Simultaneous Track and Error Propagation in Continuous Material

Esben Lund
Department of Physics
University of Oslo
Norway

Thesis submitted in partial fulfillment
of the requirements for the degree of
Philosophiae Doctor

January 2009

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
Nr. 843

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS, Oslo, 2009.

Produced in co-operation with Unipub AS.
The thesis is produced by Unipub AS merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Unipub AS is owned by
The University Foundation for Student Life (SiO)

Acknowledgements

First of all, I want to thank the Experimental Particle Physics Group for funding my studies over three years. Without their financial support I could not have completed this degree.

Moreover, I want to thank my supervisors, Professor Are Strandlie and Professor Lars Bugge, for their priceless support during my stay at the group. Without their expertize and help, this thesis would never have happened. Special thanks also go to the rest of the group, particularly to the former students Mustafa Hussain, Sigve Haug and Olav Mundal.

Furthermore, I want to mention some of the people I have had the pleasure of working with at CERN; Professor Steinar Stapnes, Andreas Salzburger, Markus Elsing, Sarka Todorova, Teddy Todorov and Igor Gavrilenko. Great and dedicated people who make CERN such an interesting place to visit.

Finally, my thoughts go to my parents Hennie and Jan Arne, my twin brother Vidar, sister Annette, brothers Jan Erik, Jørgen, Johan, Ole August, and my wife Chengetayi who came from afar to be with me.

Introduction

Experimental particle physics is on the verge of a new era, heralded by the Large Hadron Collider being commissioned at the European Center of Nuclear Research — CERN — located just outside of Geneva, Switzerland. The LHC accelerator will collide protons at a center of mass energy of 14 TeV, opening up a new window for particle discoveries and precision measurements of existing theories. Particle detectors are located at four beam crossings along the LHC, one of which is the ATLAS detector [1]. This is the largest of the LHC experiments, employing a great variety of detector and magnetic field technologies to identify a wide range of particles. The complex magnetic field and high collision rate, however, make the reconstruction of particle tracks very challenging. Things are complicated further by the relatively big amount of material within ATLAS, generating considerable disturbances to the particle tracks through material interactions such as energy loss and multiple scattering. In this thesis we present an algorithm — *Simultaneous Track and Error Propagation* (STEP) — which addresses these track reconstruction challenges.

We start out by presenting the theoretical status within the particle physics field and the experimental aspects in chapters 1 and 2, respectively. Moreover, we present the `NewTracking` framework [2] designed to streamline the ATLAS reconstruction software in Chapter 3, and the conclusion in Chapter 4. Furthermore, we present a robust method for inverting covariance matrices in Appendix A.

Following the presentation of the context of the STEP propagator, we describe the propagator in detail in four papers; starting out with the numerical method applied for transporting the track parameters and the selection process for choosing this method in the first paper [3]. Moreover, we present the semi-analytical method used for transporting the associated covariance matrix, and test the resulting transport Jacobian and covariance matrix in the second paper [4]. We then describe the energy loss and the multiple scattering, with an emphasis on muons, in the third paper [5]. Finally, we check the STEP algorithm in the realistic ATLAS reconstruction by refitting combined muon tracks through the Kalman filter formalism [6] in the last paper [7].

Contents

Chapter 1

Theory

For thousands of years people have been pondering the building blocks of matter. The Greek philosopher Democritus was the first to define a theory of elementary particles, which he called *atoma*, the Greek word for indivisible particles. According to Democritus, the universe was an empty "void", filled with a multitude of combinations of a few types of atoms. This was a remarkable achievement, and apart from the changing family of "elementary" particles — going from atoms to nucleons to quarks — the concept of the "void" filled with indivisible, point-like particles still stands.

1.1 The Standard Model

The Standard Model (SM) of particle physics is one of the most successful scientific theories of all time. In its lifetime of more than 30 years, no experimental inconsistencies have been seen. It is, however, only an effective theory; in the sense that many parameters — such as the particle masses — are put in by hand. From a theoretical viewpoint, this is not satisfying, inciting many theorists' hunt for deeper, more fundamental theories, such as supersymmetry and string theory. For now, such theories are pure speculation, any evidence being circumstantial at best.

The Standard Model divides particles into two main groups; matter particles called *fermions* which are spin-half particles, and force carriers called *bosons* which are integer-spin particles. The matter particles are divided further into *leptons* and *quarks*, having integer and fractional electric charges respectively. All the particles of the SM are listed in Table 1.1.

Ordinary matter consists of the lightest quarks and leptons; up, down and the electron. These particles — together with the electron-neutrino ν_e — are often referred to as the first *generation* of matter particles, while (c, s, μ^-, ν_μ) and (t, b, τ^-, ν_τ) are the second and third generations. All generations are similar, apart from the particle masses, and all matter particles come with an antiparticle of opposite quantum numbers and equal mass.

The SM is a quantum field theory which describes the weak, electromagnetic and strong interactions of spin-half point-like fermions. It is a gauge theory based on the $\mathrm{SU(3)}_C \times \mathrm{SU(2)}_L \times \mathrm{U(1)}_Y$ symmetry group. $\mathrm{SU(3)}_C$ being the symmetry group of strong

Table 1.1: Elementary particles of the Standard Model.

Type	Name	Spin	Electric charge
Quark	u, c, t	1/2	+2/3
	d, s, b	1/2	-1/3
Lepton	e^-, μ^-, τ^-	1/2	-1
	ν_e, ν_μ, ν_τ	1/2	0
Boson	γ	1	0
	W^+, W^-, Z^0	1	+1, -1, 0
	g_i, $i = 1, 8$	1	0
	H^0	0	0

interactions, C indicating colour, while $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ is the unified symmetry group of the electromagnetic and weak interactions — the so-called electroweak interactions — L and Y referring to the weak isospin and weak hypercharge respectively. The subscript L on the SU(2) group is to remind us that the weak isospin current only couples the left-handed fermions. These gauge symmetries give rise to several conservation laws, such as the conservation of charge and colour. Gauge symmetries are continuous dynamical symmetries.

The SM can be described by three types of fields; the matter field, which gives birth to leptons and quarks, the gauge fields, from which the gauge bosons appear, and the Higgs scalar fields, which describe the masses of the particles.

The particles

The spinor matter fields are spin-half fermions belonging to the fundamental representation of the gauge group consisting of leptons and quarks. Under $\mathrm{SU}(2)_L$ the left-handed fermion fields transform as weak isodoublets, whereas the right-handed fields are weak isosinglets. For the first generation these are

$$\begin{bmatrix} \nu_e \\ e^- \end{bmatrix}_L, \; e^-_R, \; \nu_{eR}, \; \begin{bmatrix} u \\ d \end{bmatrix}_L, \; u_R, \; d_R$$

The other generations follow the same template. The two doublets describe particles which are linked by the weak interactions. The weak transitions within the quark isospin doublet are only approximate due to the phenomenon of quark mixing, which reflects the fact that the quark mass eigenstates are not the same as their weak eigenstates. Convincing evidence of neutrino oscillations [8, 9], which are a consequence of non-zero neutrino masses and their mixing, imply the need of a right-handed neutrino, making the transitions within the lepton isospin doublet approximate as well. The quark mixing is often parametrized by a 3×3 unitary matrix $\mathbf{V}$ — the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix — operating on the charge $-1/3$ quark

mass eigenstates (d, s, b):

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

where (d', s', b') are the weak eigenstates, and the elements of the CKM matrix are related to the transition likelihood between the up-type and down-type quarks.

Particle interactions

The interactions between two particles are mediated by the exchange of bosons which couple to the conserved charges of the particles. The strength of the interaction is described by an effective constant called the *coupling constant*. Since this is a function of the energy transfer during the interaction, it is not really a constant, consequently it is often referred to as a "running" coupling constant. The SM defines three coupling constants, one for each interaction.

The $\mathrm{SU}(3)_C$ group gives rise to eight gauge fields corresponding to the eight gluons mediating the strong force, which is the source of the confinement of quarks within the strongly interacting particles, the so-called *hadrons*. The strong coupling constant is called g_s.

The electroweak interaction is carried by three $W^{\pm}, W^0$ bosons for the $\mathrm{SU}(2)_L$ sector, universally coupled with the strength g, and the neutral boson B^0 for the $\mathrm{U}(1)_Y$ sector, with coupling g'. The corresponding physical states are $W^{\pm}$, carrying the weak charged current, and γ, Z^0 carrying the electromagnetic and weak neutral currents. The latter are mixtures of the W^0 and B^0 states. The $W^{\pm}$ and Z^0 bosons are massive particles, while the photon appears to be massless. The masses of the gauge bosons are generated by spontaneous symmetry breaking of the unified $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry group through the *Higgs mechanism*. The relations between the generators of the electroweak symmetry group and the physical particles are

$$W^{\pm} = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$$

$$\begin{bmatrix} \gamma \\ Z^0 \end{bmatrix} = \begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix} \begin{bmatrix} B^0 \\ W^0 \end{bmatrix}$$

where θ_W is the weak mixing angle. Based on gauge invariance, the relation between g and g' can be expressed in terms of the unification condition;

$$e = g\sin\theta_W = g'\cos\theta_W, \quad \cos\theta_W = \frac{M_W}{M_Z}$$

The electromagnetic part of the electroweak sector describes most of the low-energy physics, while the weak part is responsible for much of the radioactive decay of unstable particles.

The Higgs mechanism

The third set of SM fields are the Higgs scalar fields introduced by the minimal Standard Model, proposed by Weinberg and Salam, and based on the Higgs mechanism developed by Higgs in the 1960s. The electroweak $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry group is spontaneously broken by the existence of a Higgs scalar field with a non-zero expectation value. By introducing this field, the masses of the quarks, leptons and intermediate weak bosons emerge. Both the $W^{\pm}$ and Z^0 become massive, while the photon remains massless. The minimal version of the SM consists of one SU(2) doublet of Higgs scalar fields with spin 0, giving one neutral scalar boson H^0 with a mass not predicted by the theory. The Higgs boson is yet to be experimentally observed, and has been excluded up to a mass of 114.4 GeV/c^2 at 95% confidence level by direct searches at the Large Electron Positron Collider (LEP) at CERN.

The Langrangian

The Lagrangian function describes the particle movement in the 4-dimensional spacetime, and determines the dynamics of the SM gauge theory. In the classical limit, it is expressed as the difference between the kinetic (T) and potential energy (U) of the system

$$L = T(\dot{x}) - U(x) = \frac{1}{2}m\dot{x}^2 - U(x)$$

where x is the position and $\dot{x} = dx/dt$ is the speed of the particle. The equation of motion can be derived from the Euler-Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0$$

$$F = \frac{dU}{dx} = ma, \quad a = \ddot{x}$$

In field theory, the definition of the Langrangian of a field, ψ, is given as a function of its density, $\mathcal{L}$,

$$L = \int \mathcal{L}(\psi, \partial_\mu \psi)$$

The Euler-Lagrange equation represents the corresponding Dirac equation solution of the wave function, and is given by

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)}\right) - \frac{\partial \mathcal{L}}{\partial \psi} = 0$$

The gauge invariant Lagrangian of the electroweak model (Glashow 1961, Weinberg 1967, Salam 1968 and 't Hooft 1971) consists of several parts and is defined by the following main components [10, 11]:

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{matter}} + \mathcal{L}_{\mathrm{gauge}} + \mathcal{L}_{\mathrm{Higgs}} + \mathcal{L}_{\mathrm{Yukawa}} + \cdots$$

The matter Lagrangian, $\mathcal{L}_{\text{matter}}$, is described by the representation of the interaction between the gauge bosons and fermions, as spinor fields ψ_j, where j represents the three generations,

$$\mathcal{L}_{\text{matter}} = i \sum_{j=1}^{3} \bar{\psi^j} \gamma^\mu D_\mu \psi_j$$

where the covariant derivative D_μ is given by the following relation

$$D_\mu = \partial_\mu - igW_\mu^A \frac{\tau^A}{2} - ig' B_\mu \frac{Y}{2}$$

with Y being the weak hypercharge, which relates the third component of the weak isospin and the electric charge as $Q = T_3 + Y/2$, and the couplings g and g' define the mixing angle $\tan\theta_W = g'/g$. The three Hermitian matrices τ_A, with $A = \{1, 2, 3\}$, are the isospin equivalents of the Pauli matrices which give the lowest representation of the rotation group.

The second part of the SM Lagrangian, $\mathcal{L}_{\text{gauge}}$, describes the gauge boson interactions as a function of the two gauge fields W_μ^a and B_μ with couplings g and g' respectively;

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}$$

The field tensors are expressed in terms of the elementary fields

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon^{abc} W_{\mu\nu}^b W_\nu^c$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

where ϵ^{abc} are the $\mathrm{SU}(2)_L$ group structure constants. This gauge Lagrangian describes both the kinetic and interaction part of the gauge bosons.

The Lagrangian contribution from the Higgs fields is given by

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi^\dagger \phi)$$

where V is the Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + |\lambda| (\phi^\dagger \phi)^2$$

and ϕ is the complex Higgs field, μ is an arbitrary value and λ is the Higgs self-coupling.

The interaction of the Higgs doublet with the matter fields, which gives masses to the fermions, is given by the Yukawa Lagrangian,

$$\mathcal{L}_{\text{Yukawa}} = \sum_{i=1}^{3} (f_i^l \bar{l}_{iL} \phi e_{iR} + f_i^u \bar{q}_{iL} \tilde{\phi} u_{iR} + f_i^d \bar{q}_{iL} \phi d_{iR})$$

where f_i^j are the Yukawa couplings, j indicating the leptons, up- or down-type quarks, summing over the three generations i.

1.2 Theoretical challenges

Even though the Standard Model has been highly successful, it has many shortcomings; the total lack of gravity being one of them. At the distances relevant to the electroweak and strong interactions, gravity becomes comparatively small. Only at larger distances, and by accumulating the gravitational forces of billions of elementary particles, does gravity become a force to be reckoned with. The main problem in combining gravity and the SM lies in the different nature of the theories; SM being a quantum field theory, whereas the general theory of relativity is a classical theory.

Furthermore, the SM needs extreme fine-tuning to avoid huge radiative corrections to the Higgs mass, which would push it many orders of magnitude above the experimentally favored value of 76^{+33}_{-24} GeV/c^2 [12], Fig. 1.1. This problem can be solved by introducing an extension to the SM called supersymmetry (SUSY) which establishes an additional symmetry between fermions and bosons, or a particle-force duality. Every fermion is given a bosonic superpartner, and every boson a fermionic superpartner at the SUSY scale of around 1 TeV. These superpartners enter the radiative corrections to the Higgs mass — and other masses — with opposite signs to their regular SM partners, thereby bringing the corrections down to a "reasonable" level. For this to work, the masses of the superpartners cannot be much higher than the masses of the SM particles, hence the 1 TeV SUSY scale.

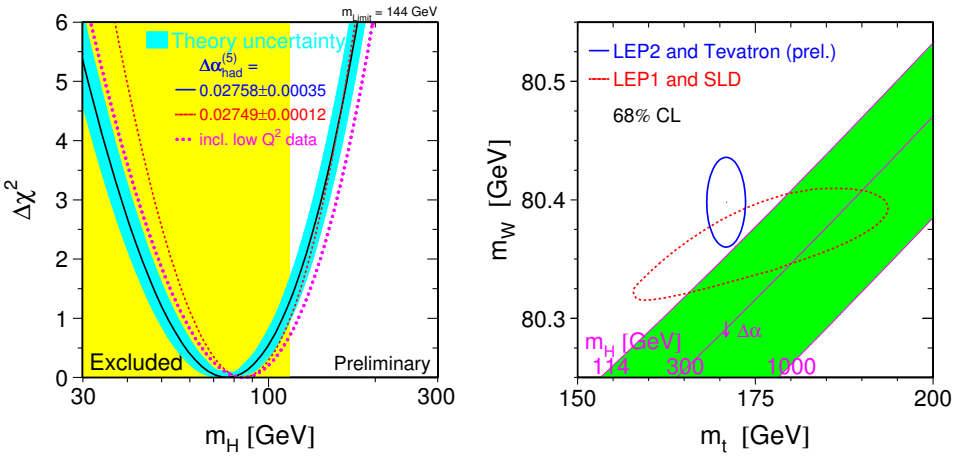


Figure 1.1: Higgs mass fits at 68% confidence level from precision electroweak measurements at LEP and the Tevatron, as of 2007 [12]. The minimum of the black curve in the left-hand figure and the center of the oval in the right-hand plot represent the preferred mass fit of 76^{+33}_{-24} GeV/c^2, considerably lower than the LEP exclusion limit of 114.4 GeV/c^2.

Another attractive feature of SUSY is the apparent unification of the coupling constants when going towards the grand unification scale ($\sim 10^{16}$ GeV), Fig. 1.2. At this energy the coupling constants of the Minimal Supersymmetric extension of the Standard Model (MSSM) seem to unite at the same strength, implying the existence of a unified force described by a *Grand Unified Theory* (GUT).

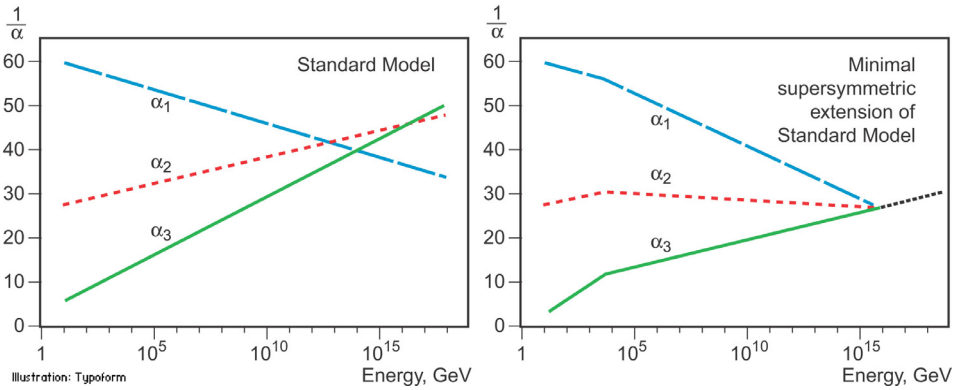


Figure 1.2: The inverse of the coupling constants of the electromagnetic (α_1), weak (α_2) and strong (α_3) interactions as a function of the energy transfer during the interaction. The left-hand figure shows the Standard Model couplings, while the right-hand plot shows the couplings of the Minimal Supersymmetric Standard Model, MSSM. (From Ref. [13], ©Nobelprize.org).

