v. Hodnoty velikostí ve směru u a v a celkové velikosti clusteru jsem rozdělila do několika kategorií (sloučila jsem případy velkých velikostí, protože zde již bylo málo hodnot a také jsme chtěli chybu odhadnout pro větší velikosti clusterů jako pro jednu skupinu), použila jsem kategorie 1 až 7 jednotlivě, pak 8–9, 10–15 a 16 a více.

1% dat, která měla největší chybu v určení polohy, jsem nebrala v úvahu. Na toto ořezání jsem napsala vlastní funkci pro program R, funkce_orezani.R. Lze volit velikost ořezání, v našem případě tedy 0,5% z každé strany. (Všechny zde uvedené vlastní funkce jsou obsaženy v Příloze 2.)

Poté jsem provedla transformaci dat, tedy převedla jsem hodnoty chyb určení polohy. Na toto jsem napsala vlastní funkci funkce_transformace.R, která provede transformaci popsanou výše.

Takto připravená data jsem dále zpracovávala pomocí regresních stromů funkcí rpart. Výsledky jsou reprezentovány pomocí binárních stromů. Zde jsem ještě mohla nastavovat jednotlivé parametry funkce rpart pro jemnost dělení (použila jsem hodnoty cp=0,0001, tento parametr je míra, s kterou se přidává dělení podle dalšího parametru dat, hodnota 0 značí, že se přidávají všechny, a minbucket=1000, tj. nejmenší počet událostí ve vytvořeném uzlu nebo listu je 1000).

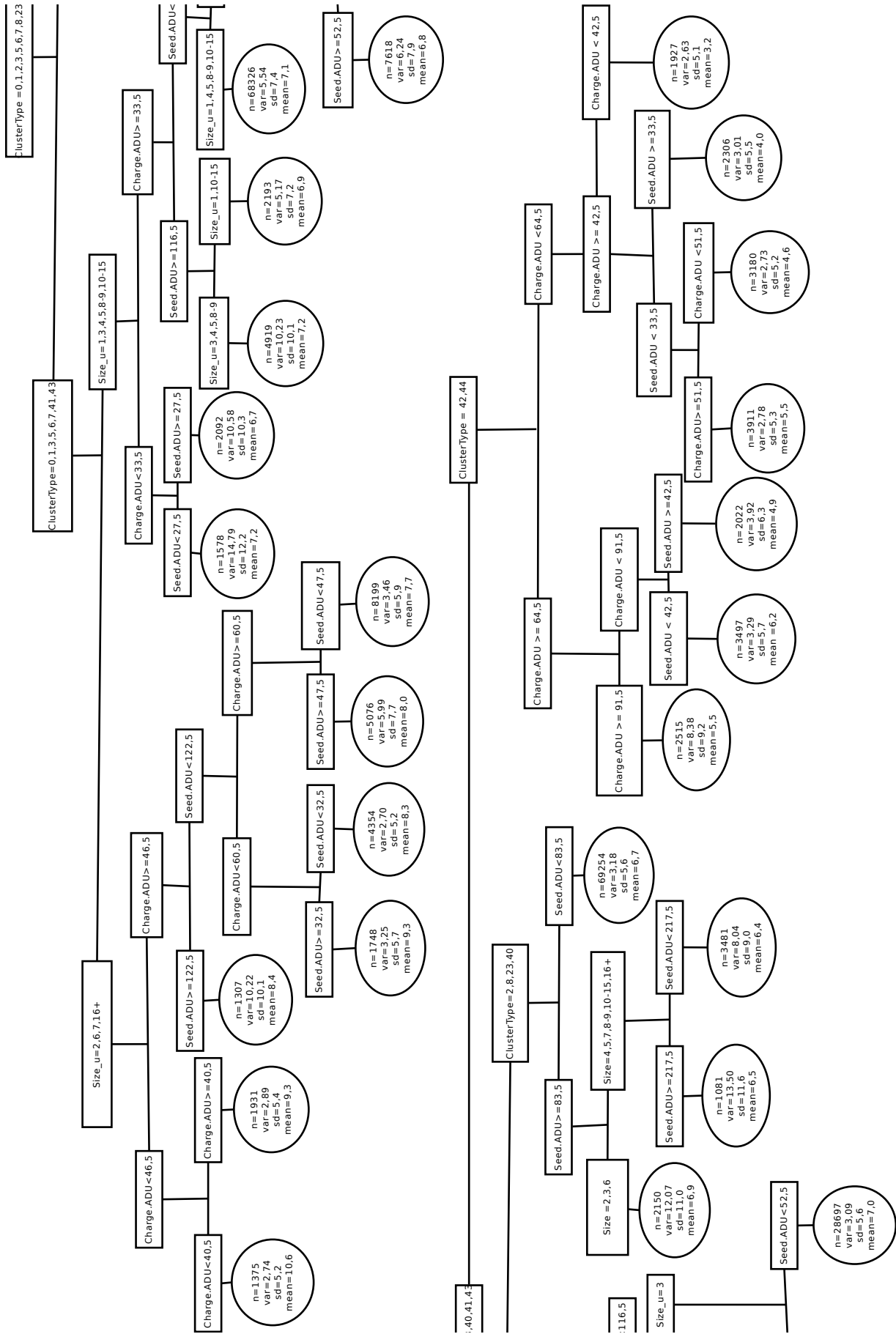
Pro výpočet rozptylu a biasu z dat v jednotlivých listech již hotového stromu jsem napsala vlastní funkci funkce_leaf.R, která spočítá tyto hodnoty (je potřeba je počítat z původních netransformovaných dat) a vypíše je v tabulce společně s hodnotami jednotlivých parametrů příslušejících danému listu.

Pracovala jsem se soubory dat obsahujícími alespoň 200 000 událostí. Nejprve jsem pro parametr Pitch.v = 55 μm vytvořila stromy pro chyby ve směru u opakovaně pro pět různých vzorků dat pro porovnání. Výsledek jednoho z těchto stromů uvádím v tabulce 4.1 (nejsou-li u některého parametru napsány hodnoty, znamená to, že v tomto listu není provedeno dělení podle tohoto parametru). Střední hodnota je vážený průměr daného rozdělení. Má-li náhodná veličina X spojité rozdělení s hustotou pravděpodobnosti $f(x)$, je střední hodnota $EX = \int f(x)xdx$, má-li diskrétní rozdělení pak je střední hodnota $EX = \sum s_i p_i$, kde p_i jsou pravděpodobnosti nabývání hodnot s_i . Rozptyl $var(X) = \sigma^2$ je druhý centrální moment $\sigma^2 = EX^2 - (EX)^2$. Směrodatná odchylka σ je odmocnina z rozptylu $\sigma = \sqrt{var(X)}$. Střední hodnota, směrodatná odchylka a rozptyl jsou v tabulce uvedeny v mm. Graficky jsem tento strom znázornila na obrázku 4.1 pomocí programu dia-gnome. V listech jsem zapsala počet událostí n připadajících na daný list, rozptyl var v jednotkách 10^{-5}mm , směrodatnou odchylku sd v jednotkách μm a střední hodnotu $mean$ taktéž v μm .

Výstupní tabulky ze zbylých čtyř stromů pro další vzorky dat pro parametr Pitch.v = 55 μm pro chyby ve směru u uvádím v příloze 3.