SUSY may also shed light on one of the greatest mysteries in Cosmology, the missing dark matter of the universe. This is estimated to be around five times the amount of the ordinary, visible matter. The dark matter is only observed through its gravitational effect, most easily seen in the speed of the stars at the edge of the galaxies. Including only the visible matter, these stars move too fast to stay within the galaxy. Adding a halo of dark matter is one way of solving this problem. SUSY provides a good candidate for this dark matter, the lightest supersymmetric particle (LSP). All supersymmetric particles have a special quantum number called R-parity which is conserved in all interactions. This implies that supersymmetric particles are always produced in pairs, and that they cannot decay into a final state without any supersymmetric particles. Because of this, the LSP becomes a stable, weakly interacting, massive particle (WIMP), which is one of the best dark matter candidates.

The hope of finding a host of superpartners around the 1 TeV scale, possibly solving the mystery of the dark matter, and the chance of finding the Higgs boson — or dealing a severe blow to the Higgs mechanism — are the main motivations for building new accelerators capable of delving deeper into the nature of matter, such as the LHC at CERN.

Chapter 2

Experiment

To test new physics theories, and improve existing ones, we need to perform experiments. Within particle physics, such experiments usually involve a "microscope" — or accelerator — and a "camera", the detector. The accelerator sends a beam of particles towards a fixed target or a meeting beam, while the detector takes a snapshot of the collision. This snapshot is then analyzed to look for new particles, such as the Higgs boson or supersymmetric particles, or to make precision measurements of parameters of an established theory, such as the Standard Model.

To probe further into the mysteries of matter, better accelerators and detectors are needed. This involves building longer accelerators, using stronger and often superconducting bending magnets and denser beams to increase the collision energy and frequency. In response, the detector must be quick enough to separate the collisions, and able to identify single particles within each collision.

The most powerful accelerator and detectors ever seen are being built at CERN, Geneva, with a planned start-up in 2009. This is the LHC and the four accompanying detectors; ATLAS, CMS, ALICE and LHCb. Two of these detectors are general purpose detectors; ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid), whereas the other two are a dedicated B-physics detector, LHCb, and A Large Ion Collider Experiment (ALICE), which is a heavy-ion detector designed to investigate Lead on Lead collisions.

2.1 The Large Hadron Collider

The LHC is a proton-proton collider, built in the existing 26.7 km Large Electron-Positron (LEP) collider tunnel at CERN. Protons are accelerated in opposite directions in two separate rings, located within the same superconducting magnet structure, to a center of mass energy of 14 TeV, and brought to collide within the four detectors located at beam crossing points along the ring. Alternatively, the beams can be Lead-ions colliding at an even higher center of mass energy. The superconducting magnets are cooled by liquid Helium to 2 K — producing a field of 8 T — and are mostly used for bending the protons, keeping them within the ring. Using one magnet structure for both rings is not optimal, causing interference between the two beams and complicating

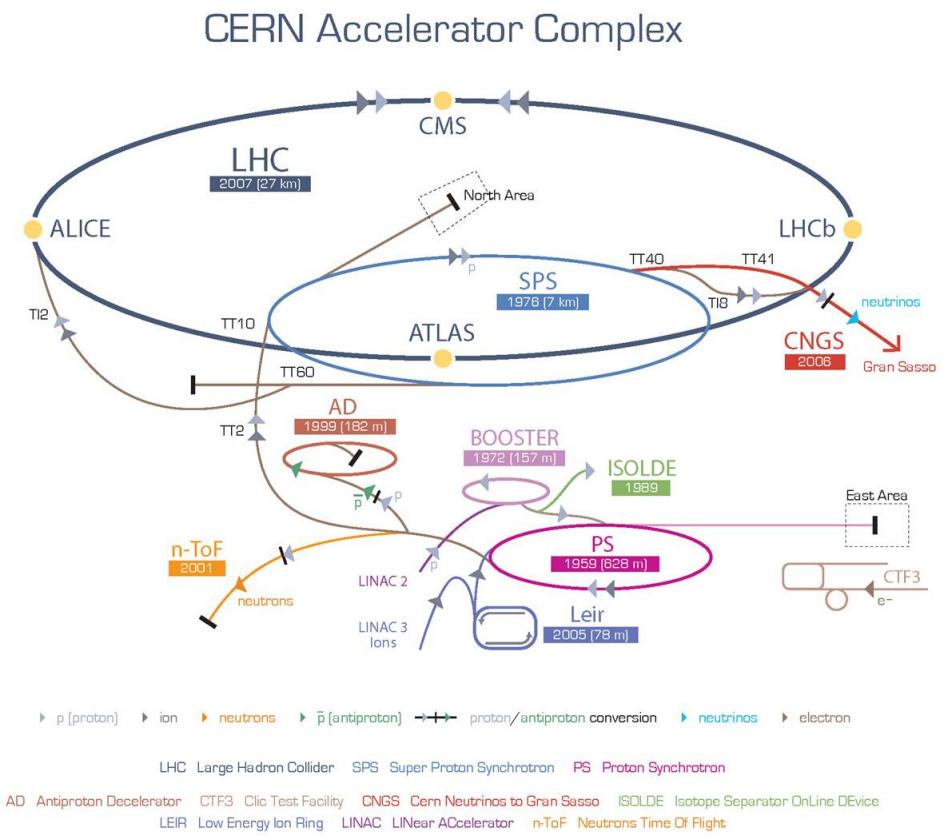


Figure 2.1: The accelerator complex at CERN. (From the CERN web pages).

the LHC operation, but the relatively small diameter of the tunnel only allows for one magnet and cryogenic line.

Most colliders — such as the proton-anti-proton Tevatron — collide particles and anti-particles to allow acceleration in a single ring, thus reducing their cost and complexity. Unfortunately, anti-protons are difficult to produce in the amounts needed at LHC, thus prompting the choice of protons in both directions. This has, however, little impact on the physics studied at the LHC.

Accelerating particles is a very complex operation, and it has proven prohibitively difficult to make an all-in-one accelerator, taking the particles from standstill to the collision energy. Thus, it is necessary to build a chain of accelerators, usually increasing the energy 10 to 100 times in each link. To reach the 14 TeV center of mass energy aimed for at the LHC, CERN uses a long injection chain of accelerators, some of which dates back to the early fifties. The CERN accelerator complex, Fig. 2.1, involves five stages of acceleration, starting with the extraction of protons from a Hydrogen plasma

in a duoplasmatron. These protons are accelerated to a kinetic energy of 50 MeV in a linear accelerator (LINAC2), before speeding to 1.4 GeV in the Proton Synchrotron Booster (PSB). Furthermore, they are accelerated to 25 GeV in the Proton Synchrotron (PS) where the LHC bunch-train structure is created. Finally, they are kicked up to 450 GeV by the Super Proton Synchrotron (SPS) before being injected — in both directions — into the LHC.

Using particles with an internal structure, such as protons, as opposed to point-like particles, such as electrons, complicates the physics of the collisions significantly, and reduces the amount of energy released in each particle collision. Most of the energy stays with the parts of the protons that are not directly involved in the collision. Protons are, however, very stable and easy to produce, and most important of all; quite heavy. This greatly reduces the energy loss from synchrotron radiation when circling the collider — compared to electrons and positrons — thus allowing center of mass energies of 14 TeV. Moreover, protons, with their constituent quarks and gluons, are sensitive to all forms of interactions, and hence produce a great variety of particles when colliding. This is essential to a discovery machine, such as the LHC. Although the internal structure of the protons lower the center of mass energy of the collisions, it is still sufficient to probe the electroweak symmetry breaking believed to be caused by the Higgs mechanism, and to reveal supersymmetric particles if existing.

At design luminosity, each bunch contains $n_b \approx 10^{11}$ protons, and the spacing between bunches is 25 ns, corresponding to a bunch rate of $f_b = 40$ MHz. The transverse bunch width is $\sigma_b = 15.9$ μm, while the bunch length is 7.5 cm. This gives a design luminosity of

$$\mathcal{L} = \frac{n_b^2 f_b}{4\pi\sigma_b^2} \approx 10^{34}\text{cm}^{-2}\text{s}^{-1}$$

Initially the luminosity will be one order of magnitude less.

2.2 The ATLAS detector

The ATLAS detector [1], Fig. 2.2, is one of the two multipurpose detectors at the LHC; in the sense that ATLAS is not geared towards looking at any particular type of events or particles. It is built to search for new, unknown particles and for doing precision measurements on established particles and models. Generally speaking, ATLAS estimates particle types and energies from material interactions — or lack thereof — while measuring particle momenta through magnetic field interactions.

The ATLAS detector is shaped like a barrel with two endcaps, providing forward-backward symmetry with respect to the interaction point. The innermost part contains the inner detector, mostly built from highly granular layers of semiconducting Silicon, providing good resolution while keeping the amount of material low. The inner detector is used for reconstructing the trajectories of charged tracks, hence it is sometimes referred to as the tracking detector. It is placed inside a superconducting solenoid magnet — providing a relatively homogeneous magnetic field of 2 T — to bend the tracks and hence facilitate the measurement of particle charges and momenta. Covering

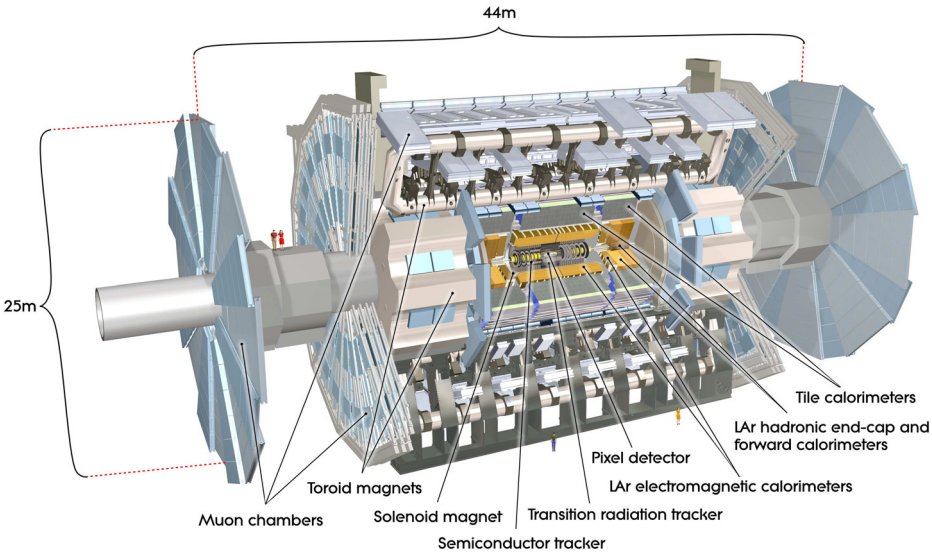


Figure 2.2: The ATLAS detector at CERN. (From Ref. [1]).

the solenoid, we find the electromagnetic and hadronic calorimeters, measuring the deposited energy of isolated particles or groups of particles — the so-called jets — by stopping them in layers of dense material, such as Iron or Lead. The calorimeters are surrounded by the muon spectrometer, intertwined with eight superconducting air-core toroidal magnets in each end-cap and eight magnets in the barrel, 24 in total. The barrel toroids also double as part of the support structure of the 7000 tonnes ATLAS detector. Since the detector is built from a barrel section and two end-caps, we see some degradation of the detector performance in the gaps between these parts. The performance goals of the separate ATLAS detector parts are listed in Table 2.1.

In the ATLAS barrel configuration, using cylindrical coordinates for the position of the detector parts and the particle momenta becomes natural. The direction out from the interaction point is given in terms of an azimuthal angle ϕ and a polar angle θ, which is often traded for the pseudo-rapidity $\eta = -\ln\tan(\theta/2)$. Pseudo-rapidity is preferred in high-energy hadron colliders because the particle production is almost constant as a function of η. This is especially useful when deciding on the detector granularity during the design phase. An even distribution of hits in the detector elements and data readout channels is guaranteed by letting every detector element cover an equal sector of η. A complete position in the detector is given by ϕ, θ and an additional z or R value, whereas a particle momentum usually is completed by giving the transverse momentum p_T.

Many of the physics processes hoped to be seen at ATLAS involve heavy short-

Table 2.1: General performance goals of the ATLAS detector, with $\oplus$ denoting a quadratic sum, and E and p_T given in GeV. (From Ref. [1]).

Detector component	Resolution goal	η coverage
Inner detector	$\sigma_{p_T}/p_T = 0.05\% \times p_T \oplus 1\%$	± 2.5
EM calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2
Hadronic calorimeter:		
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	± 2.7

lived particles, such as the Z^0, $W^\pm$ and H bosons, decaying so quickly that their decay products seem to originate from the interaction point. This is true even when the decay chains involve several stages, such as in the supersymmetric cascade decays. Important exceptions to this are the intermediate taus and B-hadrons — and to a lesser extent the D-hadrons — which typically move a few hundred μm before decaying. The particles stable enough to reach the active detector elements are; electrons, muons, photons and hadronic jets consisting of $p/\bar{p}$, $n/\bar{n}$, $\pi^\pm$, $K^\pm$ and K^0_L as well as e, μ and γ. Neutrinos pass through ATLAS undetected. However, since the total transverse momentum of an event should sum up to zero — assuming no cracks in the detector — the combined transverse momentum of the neutrinos and other invisible particles, such as the lightest supersymmetric particle, can be deduced from the missing transverse momentum.

The ATLAS detector faces many challenges, especially from the high collision rate foreseen at the design luminosity of LHC. This high collision rate is necessary because very few collisions are head-on, generating the center of mass energy needed for discovering new physics. At this luminosity, every bunch-crossing produces an average of 24 inelastic proton-proton collisions — along with interactions between the beam and the beam-pipe — flooding the detector with low-energy particles, producing the so-called pileup. This increases the amount of noise in the detector, complicates the track reconstruction and requires higher detector granularity than otherwise needed. Worse still, the high radiation rate physically breaks down the active detector elements — especially those close to the beam-pipe — degrading their performance over time. Some elements, such as the innermost pixel layer of the inner detector, are scheduled for replacement after a few years.

2.2.1 Inner detector

The inner detector, Figs. 2.3 and 2.4, is used to measure the position of the interaction point along with secondary vertices from long-lived particles, such as τs and B-hadrons, and to measure the momenta of charged particles from the curvature of their trajectories in the magnetic field.

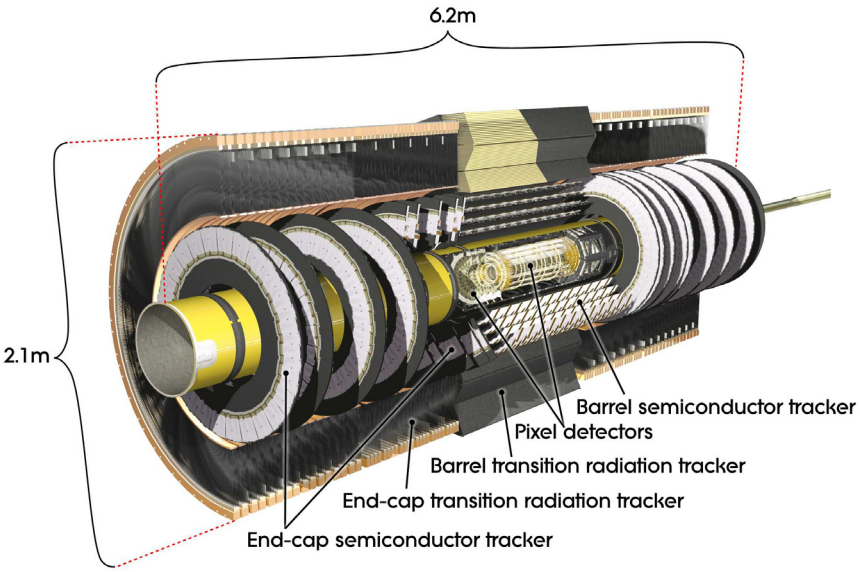


Figure 2.3: The ATLAS inner detector. The innermost part is the pixel detector, followed by the semiconductor tracker (SCT) and the transition radiation tracker (TRT). (From Ref. [1]).

The inner detector is made from three subsystems; at the beam pipe lies the pixel detector, which consists of three barrel layers and three discs on each side, ensuring three hits over the full acceptance range defined by $|\eta| < 2.5$. The layers of the barrel part are situated at radii of 5, 9 and 12 cm. Each pixel layer consists of a lattice of tiny semiconducting Silicon dots, 80 million channels in all, providing the high granularity needed to pinpoint the location of the interaction point down to the necessary μm precision. The innermost pixel layer is critical to B physics, hence often referred to as the B layer. Due to its proximity to the beam pipe, it experiences a very tough radiation climate and is scheduled to be replaced after a few years, depending on the luminosity profile.

Outside of the pixel detector, in the radial range between 30 and 52 cm, we find the semiconductor tracker (SCT). It consists of four barrel layers made from overlapping sandwiched Silicon strip sensors, Fig. 3.9, and nine end-cap layers on each side (in the $|z|$ range of 80–280 cm), providing eight precision measurements over the full acceptance range. Silicon strips provide less resolution than pixels, but are cheaper to produce and satisfy the demands of precision measurements at the radii occupied by the SCT layers. The pixel and SCT detectors are kept at $-7\,^{\circ}$C to reduce the radiation damage.