| číslo listu | číslo uzlu | rozpítl | směrodatná odchylka | střední hodnota | ClusterShapeType | Charge.ADU | Seed.ADU | Size:u | Size |
|-------------|------------|-----------|---------------------|-----------------|-------------------|--------------|---------------|-------------------|---------------------|
| 4 | 8 | 2,634E-05 | 0,0051 | 0,0032 | 42,44 | (-∞, 42,5) | (33,5, ∞) | | |
| 6 | 18 | 3,008E-05 | 0,0055 | 0,004 | 42,44 | (42,5, 64,5) | (33,5, ∞) | | |
| 8 | 38 | 2,731E-05 | 0,0052 | 0,0046 | 42,44 | (42,5, 51,5) | (-∞, 33,5) | | |
| 9 | 39 | 2,781E-05 | 0,0053 | 0,0055 | 42,44 | (51,5, 64,5) | (-∞, 33,5) | | |
| 12 | 20 | 3,922E-05 | 0,0063 | 0,0049 | 42,44 | (64,5, 91,5) | (42,5, ∞) | | |
| 13 | 21 | 3,289E-05 | 0,0057 | 0,0062 | 42,44 | (64,5, 91,5) | (-∞, 42,5) | | |
| 14 | 11 | 8,379E-05 | 0,0092 | 0,0055 | 42,44 | (91,5, ∞) | | | |
| 17 | 12 | 3,177E-05 | 0,0056 | 0,0067 | 2,8,23,40 | | (-∞, 83,5) | | |
| 20 | 52 | 8,036E-05 | 0,009 | 0,0064 | 2,8,23,40 | | (83,5, 217,5) | | 4,5,7,8-9,10-15,16+ |
| 21 | 53 | 1,350E-04 | 0,0116 | 0,0065 | 2,8,23,40 | | (217,5, ∞) | | 4,5,7,8-9,10-15,16+ |
| 22 | 27 | 1,207E-04 | 0,011 | 0,0069 | 2,8,23,40 | | (83,5, ∞) | | 2,3,6 |
| 28 | 224 | 3,094E-05 | 0,0056 | 0,007 | 0,1,3,5,6,7,41,43 | (33,5, ∞) | (-∞, 52,5) | 3 | |
| 29 | 225 | 6,244E-05 | 0,0079 | 0,0068 | 0,1,3,5,6,7,41,43 | (33,5, ∞) | (52,5, 116,5) | 3 | |
| 30 | 113 | 5,536E-05 | 0,0074 | 0,0071 | 0,1,3,5,6,7,41,43 | (33,5, ∞) | (-∞, 116,5) | 1,4,5,8-9, 10-15 | |
| 32 | 114 | 5,170E-05 | 0,0072 | 0,0069 | 0,1,3,5,6,7,41,43 | (33,5, ∞) | (116,5, ∞) | 1,10-15 | |
| 33 | 115 | 1,023E-04 | 0,0101 | 0,0072 | 0,1,3,5,6,7,41,43 | (33,5, ∞) | (116,5, ∞) | 3,4,5,8-9 | |
| 35 | 58 | 1,058E-04 | 0,0103 | 0,0067 | 0,1,3,5,6,7,41,43 | (-∞, 33,5) | (27,5, ∞) | 1,3,4,5,8-9,10-15 | |
| 36 | 59 | 1,479E-04 | 0,0122 | 0,0072 | 0,1,3,5,6,7,41,43 | (-∞, 33,5) | (-∞, 27,5) | 1,3,4,5,8-9,10-15 | |
| 41 | 240 | 3,456E-05 | 0,0059 | 0,0077 | 0,1,3,5,6,7,41,43 | (60,5, ∞) | (-∞, 47,5) | 2,6,7,16+ | |
| 42 | 241 | 5,993E-05 | 0,0077 | 0,008 | 0,1,3,5,6,7,41,43 | (60,5, ∞) | (47,5, 122,5) | 2,6,7,16+ | |
| 44 | 242 | 2,697E-05 | 0,0052 | 0,0083 | 0,1,3,5,6,7,41,43 | (46,5, 60,5) | (-∞, 32,5) | 2,6,7,16+ | |
| 45 | 243 | 3,246E-05 | 0,0057 | 0,0093 | 0,1,3,5,6,7,41,43 | (46,5, 60,5) | (32,5, 122,5) | 2,6,7,16+ | |
| 46 | 61 | 1,022E-04 | 0,0101 | 0,0084 | 0,1,3,5,6,7,41,43 | (46,5, ∞) | (122,5, ∞) | 2,6,7,16+ | |
| 48 | 62 | 2,893E-05 | 0,0054 | 0,0093 | 0,1,3,5,6,7,41,43 | (40,5, 46,5) | | 2,6,7,16+ | |
| 49 | 63 | 2,741E-05 | 0,0052 | 0,0106 | 0,1,3,5,6,7,41,43 | (-∞, 40,5) | | 2,6,7,16+ | |

Tabulka 4.1: Výstupní hodnoty pro strom vytvořený pro vzorek dat pro Pitch.v = 55 μ m, směr u.



Obrázek 4.1: Regresní strom vytvořený pro vzorek dat pro Pitch.v = 55µm, směr u.

| Pitch.v | směr | vzorek | ClusterSType | Charge.ADU | Seed.ADU | Size.u(v) | Size |
|---------|------|--------|--------------|------------|----------|-----------|------|
| 55 | u | 1 | 2 | 8 | 10 | 3 | 1 |
| 55 | u | 2 | 5 | 7 | 7 | 2 | 0 |
| 55 | u | 3 | 3 | 9 | 6 | 2 | 3 |
| 55 | u | 4 | 4 | 9 | 8 | 1 | 1 |
| 55 | u | 5 | 3 | 6 | 7 | 1 | 1 |
| 60 | u | 1 | 3 | 6 | 9 | 2 | 3 |
| 70 | u | 1 | 4 | 7 | 6 | 0 | 0 |
| 85 | u | 1 | 2 | 6 | 8 | 2 | 1 |
| 55 | v | 1 | 2 | 7 | 5 | 1 | 1 |
| 60 | v | 1 | 4 | 7 | 5 | 1 | 0 |
| 75 | v | 1 | 2 | 8 | 3 | 1 | 0 |
| 85 | v | 1 | 4 | 8 | 3 | 1 | 0 |

Tabulka 4.2: Počty dělení podle jednotlivých parametrů pro různé typy stromů (různé hodnoty Pitch.v, směru a vzorku.

Dále jsem stejným způsobem zpracovala vždy po jednom vzorku dat pro zbylé hodnoty parametru Pitch.v = 60, 70, 85 μm pro chyby ve směru u a po jednom vzorku pro všechny čtyři hodnoty parametru Pitch.v pro chyby ve směru v (výstupní tabulky taktéž uvedeny v příloze 3).

Jako kritéria pro zhodnocení vzniklých stromů jsem zvolila počet dělení podle daného parametru. Tyto počty, kolikrát se dělilo podle jednotlivých kritérií, jsou uvedeny v tabulce 4.2.

5. Diskuse

Pro tvorbu regresních stromů jsem nakonec zvolila metodu dělení pomocí analýzy rozptylu (ANOVA), protože jsem tím pádem mohla využít již hotové funkce (splitovací, inicializační a vyhodnocující) v programu R. Jelikož toto předpokládá odlišné střední hodnoty, provedla jsem transformaci dat. Zajímavé by bylo použít metodu dělení pomocí entropie, jelikož by se mohla používat přímo simulovaná data.

Při tvorbě stromů pomocí balíku RPART se dala nastavovat jemnost dělení. Otázkou je, jak podrobně jednotlivé parametry rozdělovat. Připadalo by mi dobré více rozdělovat podle parametru ClusterShapeType, ale pak jsme dostávali příliš jemné dělení v energiích (v Charge.ADU i v Seed.ADU) a obrovský, nepřehledný strom. Nakonec jsem parametry zvolila tak, aby výsledné stromy měly kolem 20 listů.

U vzniklých regresních stromů nás zajímala jejich stabilita, tj. jak moc se budou lišit dělicí parametry pro různé vzorky dat pro stejné vstupní parametry (Pitch.v, směr). Počty dělení podle jednotlivých parametrů jsem shrnula v tabulce 4.2. Vidíme, že pro několik vzorků dat pro vstupní parametry Pitch.v = 55 μm a směr u dochází vždy k největšímu počtu dělení v parametrech Charge.ADU a Seed.ADU. Počty dělení v parametrech ClusterShapeType a Size.u jsou menší a nejméně (s výjimkou jednoho případu) dělení se provádí v parametru Size. Pro lepší shodnocení stability stromů bychom potřebovali více vzorků dat a příp. toto provést i pro jiné vstupní parametry. Pro větší stabilitu stromů by bylo potřeba zpracovávat větší vzorky dat.

Znatelný rozdíl nastává pro stromy pro chyby ve směru u a v. Pro směr v dochází ještě méně k dělení podle velikosti clusterů, jak v Size.v, tak v Size. Také dochází méně k dělení pro parametr Seed.ADU. Hlavním dělicím parametrem je v obou případech energie částice Charge.ADU.

Závěr

V této práci byly popsány základní rysy experimentu Belle a Belle II v laboratoři KEK, byly zde sepsány základní informace ohledně křemíkových detektorů, jejich přesnosti, významu odhadu jejich přesnosti, simulace, podrobněji byly popsány stripové a pixelové detektory a detektor DEPFET. Dále zde byly shrnuty některé statistické metody a to konkrétně regresní stromy, kde lze využít dělení podle analýzy rozptylu nebo dělení podle entropie.

Poté bylo popsáno zpracování simulovaných dat pomocí regresních stromů v programu R s využitím funkce RPART. Byly zde představeny nové funkce pro program R, které byly využity při zpracování, dále byly popsány výsledky tohoto zpracování.

Pro jednotlivé hodnoty parametru $Pitch.v$ byly zpracovány vzorky dat o velikosti alespoň 200 tisíc událostí a nalezeny hodnoty rozptylu, standardní odchylky a systematického posunu (biasu) pro kategorie parametrů vytvořené pomocí regresních stromů. Zvlášť byly vyšetřovány chyby ve směru u a v .

Seznam použité literatury

- [1] Doležal, Z. Uno, S. et al. *Belle II Technical Design Report*. High Energy Accelerator Research Organization (KEK), 2010. KEK Report 2010-1.
- [2] Doležal, Z. *Polovodičové detektory v jaderné a subjaderné fyzice*. Text k přednášce JSF101p1a. 20. 12. 2007. Dostupné z: <http://www-ucjf.troja.mff.cuni.cz/~dolezal/teach/semicon>
- [3] Turchetta, R. Spatial resolution of silicon microstrip detectors. *Nuclear Instruments and Methods in Physics Research*. 1993, A 335, 44-58. doi:10.1016/0168-9002(93)90255-G.
- [4] Shalizi, C. *Lecture 10: Regression Trees*. Text k přednášce 36-350, Statistical Computing. 11. 10. 2006. Dostupné z: <http://www.stat.cmu.edu/~cshalizi/statcomp/lectures/10/lecture-10.pdf>
- [5] Analysis of variance. *Wikipedia, The Free Encyclopedia*. 7. 5. 2013, 7.32 UTC. Dostupné z: http://en.wikipedia.org/w/index.php?title=Analysis_of_variance&oldid=553914265
- [6] Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York: Springer Science+Business Media, LLC, 2006. ISBN 978-0387-31073-2.
- [7] Log-normal distribution. *Wikipedia, The Free Encyclopedia*. 17. 5. 2013, 8.02 UTC. Dostupné z: http://en.wikipedia.org/w/index.php?title=Log-normal_distribution&oldid=555480426
- [8] Therneau, Terry M. Atkinson, Eliabeth J. *An Introduction to Recursive Partitioning Using the RPART Routines*. Dokumentace k balíku RPART. 3. 9. 1997. Dostupné z: <http://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>
- [9] Therneau, Terry. *User written splitting functions for RPART*. Dokumentace z balíku RPART. 20. 3. 2013. Dostupné z: <http://cran.r-project.org/web/packages/rpart/vignettes/usercode.pdf>

Příloha 1 – Dodatky k teorii

Splitovací kritéria pro regresní stromy

Při stavbě regresního stromu mohou být pro dělení využita různá kritéria. Lze využít např. odchylku měření, jakožto kritérium pro vyhodnocování jednotlivých dělení. Pro naše účely by mohla být využita entropie, jakožto veličina, kterou bychom porovnávali.

Statistická entropie

Úvod do využití statistické entropie je uveden např. v [6]. Máme-li dva statisticky nezávislé jevy x a y , pak pro pravděpodobnost pozorování obou platí $p(x, y) = p(x)p(y)$. Míra informačního obsahu I závisí na pravděpodobnostním rozdělení p a její velikost je monotónní funkce pravděpodobnosti p . Informace pozorování dvou navzájem nezávislých jevů je stejná jako součet informací pozorování každého jevu zvlášť $I(x, y) = I(x) + I(y)$. Což ukazuje, že informační obsah I musí záviset na $\log p$. Záporné znaménko zajišťuje, že informační obsah I je kladný nebo nula. Nízká pravděpodobnost události x značí vysokou hodnotu informačního obsahu. Výběr základu logaritmu je libovolný, např. v informační teorii se nejčastěji používá základ 2. Pro naše účely bychom používali přirozený logaritmus. Zajímá-li nás celkové množství informačního obsahu nějakého procesu, musíme vycházet s předchozího vztahu a vzít v úvahu hustotu pravděpodobnosti p a tedy dostáváme $H(x) = -\sum p(x) \log p(x)$. Tato důležitá veličina se nazývá entropie náhodné proměnné x .

Odhad entropie a jeho vlastnosti

Statistická entropie je definována jako $H = -\sum p_i \log p_i$ z pravděpodobností pozorování p_i . Neznáme-li tyto pravděpodobnosti, můžeme pro ně použít odhad $f_i = \frac{n_i}{n}$, pak je odhad entropie roven $\hat{H} = -\sum f_i \log f_i$. Rozdíl těchto dvou veličin (zapsáno pomocí rozvoje) je

$$\hat{H} - H = \sum \frac{dH}{dp_i} \delta p_i + \frac{1}{2} \sum \frac{d^2H}{dp_i dp_j} \delta p_i \delta p_j = -\sum (\log p_i + 1) \delta p_i - \frac{1}{2} \sum \frac{1}{p_i} \delta_{ij} \delta p_i^2.$$

Pro odhad biasu vezmeme jen druhý člen a provedeme další úpravy $-\frac{1}{2} \sum \frac{1}{p_i} \delta_{ij} \delta p_i^2 = -\frac{1}{2} \sum \frac{\delta p_i^2}{p_i}$. Místo p_i použijeme odhad f_i a dostaneme (vyjádřeno pomocí n, n_i)

$$-\frac{1}{2} \sum \frac{\delta p_i^2}{p_i} = -\frac{1}{2n} \sum \frac{(n_i - \langle n_i \rangle)^2}{n_i} = -\frac{1}{2n} \langle \chi_{m-1}^2 \rangle = -\frac{m-1}{2n}.$$

Pro odhad rozptylu vezmeme $\text{var}(H) = \langle (\hat{H} - H - B(H))^2 \rangle$, kde použijeme rozvoj a vezmeme jen první dva členy. Pak

$$\text{var}(H) = \langle \sum (\log p_i + 1)(\log p_j + 1) \delta p_i \delta p_j \rangle,$$

což je zhruba rovno

$$\text{var}(H) \approx \langle \sum (\log p_i + 1)^2 \delta p_i^2 \rangle \approx \sum (\log p_i + 1)^2 n p_i (1 - p_i).$$

Po dosazení $\hat{p}_i \approx f_i = \frac{n_i}{n}$ dostáváme

$$\frac{1}{n} \sum (\log f_i + 1)^2 f_i (1 - f_i).$$

Pro odhad entropie tedy dostáváme

$$\langle \hat{H} - H \rangle = -\frac{m+1}{2n},$$

$$\text{var} \hat{H} = \frac{1}{n} \sum (1 + \log f_i) f_i (1 - f_i).$$

Log-normální rozdělení

Jedná se o spojité rozdělení pravděpodobnosti náhodné veličiny x takové, že pak veličina $\log(x)$ má normální rozdělení. Střední hodnota μ a směrodatná odchylka σ tohoto normálního rozdělení jsou pak parametry μ a σ log-normálního rozdělení veličiny x .

Střední hodnota log-normálního rozdělení je (podle [7]) $E(x) = e^{\mu + \frac{\sigma^2}{2}}$, tedy závisí jak na velikosti střední hodnoty, tak na velikosti směrodatné odchylky.

Medián log-normálního rozdělení závisí stále jen na velikosti střední hodnoty $x_{0,5} = e^{\mu}$.

Příloha 2 – Funkce pro program R

Funkce pro ořezávání dat

Před samotným zpracováním jsem vynechávala 1% dat s největší chybou, k čemuž jsem si napsala vlastní funkci.

```
orez <- function(x, y, p){  
#x je celý datový soubor  
#y je sloupec podle kterého se ořezává  
#p je násobek 1% kolik se má ořezat z každé strany  
  cuts <- quantile(y, c(0.01*p, 1-0.01*p))  
  x <- subset(x, y>cuts[1] & y<cuts[2])  
}
```

Funkce pro transformaci dat

Tato funkce transformuje data mnou zvolenou transformací.

```
transformuj <- function(x){  
  exp(1+0.25/(mad(x))*x)  
}
```