The outermost part of the inner detector is occupied by the transition radiation tracker (TRT). This is a continuous analog tracking device, as opposed to the pixel

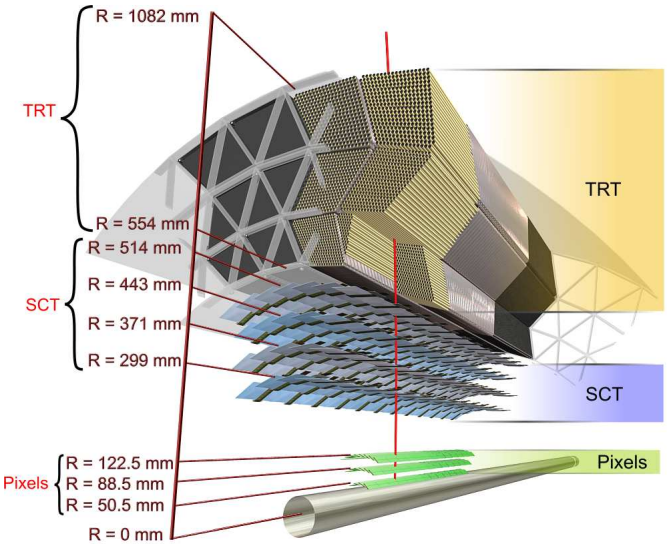


Figure 2.4: Illustration of the inner detector barrel section. (From Ref. [1]).

and SCT detectors which produce discrete digital measurements. The TRT is based on 4 mm diameter straw tubes, made up of a cylinder — serving as the cathode — filled with a Xenon-based gas mixture and a central anode wire. When a charged particle traverses the straw the gas is ionized, generating a measurable current. In the barrel region at radii of 55–108 cm, and in the end-caps at $|z|$ between 80 and 280 cm, a total number of 350 000 straws assure approximately 36 hits per track, with an intrinsic accuracy of 130 μm per straw. The barrel TRT straws are mounted parallel to the beam axis, whereas the end-cap straws fan out on disc structures. The transition radiation generated by the radiator material interspaced between the straws can be used to separate electrons from charged pions up to momenta of around 100 GeV/c.

2.2.2 Calorimeters

The ATLAS calorimeters, Fig. 2.5, are primarily used for measuring the energy of electrons, photons and jets. While the inner detector is built to let particles pass undisturbed — estimating momenta by studying trajectory curvatures — the calorimeters measure energy deposits by stopping particles. This allows measuring the energy of jets and particles insensitive to the magnetic field, such as photons and neutrons. Ideally, the calorimeters should stop everything but muons — and neutrinos — from entering the muon spectrometer, which surrounds the calorimeters. Very energetic hadrons, such as protons and neutrons, might, however, punch through the calorimeters, contaminating the muon spectrometer. Such punch-through is, however, unlikely to produce fake muon tracks apparently originating from the interaction point. Fake muon tracks can be suppressed further by doing a combined reconstruction of inner detector and muon spectrometer tracks. The ATLAS calorimeters cover an exceptional

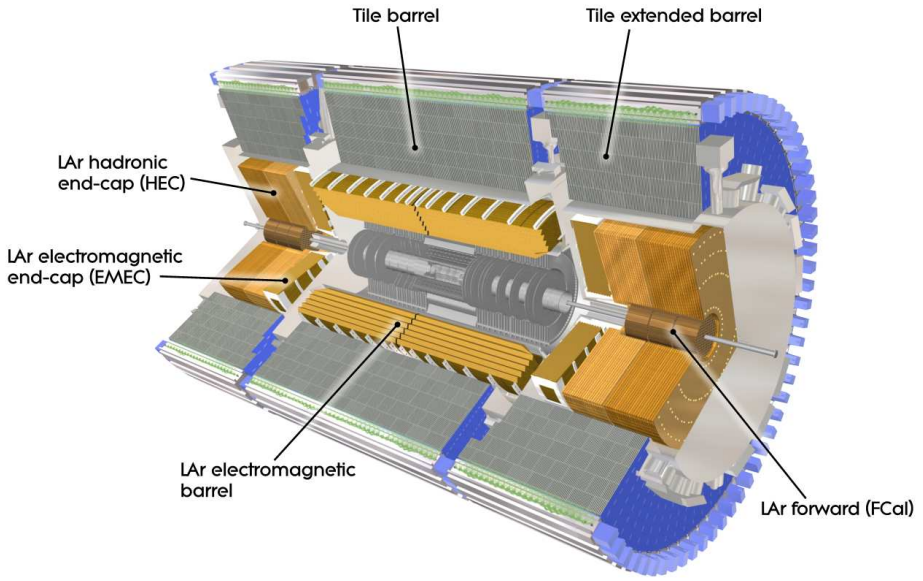


Figure 2.5: The ATLAS electromagnetic and hadronic calorimeters. (From Ref. [1]).

range ($|\eta| < 4.9$) to allow a detailed study of the missing transverse energy E_T^{miss} related to particles going undetected through ATLAS, such as neutrinos and the lightest supersymmetric particle.

Generally speaking, the calorimeters are divided into two categories; electromagnetic and hadronic calorimeters. The first are tuned to measuring the energy deposited by the particle showers — made from consecutive bremsstrahlung and pair production processes — originating from electrons and photons, while the latter are designed to study showers generated by hadronic particles, such as protons and neutrons. These heavier particles penetrate deeper into the material, producing bigger and more complex showers. Hence, the hadronic calorimeters are thicker and placed outside their electromagnetic counterparts. At ATLAS, the average thickness of the EM calorimeter is 23 radiation lengths (X_0), while the hadronic calorimeter covers approximately 10 interaction lengths (λ), which translates to around 60 cm and 2 m, respectively. The total detector material in front of the muon spectrometer is between 11 and 15 λ, most of which is in the hadronic calorimeter. This is sufficient for measuring the energy of high-energetic jets with good resolution, as well as keeping the punch-through at a reasonable level.

All of the ATLAS calorimeters are of the sampling type, meaning that they are built from alternating passive and active layers. The passive — in the sense that no measurements are done within them — absorber layers are mostly made from dense

and cheap metals, such as Iron or Lead, whereas the active layers are made from scintillators or liquid Argon. Other calorimeter configurations are also possible; at the CMS experiment, the EM calorimeter is made entirely from dense crystals, serving both to generate showers and measure energy deposits.

The energy resolution of a calorimeter can be parametrized by

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.1}$$

where $\oplus$ denotes a quadratic sum. The dominant first term is due to the sampling fluctuations which are Poisson distributed in nature. The second term reflects the electronic noise of the calorimeter, while the third term is a constant term related to the calibration errors and other systematic effects. Since the energy resolution is usually dominated by the stochastic and noise terms, the resolution improves with increasing energy deposited in the calorimeter.

Electromagnetic calorimeter

The ATLAS EM calorimeter is divided into a barrel part ($|\eta| < 1.5$) and two end-caps ($1.4 < |\eta| < 3.2$), each housed in separate cryostats. To minimize the amount of dead material in the detector — such as support structures, cables, pipes and cryostat walls — the EM calorimeter shares the barrel cryostat with the central solenoid magnet, while the end-cap cryostats are shared with the hadronic calorimeter. The EM calorimeter is built from accordion-shaped kapton electrodes and Lead absorber plates, immersed in liquid Argon (LAr). The accordion shape provides complete ϕ symmetry without azimuthal cracks. Moreover, the EM calorimeter end-caps are divided into two coaxial wheels; an outer wheel covering the region $1.4 < |\eta| < 2.5$, and an inner wheel covering $2.5 < |\eta| < 3.2$. The outer wheel — along with the barrel EM calorimeter — is dedicated to high-precision physics measurements, and hence is segmented into three sections in depth. The less precise inner end-cap wheel is segmented into two sections in depth, and has a coarser lateral granularity than the rest of the EM calorimeter. In the region $|\eta| < 1.8$, a presampler — in the form of a 1 cm thick active LAr layer — is used to register initial showering by electrons and photons upstream of the EM calorimeter.

The high granularity of the EM calorimeter allows for the separation of the two showers from $\pi^0 \to \gamma\gamma$, resulting in a good γ/π^0 identification, which is vital to the Higgs discovery through the $H \to \gamma\gamma$ channel.

Hadronic calorimeter

The hadronic calorimeter is built in two separate ways depending on the location within the detector. Most of it — the so-called tile calorimeter — is split into three parts covering the EM cryostats; the tile barrel and the two extended tile barrels. The rest of the hadronic calorimeter shares the end-cap cryostats with the EM calorimeter. Keeping most of the hadronic calorimeter out of the cryostats allows for a simpler and cheaper design.

The tile calorimeter covers the region $|\eta| < 1.7$, and is built as a sampling calorimeter with alternating steel and scintillating layers. Radially, it extends from an inner radius of 2.3 m to an outer radius of 4.3 m. It is segmented in depth into three layers, approximately 1.5, 4.1 and 1.8 interaction lengths (λ) thick for the barrel, and 1.5, 2.6 and 3.3 λ for the extended barrels, totalling close to 10 λ for the whole tile calorimeter.

The LAr hadronic end-cap calorimeter consists of two independent wheels per end-cap, located directly behind the EM calorimeter. The hadronic end-cap calorimeter covers the region $1.5 < |\eta| < 3.2$, slightly overlapping the tile and forward calorimeters. Each wheel is divided into two segments in depth, for a total of four segments per end-cap. The LAr hadronic end-cap calorimeter is quite similar to the EM calorimeter, the main difference being the use of Copper, instead of Lead, in the passive absorber layers.

Forward calorimeter

The LAr forward calorimeter is integrated into the end-cap cryostats, very close to the beam pipe, covering the region $3.1 < |\eta| < 4.9$. Such a wide coverage increases the precision of the missing transverse energy measurement, which is particularly important to the SUSY studies. The forward calorimeter is built from three modules in each end-cap; the first, optimized for electromagnetic measurements, uses Copper in the passive layers, whereas the other two modules, tuned to measure hadronic interactions, uses Tungsten for the absorber layers. The denser Tungsten provides the required amount of material in the limited space allocated to the forward calorimeter.

2.2.3 Muon spectrometer

The muon spectrometer, Fig. 2.6, surrounds the calorimeter and defines the overall dimensions of the ATLAS detector. It is intertwined with the barrel and end-cap superconducting toroidal magnets, generating strong bending power and providing the support structure for the muon spectrometer. The light and open structure of the muon spectrometer minimizes the multiple scattering. Excellent muon momentum resolution is achieved with three layers of high precision tracking chambers, while triggering is achieved by a separate set of trigger chambers.

The eight barrel toroids are housed in individual cryostats, while the 16 end-cap toroids are evenly split between two end-cap cryostats. The barrel toroids provide most of the magnetic bending in the range $|\eta| < 1.4$, while the end-cap magnets generate the field in the region $1.6 < |\eta| < 2.7$. The transition region ($1.4 < |\eta| < 1.6$) is covered by both the barrel and end-cap magnets. In the barrel region, tracks are measured in chambers arranged in three cylindrical layers around the beam axis, while in the transition and end-cap regions, the chambers are installed in three planes perpendicular to the beam axis.

The muon spectrometer uses two muon chamber designs for precision tracking measurements, and two chamber designs for triggering — four in all — listed in Table 2.2. Most of the precision tracking is provided by the monitored drift tubes (MDT), which

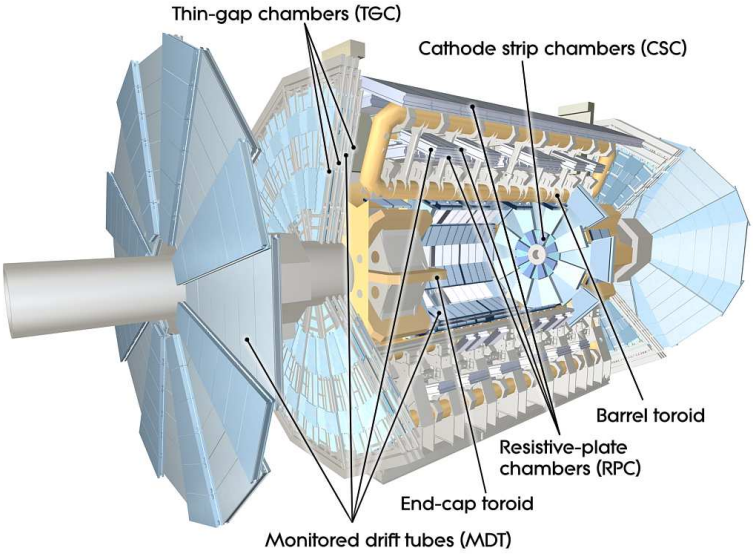


Figure 2.6: The ATLAS muon spectrometer. (From Ref. [1]).

are Aluminum tubes of a 3 cm diameter filled with an Argon-based gas mixture and a central anode wire. The mechanical isolation in the drift tubes of each sense wire from its neighbours guarantees a robust and reliable operation. A total of 354 000 tubes are used, with the length of the tubes varying from 70 to 630 cm. In the regions closer to the interaction point, with more demanding particle rates and background conditions, cathode strip chambers (CSC) are used. These are multiwire proportional chambers with shorter response time and higher granularity than the MDT chambers.

Table 2.2: Main parameters of the muon spectrometer. (From Ref. [1]).

Muon chamber type	Function	Channels	η coverage
Monitored drift tubes	Precision tracking	354 000	± 2.7
Cathode strip chambers	Precision tracking	31 000	$2.0 < \|\eta\| < 2.7$
Resistive plate chambers	Triggering, 2nd coordinate	373 000	± 1.1
Thin gap chambers	Triggering, 2nd coordinate	318 000	$1.1 < \|\eta\| < 2.7$

In the trigger system, covering the range $|\eta| < 2.4$, resistive plate chambers (RPC) and thin gap chambers (TGC) are used in the barrel and end-caps, respectively. The trigger chambers in the muon spectrometer serve a threefold purpose; providing bunch-crossing identification, giving well defined p_T thresholds, and measuring the muon

coordinate in the direction orthogonal to that determined by the precision tracking chambers.

The design requirements of the stand-alone muon momentum resolution calls for a precision of 30 μm on the relative alignment of the muon tracking chambers. These stringent requirements are met by a combination of precision mechanical-assembly techniques and optical alignment systems, both within and between the muon chambers. Approximately 12 000 alignment sensors, based on the optical monitoring of deviations from straight lines, are used to monitor the MDT chambers alone.

Magnetic field

The superconducting magnetic system of ATLAS consists of the central solenoid and the air-core barrel and end-cap toroids, Figs. 2.2 and 2.6. The central solenoid provides a nominal magnetic field of 2 T for the inner detector tracking system, while the barrel and end-cap toroids set up an average field of around 1 T, with a peak value of 4 T, in the muon spectrometer. The solenoid field is mainly directed along the z-axis, whereas the toroidal field is directed along ϕ, Fig. 2.7. Due to its position in front of the EM calorimeter, the amount of material in the solenoid is kept at a minimum. The amount of material in the toroidal magnets is, however, of less concern since they are located in the muon system. Each of the three toroidal barrel and end-cap systems are made from eight coils contained in Aluminum alloy casings. The barrel magnet casings also double as an important part of the ATLAS mechanical support structure. The solenoid flux-return yoke is integrated into the hadronic tile calorimeter support structure and weighs around 4000 tons. All of the ATLAS magnets are cooled by liquid Helium to 4.5 K, and are equipped with a quench protection system to safely dissipate the stored energy without overheating the coil windings. The magnetic field is continuously monitored by 1800 Hall sensors distributed throughout the spectrometer volume.

Trigger

The bunch crossing rate at the LHC accelerator is 40 MHz, whereas the ATLAS data acquisition system only operates at 100 Hz. A three-level trigger selects which events to store; the hardware-based first-level trigger selects events with big energy deposits in the calorimeters or hits in the muon chambers, which brings the rate down to 75 kHz. The software-based level-2 trigger then reduces this rate to 1 kHz by doing a rudimentary event analysis. Finally, the event filter applies sophisticated algorithms, shared with the offline reconstruction software, to arrive at the 100 Hz storage rate, translating to around 100 MB/s.

2.3 Particle identification

In the previous Section, we described the different parts of the ATLAS detector, most of which are dedicated to the study of electrons, photons, muons and jets, Fig. 2.8.

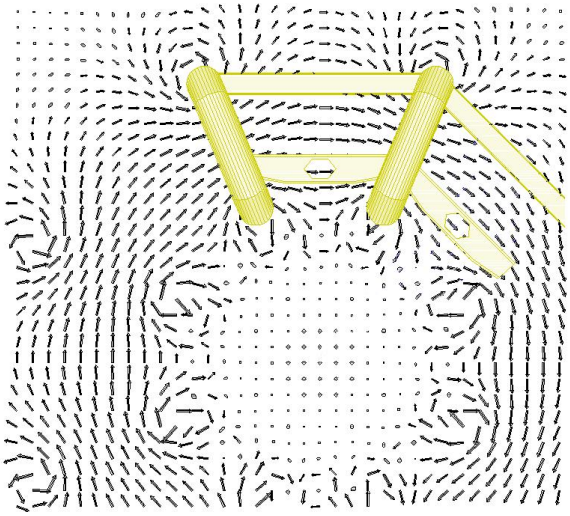


Figure 2.7: The magnetic field of the ATLAS detector in the R-ϕ plane. Two of the toroidal barrel magnets are also shown. (From the ATLAS web pages).

In the following, we discuss some techniques for identifying and reconstructing these objects, along with the detector resolution and efficiency for doing this at the LHC design luminosity.

Muons

Muons are in many ways the easiest particles to identify and reconstruct, hence large parts of the ATLAS detector are devoted to studying them. Muons have a long lifetime ($c\tau = 659$ m) and great penetration power, making them the only visible particles passing through the whole ATLAS detector. The muon momentum is determined from the track curvature, which relates to the inverse transverse momentum $1/p_T$. The momentum resolution of the track then becomes proportional to p_T;

$$\frac{\sigma_{p_T}}{p_T} \propto p_T \tag{2.2}$$

Hence, the momentum resolution deteriorates as the momentum grows, contrary to the energy resolution of the calorimeter readings, Eq. (2.1). The muon momentum is measured in the inner detector and the muon spectrometer, which complement each other due to their different regions of sensitivity. The inner detector provides the most accurate measurements below transverse momenta of around 40 GeV/c, while the muon spectrometer performs better for momenta above this value. The expected resolutions at $p_T = 10/100/1000$ GeV/c are $\sigma_{p_T}/p_T \approx 1.5/2.5/8\%$, showing the reduced measurement accuracy for increased momenta. The expected reconstruction efficiency for the same transverse momenta is 97/95/85%.