Funkce pro výstupní hodnoty pro strom vytvořený funkcí RPART

Pro reprezentaci výstupních hodnot stromu jsem zvolila tabulku, kde jsem vypsala číslo listu, jeho číslo v zápisu, který používá funkce RPART, rozptyl a střední hodnotu (ty jsem musela spočítat s původních, tedy netransformovaných dat) a hodnoty jednotlivých parametrů. Část kódu jsem převzala ze souboru labels.rpart.s z balíku RPART.

```
varmeans <- function(strom, data){  
#data jsou sloupek, ze kterého chceme počítat var a mean  
  ff <- strom$frame  
  x <- ff$var=="<leaf>"  
  nn <- length(x)  
  x <- (1:nn)[x]  
  n <- length(x)  
  node <- as.integer(row.names(ff)) #dvojková čísla všech uzlů  
#převzato z labels.rpart.s  
  if (nn==1) return("root") #special case of no splits  
  
  is.leaf <- (ff$var == "<leaf>")  
  whichrow <- !is.leaf  
  vnames <- ff$var[whichrow]  
  #the variable names for the primary splits
```

```

index <- cumsum(c(1, ff$ncompete + ff$nsurrogate + 1*(!is.leaf)))
irow <- index[c(whichrow, FALSE)]
#we only care about the primary splits
#konec převzaté části
dlouhyirow <- irow[cumsum(whichrow)]*whichrow
#přiřazuje správný index ze splits k číslům řádku (0 je list)
tabulka=data.frame(cislo_listu=vector(mode="integer", length=n),
node=node[x], var=vector(mode="numeric", length=n),
mean=vector(mode="numeric", length=n))
tmnoziny = vector(mode="character", length=n)
csplitncol <- ncol(strom$csplit)
#maximální počet hodnot kategoriální proměnné
for(i in 1:n){
  uzel <- node[x[i]]
  promenne=list() #seznam dělicích proměnných
  typ=list() #seznam, zda jde o kategoriální proměnnou
  while(uzel != 1){
    modulo <- uzel %% 2
    uzel <- uzel %/% 2
    a <- (1:n)[uzel == node] #a je číslo řádku v frame
    split <- strom$splits[dlouhyirow[a],]
    pr <- as.character(ff$var[a])
    if (split[2] < 2){ #případ spojitě proměnné
if (is.null(promenne[[pr]])){
  promenne[[pr]] <- c(-Inf, Inf)
  typ[[pr]] <- FALSE
}
if (xor(split[2] == -1, modulo == 0)){ #jsme nalevo od dělicí hranice
  if (promenne[[pr]][1] < split[4]){
    promenne[[pr]][1] <- split[4]
  }
} else { #případ vpravo od dělicí hranice
  if (promenne[[pr]][2] > split[4]){
    promenne[[pr]][2] <- split[4]
  }
}
} else { #případ kategoriální proměnné
if (is.null(promenne[[pr]])){
  typ[[pr]] <- TRUE
  promenne[[pr]] <- if (modulo == 0) {
    (1:csplitncol)[strom$csplit[split[4],] == 1]
  } else {
    (1:csplitncol)[strom$csplit[split[4],] == 3]
  }
}
}
}
}

```

```

    }
  }
  xlevels <- attr(strom, "xlevels")
  mnoziny=character()
  for(pr in labels(promenne)){
    mnozina <- paste(if (typ[[pr]]) xlevels[[pr]][promenne[[pr]]
      else promenne[[pr]], collapse=",")
    mnoziny <- paste(c(mnoziny, paste(c(pr, " ", if(typ[[pr]]) '{'
      else '(', mnozina, if(typ[[pr]]) '}' else ')'),collapse='')), collapse="; ")
  }
  tabulka$cislo_listu[i] = x[i]
  y = data[x[i]==strom$where]
  tabulka$var[i] = var(y)
  tabulka$mean[i] = mean(y)
  tmnoziny[i] = mnoziny
}
tabulka$mnoziny=tmnoziny
tabulka
}

```


Příloha 3 – Výstupní tabulky pro vytvořené regresní stromy

| # | node | var | mean | mnoziny |
|----|------|-----------|-----------|--|
| 1 | 4 | 2,587E-05 | 3,277E-03 | charge.ADU \in ($-\infty$,44.5); ClusterType \in 42,44 |
| 2 | 5 | 2,660E-05 | 4,523E-03 | charge.ADU \in (44,5,49.5); ClusterType \in 42,44 |
| 3 | 8 | 3,132E-05 | 4,561E-03 | seed.ADU \in (35,5, ∞); charge.ADU \in (49,5,71.5); ClusterType \in 42,44 |
| 4 | 9 | 2,789E-05 | 5,477E-03 | seed.ADU \in ($-\infty$,35.5); charge.ADU \in (49,5,71.5); ClusterType \in 42,44 |
| 5 | 10 | 5,793E-05 | 5,560E-03 | charge.ADU \in (71,5, ∞); ClusterType \in 42,44 |
| 6 | 16 | 2,888E-05 | 6,703E-03 | ClusterType \in 0,2,8,23,40; seed.ADU \in ($-\infty$,55.5); f_size_u \in 2,3 |
| 7 | 17 | 3,595E-05 | 6,923E-03 | ClusterType \in 5,7; seed.ADU \in ($-\infty$,55.5); f_size_u \in 2,3 |
| 8 | 19 | 4,987E-05 | 6,479E-03 | ClusterType \in 8,40; seed.ADU \in (55,5,105.5); f_size_u \in 2,3 |
| 9 | 20 | 6,351E-05 | 6,783E-03 | ClusterType \in 0,2,5,7,23; seed.ADU \in (55,5,105.5); f_size_u \in 2,3 |
| 10 | 21 | 4,655E-05 | 7,204E-03 | f_size_u \in 1,10-15,16+4,5,6,7,8-9; seed.ADU \in ($-\infty$,105.5); ClusterType \in 0,2,5,6,7,8,23,40 |
| 11 | 24 | 8,289E-05 | 6,452E-03 | seed.ADU \in (105,5,209.5); ClusterType \in 5,6,8,40 |
| 12 | 25 | 1,248E-04 | 6,853E-03 | seed.ADU \in (209,5, ∞); ClusterType \in 5,6,8,40 |
| 13 | 26 | 1,013E-04 | 7,310E-03 | ClusterType \in 0,2,7,23; seed.ADU \in (105,5, ∞) |
| 14 | 30 | 5,896E-05 | 6,995E-03 | charge.ADU \in (39,5, ∞); f_size_u \in 1; ClusterType \in 1,3,41,43 |
| 15 | 31 | 8,840E-05 | 6,951E-03 | charge.ADU \in (29,5,39.5); f_size_u \in 1; ClusterType \in 1,3,41,43 |
| 16 | 32 | 1,497E-04 | 7,515E-03 | charge.ADU \in ($-\infty$,29.5); f_size_u \in 1; ClusterType \in 1,3,41,43 |
| 17 | 37 | 3,017E-05 | 7,571E-03 | seed.ADU \in ($-\infty$,40.5); charge.ADU \in (56,5, ∞); f_size_u \in 2; ClusterType \in 1,3,41,43 |
| 18 | 38 | 4,340E-05 | 8,210E-03 | seed.ADU \in (40,5,62.5); charge.ADU \in (56,5, ∞); f_size_u \in 2; ClusterType \in 1,3,41,43 |
| 19 | 40 | 2,558E-05 | 8,291E-03 | seed.ADU \in ($-\infty$,30.5); charge.ADU \in (45,5,56.5); f_size_u \in 2; ClusterType \in 1,3,41,43 |
| 20 | 41 | 3,149E-05 | 9,664E-03 | seed.ADU \in (30,5,62.5); charge.ADU \in (45,5,56.5); f_size_u \in 2; ClusterType \in 1,3,41,43 |
| 21 | 42 | 9,677E-05 | 8,682E-03 | seed.ADU \in (62,5, ∞); charge.ADU \in (45,5, ∞); f_size_u \in 2; ClusterType \in 1,3,41,43 |
| 22 | 43 | 2,648E-05 | 1,008E-02 | charge.ADU \in ($-\infty$,45.5); f_size_u \in 2; ClusterType \in 1,3,41,43 |

Tabulka 5.1: Pitch.v = 55 μm , chyby ve směru u, 2. vzorek dat.

| # | node | var | mean | mnoziny |
|----|------|-----|-----------|---|
| 1 | 4 | 8 | 2,729E-05 | 3,229E-03 |
| 2 | 5 | 9 | 2,768E-05 | 4,549E-03 |
| 3 | 8 | 20 | 3,142E-05 | 4,584E-03 |
| 4 | 9 | 21 | 2,899E-05 | 5,729E-03 |
| 5 | 10 | 11 | 6,317E-05 | 5,542E-03 |
| 6 | 14 | 24 | 2,876E-05 | 6,669E-03 |
| 7 | 15 | 25 | 5,214E-05 | 6,664E-03 |
| 8 | 17 | 26 | 9,778E-05 | 6,349E-03 |
| 9 | 18 | 27 | 1,371E-04 | 6,927E-03 |
| 10 | 23 | 112 | 2,748E-05 | 6,858E-03 |
| 11 | 25 | 226 | 3,119E-05 | 6,868E-03 |
| 12 | 26 | 227 | 6,482E-05 | 6,915E-03 |
| 13 | 28 | 114 | 6,847E-05 | 6,529E-03 |
| 14 | 30 | 230 | 6,485E-05 | 6,991E-03 |
| 15 | 31 | 231 | 1,093E-04 | 7,359E-03 |
| 16 | 34 | 116 | 5,578E-05 | 7,107E-03 |
| 17 | 36 | 234 | 5,231E-05 | 7,335E-03 |
| 18 | 37 | 235 | 9,266E-05 | 7,843E-03 |
| 19 | 38 | 59 | 1,129E-04 | 7,061E-03 |
| 20 | 43 | 240 | 2,969E-05 | 7,514E-03 |
| 21 | 44 | 241 | 4,759E-05 | 8,003E-03 |
| 22 | 45 | 121 | 2,782E-05 | 8,566E-03 |
| 23 | 46 | 61 | 1,033E-04 | 8,497E-03 |
| 24 | 47 | 31 | 2,790E-05 | 9,777E-03 |
| | | | | charge.ADU $\in(-\infty, 40.5)$; ClusterType $\in 9, 42, 44$ |
| | | | | charge.ADU $\in(40.5, 56.5)$; ClusterType $\in 9, 42, 44$ |
| | | | | seed.ADU $\in(39.5, \infty)$; charge.ADU $\in(56.5, 75.5)$; ClusterType $\in 9, 42, 44$ |
| | | | | seed.ADU $\in(-\infty, 39.5)$; charge.ADU $\in(56.5, 75.5)$; ClusterType $\in 9, 42, 44$ |
| | | | | charge.ADU $\in(75.5, \infty)$; ClusterType $\in 9, 42, 44$ |
| | | | | seed.ADU $\in(-\infty, 50.5)$; ClusterType $\in 2, 8, 23, 40$ |
| | | | | seed.ADU $\in(50.5, 118.5)$; ClusterType $\in 2, 8, 23, 40$ |
| | | | | f_size $\in 5, 6, 7, 8-9, 10-15$; seed.ADU $\in(118.5, \infty)$; ClusterType $\in 2, 8, 23, 40$ |
| | | | | f_size $\in 2, 3, 4, 16+$; seed.ADU $\in(118.5, \infty)$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | charge.ADU $\in(-\infty, 89.5)$; seed.ADU $\in(-\infty, 72.5)$; f_size_u $\in 3$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | ClusterType $\in 0$; charge.ADU $\in(89.5, \infty)$; seed.ADU $\in(-\infty, 72.5)$; f_size_u $\in 3$ |
| | | | | ClusterType $\in 5, 7$; charge.ADU $\in(89.5, \infty)$; seed.ADU $\in(-\infty, 72.5)$; f_size_u $\in 3$ |
| | | | | f_size $\in 8-9, 10-15, 16+$; seed.ADU $\in(72.5, \infty)$; f_size_u $\in 3$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | charge.ADU $\in(-\infty, 182.5)$; f_size $\in 3, 4, 5, 6, 7$; seed.ADU $\in(72.5, \infty)$; f_size_u $\in 3$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | charge.ADU $\in(182.5, \infty)$; f_size $\in 3, 4, 5, 6, 7$; seed.ADU $\in(72.5, \infty)$; f_size_u $\in 3$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | charge.ADU $\in(34.5, 286.5)$; f_size_u $\in 1, 4, 5, 6, 8-9, 10-15$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | f_size $\in 1, 2, 3, 4, 5, 10-15, 16+$; charge.ADU $\in(286.5, \infty)$; f_size_u $\in 1, 4, 5, 6, 8-9, 10-15$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | f_size $\in 6, 7, 8-9$; charge.ADU $\in(286.5, \infty)$; f_size_u $\in 1, 4, 5, 6, 8-9, 10-15$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | charge.ADU $\in(-\infty, 34.5)$; f_size_u $\in 1, 4, 5, 6, 8-9, 10-15$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | seed.ADU $\in(-\infty, 39.5)$; charge.ADU $\in(56.5, \infty)$; f_size_u $\in 2, 7, 16+$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | seed.ADU $\in(39.5, 85.5)$; charge.ADU $\in(56.5, \infty)$; f_size_u $\in 2, 7, 16+$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | charge.ADU $\in(47.5, 56.5)$; seed.ADU $\in(-\infty, 85.5)$; f_size_u $\in 2, 7, 16+$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | seed.ADU $\in(85.5, \infty)$; charge.ADU $\in(47.5, \infty)$; f_size_u $\in 2, 7, 16+$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |
| | | | | charge.ADU $\in(-\infty, 47.5)$; f_size_u $\in 2, 7, 16+$; ClusterType $\in 0, 1, 3, 5, 6, 7, 41, 43$ |

Tabulka 5.2: Pitch.v = 55 μm , chyby ve směru u, 3. vzorek dat.

| # | node | var | mean | mnoziny |
|----|------|-----|-----------|-----------|
| 1 | 4 | 8 | 2,536E-05 | 2,734E-03 |
| 2 | 6 | 18 | 3,044E-05 | 3,962E-03 |
| 3 | 8 | 38 | 2,554E-05 | 4,417E-03 |
| 4 | 9 | 39 | 2,678E-05 | 5,624E-03 |
| 5 | 11 | 10 | 3,456E-05 | 5,558E-03 |
| 6 | 12 | 11 | 8,909E-05 | 5,564E-03 |
| 7 | 18 | 96 | 2,845E-05 | 6,719E-03 |
| 8 | 20 | 194 | 3,302E-05 | 6,916E-03 |
| 9 | 21 | 195 | 7,473E-05 | 7,210E-03 |
| 10 | 22 | 49 | 5,257E-05 | 6,813E-03 |
| 11 | 25 | 100 | 8,123E-05 | 6,622E-03 |
| 12 | 26 | 101 | 8,885E-05 | 7,342E-03 |
| 13 | 28 | 102 | 1,089E-04 | 6,642E-03 |
| 14 | 29 | 103 | 1,279E-04 | 7,429E-03 |
| 15 | 32 | 52 | 5,454E-05 | 7,097E-03 |
| 16 | 33 | 53 | 7,782E-05 | 7,927E-03 |
| 17 | 35 | 54 | 9,433E-05 | 6,597E-03 |
| 18 | 36 | 55 | 1,223E-04 | 7,002E-03 |
| 19 | 41 | 112 | 2,561E-05 | 7,179E-03 |
| 20 | 42 | 113 | 3,426E-05 | 8,110E-03 |
| 21 | 44 | 114 | 2,622E-05 | 8,479E-03 |
| 22 | 45 | 115 | 3,050E-05 | 9,557E-03 |
| 23 | 46 | 29 | 9,033E-05 | 8,795E-03 |
| 24 | 47 | 15 | 2,733E-05 | 1,012E-02 |

Tabulka 5.3: Pitch.v = 55 μm , chyby ve směru u, 4. vzorek dat.