When studying muons, we are mostly interested in those coming from the hard process at the interaction point, particularly muons produced by decaying Z^0 bosons or SUSY particles. There are, however, other sources of muon production at ATLAS,

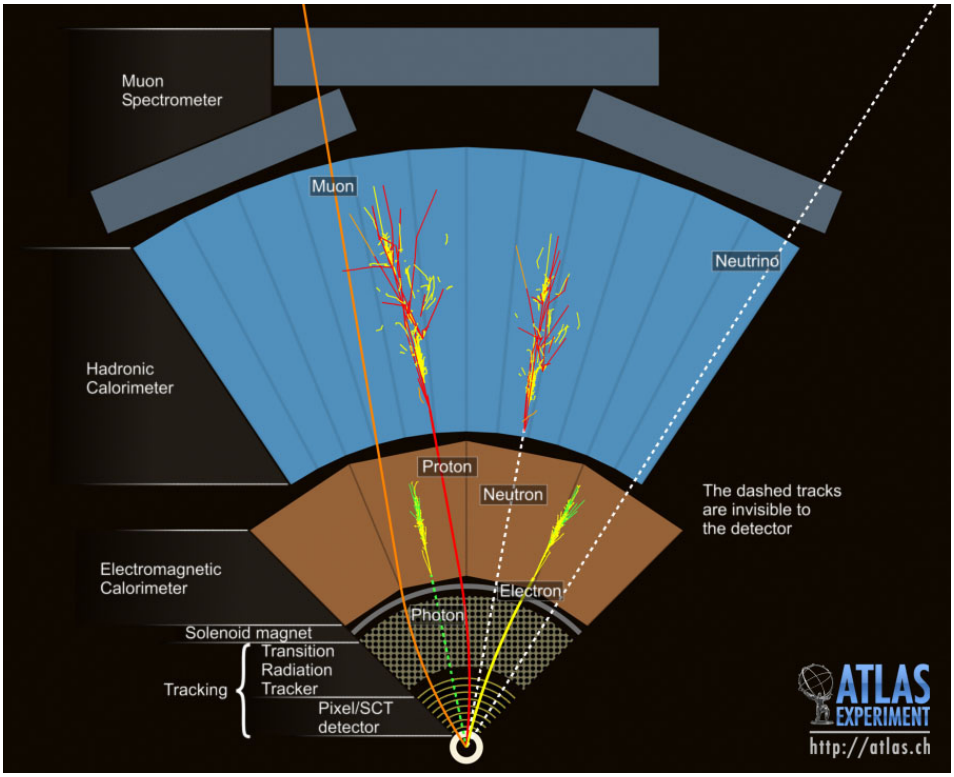


Figure 2.8: Typical tracks in the ATLAS detector (not to scale). (From the ATLAS web pages).

such as the muonic decay of $\pi^{\pm}/K^{\pm}$ ($c\tau$ = 7.8/3.6 m). These particles have a fair chance of decaying to low-p_T muons in the inner parts of the detector. Such muons are distinguished from the centrally produced muons by looking for abnormal discrepancies in the momenta found by the inner detector and those seen in the muon spectrometer. Pion and kaon decays also produce kinks in the combined inner detector and muon spectrometer track which ruin the track fit. Another source of "intermediate" muons is the semileptonic decay of B- and D-hadrons. These muons are removed by using an isolation cut, i.e. by looking at the calorimetric energy deposits in a cone surrounding the muon track. Muons coming from semileptonic decays are often part of jets, hence not isolated in the calorimeters.

Electrons

The passage of electrons through the inner detector is registered through hits in the pixel, SCT and TRT subdetectors. Electrons are separated from the muons through transition radiation in the TRT subdetector and energy deposits in the EM calorimeter, bringing them to a full stop. Electrons are prone to bremsstrahlung, which is proportional to $1/m^2$, giving them an energy loss from bremsstrahlung 40 000 times greater than that of muons. This is the main reason why muons pass the calorimeters relatively undisturbed. The strong bremsstrahlung by electrons degrades the quality of the track reconstruction and momentum measurement, and is — along with photon conversion — the main motivation for keeping the material of the inner detector at a minimum.

Electron reconstruction is based on the combined inner detector and EM calorimeter readings. From Eqs. (2.1) and (2.2), we know that the accuracy of the energy measurement increases with energy in the calorimeter, while decreasing with momentum in the inner detector, with the crossing point around $p_T = 15$ GeV/c. The expected resolutions at $E = 20/50/200$ GeV are $\sigma_E/E \approx 3/2/1\%$. Several factors are taken into account to achieve sufficient rejection against jets, such as the shape, structure and energy of the EM shower. Furthermore, the deposited EM calorimeter energy should match the momentum found in the inner detector in the case of an electron ($E \approx p$). Through the combination of cuts, a rejection factor of 10^5, while keeping the electron efficiency close to 70%, is achieved.

Photons

Ideally, the neutral photons leave no tracks in the inner detector, they are only spotted in the EM calorimeter and reconstructed with an energy resolution similar to that of the electrons. However, due to interactions with the material in the inner detector, around 30% of the photons convert to e^+e^- pairs before reaching the calorimeter. By reconstructing opposite-charged tracks, requiring that they originate from a common vertex with no opening angle, and that the reconstructed photon points back to the main interaction point, conversions can be recovered with an efficiency of around 60%. The energy resolution of converted photons is somewhat less than for the unconverted photons, which are seen only in the EM calorimeter and reconstructed from the shower deposited therein.

As in the case of the electrons, which also rely heavily on calorimeter deposits, good rejection against jets is crucial. This is, however, easier for photons since most of them leave no tracks in the inner detector. Furthermore, the shower shape and leakage into the hadron calorimeter is investigated to reach the desired rejection of 10^3 at around 80% efficiency.

At low Higgs masses the most common decay channel by far is $H \to b\bar{b}$. However, due to the high QCD background at the LHC, the best discovery channel is $H \to \gamma\gamma$, prompting good directional resolution of the EM calorimeter to reconstruct the Higgs mass accurately.

Jets; b-jets, τ-jets

When protons collide at the interaction point, expelling partons (quarks and gluons), a long and complex chain of events is set into motion. First, the hard process produces additional partons through initial and final state radiation before fragmentation (confinement) turns them into colourless hadrons. These hadrons decay into new hadrons, leptons and photons until the partons are turned into narrow jets (cones) of relatively stable particles. Charged particles are seen in the inner detector, having their tracks bent by the magnetic field before stopping in the calorimeter, whereas the neutral particles head directly for the calorimeter. This splitting of the jets is spotted through reconstruction of the charged inner detector tracks, allowing separate calorimeter deposits to be attributed to the same jet. Since hadronic showers produced by jets are more complex in nature than the EM showers coming from electrons and photons, their energy resolution in the calorimeter is worse. Eq. (2.1) is still valid for jets, though with larger coefficients than for electrons and photons. The expected resolutions at $E = 50/100/200/1000$ GeV are $\sigma_E/E \approx 15/9/7/3\%$.

The identity of the initial particle of a jet is usually lost, with the exception of jets originating from b-quarks and hadronically decaying τs. B-hadrons are tagged through their noticeable lifetime ($c\tau \approx 450$ μm), producing an observable secondary vertex, giving a tagging efficiency of around 50%, and rejection factors of 10 against c-jets and 100 against gluon-, u- and d-jets. Rejecting c-jets is complicated by the fact that they also generate secondary vertices, though closer to the beam line. Hadronic decays of τs produce lean jets containing few tracks, typically one or three charged pions. Looking for these tracks and a matching shower shape in the calorimeter allows for a useful τ-tagging efficiency, which becomes better with increasing p_T.

Chapter 3

The NewTracking framework

Historically, the track reconstruction at ATLAS has been handled by monolithic legacy packages, such as `xKalman` [14] and `iPatRec` [15] for the inner detector, and `Muonboy` [16] for the muon spectrometer. These packages contain everything, from detector and material descriptions to pattern recognition and track fitting. Due to their complexity and lacking transparency, it was decided by the *Reconstruction Task Force* [17] to split them into *modules*, each performing a well-defined and specific task. This was the birth of the `NewTracking` framework [2]. The basis of the `NewTracking` framework is the common *event data model* (EDM) [18] — which defines the *track parameters* and objects to be used throughout the ATLAS inner detector, calorimeter and muon spectrometer reconstruction software — and the common *tracking geometry* [19] describing the whole detector in a manner suitable to fast and efficient track reconstruction. The specific tools for the actual track reconstruction are placed on top of the common EDM and tracking geometry, and must adhere to the rules given by the framework. Some of these tools have been lifted out of the legacy packages, while others are written specifically for `NewTracking`. Among the latter, we find the *track extrapolation* package [20] containing the *simultaneous track and error propagation* algorithm — or STEP propagator — described in this thesis.

3.1 The event data model

The event data model [18] describes the common track parameters and track objects — such as calibration corrections, measured hits, estimated holes, material interactions and fit quality — used throughout the `NewTracking` framework. Since most of the track information is somehow related to the position and momentum of the track, the track parameters are given special attention in the EDM. The track parameters are given at, and defined according to, the surfaces crossed by the track. These surfaces may be helper surfaces only found in the tracking geometry, or representations of real detector elements. In addition to the track positions and covariance matrices given locally at these associated surfaces, the track parameters contain track positions and momenta given in the global ATLAS coordinate system.

The local track position at a surface is given by two coordinates (l_0, l_1) specific to

each of the five surface types defined in the EDM; planes, discs, cylinders, straight lines and perigees which are a special case of straight line used to describe the beam line. Further details on these surface types are found in Refs. [18,19]. The global track position is given by the (x, y, z) coordinates in the global Cartesian coordinate system of the ATLAS detector, whereas the momentum is represented by the azimuthal angle ϕ, the polar angle θ and the signed inverse momentum λ within the same system, Fig. 3.1. The signed inverse momentum is defined as $\lambda \equiv q/p$, where q is the charge of the particle and p is the momentum. Together, the local track position at a surface (l_0, l_1) and the global track momentum (ϕ, θ, λ) form the *local track parameters*

$$(l_0, l_1, \phi, \theta, \lambda)$$

at a surface.

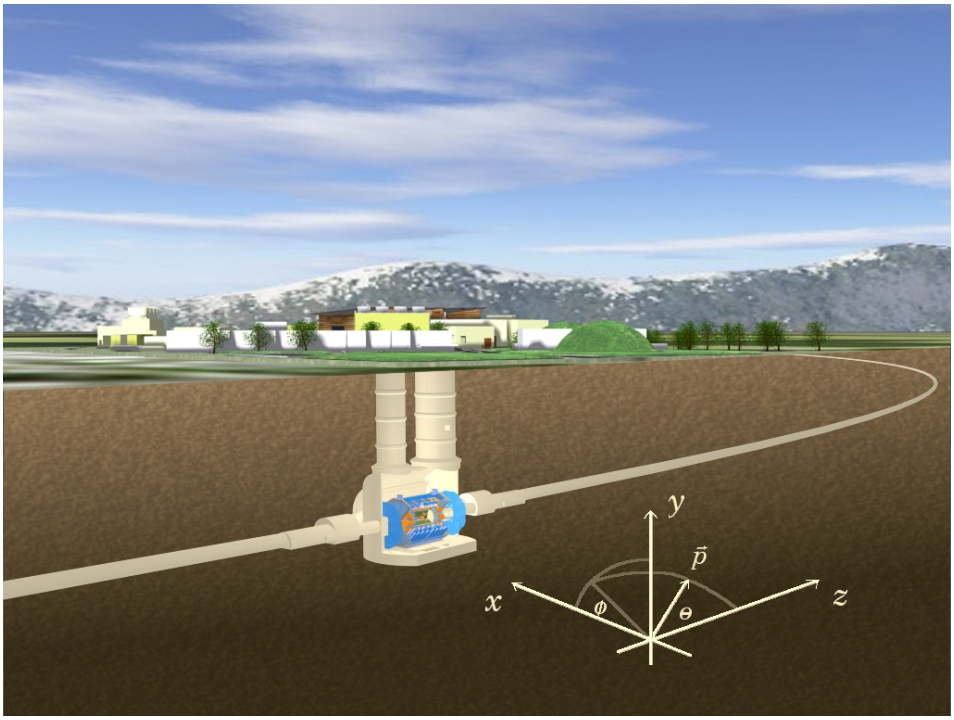


Figure 3.1: The global ATLAS coordinate system given by the beam pipe (z), the LHC tunnel centre (x) and the earth's surface (y), and the definition of the spherical polar coordinates ϕ and θ within the same system. The azimuthal angle ϕ is formed by the projection of the particle momentum into the x-y plane and the x-axis, while the polar angle θ is the opening between the momentum and the z-axis. Together they define the direction of the particle momentum unambiguously. (From the ATLAS web pages).

3.2 The tracking geometry

To estimate the particle trajectory from the initial track parameters, the detector geometry and the magnetic field need to be modeled. The granularity and complexity of the detector geometry model greatly affect the speed of the track extrapolation, therefore a simpler model is used for the more time-critical reconstruction than for the event simulation. The ATLAS reconstruction detector geometry model [19] — often referred to as the *tracking geometry* — is built by *tracking volumes* confined by a choice of three types of surfaces defined in the EDM; planes, discs and cylinders. These volumes may contain any variety of the five EDM surface types, describing anything from the material layers of the inner detector, Fig. 3.2, to the beam pipe. In addition to the different surfaces, volumes may contain a magnetic field map and a dense material description, which is useful when the material interactions are not easily approximated by point-like updates in layers. This applies especially to the calorimeters, where much of the ATLAS material is found.

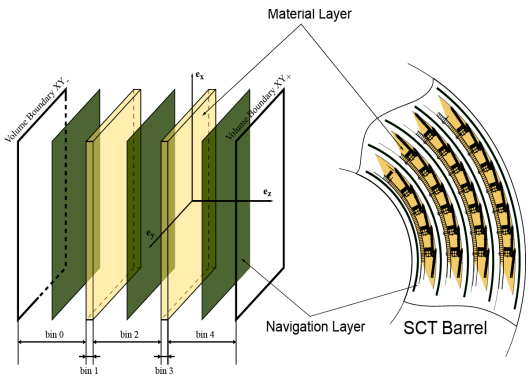


Figure 3.2: The navigation and material layers of the SCT barrel tracking geometry. The navigational helper surfaces (dark) are place-holders with the single purpose of pointing to the next, respectively previous, material layer (bright). (From Ref. [19]).

The tracking geometry is built from the geometry developed for the `GEANT4` simulation tool [21] during the initialization of the reconstruction software to assure consistency and ease maintenance. It can be built to different levels of detail, depending on the use. The inner detector and calorimeter tracking geometry is illustrated in Fig. 3.3, while the muon tracking geometry is presented in Fig. 3.4.

3.3 Track extrapolation

The track extrapolation package [20] deals with the navigation and propagation of particles through the volumes of the tracking geometry. Transporting the track parameters and the associated covariance matrix is a very frequent and important part of the track reconstruction. Most progressive fitting techniques, such as the *Kalman filter* formalism [6], Fig. 3.5, rely on the prediction of the gathered track information on the successive measurement surface. In global fitting techniques on the other hand, the prediction of the track depending on the initial parameters (i.e. the fitted parameters)

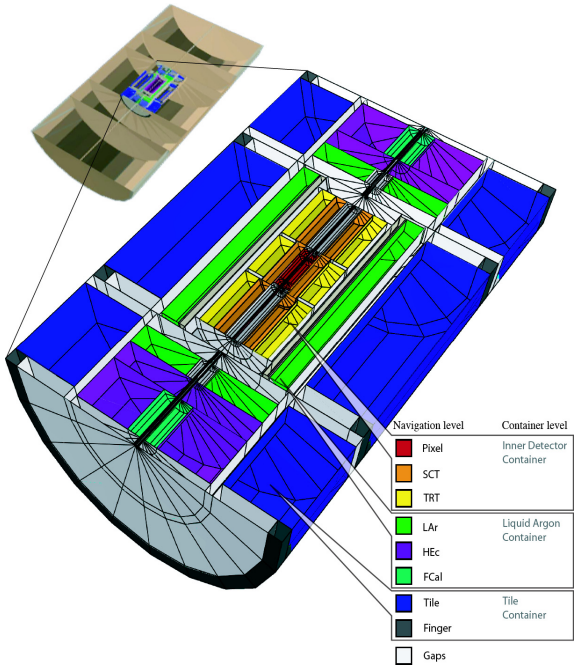


Figure 3.3: The volumes of the inner detector and calorimeter tracking geometry. The internal layers of the volumes are omitted. (From Ref. [19]).

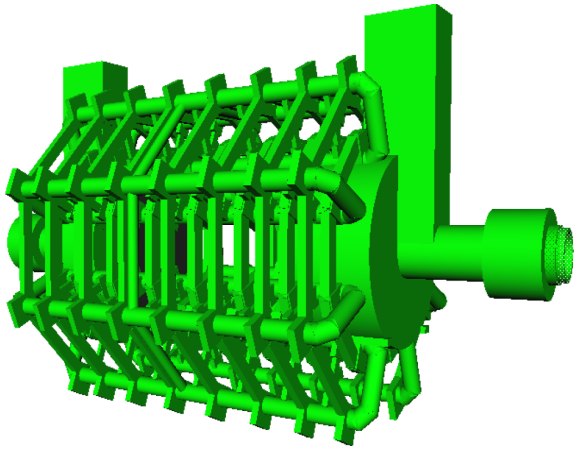


Figure 3.4: Some of the volumes of the muon tracking geometry, built at a low level of detail. (From Ref. [19]).

enters the global χ^2 function to be minimized. For both, global and sequential track fitting algorithms, the correct treatment of material effects, such as the energy loss and multiple scattering, is essential.