| # | node | var | mean | mnoziny |
|----|------|-----------|-----------|---|
| 1 | 4 | 2,706E-05 | 3,252E-03 | charge.ADU $\in(-\infty,41.5)$; ClusterType $\in 9,42,44$ |
| 2 | 5 | 2,729E-05 | 4,531E-03 | charge.ADU $\in(41.5,56.5)$; ClusterType $\in 9,42,44$ |
| 3 | 8 | 2,925E-05 | 4,806E-03 | seed.ADU $\in(33.5,\infty)$; charge.ADU $\in(56.5,69.5)$; ClusterType $\in 9,42,44$ |
| 4 | 9 | 2,981E-05 | 5,902E-03 | seed.ADU $\in(-\infty,33.5)$; charge.ADU $\in(56.5,69.5)$; ClusterType $\in 9,42,44$ |
| 5 | 10 | 5,402E-05 | 5,597E-03 | charge.ADU $\in(69.5,\infty)$; ClusterType $\in 9,42,44$ |
| 6 | 14 | 2,725E-05 | 6,630E-03 | seed.ADU $\in(-\infty,36.5)$; ClusterType $\in 2,8,23,40$ |
| 7 | 15 | 4,157E-05 | 6,660E-03 | seed.ADU $\in(36.5,128.5)$; ClusterType $\in 2,8,23,40$ |
| 8 | 17 | 1,082E-04 | 6,078E-03 | f_size $\in 3,5,6,8-9,10-15,16+$; seed.ADU $\in(128.5,\infty)$; ClusterType $\in 2,8,23,40$ |
| 9 | 18 | 1,367E-04 | 7,054E-03 | f_size $\in 2,4,7$; seed.ADU $\in(128.5,\infty)$; ClusterType $\in 2,8,23,40$ |
| 10 | 23 | 2,995E-05 | 6,935E-03 | seed.ADU $\in(-\infty,48.5)$; f_size_u $\in 3$; ClusterType $\in 0,1,3,5,6,7$ |
| 11 | 24 | 5,790E-05 | 6,886E-03 | seed.ADU $\in(48.5,133.5)$; f_size_u $\in 3$; ClusterType $\in 0,1,3,5,6,7$ |
| 12 | 25 | 1,102E-04 | 7,119E-03 | seed.ADU $\in(133.5,\infty)$; f_size_u $\in 3$; ClusterType $\in 0,1,3,5,6,7$ |
| 13 | 27 | 5,653E-05 | 7,132E-03 | charge.ADU $\in(35.5,\infty)$; f_size_u $\in 1,4,5,6,7,8-9,10-15,16+$; ClusterType $\in 0,1,3,5,6,7$ |
| 14 | 28 | 1,079E-04 | 6,853E-03 | charge.ADU $\in(-\infty,35.5)$; f_size_u $\in 1,4,5,6,7,8-9,10-15,16+$; ClusterType $\in 0,1,3,5,6,7$ |
| 15 | 32 | 3,302E-05 | 7,458E-03 | charge.ADU $\in(62.5,\infty)$; seed.ADU $\in(-\infty,49.5)$; ClusterType $\in 41,43$ |
| 16 | 34 | 2,761E-05 | 8,006E-03 | seed.ADU $\in(-\infty,29.5)$; charge.ADU $\in(46.5,62.5)$; ClusterType $\in 41,43$ |
| 17 | 35 | 3,001E-05 | 9,046E-03 | seed.ADU $\in(29.5,49.5)$; charge.ADU $\in(46.5,62.5)$; ClusterType $\in 41,43$ |
| 18 | 36 | 7,690E-05 | 8,379E-03 | seed.ADU $\in(49.5,\infty)$; charge.ADU $\in(46.5,\infty)$; ClusterType $\in 41,43$ |
| 19 | 37 | 2,753E-05 | 9,844E-03 | charge.ADU $\in(-\infty,46.5)$; ClusterType $\in 41,43$ |

Tabulka 5.4: Pitch.v = 55 μm , chyby ve směru u, 5. vzorek dat.

| # | node | var | mean | mnoziny |
|----|------|-----------|-----------|---|
| 1 | 4 | 2,630E-05 | 3,782E-03 | |
| 2 | 5 | 2,793E-05 | 5,040E-03 | charge.ADU∈(-∞,50.5); ClusterType∈42,44 |
| 3 | 6 | 5,339E-05 | 5,581E-03 | charge.ADU∈(50.5,70.5); ClusterType∈42,44 charge.ADU∈(70.5,∞); ClusterType∈42,44 |
| 4 | 10 | 2,950E-05 | 6,645E-03 | seed.ADU∈(-∞,64.5); ClusterType∈2,8,23,40 |
| 5 | 11 | 6,326E-05 | 6,578E-03 | seed.ADU∈(64.5,149.5); ClusterType∈2,8,23,40 |
| 6 | 13 | 1,229E-04 | 6,411E-03 | f_size∈4,6,7,8-9,10-15; seed.ADU∈(149.5,∞); ClusterType∈2,8,23,40 |
| 7 | 14 | 1,446E-04 | 7,462E-03 | f_size∈2,3,5,16+; seed.ADU∈(149.5,∞); ClusterType∈2,8,23,40 |
| 8 | 20 | 2,474E-05 | 6,816E-03 | ClusterType∈0; seed.ADU∈(-∞,59.5); f_size_u∈3 |
| 9 | 21 | 3,868E-05 | 6,936E-03 | ClusterType∈5,7; seed.ADU∈(-∞,59.5); f_size_u∈3 |
| 10 | 22 | 6,211E-05 | 6,851E-03 | seed.ADU∈(59.5,127.5); f_size_u∈3; ClusterType∈0,1,3,5,6,7,41,43 |
| 11 | 23 | 1,082E-04 | 7,092E-03 | seed.ADU∈(127.5,∞); f_size_u∈3; ClusterType∈0,1,3,5,6,7,41,43 |
| 12 | 27 | 5,418E-05 | 7,032E-03 | seed.ADU∈(32.5,97.5); charge.ADU∈(29.5,∞); f_size_u∈1,4,5,6,7,8-9; ClusterType∈0,1,3,5,6,7,41,43 |
| 13 | 28 | 6,875E-05 | 7,129E-03 | seed.ADU∈(-∞,32.5); charge.ADU∈(29.5,∞); f_size_u∈1,4,5,6,7,8-9; ClusterType∈0,1,3,5,6,7,41,43 |
| 14 | 30 | 5,491E-05 | 6,928E-03 | f_size∈1,2,3,4,16+; seed.ADU∈(97.5,∞); charge.ADU∈(29.5,∞); f_size_u∈1,4,5,6,7,8-9; ClusterType∈0,1,3,5,6,7,41,43 |
| 15 | 31 | 8,774E-05 | 7,603E-03 | f_size∈5,6,7,8-9,10-15; seed.ADU∈(97.5,∞); charge.ADU∈(29.5,∞); f_size_u∈1,4,5,6,7,8-9; ClusterType∈0,1,3,5,6,7,41,43 |
| 16 | 32 | 1,518E-04 | 6,708E-03 | charge.ADU∈(-∞,29.5); f_size_u∈1,4,5,6,7,8-9; ClusterType∈0,1,3,5,6,7,41,43 |
| 17 | 36 | 3,076E-05 | 7,673E-03 | charge.ADU∈(57.5,∞); seed.ADU∈(-∞,41.5); f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |
| 18 | 38 | 2,579E-05 | 7,990E-03 | seed.ADU∈(-∞,26.5); charge.ADU∈(46.5,57.5); f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |
| 19 | 39 | 3,036E-05 | 9,209E-03 | seed.ADU∈(26.5,41.5); charge.ADU∈(46.5,57.5); f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |
| 20 | 41 | 6,606E-05 | 7,479E-03 | f_size∈16+; seed.ADU∈(41.5,∞); charge.ADU∈(46.5,∞); f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |
| 21 | 44 | 4,367E-05 | 8,165E-03 | charge.ADU∈(71.5,∞); seed.ADU∈(41.5,66.5); f_size∈3,10-15; f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |
| 22 | 45 | 3,509E-05 | 9,585E-03 | charge.ADU∈(46.5,71.5); seed.ADU∈(41.5,66.5); f_size∈3,10-15; f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |
| 23 | 46 | 1,009E-04 | 8,440E-03 | seed.ADU∈(66.5,∞); f_size∈3,10-15; charge.ADU∈(46.5,∞); f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |
| 24 | 47 | 2,590E-05 | 9,920E-03 | charge.ADU∈(-∞,46.5); f_size_u∈2,10-15,16+; ClusterType∈0,1,3,5,6,7,41,43 |

Tabuľka 5.5: Pitch.v = 60 μm , chyby ve směru u, 1. vzorek dat.

| # | node | var | mean | mnoziny | |
|----|------|-----|-----------|-----------|--|
| 1 | 3 | 4 | 2,292E-05 | 3,760E-03 | charge.ADU $\in(-\infty,47.5)$; ClusterType $\in 42,44$ |
| 2 | 5 | 10 | 2,497E-05 | 4,657E-03 | charge.ADU $\in(47.5,63.5)$; ClusterType $\in 42,44$ |
| 3 | 6 | 11 | 3,747E-05 | 5,021E-03 | charge.ADU $\in(63.5,\infty)$; ClusterType $\in 42,44$ |
| 4 | 11 | 48 | 2,223E-05 | 6,678E-03 | seed.ADU $\in(-\infty,34.5)$; ClusterType $\in 0,2,8,23,40$ |
| 5 | 13 | 98 | 2,810E-05 | 6,697E-03 | ClusterType $\in 8,23,40$; seed.ADU $\in(34.5,60.5)$ |
| 6 | 14 | 99 | 3,791E-05 | 6,838E-03 | ClusterType $\in 0,2$; seed.ADU $\in(34.5,60.5)$ |
| 7 | 16 | 50 | 4,807E-05 | 6,830E-03 | seed.ADU $\in(60.5,88.5)$; ClusterType $\in 0,2,8,23,40$ |
| 8 | 17 | 51 | 7,691E-05 | 6,844E-03 | seed.ADU $\in(88.5,\infty)$; ClusterType $\in 0,2,8,23,40$ |
| 9 | 20 | 52 | 4,851E-05 | 6,939E-03 | charge.ADU $\in(53.5,\infty)$; ClusterType $\in 1,3,5,6,7$ |
| 10 | 21 | 53 | 6,136E-05 | 6,974E-03 | charge.ADU $\in(38.5,53.5)$; ClusterType $\in 1,3,5,6,7$ |
| 11 | 22 | 27 | 8,557E-05 | 7,002E-03 | charge.ADU $\in(-\infty,38.5)$; ClusterType $\in 1,3,5,6,7$ |
| 12 | 27 | 112 | 2,204E-05 | 7,949E-03 | seed.ADU $\in(-\infty,31.5)$; charge.ADU $\in(53.5,\infty)$; ClusterType $\in 41,43$ |
| 13 | 29 | 226 | 3,661E-05 | 7,930E-03 | charge.ADU $\in(84.5,\infty)$; seed.ADU $\in(31.5,60.5)$; ClusterType $\in 41,43$ |
| 14 | 30 | 227 | 2,782E-05 | 8,948E-03 | charge.ADU $\in(53.5,84.5)$; seed.ADU $\in(31.5,60.5)$; ClusterType $\in 41,43$ |
| 15 | 32 | 114 | 2,066E-05 | 8,595E-03 | seed.ADU $\in(-\infty,25.5)$; charge.ADU $\in(42.5,53.5)$; ClusterType $\in 41,43$ |
| 16 | 33 | 115 | 2,803E-05 | 9,717E-03 | seed.ADU $\in(25.5,60.5)$; charge.ADU $\in(42.5,53.5)$; ClusterType $\in 41,43$ |
| 17 | 34 | 29 | 6,115E-05 | 9,308E-03 | seed.ADU $\in(60.5,\infty)$; charge.ADU $\in(42.5,\infty)$; ClusterType $\in 41,43$ |
| 18 | 35 | 15 | 2,454E-05 | 1,005E-02 | charge.ADU $\in(-\infty,42.5)$; ClusterType $\in 41,43$ |

Tabulka 5.6: Pitch.v = 70 μm , chyby ve směru u, 1. vzorek dat.

| # | node | var | mean | mnoziny |
|----|------|-----|-----------|-----------|
| 1 | 3 | 4 | 2,270E-05 | 3,505E-03 |
| 2 | 6 | 20 | 3,305E-05 | 3,774E-03 |
| 3 | 7 | 21 | 2,583E-05 | 4,764E-03 |
| 4 | 8 | 11 | 5,841E-05 | 4,713E-03 |
| 5 | 13 | 48 | 1,560E-05 | 6,639E-03 |
| 6 | 15 | 98 | 2,776E-05 | 6,652E-03 |
| 7 | 17 | 198 | 3,290E-05 | 6,740E-03 |
| 8 | 18 | 199 | 5,201E-05 | 6,872E-03 |
| 9 | 20 | 50 | 7,527E-05 | 6,719E-03 |
| 10 | 21 | 51 | 1,118E-04 | 7,314E-03 |
| 11 | 24 | 52 | 4,921E-05 | 6,999E-03 |
| 12 | 25 | 53 | 1,026E-04 | 7,372E-03 |
| 13 | 26 | 27 | 7,718E-05 | 6,964E-03 |
| 14 | 31 | 112 | 2,435E-05 | 8,324E-03 |
| 15 | 33 | 226 | 2,211E-05 | 8,543E-03 |
| 16 | 34 | 227 | 2,893E-05 | 9,624E-03 |
| 17 | 36 | 114 | 3,934E-05 | 8,765E-03 |
| 18 | 37 | 115 | 3,115E-05 | 9,826E-03 |
| 19 | 38 | 29 | 2,344E-05 | 1,025E-02 |
| 20 | 39 | 15 | 8,256E-05 | 1,034E-02 |

Tabuľka 5.7: Pitch.v = 85 μm , chybby ve směru u, 1. vzorek dat.

| # | node | var | mean | mnoziny |
|----|------|-----------|------------|---|
| 1 | 4 | 3,883E-05 | -3,935E-03 | charge.ADU $\in(-\infty, 43.5)$; ClusterType $\in 43, 44$ |
| 2 | 5 | 3,479E-05 | -2,503E-03 | charge.ADU $\in(43.5, 50.5)$; ClusterType $\in 43, 44$ |
| 3 | 7 | 4,026E-05 | -1,321E-03 | charge.ADU $\in(50.5, 79.5)$; ClusterType $\in 43, 44$ |
| 4 | 8 | 9,186E-05 | -1,051E-03 | charge.ADU $\in(79.5, \infty)$; ClusterType $\in 43, 44$ |
| 5 | 12 | 3,787E-05 | -7,256E-05 | seed.ADU $\in(-\infty, 35.5)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 6 | 13 | 5,331E-05 | -3,206E-05 | seed.ADU $\in(35.5, 63.5)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 7 | 15 | 9,366E-05 | 6,074E-05 | seed.ADU $\in(63.5, 160.5)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 8 | 17 | 1,218E-04 | 6,545E-05 | f_size $\in 3, 6, 7, 8-9, 10-15, 16+$; seed.ADU $\in(160.5, \infty)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 9 | 18 | 1,748E-04 | 6,041E-04 | f_size $\in 2, 4, 5$; seed.ADU $\in(160.5, \infty)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 10 | 22 | 8,511E-05 | 6,039E-05 | charge.ADU $\in(43.5, \infty)$; f_size_v $\in 1$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |
| 11 | 23 | 1,200E-04 | 4,761E-05 | charge.ADU $\in(28.5, 43.5)$; f_size_v $\in 1$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |
| 12 | 24 | 2,144E-04 | 6,024E-04 | charge.ADU $\in(-\infty, 28.5)$; f_size_v $\in 1$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |
| 13 | 29 | 3,769E-05 | 8,357E-04 | seed.ADU $\in(-\infty, 40.5)$; charge.ADU $\in(52.5, \infty)$; f_size_v $\in 2, 16+$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |
| 14 | 30 | 6,063E-05 | 1,429E-03 | seed.ADU $\in(40.5, 79.5)$; charge.ADU $\in(52.5, \infty)$; f_size_v $\in 2, 16+$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |
| 15 | 31 | 3,651E-05 | 2,296E-03 | charge.ADU $\in(44.5, 52.5)$; seed.ADU $\in(-\infty, 79.5)$; f_size_v $\in 2, 16+$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |
| 16 | 32 | 1,531E-04 | 2,439E-03 | seed.ADU $\in(79.5, \infty)$; charge.ADU $\in(44.5, \infty)$; f_size_v $\in 2, 16+$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |
| 17 | 33 | 3,637E-05 | 3,874E-03 | charge.ADU $\in(-\infty, 44.5)$; f_size_v $\in 2, 16+$; ClusterType $\in 1, 2, 5, 9, 41, 42$ |

Tabulka 5.8: Pitch.v = 55 μm , chyby ve směru v, 1. vzorek dat.