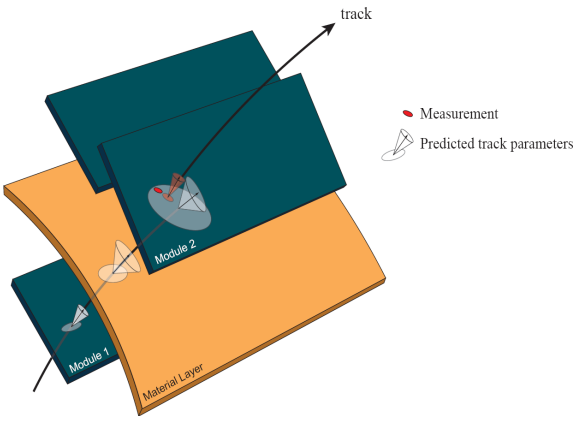


Figure 3.5: Illustration of a typical Kalman filter step. The track parameters associated to module 1 are propagated through a material layer onto module 2 to find the *predicted* track parameters at this module. In the Kalman filter formalism, the weighted mean between the prediction and the associated measurement form the updated track parameters, which are the starting point of the next filter step. (From Ref. [20]).

Propagation

Track propagation is the process of transporting the track parameters and the associated covariance matrix from one surface of the tracking geometry to another. Five types of surfaces are defined by the EDM, and they form the walls of, and objects within, the tracking volumes of the geometry. The propagation is done by applying a suitable *track model* to the initial parameters, predicting the most likely trajectory. The choice of track model is heavily influenced by the magnetic field and the amount of material in the detector; varying from a linear to a helical or numerical model. The analytical linear and helical models are only applicable to geometries involving no magnetic field and a homogeneous field respectively. The track extrapolation package contains four track models, implemented in the form of algorithms; the Straight Line [20], Helix [20], RungeKutta [14] and STEP propagators. The latter propagators employ numerical track models, well suited to the inhomogeneous magnetic field of the ATLAS detector.

Navigation

The navigator is an integral part of the track extrapolation package, providing the link between the tracking geometry and the propagator. During the extrapolation process, the navigator gathers the information about the traversed volumes from the reconstruction geometry and passes it on to the propagator. This information includes target surfaces, material properties and the magnetic field map. Furthermore, the navigator uses the feedback from the propagator to determine the successive tracking volume, Fig. 3.6.

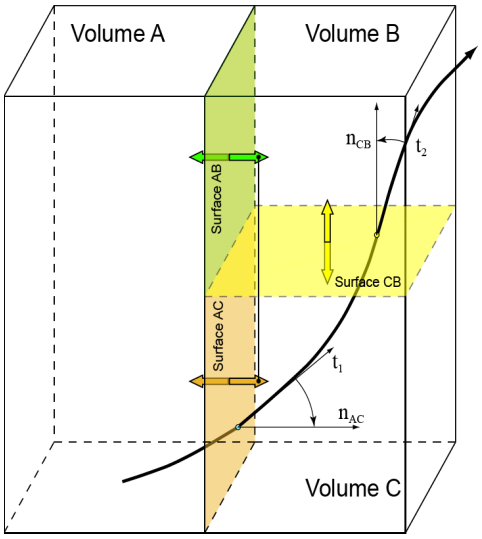


Figure 3.6: Illustration of the navigation through the tracking geometry. Target surfaces, material properties and the magnetic field map are fed into the propagator by the navigator, which determines the successive volume according to the feedback. (From Ref. [19]).

3.4 Track reconstruction

Track reconstruction is the process of turning the registered hits in the detector into tracks. This is a very complex task at the ATLAS detector due to the high track density of the experiment. ATLAS track reconstruction focuses on three strategies; inside-out and outside-in track reconstruction in the inner detector, along with the combined reconstruction within the whole ATLAS detector.

Inside-out track reconstruction

This is the main track reconstruction technique used at ATLAS. It starts out by turning the measured hits in the pixel and SCT subdetectors — referred to as the Silicon detector in the following — into three-dimensional *spacepoints* in the global ATLAS frame. Subsequently, *track seeds* are built from the spacepoints found in the pixel detector layers, and extrapolated through the SCT layers by using a Kalman filter to produce the *track candidates*. The track seed search may be constrained to the interaction point, Fig. 3.7, or not, Fig. 3.8. The unconstrained seed search is more time consuming, but allows to find loosely constrained tracks, such as the photons coming from $H \rightarrow \gamma\gamma$.

Ambiguity solving

The track finding results in a high number of track candidates, which have to be resolved before the extension into the outer TRT can be done. The track candidates are ranked according to their likelihood of describing a real particle trajectory of the underlying physics event. This likelihood is represented by a *track score*, found by combining the

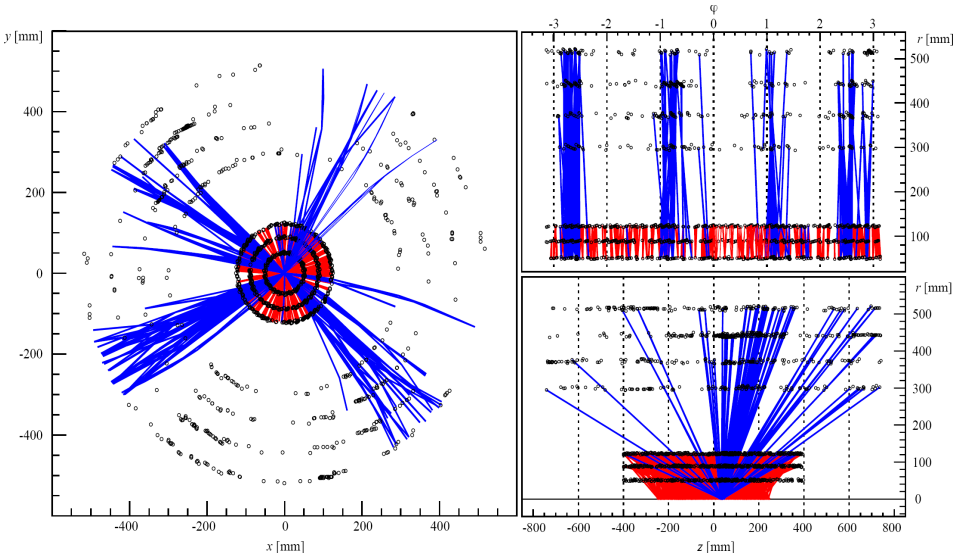


Figure 3.7: Constrained track seed search and the successive track finding in a simulated $t\bar{t}$ event. Track seeds are found in the pixel detector, constrained to the interaction point and extrapolated into the SCT detector by using a Kalman filter. (From Ref. [2]).

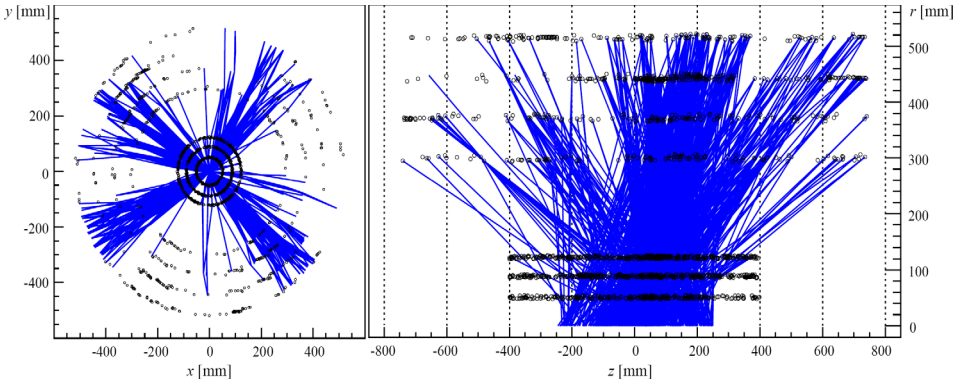


Figure 3.8: Unconstrained track seed search and the successive track finding in the same simulated $t\bar{t}$ event as shown in Fig. 3.7. Track seeds are found in the pixel layers and extrapolated into the SCT detector by using a Kalman filter. (From Ref. [2]).

χ^2/n_{dof} of the Kalman filter track extrapolation with many other factors concerning the quality of the track candidate, Fig. 3.9.

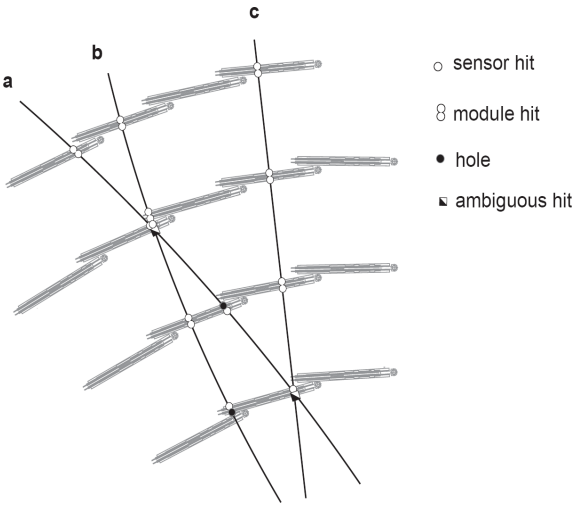


Figure 3.9: Illustration of the ambiguity solving process in the four SCT barrel layers. Each layer is built from overlapping modules made from sandwiched Silicon sensors. Track candidates are scored by combining their χ^2/n_{dof} with the amount of hits and holes on the track. Hits increase the score, whereas holes decrease it. Module and overlap region hits, as seen in track **b**, are particularly highly scored. (From Ref. [2]).

TRT track extension

Following the track resolving, the highest scoring track candidates are extended into the TRT detector and matched with successive TRT hits by using a Kalman filter, Fig. 3.10. Furthermore, the extended tracks are refitted, scored and evaluated with respect to the pure Silicon tracks. In case that the track score of the Silicon track is higher than the extended version, the Silicon track is kept and the associated TRT hits are put as *outlier* measurements to the track.

Outside-in track reconstruction

The inside-out track reconstruction process relies on initial track seeds found in the Silicon detector. Some of these seeds may, however, not be found or do not even exist; ambiguous hits can shadow the track seeds on the one hand, and on the other hand, tracks coming from secondary vertices inside the inner detector (e.g. K_S decays) or from photon conversions, may have insufficient Silicon tracker hits to comply with the inside-out track scoring demands. Moreover, substantial energy loss — mostly by electrons — at outer radii of the Silicon tracker, may deflect the particle such that no corresponding TRT hits are found. All these cases may leave track segments in the TRT overlooked by the inside-out procedure. The outside-in track reconstruction seeks to find these overlooked TRT segments and backtrack them into the Silicon detector, reconstructing secondary vertices and recovering photon conversions.

The event filter realisation

The event filter is the final stage of the three-level ATLAS trigger chain. It is a pure software trigger, analyzing the *regions of interest* (ROI) found by the first-level trigger and refined by the level-2 trigger. Each ROI incorporates a structural hypothesis,

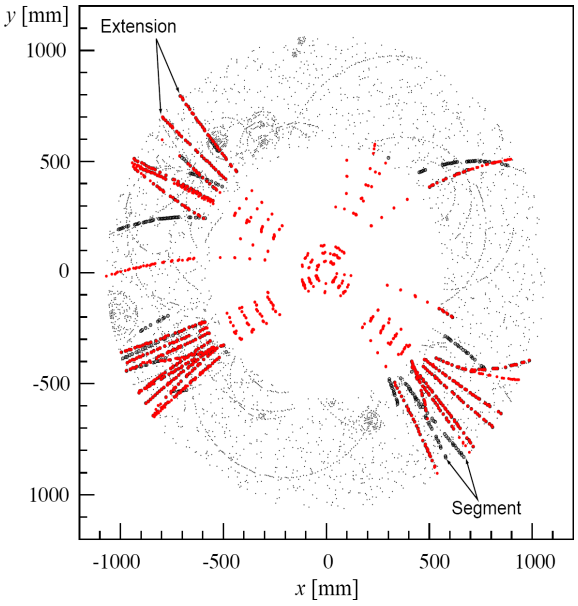


Figure 3.10: Track extensions into the TRT subdetector of the simulated $t\bar{t}$ event presented in Fig. 3.7. TRT segments found by the outside-in track reconstruction are also shown. (From Ref. [2]).

such as an electron, photon or jet, steering the event filter towards one of several predefined reconstruction modes called *slices*. Each slice contains an inside-out track reconstruction algorithm, similar to that mentioned above, but specialized to the object at hand. The algorithms used in the online event filter are taken directly from the offline reconstruction chain described above, but operated in an optimized ROI seeded mode. Fig. 3.11 shows an example of a simulated $t\bar{t}$ event reconstructed offline and online inside the `NewTracking` framework, illustrated with the ATLAS event display ATLANTIS [22].

Combined reconstruction

Initially deployed in the ATLAS inner detector, common components and modules of the `NewTracking` framework have spread widely into software applications of the muon and combined reconstruction. The common ATLAS tracking EDM has become the input event data format of all track reconstruction algorithms of the muon system, and the tracking geometry has been expanded to cover the whole ATLAS detector. Although no complete reconstruction chain based on the `NewTracking` modules has been written for the muon system and combined reconstruction yet, several specialized tools have been made. Among these are the muon backtracking tools to the calorimeter and the beam line, both based on the track extrapolation engine described in Section 3.3. The first tool associates calorimetric energy deposits to the muon tracks, while the second tool matches inner detector and muon spectrometer track segments to perform a combined muon track refit. The optimal ID and MS track segment combinations are found by representing all ID and MS segments at a common reference surface

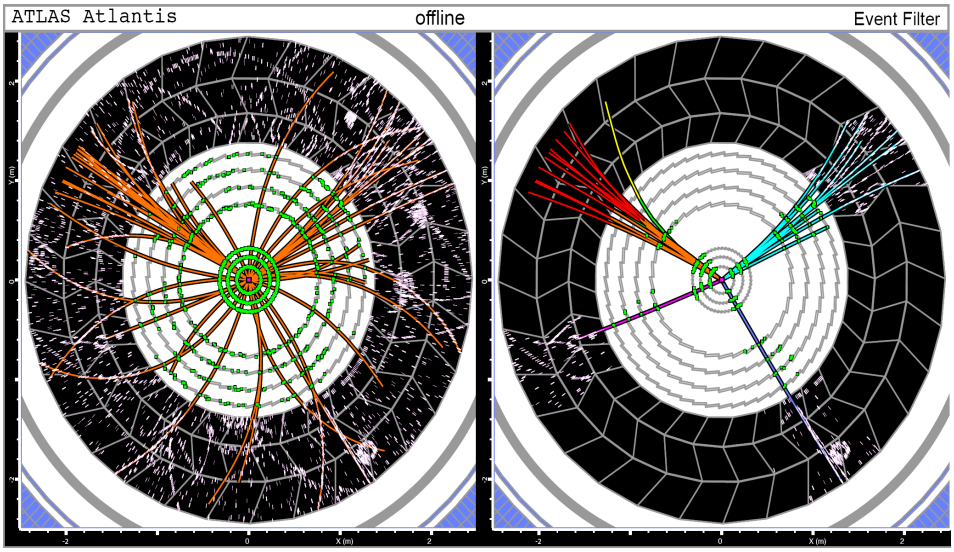


Figure 3.11: A simulated $t\bar{t}$ event reconstructed offline and online inside the NewTracking framework. The four ROIs analyzed by the event filter are all electron slices, which incorporate the full inside-out track reconstruction. (From Ref. [2]).

— by extrapolating them to the beam line — before matching them through a χ^2 minimization [23].

Recently, another combined muon track refitter — using the ATLAS Kalman fitter software package and the STEP algorithm — has been implemented and successfully tested in the full ATLAS track reconstruction software [7].

Chapter 4

Conclusion

In this thesis we have described and tested the STEP algorithm, which was developed for the ATLAS experiment in accordance with the `NewTracking` framework. Among the things achieved during this work, some deserve special mention:

- An extensive comparison of some numerical methods suitable to parameter propagation has been carried out.
- A new adaptive Runge-Kutta-Nyström method has been developed, which has potential beyond track reconstruction software.
- The impact of the magnetic field and energy loss gradients on the semi-analytical Bugge-Myrheim method of transporting covariance matrices has been studied for the first time.
- Some new solutions to the mean and most probable energy loss, its fluctuations, and the multiple scattering of particles in dense material have been presented and tested in the ATLAS calorimeter description.
- The STEP algorithm has been employed by the ATLAS Kalman fitter software package and successfully used for refitting combined inner detector and muon spectrometer tracks in the full ATLAS detector description.

Appendix A

Inverting covariance matrices by singular value decomposition

Many of the covariance matrices associated to tracks reconstructed at the ATLAS experiment — given in the local track parameters $(l_0, l_1, \phi, \theta, \lambda)$ — are *singular*, or close to singular. Singularities are rows, or columns, of zeros or relatively small values in the covariance matrix. At ATLAS, the values of the last row and column of the symmetric covariance matrix, containing the λ variance and covariances, are often very small compared to the rest of the covariance matrix. This is mainly due to the ATLAS choice of MeV and MeV/c as the units of energy and momentum. Measured in $\text{MeV}^{-1}c$, the inverse momentum λ often becomes very small at the high momenta found at the LHC, especially compared to the positional parameters l_0 and l_1 which are measured in mm.

To invert these singular, or almost singular, matrices correctly, we apply the *singular value decomposition* (SVD). The main idea behind this method is to decompose the covariance matrix into three parts

$$\mathbf{\Sigma} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^{\mathrm{T}} = \mathbf{U} \cdot \begin{bmatrix} w_1 & & \\ & \ddots & \\ & & w_5 \end{bmatrix} \cdot \mathbf{V}^{\mathrm{T}} \tag{A.1}$$

where the diagonal matrix $\mathbf{W}$ contains the singular values w_i. These go from zero and up, and can be thought of as weights when forming the covariance matrix from linear combinations of the 5×5 orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$. Given the above definition of the covariance matrix, the inverse becomes

$$\mathbf{\Sigma}^{-1} = \mathbf{V} \cdot \mathbf{W}^{-1} \cdot \mathbf{U}^{\mathrm{T}} = \mathbf{V} \cdot \begin{bmatrix} 1/w_1 & & \\ & \ddots & \\ & & 1/w_5 \end{bmatrix} \cdot \mathbf{U}^{\mathrm{T}} \tag{A.2}$$

The weights w_i tell us the singularity, or *condition number* of the matrix. This is defined as the ratio of the largest weight to the smallest. A matrix is considered singular if its condition number is infinite, and *ill-conditioned* if the reciprocal of the condition

number approaches, or exceeds the machine's floating-point precision. Almost all of the covariance matrices found at ATLAS are ill-conditioned, with condition numbers around 10^{20} and reciprocals far less than the machine's floating-point precision for doubles at around 10^{-12}. Looking at the definition of the inverse (A.2), containing the reciprocals of the weights, it is clear that the smallest weights of these ill-conditioned matrices contribute most to the inverse. In fact, so much that any small numerical error in their calculation destabilizes the whole solution.