| # | node | var | mean | mmoziiny |
|----|------|-----------|------------|--|
| 1 | 5 | 4,455E-05 | -5,500E-03 | charge.ADU $\in(-\infty, 39.5)$; ClusterType $\in 43, 44$ |
| 2 | 6 | 4,329E-05 | -3,477E-03 | charge.ADU $\in(39.5, 49.5)$; ClusterType $\in 43, 44$ |
| 3 | 8 | 4,422E-05 | -1,851E-03 | charge.ADU $\in(49.5, 60.5)$; ClusterType $\in 43, 44$ |
| 4 | 9 | 7,288E-05 | -9,160E-04 | charge.ADU $\in(60.5, \infty)$; ClusterType $\in 43, 44$ |
| 5 | 12 | 4,434E-05 | -5,451E-05 | seed.ADU $\in(-\infty, 37.5)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 6 | 13 | 6,073E-05 | 2,096E-05 | seed.ADU $\in(37.5, 63.5)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 7 | 15 | 9,983E-05 | -1,282E-04 | f_size_v $\in 4, 5, 6, 7, 8-9, 10-15$; seed.ADU $\in(63.5, \infty)$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 8 | 17 | 1,117E-04 | 7,275E-05 | seed.ADU $\in(63.5, 132.5)$; f_size_v $\in 2, 3, 16+$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 9 | 18 | 1,877E-04 | 1,192E-04 | seed.ADU $\in(132.5, \infty)$; f_size_v $\in 2, 3, 16+$; ClusterType $\in 0, 3, 6, 7, 8, 23, 40$ |
| 10 | 21 | 1,086E-04 | -1,227E-05 | ClusterType $\in 1, 2, 5$; charge.ADU $\in(39.5, \infty)$ |
| 11 | 26 | 4,319E-05 | 6,287E-04 | seed.ADU $\in(-\infty, 40.5)$; charge.ADU $\in(56.5, \infty)$; ClusterType $\in 41, 42$ |
| 12 | 27 | 6,727E-05 | 1,307E-03 | seed.ADU $\in(40.5, 78.5)$; charge.ADU $\in(56.5, \infty)$; ClusterType $\in 41, 42$ |
| 13 | 28 | 4,371E-05 | 2,221E-03 | charge.ADU $\in(47.5, 56.5)$; seed.ADU $\in(-\infty, 78.5)$; ClusterType $\in 41, 42$ |
| 14 | 29 | 1,822E-04 | 2,147E-03 | seed.ADU $\in(78.5, \infty)$; charge.ADU $\in(47.5, \infty)$; ClusterType $\in 41, 42$ |
| 15 | 30 | 4,259E-05 | 3,880E-03 | charge.ADU $\in(39.5, 47.5)$; ClusterType $\in 1, 2$ |
| 16 | 33 | 1,551E-04 | -7,189E-05 | charge.ADU $\in(33.5, 39.5)$; ClusterType $\in 1, 2$ |
| 17 | 34 | 2,143E-04 | 3,828E-04 | charge.ADU $\in(-\infty, 33.5)$; ClusterType $\in 1, 2$ |
| 18 | 35 | 4,703E-05 | 5,710E-03 | ClusterType $\in 5, 41, 42$; charge.ADU $\in(-\infty, 39.5)$ |

Tabulka 5.9: Pitch.v = 60 μm , chyby ve směru v, 1. vzorek dat.

| | # | node | var | mean | mnoziny |
|----|----|------|-----------|------------|---|
| 1 | 4 | 8 | 5,369E-05 | -7,082E-03 | charge.ADU $\in(-\infty,39.5)$; ClusterType $\in 43,44$ |
| 2 | 5 | 9 | 5,798E-05 | -4,601E-03 | charge.ADU $\in(39.5,50.5)$; ClusterType $\in 43,44$ |
| 3 | 7 | 10 | 5,492E-05 | -2,179E-03 | charge.ADU $\in(50.5,76.5)$; ClusterType $\in 43,44$ |
| 4 | 8 | 11 | 1,087E-04 | -1,637E-03 | charge.ADU $\in(76.5,\infty)$; ClusterType $\in 43,44$ |
| 5 | 12 | 24 | 3,392E-05 | -1,480E-04 | seed.ADU $\in(-\infty,24.5)$; ClusterType $\in 0,3,5,6,7,8,23,40$ |
| 6 | 13 | 25 | 7,197E-05 | 8,286E-06 | seed.ADU $\in(24.5,89.5)$; ClusterType $\in 0,3,5,6,7,8,23,40$ |
| 7 | 14 | 13 | 1,729E-04 | -8,323E-05 | seed.ADU $\in(89.5,\infty)$; ClusterType $\in 0,3,5,6,7,8,23,40$ |
| 8 | 17 | 28 | 1,431E-04 | -1,825E-05 | charge.ADU $\in(38.5,\infty)$; f_size_v $\in 1,10-15$; ClusterType $\in 1,2,9,41,42$ |
| 9 | 19 | 58 | 2,187E-04 | 1,422E-04 | charge.ADU $\in(28.5,38.5)$; f_size_v $\in 1,10-15$; ClusterType $\in 1,2,9,41,42$ |
| 10 | 20 | 59 | 3,540E-04 | 4,242E-04 | charge.ADU $\in(-\infty,28.5)$; f_size_v $\in 1,10-15$; ClusterType $\in 1,2,9,41,42$ |
| 11 | 24 | 120 | 6,937E-05 | 1,208E-03 | charge.ADU $\in(62.5,\infty)$; seed.ADU $\in(-\infty,81.5)$; f_size_v $\in 2,6$; ClusterType $\in 1,2,9,41,42$ |
| 12 | 25 | 121 | 5,239E-05 | 2,798E-03 | charge.ADU $\in(47.5,62.5)$; seed.ADU $\in(-\infty,81.5)$; f_size_v $\in 2,6$; ClusterType $\in 1,2,9,41,42$ |
| 13 | 26 | 61 | 1,979E-04 | 3,257E-03 | seed.ADU $\in(81.5,\infty)$; charge.ADU $\in(47.5,\infty)$; f_size_v $\in 2,6$; ClusterType $\in 1,2,9,41,42$ |
| 14 | 28 | 62 | 5,765E-05 | 4,738E-03 | charge.ADU $\in(40.5,47.5)$; f_size_v $\in 2,6$; ClusterType $\in 1,2,9,41,42$ |
| 15 | 29 | 63 | 5,741E-05 | 6,639E-03 | charge.ADU $\in(-\infty,40.5)$; f_size_v $\in 2,6$; ClusterType $\in 1,2,9,41,42$ |

Tabulka 5.10: Pitch.v = 70 μm , chyby ve směru v, 1. vzorek dat.

| | # | node | var | mean | mnoziny |
|----|----|------|-----------|------------|--|
| 1 | 4 | 8 | 9,099E-05 | -8,855E-03 | charge.ADU \in ($-\infty$,47.5); ClusterType \in 43,44 |
| 2 | 5 | 9 | 8,518E-05 | -6,162E-03 | charge.ADU \in (47.5,55.5); ClusterType \in 43,44 |
| 3 | 7 | 10 | 7,886E-05 | -3,442E-03 | charge.ADU \in (55.5,70.5); ClusterType \in 43,44 |
| 4 | 8 | 11 | 1,326E-04 | -1,684E-03 | charge.ADU \in (70.5, ∞); ClusterType \in 43,44 |
| 5 | 12 | 24 | 8,718E-05 | -1,948E-05 | ClusterType \in 6,8,23,40; seed.ADU \in ($-\infty$,97.5) |
| 6 | 13 | 25 | 1,242E-04 | 3,971E-05 | ClusterType \in 0,3,5,7; seed.ADU \in ($-\infty$,97.5) |
| 7 | 15 | 26 | 2,247E-04 | 1,855E-04 | ClusterType \in 0,5,6,7,8,23,40; seed.ADU \in (97.5, ∞) |
| 8 | 16 | 27 | 3,515E-04 | 1,335E-03 | ClusterType \in 3; seed.ADU \in (97.5, ∞) |
| 9 | 19 | 28 | 2,163E-04 | -4,476E-05 | charge.ADU \in (46.5, ∞); f_size_v \in 1; ClusterType \in 1,2,9,41,42 |
| 10 | 21 | 58 | 3,076E-04 | 2,394E-04 | charge.ADU \in (33.5,46.5); f_size_v \in 1; ClusterType \in 1,2,9,41,42 |
| 11 | 22 | 59 | 4,652E-04 | 5,708E-04 | charge.ADU \in ($-\infty$,33.5); f_size_v \in 1; ClusterType \in 1,2,9,41,42 |
| 12 | 27 | 240 | 8,044E-05 | 1,207E-03 | seed.ADU \in ($-\infty$,50.5); charge.ADU \in (65.5, ∞); f_size_v \in 2,10-15; ClusterType \in 1,2,9,41,42 |
| 13 | 28 | 241 | 1,408E-04 | 2,165E-03 | seed.ADU \in (50.5,102.5); charge.ADU \in (65.5, ∞); f_size_v \in 2,10-15; ClusterType \in 1,2,9,41,42 |
| 14 | 29 | 121 | 7,689E-05 | 3,826E-03 | charge.ADU \in (55.5,65.5); seed.ADU \in ($-\infty$,102.5); f_size_v \in 2,10-15; ClusterType \in 1,2,9,41,42 |
| 15 | 30 | 61 | 3,378E-04 | 3,872E-03 | seed.ADU \in (102.5, ∞); charge.ADU \in (55.5, ∞); f_size_v \in 2,10-15; ClusterType \in 1,2,9,41,42 |
| 16 | 32 | 62 | 8,661E-05 | 6,542E-03 | charge.ADU \in (42.5,55.5); f_size_v \in 2,10-15; ClusterType \in 1,2,9,41,42 |
| 17 | 33 | 63 | 8,677E-05 | 9,619E-03 | charge.ADU \in ($-\infty$,42.5); f_size_v \in 2,10-15; ClusterType \in 1,2,9,41,42 |

Tabulka 5.11: Pitch. v = 85 μ m, chyby ve směru v, 1. vzorek dat.