The choice of MeV and mm as units in ATLAS contributes a lot to the high condition numbers of the covariance matrices. Compared to using GeV and m, the ratio between the biggest and smallest variances is increased by twelve orders of magnitude. Half of this increase comes from the choice of units, the other half from squaring in the variances.

Since the inverse is made from linear combinations of the different weight reciprocals and their orthogonal matrices, it is possible to remove one, or more, combinations and still end up with a good approximation of the inverse. This is the strength of the singular value decomposition, simply remove combinations that do more harm than good. Any weights with a ratio less than 10^{-10} to the largest weight are set to zero, as are their reciprocals (A.2). By doing this approximation, we have given up one, or more, degrees of freedom to stabilize the inverse solution. The number of degrees of freedom of the inverse covariance matrix is no longer always 5, but is given by the number of non-zero weight reciprocals in $\mathbf{W}^{-1}$.

An extensive description of, and algorithm for doing the singular value decomposition is found in Ref. [24].

Bibliography

[1] The ATLAS Collaboration, G. Aad *et al.*, JINST **3** (2008) S08003.

[2] T. Cornelissen *et al.*, Concepts, Design and Implemenentation of the ATLAS New Tracking (NEWT), ATLAS Public Note ATL-SOFT-PUB-2007-007, 2007.

[3] E. Lund *et al.*, Track parameter propagation through the application of a new adaptive Runge-Kutta-Nyström method in the ATLAS experiment, ATLAS Public Note ATL-SOFT-PUB-2009-001, submitted to Journal of Instrumentation, 2009.

[4] E. Lund *et al.*, Transport of covariance matrices in the inhomogeneous magnetic field of the ATLAS experiment by the application of a semi-analytical method, ATLAS Public Note ATL-SOFT-PUB-2009-002, submitted to Journal of Instrumentation, 2009.

[5] E. Lund *et al.*, Treatment of energy loss and multiple scattering in the context of track parameter and covariance matrix propagation in continuous material in the ATLAS experiment, ATLAS Public Note ATL-SOFT-PUB-2008-003, 2008.

[6] R. Frühwirth, Nucl. Inst. and Meth. A **262** (1987) 444.

[7] E. Lund *et al.*, Refitting of combined inner detector and muon spectrometer tracks from Monte Carlo samples by using the Kalman fitter and the STEP algorithm in the ATLAS experiment, ATLAS Public Note ATL-SOFT-PUB-2008-005, 2008.

[8] S. Ahmed *et al.* (SNO Collaboration), Phys. Rev. Lett. **92** (2004).

[9] K. Eguchi *et al.* (KamLAND Collaboration), Phys. Rev. Lett. **90** (2003).

[10] S. Pokorski, Gauge Field Theories, (Cambridge University Press, Cambridge, 1987).

[11] C. Quigg, The Electroweak Theory, TASI lectures on Electroweak Theory, hp/0204104, 2002.

[12] The LEP Electroweak Working Group, The Electroweak Theory, http://lepewwg.web.cern.ch/LEPEWWG/plots/winter2007, 2007.

[13] Presentation of the 2004 Nobel prize in physics, Nobelprize.org, 2004.

[14] I. Gavrilenko, Description of Global Pattern Recognition Program (XKALMAN), ATLAS Note ATL-INDET-97-165, 1997.

[15] R. Clifft and A. Poppleton, IPATREC: Inner detector pattern-recognition and track-fitting, ATLAS Internal Note Soft-94-009, 1994.

[16] http://cern.ch/atlas-samusog/muonboy/Muonboy.htm.

[17] V. Boisvert *et al.*, Final Report of the ATLAS Reconstruction Task Force, ATLAS Note ATL-SOFT-2003-010, 2003.

[18] F. Åkesson *et al.*, The ATLAS Tracking Event Data Model, ATLAS Public Note ATL-SOFT-PUB-2006-004, 2006.

[19] A. Salzburger *et al.*, The ATLAS Tracking Geometry Description, ATLAS Public Note ATL-SOFT-PUB-2007-004, 2007.

[20] A. Salzburger, The ATLAS Track Extrapolation Package, ATLAS Public Note ATL-SOFT-PUB-2007-005, 2007.

[21] S. Agostinelli *et al.*, Nucl. Inst. and Meth. A **506** (2003) 250.

[22] http://www.hep.ucl.ac.uk/atlas/atlantis/.

[23] T. Cornelissen *et al.*, J. Phys.: Conf. Ser. **119** (2008).

[24] W. Press *et al.*, Numerical Recipes in C, (Cambridge University Press, Cambridge, 1999).

List of papers

This thesis is based on the following papers:

1. E. Lund, L. Bugge, I. Gavrilenko, A. Strandlie,
   Track parameter propagation through the application of a new adaptive Runge-Kutta-Nyström method in the ATLAS experiment,
   ATLAS Public Note ATL-SOFT-PUB-2009-001, submitted to Journal of Instrumentation, 2009.

2. E. Lund, L. Bugge, I. Gavrilenko, A. Strandlie,
   Transport of covariance matrices in the inhomogeneous magnetic field of the ATLAS experiment by the application of a semi-analytical method,
   ATLAS Public Note ATL-SOFT-PUB-2009-002, submitted to Journal of Instrumentation, 2009.

3. E. Lund, L. Bugge, E. W. Hughes, D. López Mateos, A. Salzburger, A. Strandlie,
   Treatment of energy loss and multiple scattering in the context of track parameter and covariance matrix propagation in continuous material in the ATLAS experiment,
   ATLAS Public Note ATL-SOFT-PUB-2008-003, 2008.

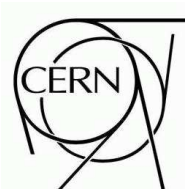
4. E. Lund, L. Bugge, T. G. Cornelissen, M. Elsing, I. Gavrilenko,
   W. Liebig, D. López Mateos, A. Salzburger, A. Strandlie, S. Todorova,
   Refitting of combined inner detector and muon spectrometer tracks from Monte Carlo samples by using the Kalman fitter and the STEP algorithm in the ATLAS experiment,
   ATLAS Public Note ATL-SOFT-PUB-2008-005, 2008.

I

ATLAS NOTE

ATL-SOFT-PUB-2009-001

January 23, 2009



Track parameter propagation through the application of a new adaptive Runge-Kutta-Nyström method in the ATLAS experiment

E. Lund*[,1], L. Bugge[1], I. Gavrilenko[2,3], A. Strandlie[1,4]

[1] *University of Oslo, Oslo, Norway*
[2] *CERN, Geneva, Switzerland*
[3] *P. N. Lebedev Institute of Physics, Moscow, Russia*
[4] *Gjøvik University College, Gjøvik, Norway*
* corresponding author (esben.lund@fys.uio.no)

Abstract

In this paper we study several fixed step and adaptive Runge-Kutta methods suitable for transporting track parameters through an inhomogeneous magnetic field. Moreover, we present a new adaptive Runge-Kutta-Nyström method which estimates the local error of the extrapolation without introducing extra stages to the original Runge-Kutta-Nyström method. Furthermore, these methods are compared for propagation accuracy and computing cost efficiency in the *simultaneous track and error propagation* (STEP) algorithm of the common ATLAS tracking software. The tests show the new adaptive Runge-Kutta-Nyström method to be the most computing cost efficient.

1 Introduction

Experimental particle physics is on the verge of a new era, heralded by the Large Hadron Collider being commissioned at the European Organization for Nuclear Research, CERN, located just outside Geneva, Switzerland. The LHC accelerator will collide protons at a center of mass energy of 14 TeV, opening up a new window for particle discoveries and precision measurements of existing theories. Particle detectors are located at four beam crossings along the LHC, one of which houses the ATLAS detector [1]. This is the largest of the LHC experiments, employing a great variety of detector and magnetic field technologies to identify a wide range of particles. The complex magnetic field and high collision rate, however, make the reconstruction of particle tracks very challenging. Things are further complicated by the relatively big amount of material within ATLAS, generating considerable disturbances to the particle tracks through material interactions such as energy loss and multiple scattering.

Particle trajectories are often parametrized with respect to surfaces crossed by the track. At ATLAS, there are five standard track parameters [2]; two related to the track position, another two giving the track direction, while the last is related to the particle charge and momentum. Track parameter propagation — which is the process of transporting track parameters and their associated covariance matrices through the magnetic field and material of the detector — is an important part of any track reconstruction algorithm, hence high accuracy and speed are essential to the propagation.

Historically, track parameter propagation has often been handled by the fixed step Runge-Kutta-Nyström method [3], which is designed to solve second-order differential equations, such as the equation of motion of a particle in a magnetic field. The basic idea of this method is to divide the integration interval into steps and solve these independently in an iterative procedure. The solution to each step is estimated by evaluating the equation of motion at different points — often referred to as *stages* — along the step. Fixed step methods are, however, not optimal in case of a very inhomogeneous magnetic field, such as in the ATLAS detector. To accomodate the most challenging parts of the ATLAS magnetic field, the fixed step size has to be small, thereby wasting time in the smooth parts of the field. Adaptive methods, such as the Runge-Kutta-Fehlberg method [4], solve this problem. These adaptive methods, however, require at least one extra stage in every step compared to the fixed step methods, but the amount of steps taken during the propagation is often reduced significantly.

In this paper, we study several adaptive Runge-Kutta methods along with the classical fixed step Runge-Kutta (RK) and Runge-Kutta-Nyström (RKN) methods. We also introduce a new adaptive Runge-Kutta-Nyström method, which combines the simplicity of the fixed step method with the efficiency of the adaptive methods without introducing extra stages. This is the method employed by the STEP algorithm, which is used for both stand-alone and combined muon reconstruction — along with fast track simulation in the muon spectrometer — within the ATLAS tracking framework [5].

In Section 2, we present the equation of motion along with some fixed step and adaptive Runge-Kutta methods for solving the track parameter propagation, followed by the new adaptive Runge-Kutta-Nyström method. In Section 3, we give a short presentation of the STEP algorithm before validating and comparing the Runge-Kutta methods in Section 4. Finally we present a brief conclusion in Section 5.

Natural units ($\hbar = c = 1$) are used throughout this paper.

2 Integrating the equation of motion

There are many representations of the equation of motion of a charged particle moving through a magnetic field [6], primarily separated by the free spatial parameter, which is mostly dictated by the geometry of the experiment. For a fixed target experiment, involving tracks going mainly along the beam axis, using a fixed Cartesian coordinate system and z as the free parameter is often satisfactory. In a beam collision, particles move outwards from the interaction point in all directions, facilitating r as the free

parameter, either in a fixed spherical or cylindrical coordinate system. However, the introduction of a magnetic field makes the integration along a fixed axis difficult and error prone. Using the arc length s and integrating along the path itself solves this problem.

Technically, the main difference between using z, r and s as the free parameter is the need to constantly check the distance to the destination surface when applying s as the free parameter. This is because the s propagation sets up a curvilinear coordinate system which follows the track at all times, whereas the z and r propagations use a static coordinate system defined by the detector geometry. The increased complexity and computing cost of using a curvilinear system is, however, outweighed by its robustness and ability to handle sharply bending tracks.

By using s as the free parameter, the equation of motion, given by the Lorentz force, becomes

$$\frac{d^2\boldsymbol{r}}{ds^2} = \frac{q}{p}\left(\frac{d\boldsymbol{r}}{ds} \times \boldsymbol{B}(\boldsymbol{r})\right) = \lambda(\boldsymbol{T} \times \boldsymbol{B}(\boldsymbol{r})) \tag{1}$$

where $\boldsymbol{T} = d\boldsymbol{r}/ds$ is the normalized tangent vector to the track, $\boldsymbol{B}(\boldsymbol{r})$ is the magnetic field and $\lambda \equiv q/p$. To see that $\boldsymbol{T}$ is the normalized tangent vector it is useful to look at the definition of the speed $d\boldsymbol{r}/dt = \boldsymbol{v}$, and the arc length $ds = vdt$. The derivative $d\boldsymbol{r}/ds$ then becomes $d\boldsymbol{r}/vdt = \boldsymbol{v}/v = \boldsymbol{T}$.

The electric field component of the Lorentz force is left out of the equation of motion because the electric field inside most of the detector is small compared to the magnetic field.

Given the inhomogeneous nature of the magnetic field, finding an analytical solution to the second-order differential equation of motion (1) becomes impossible. It has to be solved numerically. Many methods are available, but one group of methods proves especially useful in tracking. These are the recursive Runge-Kutta methods [7], most of which are developed for solving first-order differential equations, since higher order equations can always be reduced to sets of first-order equations by introducing auxiliary variables.

The basic idea of the recursive Runge-Kutta method is to divide the integration interval into steps and solve these independently in an iterative procedure. Every step becomes an initial value problem and can be solved as best suited for that particular part of the integration interval. This is especially useful when varying the step length to make the procedure adaptive. The solution to each step is estimated by evaluating the equation of motion at different points k — often referred to as *stages* — along the step. Every stage, except the first, is based on the previous stages of the step. In the end, all stages are weighted and summed to find the solution to the step.

Generally speaking, every evaluation of the equation cancels an error term, producing a solution correct to the order of the number of evaluations. The four stages of the original Runge-Kutta method produce a result correct to the fourth order. Using higher-order solutions gives better accuracy per step and allows longer steps. However, if the integration environment is not known to the same order as the Runge-Kutta method, the extra evaluations may be better spent on more steps. Six evaluations of the equation of motion in a given interval can produce one sixth-order step, two third-order steps or three second-order steps. Which gives the best overall accuracy is hard to tell without testing. Later in this paper, it is shown that the ATLAS magnetic field does not warrant the use of Runge-Kutta methods beyond fourth order.

2.1 The Runge-Kutta-Nyström method

These are the four stages and the solutions to the fourth-order Runge-Kutta-Nyström method;

$$\begin{aligned} k_1 &= f(x_n, y_n, y'_n) \\ k_2 &= f(x_n + \frac{h}{2}, y_n + \frac{h}{2}y'_n + \frac{h^2}{8}k_1, y'_n + \frac{h}{2}k_1) \end{aligned}$$

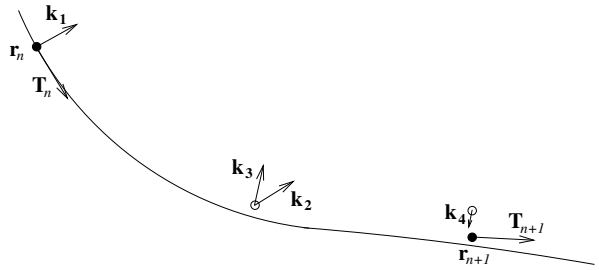


Figure 1: Integrating the equation of motion using the fourth-order Runge-Kutta-Nyström method. The equation of motion is evaluated four times in each step; once at the initial point, twice at a trial midpoint and once at a trial endpoint. From these stages $\boldsymbol{k}$, the position $\boldsymbol{r}_{n+1}$ and the normalized tangent vector to the track $\boldsymbol{T}_{n+1}$ are calculated.

$$\begin{aligned} k_3 &= f(x_n+\frac{h}{2}, y_n+\frac{h}{2}y'_n+\frac{h^2}{8}k_1, y'_n+\frac{h}{2}k_2) \\ k_4 &= f(x_n+h, y_n+hy'_n+\frac{h^2}{2}k_3, y'_n+hk_3) \\ y'_{n+1} &= y'_n+\frac{h}{6}(k_1+2k_2+2k_3+k_4) \\ y_{n+1} &= y_n+hy'_n+\frac{h^2}{6}(k_1+k_2+k_3) \end{aligned} \tag{2}$$

where h is the step length, x is the integration variable, n is the step number and f is the second-order differential equation

$$\frac{d^2y(x)}{dx^2} = f(x,y,y') \tag{3}$$

which is integrated twice; first to produce the y'_{n+1} solution, second to produce y_{n+1}. This method was specifically designed for solving second-order differential equations such as the equation of motion. Integrating the equation of motion (1), Fig. 1, the first integration gives the normalized tangent vector $\boldsymbol{T}_{n+1}$, while the second integration gives the position of the particle $\boldsymbol{r}_{n+1}$.

An appealing aspect of the Runge-Kutta-Nyström method is that stage two and three share the same point in space; $(x_n+\frac{h}{2}, y_n+\frac{h}{2}y'_n+\frac{h^2}{8}k_1)$, and hence magnetic field. To reduce calls to the magnetic field service further, it is possible to use the magnetic field found in the last stage of the current step as the starting field of the next step, because the position of the last stage $(x_n+h, y_n+hy'_n+\frac{h^2}{2}k_3)$ is very close to the final solution of the step $(x_n+h, y_n+hy'_n+\frac{h^2}{6}(k_1+k_2+k_3))$, i.e. the starting point of the next step. By using these optimizations, the Runge-Kutta-Nyström method only needs to evaluate the magnetic field twice per step.

It is, however, a problem of the Runge-Kutta-Nyström method that it only gives an idea of the quality of the integration, i.e. correct to the fourth order. The actual integration error of the step is unknown. Furthermore, the accuracy of the integration can only be adjusted by setting the fixed step length h, which has to accommodate the most inhomogeneous parts of the magnetic field. In ATLAS, this step length is a lot smaller than that needed for the more well-behaved parts of the field, making a fixed step length inefficient.

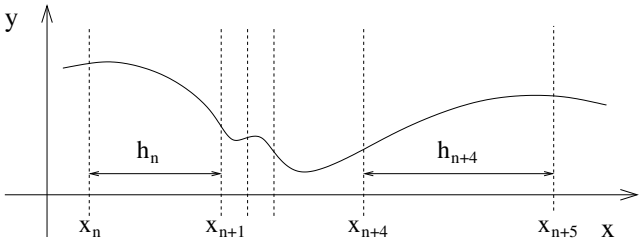


Figure 2: Illustration of the varying step size of the adaptive Runge-Kutta methods. The step size algorithm adjusts to the roughness of the function by using small steps in the rocky sections and longer steps in the smooth parts.

2.2 Adaptive Runge-Kutta methods

To avoid the problems of the fixed step length methods, and to adjust the accuracy in a predictable way, some adaptive Runge-Kutta method is needed. The most straightforward technique is step doubling. Every step is taken twice, first as a full step, and then as two half-steps producing a solution of higher order. The difference between these solutions gives an estimate of the *local error* ε of each step of the lower-order full step solution. The local error is checked against a user specified criterion, the *error tolerance* τ, and if the error is bigger than the error tolerance the step fails. It is then shortened and redone. If the local error is smaller than the error tolerance, the step is successful and the step size of the following step is increased to minimize the number of steps needed for the integration. A good step size selection algorithm provides *error tolerance proportionality*, which means that the logarithm of the global error of the whole integration is proportional to the logarithm of the user specified error tolerance. This is achieved by keeping the local error close to the error tolerance during the integration by adjusting the step size, Fig. 2. Managing this with a minimum number of failed steps is the main challenge of the step size selection algorithm.

2.3 Embedded Runge-Kutta pairs

An alternative way of estimating the local error of a step is based on the embedded Runge-Kutta formulae — the so-called embedded pairs — invented by Fehlberg [4]. The most well-known Runge-Kutta-Fehlberg method is a fifth-order method with six stages, where another combination of these six stages gives a fourth-order method. As in step doubling, the difference between these two solutions is used to estimate the local error of the lower-order solution of the step. Since the embedded formulae provide error estimation at almost no extra cost, many additional embedded pairs of varying orders have been created since Fehlberg's original work. Most of which contain only one extra stage compared to the classical Runge-Kutta methods of the same order. The Bogacki-Shampine 3(2)FSAL (BS32) embedded pair [8] provides an illustrative example;

$$
\begin{aligned}
k_1 &= hf(x_n, y_n) \\
k_2 &= hf(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}) \\
k_3 &= hf(x_n + \frac{3}{4}h, y_n + \frac{3}{4}k_2) \\
k_4 &= hf(x_n + h, y_n + \frac{2}{9}k_1 + \frac{1}{3}k_2 + \frac{4}{9}k_3)
\end{aligned}
$$

$$\begin{aligned} \text{3rd-order solution: } y_{n+1} &= y_n + \frac{2}{9}k_1 + \frac{1}{3}k_2 + \frac{4}{9}k_3 \\ \text{2nd-order solution: } \hat{y}_{n+1} &= y_n + \frac{7}{24}k_1 + \frac{1}{4}k_2 + \frac{1}{3}k_3 + \frac{1}{8}k_4 \end{aligned} \tag{4}$$

This is a "first-same-as-last" (FSAL) pair, meaning that the last stage k_4 is evaluated at the final higher-order solution of the step y_{n+1}. Since the next step starts out where the current ends, we get $k_1(n+1) = k_4(n)$, hence the name "first-same-as-last". This property reduces the number of evaluations of the differential equation, thereby increasing the efficiency. The 3(2) in the naming of the pair indicates the orders of the pair, the lower-order solution $\hat{y}_{n+1}$ being used for error estimation only.

Since the error estimate applies to the lower-order solution, the higher-order solution is often more accurate than the error estimate suggests. It has therefore become a standard procedure to use the higher-order solution along with the error estimate of the lower-order solution unless very precise error estimates are needed. This is the so-called local extrapolation, which is applied throughout this paper.

2.4 Adjusting the step size according to the error tolerance

As mentioned earlier, a step size selection algorithm using the local error estimate of each step

$$\varepsilon = y_{n+1} - \hat{y}_{n+1} \tag{5}$$

is needed to make the Runge-Kutta methods adaptive. The most common step size algorithm is

$$h_{n+1} = h_n \left(\frac{\tau}{|\varepsilon|} \right)^{\frac{1}{q+1}} \tag{6}$$

where the new step length h_{n+1} is given by the local error of the current step ε, the current step length h_n, the user specified error tolerance τ, and the order q of the lower-order solution $\hat{y}_{n+1}$. The derivation of this algorithm is left to Refs. [7,9]. The core of the algorithm is the fraction $\tau/|\varepsilon|$ which becomes less than one if the local error is bigger than the error tolerance, reducing the step length, and vice versa.

After "trimming" by a limitation criterion

$$\frac{1}{4}h_n \leq h_{n+1} \leq 4h_n \tag{7}$$

the new step size h_{n+1} is used for the next step, or for retrying the current step if it fails. The limitation criterion is introduced to prevent extreme changes to the step size during the integration. Without the limitation criterion, a tiny bump or smooth region in the integration interval may change the step length by a factor of hundred or more. This may compromise the stability of the solution and the error tolerance proportionality.

After finding the new step length (6) and trimming it (7), the quality of the current step is evaluated by an acceptance criterion

$$|\varepsilon| < 4\tau \tag{8}$$

to decide whether the step is accepted. The safety factor of 4 is introduced to allow for the local error to oscillate around the optimal value of $|\varepsilon| = \tau$.

Given the above step size algorithm (6) and acceptance criterion (8), the estimated global error of the integration becomes

$$g_\varepsilon = c\tau^a \tag{9}$$

For a given Runge-Kutta method and differential equation, c and a are constants. The logarithm of the global error g_ε — and hence integration accuracy — is therefore proportional to the logarithm of the user specified error tolerance. The error tolerance proportionality of the global error is very useful in the reconstruction software, allowing the same propagator to be used for a slow, accurate reconstruction, or a fast, less precise high-level trigger by only adjusting the error tolerance.

2.5 An adaptive Runge-Kutta-Nyström method

The fourth-order Runge-Kutta-Nyström method (2) is the simplest and most attractive of the Runge-Kutta methods discussed above. Estimating the local error of each step, enabling step size adjustment, is all that separates this fixed step method from the adaptive methods.

To estimate the local error of a step it is useful to look at the Taylor expansion around x_n that forms the basis of the Runge-Kutta methods;

$$g(x_n+h) = g(x_n) + hg'(x_n) + \frac{1}{2}h^2 g''(x_n) + \frac{1}{6}h^3 g^{(3)}(x_n) + \frac{1}{24}h^4 g^{(4)}(x_n) + \cdots \tag{10}$$

The main part of the local error is the truncation error from higher-order terms that are excluded, i.e. all terms above fourth order for the Runge-Kutta-Nyström method. This leaves out the possibility to calculate the local error directly, but it makes sense to assume that the last known term indicates how quickly the Taylor expansion converges. A small term indicates a quick convergence and small truncation (local) error, whereas a big term indicates the opposite. This assumption is confirmed by the widely used local error estimate (5) of the embedded pairs. For a 4(3) pair, the local error becomes

$$\begin{aligned} &\varepsilon = y_{n+1} - \hat{y}_{n+1} = g(x_n+h)_{\text{fourth order}} - g(x_n+h)_{\text{third order}} = \\ &g(x_n+h)_{\text{fourth-order term}} = \tfrac{1}{24}h^4 g^{(4)}(x_n) \end{aligned} \tag{11}$$

which is the last known term. Calculating this term directly from the four stages of the Runge-Kutta-Nyström method is not possible. However, since any error estimate along the step will suffice, we can calculate it in the middle of the step ($x_n + \frac{1}{2}h$), instead of at the beginning (x_n). This allows us to find the local error ε by using the symmetric derivative

$$g'(x_n) \approx \frac{g(x_n+h) - g(x_n-h)}{2h} \tag{12}$$

and the four stages given by the second-order derivatives k_l ($= g_l''$) already calculated by the Runge-Kutta-Nyström method.

First, we find the third-order symmetric derivative of the first half of the step by using a step size of $\frac{1}{4}h$, Fig. 3,

$$\begin{aligned} g^{(3)}(x_n + \frac{1}{4}h) &= \frac{g''(x_n + \frac{1}{4}h + \frac{1}{4}h) - g''(x_n + \frac{1}{4}h - \frac{1}{4}h)}{\frac{1}{2}h} = \\ \frac{g_2''(x_n + \frac{1}{2}h) - g_1''(x_n)}{\frac{1}{2}h} &= \frac{k_2 - k_1}{\frac{1}{2}h} \end{aligned} \tag{13}$$

Moreover, we calculate the third-order symmetric derivative of the last half of the step. Again, using a step size of $\frac{1}{4}h$,

$$\begin{aligned} g^{(3)}(x_n + \frac{3}{4}h) &= \frac{g''(x_n + \frac{3}{4}h + \frac{1}{4}h) - g''(x_n + \frac{3}{4}h - \frac{1}{4}h)}{\frac{1}{2}h} = \\ \frac{g_4''(x_n + h) - g_3''(x_n + \frac{1}{2}h)}{\frac{1}{2}h} &= \frac{k_4 - k_3}{\frac{1}{2}h} \end{aligned} \tag{14}$$

These two derivatives are then used to calculate the fourth-order symmetric derivative at the middle of the step

$$\begin{aligned} g^{(4)}(x_n + \frac{1}{2}h) &= \frac{g^{(3)}(x_n + \frac{1}{2}h + \frac{1}{4}h) - g^{(3)}(x_n + \frac{1}{2}h - \frac{1}{4}h)}{\frac{1}{2}h} = \\ \frac{g^{(3)}(x_n + \frac{3}{4}h) - g^{(3)}(x_n + \frac{1}{4}h)}{\frac{1}{2}h} &= \frac{\frac{k_4 - k_3}{\frac{1}{2}h} - \frac{k_2 - k_1}{\frac{1}{2}h}}{\frac{1}{2}h} = \frac{k_1 - k_2 - k_3 + k_4}{\frac{1}{4}h^2} \end{aligned} \tag{15}$$

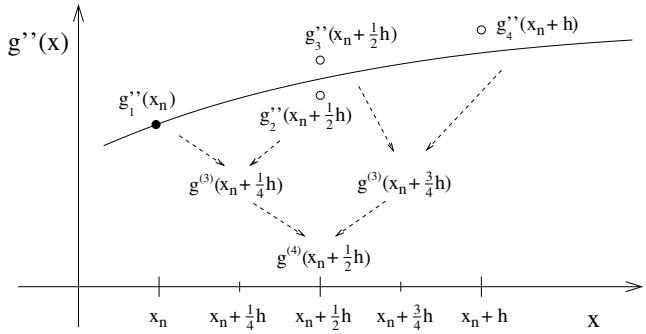


Figure 3: Using the four stages k_l ($= g_l''$) of the fourth-order Runge-Kutta-Nyström method to estimate the local error ε of a step. The four stages are used to produce the third-order symmetric derivatives at the intermediate points $(x_n + \frac{1}{4}h)$ and $(x_n + \frac{3}{4}h)$, which go into the calculation of the fourth-order symmetric derivative in $(x_n + \frac{1}{2}h)$, corresponding to the local error ε.

corresponding to the error estimate

$$\varepsilon = \frac{1}{24}h^4 g^{(4)}(x_n + \frac{1}{2}h) = \frac{1}{24}h^4 \frac{k_1 - k_2 - k_3 + k_4}{\frac{1}{4}h^2} = \frac{h^2}{6}(k_1 - k_2 - k_3 + k_4) \tag{16}$$

Since we are always looking at the local error estimate relative to some arbitrarily chosen error tolerance (6, 8), we can absorb the $1/6$ into the error tolerance, leaving only

$$\varepsilon = h^2(k_1 - k_2 - k_3 + k_4) \tag{17}$$

Whereas the embedded Runge-Kutta pairs introduce extra stages to estimate the local error, this approach only recycles the four Runge-Kutta-Nyström stages already known.

Equipped with the local error estimate (17), we can apply the step size algorithm (6) with $q = 3$ — assuming an error estimate correct to one order less than the Runge-Kutta-Nyström method — followed by the limitation criterion (7) and the acceptance criterion (8) to create an adaptive Runge-Kutta-Nyström method. In the following sections, we show that this method is very computing cost efficient, and hence has become the standard STEP method.

3 The STEP algorithm

The STEP algorithm transports the track parameters from one surface of the tracking geometry to another. To find the crossing point with the target surface, the algorithm starts off with finding the distance from the starting point to the surface in the direction of the momentum. If it is less than 10 cm the propagation is done in a single step, else the propagation starts off with a 10 cm step. The STEP algorithm then does the initial step applying the chosen Runge-Kutta method. From the step size and local error estimate of the first step, a new step size is calculated by the step size algorithm. This new step size is then "trimmed" according to the limitation criterion before a decision to go on, or retry the current step is made by the acceptance criterion.

The whole procedure, starting with finding the distance to the target surface is then repeated. This time, and for all the following steps, the distance is only used as a maximum limit to the new step size to avoid stepping through the target surface. When the distance to the target surface is down to 10 μm,

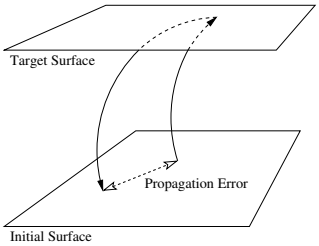


Figure 4: The definition of the propagation error in the test setup. Dividing the error by the total path length — back and forth — produces the relative propagation error.

or less, the Runge-Kutta propagation is stopped, leaving the remaining propagation to a simple Taylor expansion

$$\begin{aligned} \boldsymbol{r}_{n+1} &= \boldsymbol{r}_{\mathrm{RK}} + h\boldsymbol{r}'_{\mathrm{RK}} + \tfrac{1}{2}h^2\boldsymbol{r}''_{\mathrm{RK}} &&= \boldsymbol{r}_{\mathrm{RK}} + h(\tfrac{d\boldsymbol{r}}{ds})_{\mathrm{RK}} + \tfrac{1}{2}h^2(\tfrac{d^2\boldsymbol{r}}{ds^2})_{\mathrm{RK}} \\ \boldsymbol{T}_{n+1} &= \boldsymbol{T}_{\mathrm{RK}} + h\boldsymbol{T}'_{\mathrm{RK}} &&= (\tfrac{d\boldsymbol{r}}{ds})_{\mathrm{RK}} + h(\tfrac{d^2\boldsymbol{r}}{ds^2})_{\mathrm{RK}} \end{aligned} \tag{18}$$

where h is the remaining distance to the target surface and the derivatives are given by the equation of motion (1). The parameters indexed by RK are those found by the last step of the Runge-Kutta propagation.

Since the standard STEP algorithm uses an adaptive — instead of a fixed step — Runge-Kutta method, the initial step length is of less importance, and hence cannot be set by the user.

4 Validating the parameter propagation

To test the STEP algorithm and compare Runge-Kutta methods, we propagate the same set of 50 000 charged particles through the realistic ATLAS magnetic field by using the different methods. The particles are sent in random directions, covering all azimuthal and polar angles at momenta ranging from 500 MeV to 500 GeV, starting off from an initial surface located at the interaction point. The particles are then propagated towards a target surface randomly placed and rotated in a cube with sides of 20 m centered in the detector. Since the particles are sent in all directions, less than half of them hit their target surface, thereby triggering their propagation straight back to the initial surface, which they should ideally hit exactly where they originally started out. However, due to errors introduced by the propagation, the final position is slightly displaced, Fig. 4, giving the global propagation error of the track. Due to the great variation of the distance to the target surface, the global propagation error is divided by the total path length — back and forth — to produce a global propagation error per unit distance. To avoid smearing of the global propagation errors by the orientation of the initial surface, it is always set normal to the initial particle momentum. For simplicity, the global propagation error per unit distance is hereafter referred to as the *relative propagation error*.

This test does not represent typical ATLAS situations in the sense that the actual propagation distances are not randomly distributed in the detector volume; the majority being short range propagations in the inner detector. It is, however, sufficient for comparing and qualifying propagation algorithms.

4.1 Relative propagation errors

Figure 5 shows distributions of the logarithms of relative propagation errors at different fixed step lengths and error tolerances. Distributions are presented for the fixed step, step doubling and adaptive (STEP)

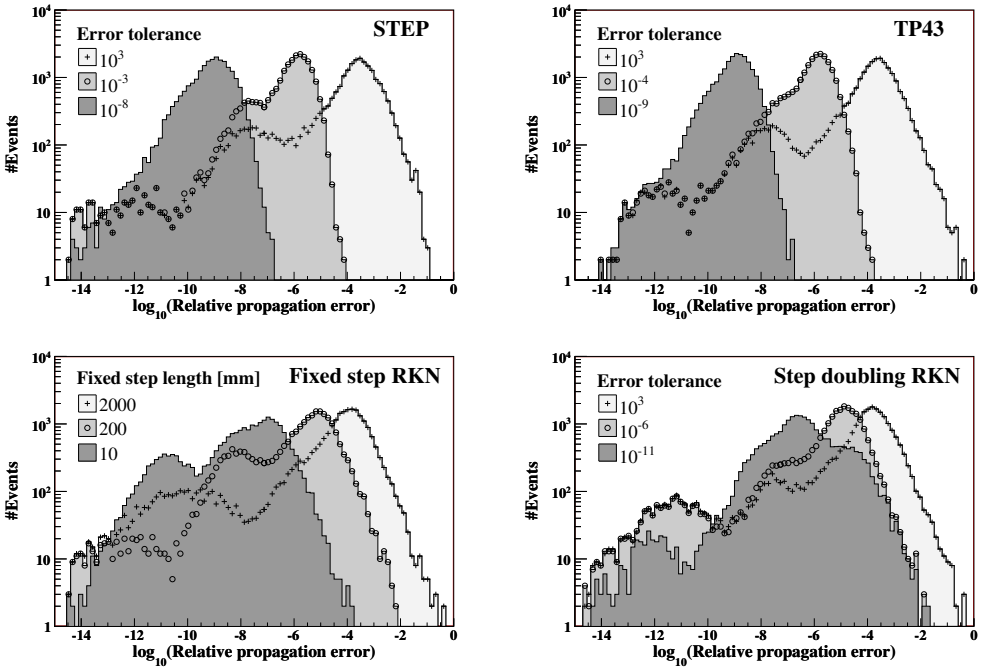


Figure 5: Distributions of the logarithms of relative propagation errors of the fixed step, step doubling and adaptive (STEP) Runge-Kutta-Nyström methods, and for the embedded pair TP43, for three fixed step lengths and error tolerances.

Runge-Kutta-Nyström methods, and for the embedded pair Tsitouras-Papakostas 4(3)FSAL (TP43) [9]. In this pair, the last two stages are both calculated at the full step length, making them closely positioned in space, almost sharing magnetic field values. Here, we use identical field values for both stages to increase the efficiency of the pair. The classical Runge-Kutta distributions (not shown) are similar to those of the Runge-Kutta-Nyström method, both in the fixed step and step doubling case.

The distributions of Fig. 5 show several peaks; one at -12 which is made up of very short tracks, typically 10 cm. Another at -8 which is made up of tracks within the ATLAS solenoid, and the main peak at -4 which contains the tracks moving into the ATLAS toroidal magnetic field, which is the most challenging part to maneuver accurately.

Figure 6 shows the mean relative propagation errors as a function of the error tolerance for the above-mentioned adaptive Runge-Kutta methods, along with three other embedded pairs; Cash-Karp 5(4) (CK54) [7], Dormand-Prince 5(4)FSAL (DP54) [10] and Tsitouras-Papakostas 7(5) (TP75) [9]. Similarly to the TP43 pair, the last two stages of DP54 are both calculated at the full step length, sharing the same magnetic field values to increase the efficiency of the pair. Furthermore, the fixed step methods are presented as functions of the step length. The "ATLAS RungeKutta" is another ATLAS propagator — originally part of the xKalman package [11] — while GEANT4 [12] is a simulation package developed for tracking particles through material.

As mentioned in Section 2.1, it is possible to use the same magnetic field values for calculating the last stage of the current step and the first stage of the next step in the Runge-Kutta-Nyström method because both stages are closely located in space. This optimization is implemented in the adaptive

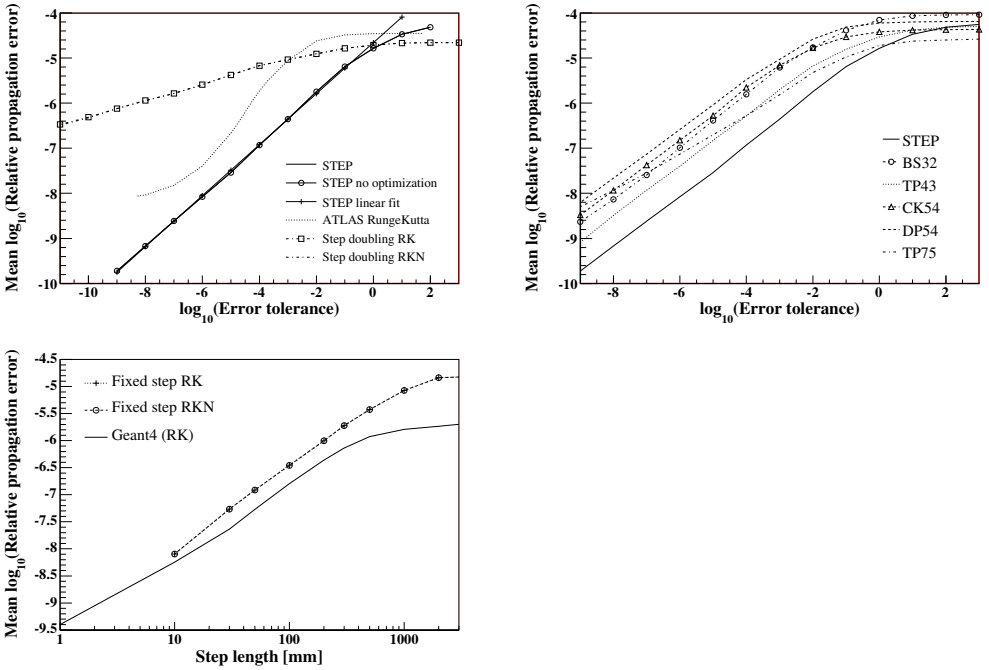


Figure 6: Mean values of the logarithms of relative propagation errors of some fixed step and adaptive Runge-Kutta methods.

Runge-Kutta-Nyström method employed by STEP, and it is validated by the "STEP no optimization" curve of the top left plot of Fig. 6.

In Fig. 6, we see that all of the methods reach a plateau at high error tolerance, which is the inherent accuracy of the method. On the plateau, the error control is practically disabled and the propagation goes "unchecked", similarly to the classical fixed step methods. Besides, the computing time also levels out, gaining nothing by going to very high error tolerances.

Surprisingly, the step doubling methods perform worse than the fixed step methods, showing little correlation between the error tolerance and the relative propagation error. This is because halving the step length does not necessarily increase the accuracy of the step sufficiently to give a good local error estimate. Moreover, in many cases the local error estimate of the step doubling is too low, giving an overly optimistic view of the quality of the propagation. This prevents the step size algorithm from working properly, making the steps too long to produce the required accuracy. These tracks show up as an extra shoulder on the right-hand side of the step doubling distribution of the lowest error tolerance of Fig. 5. Trials by using step tripling give a slightly better error estimate, but still worse than that of the other adaptive methods.

Figure 6 clearly shows the error tolerance proportionality of the relative propagation error. Doing a linear fit to the adaptive Runge-Kutta-Nyström method (STEP), Eq. (9) becomes

$$\text{mean relative propagation error} = 2.2 \times 10^{-5}\, \tau^{0.57} \tag{19}$$

This fit, shown in the top left plot of Fig. 6, is only valid when the error control is in full effect; at error tolerances below 10^{-1}. Using Eq. (19), the user can calculate the mean relative propagation error at a

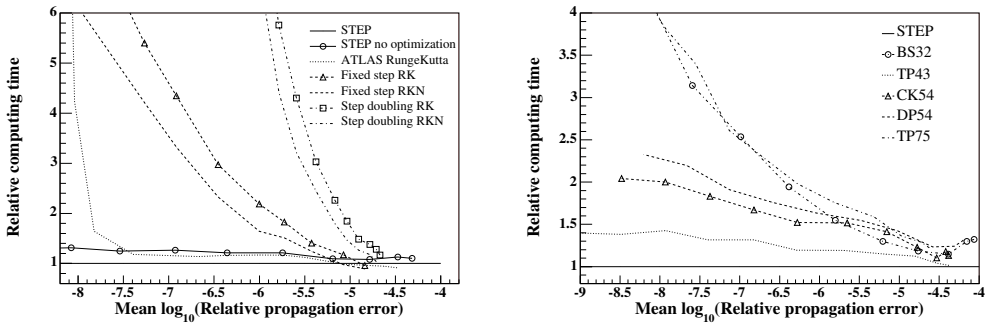


Figure 7: Computing cost efficiencies relative to the efficiency of the adaptive Runge-Kutta-Nyström method (STEP).

given error tolerance, or estimate the error tolerance needed to reach a given mean relative propagation error.

4.2 Efficiencies

All of the above-mentioned Runge-Kutta methods can produce a sufficient relative propagation error, even though the error tolerances needed to produce the desired accuracy differs. To compare methods, we need to study the computing cost of reaching some given accuracy; the so-called efficiency. Figure 7 shows efficiencies relative to the most efficient method, the new adaptive Runge-Kutta-Nyström method (STEP). From this figure, we see that the optimized STEP method is more efficient than the non-optimized STEP method. Moreover, we notice that the step doubling methods perform badly because every step is actually made up of three steps; two steps for the solution, and one for the error estimation, producing a significant overhead for failed steps.

Even though the new adaptive Runge-Kutta-Nyström method is the most efficient in ATLAS, the TP43 efficiency becomes comparable when the computing cost of polling the magnetic field is negligible, such as in the case of a parameterized magnetic field.

5 Conclusion and outlook

In this paper, we have presented a new adaptive Runge-Kutta-Nyström method developed for the STEP algorithm within the ATLAS tracking framework. The STEP algorithm transports track parameters and their associated covariance matrices through the ATLAS detector, taking the magnetic field and material interactions into account. A relatively extensive computing cost comparison with existing numerical methods — suited for parameter propagation — shows the new adaptive Runge-Kutta-Nyström method to be the most efficient, having potential beyond the STEP application presented here.

In addition to the track parameter propagation discussed here, the transport of the associated covariance matrix — the so-called error propagation — is necessary to provide a full description of the track. The covariance matrix provides information about the uncertainties related to the reconstruction of the track parameters from the empirical data, and its transport is treated in a separate paper [13].

6 Acknowledgements

This work has been carried out as part of the developments of the ATLAS tracking group. We would like to thank our colleagues for their help in integrating the software and for their support in preparing this note.

References

[1] The ATLAS Collaboration, G. Aad *et al.*, JINST **3** (2008) S08003.

[2] F. Åkesson *et al.*, The ATLAS Tracking Event Data Model, ATLAS Public Note ATL-SOFT-PUB-2006-004, 2006.

[3] E. J. Nyström, Acta Soc. Sci. Fenn. **50** (1925) 1–55.

[4] E. Fehlberg, Classical fifth-, sixth-, seventh-, and eighth-order Runge-Kutta formulas with stepsize control, NASA-TR-R-287, 1968.

[5] T. Cornelissen *et al.*, Concepts, Design and Implemenentation of the ATLAS New Tracking (NEWT), ATLAS Public Note ATL-SOFT-PUB-2007-007, 2007.

[6] L. Bugge and J. Myrheim, Nucl. Inst. and Meth. **179** (1981) 365.

[7] W. Press *et al.*, Numerical Recipes in C, (Cambridge University Press, Cambridge, 1999).

[8] P. Bogacki and L. F. Shampine, Appl. Math. Lett. **2** (1989) 321.

[9] C. Tsitouras and S. N. Papakostas, SIAM J. Sci. Comput. **20** (1999) 2067.

[10] J. R. Dormand and P. J. Prince, J. Comp. Appl. Math. **6** (1980) 19.

[11] I. Gavrilenko, Description of Global Pattern Recognition Program (XKALMAN), ATLAS Note ATL-INDET-97-165, 1997.

[12] S. Agostinelli *et al.*, Nucl. Inst. and Meth. A **506** (2003) 250.

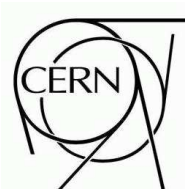
[13] E. Lund *et al.*, Transport of covariance matrices in the inhomogeneous magnetic field of the ATLAS experiment by the application of a semi-analytical method, ATLAS Public Note ATL-SOFT-PUB-2009-002, submitted to Journal of Instrumentation, 2009.

II

ATLAS NOTE

ATL-SOFT-PUB-2009-002

January 23, 2009



Transport of covariance matrices in the inhomogeneous magnetic field of the ATLAS experiment by the application of a semi-analytical method

E. Lund[*,1], L. Bugge[1], I. Gavrilenko[2,3], A. Strandlie[1,4]

[1] *University of Oslo, Oslo, Norway*
[2] *CERN, Geneva, Switzerland*
[3] *P. N. Lebedev Institute of Physics, Moscow, Russia*
[4] *Gjøvik University College, Gjøvik, Norway*
[*] corresponding author (esben.lund@fys.uio.no)

ATL-SOFT-PUB-2009-002
23 January 2009

Abstract

In this paper we study the transport of track parameter covariance matrices — the so-called error propagation — in the inhomogeneous magnetic field of the ATLAS experiment. The Jacobian elements are transported in parallel with the track parameters, avoiding the inherent need of any purely numerical scheme of propagating a set of auxiliary tracks. We evaluate the quality of the transported Jacobians by a very thorough, purely numerical approach of obtaining the same derivatives, providing a quantitative understanding of the effects of including gradients of energy loss and the magnetic field on the accuracy of the error propagation. Irrespective of the accuracy of the underlying track parameter propagation, the method of parallel integration of the derivatives is demonstrated to be significantly faster than even the simplest numerical scheme. The error propagation presented in this paper is part of the *simultaneous track and error propagation* (STEP) algorithm of the common ATLAS tracking software.

1 Introduction

Experimental particle physics is opening a new window for particle discoveries and precision measurements of existing theories by the startup of the Large Hadron Collider being commissioned at the European Organization for Nuclear Research — CERN — located just outside Geneva, Switzerland. The LHC accelerator will collide protons at a center of mass energy of 14 TeV at four beam crossings, one of which houses the ATLAS detector [1]. This is the largest of the LHC experiments, employing a great variety of detector technologies to identify and measure the properties of a wide range of particles. The complex magnetic field and big amount of material within the ATLAS detector, along with the high collision rate of the accelerator, make track reconstruction very challenging. Track parameter and the associated error propagation is at the heart of almost any reconstruction algorithm, hence good accuracy and high speed — along with the consideration of material effects, such as energy loss and multiple scattering — are essential to the ATLAS tracking algorithms, such as the *simultaneous track and error propagation* (STEP) algorithm presented here. This algorithm transports the track parameters and associated covariance matrices through the dense volumes of the simplified ATLAS material description — the so-called tracking geometry [2] — which approximates the material distribution of the ATLAS calorimeter and muon spectrometer by a set of blended dense volumes to speed up the tracking process. This paper describes the error propagation of the STEP algorithm, while the transport of the track parameters is found in Ref. [3].

Error propagation is usually handled in a purely analytical or numerical way. The first case is possible when the track model is explicitly given, thus allowing a direct derivation from the track model of the Jacobian needed to transport the covariance matrix. Unfortunately, explicit track models are limited to straight lines or helices, only useful in a vanishing or homogeneous magnetic field. Within the inhomogeneous ATLAS magnetic field, a numerical approach is necessary. The simplest numerical way of finding the derivatives of the transport Jacobian involves the propagation of one auxiliary track for every track parameter, which usually amounts to five additional tracks.

There is, however, a third alternative to the error propagation; the semi-analytical *Bugge-Myrheim method* [4]. This method propagates the transport Jacobian in parallel with the track parameters at little extra cost. Although this method has been known for many years, its accuracy and speed are not well documented in the scientific literature. In this paper we study the quality and speed of the Bugge-Myrheim method as a function of the accuracy of the underlying track parameter propagation. Furthermore, we look at the impact of the magnetic field and energy loss gradients on the accuracy and speed of the error propagation. We also show that the Bugge-Myrheim method is significantly faster than any purely numerical approach.

In Section 2 we describe the error propagation in general before going into detail on the Bugge-Myrheim method in Section 3. Furthermore, we look at the numerical error propagation — used for validating the semi-analytical error propagation — in Section 4. In Section 5 we compare the elements of the transport Jacobian obtained by the semi-analytical and the numerical error propagation. Moreover, in Section 6 we perform a statistical test of the semi-analytically transported covariance matrix. Finally, we present a short conclusion in Section 7.

Natural units ($\hbar = c = 1$) are used throughout this paper, and vectors and matrices are generally given in bold italic and bold capital letters, respectively.

2 Error propagation

The track parameters are often reconstructed from empirical data with associated uncertainties introduced by noise from the material interactions during the parameter transport, and uncertainties related to the misalignment and limited resolution of the detector. Here we focus on transporting the intrinsic

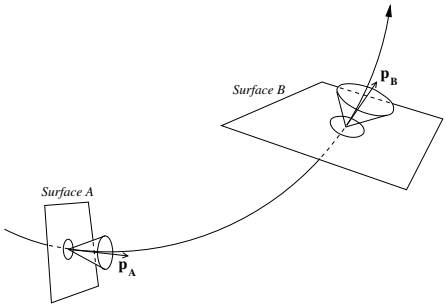


Figure 1: Transporting the track parameters and the associated uncertainties (covariance matrix) — indicated by the ellipses and cones — from one surface to another. The uncertainty of the momentum magnitude is omitted in this illustration.

measurement errors arising from the limited detector resolution, Fig. 1, given by the symmetric 5×5 covariance matrix with entry ij;

$$\Sigma_{ij} = \langle (\xi_i - \langle \xi_i \rangle)(\xi_j - \langle \xi_j \rangle) \rangle \tag{1}$$

where $\boldsymbol{\xi}$ is a vector of the *local track parameters*

$$\boldsymbol{\xi} = \begin{bmatrix} l_0 \\ l_1 \\ \phi \\ \theta \\ \lambda \end{bmatrix} \tag{2}$$

and $\langle \xi_i \rangle$ are the expectation values of these parameters. The local track parameters [5] are defined by the local track position at a surface (l_0, l_1), the direction of the track momentum in the global ATLAS coordinate system (ϕ, θ) and the signed inverse of the momentum ($\lambda \equiv q/p$). The global Cartesian, right-handed ATLAS coordinate system is approximately given by the LHC tunnel centre (x), the earth's surface (y) and the LHC beam pipe (z). The spherical polar coordinates ϕ and θ are defined as follows within this coordinate system; the azimuthal angle ϕ is given by the opening between the projection of the momentum into the x-y plane, and the x-axis, while the polar angle θ is given by the opening between the momentum and the z-axis. Together they define the direction of the momentum in the ATLAS coordinate system unambiguously, giving the following relations between the momentum components (p_x, p_y, p_z) and the angles (ϕ, θ);

$$\begin{aligned} p_x &= p \cos\phi \sin\theta \\ p_y &= p \sin\phi \sin\theta \\ p_z &= p \cos\theta \end{aligned} \tag{3}$$

The *global track parameters* used within STEP — to accommodate the propagation of bending tracks — are given in the above-mentioned global ATLAS coordinate system by

$$(x, y, z, T^x, T^y, T^z, \lambda)$$

where $\boldsymbol{T} = \boldsymbol{p}/p$ is the normalized tangent vector to the track, and (x, y, z) is the track position.

If the track model is explicitly given, the common approach to transporting the covariance matrix is to expand the analytical parameter propagation functions to first order in a Taylor series, and use these derivatives to propagate the covariance matrix in an approximate way. This is called *linear error propagation*. The availability of the derivatives of the propagated track parameters with respect to those at the starting point of the propagation — the so-called *Jacobian* $\mathbf{J}$ — is therefore essential to the linear error propagation. In our case, the Jacobian becomes a 5×5 matrix

$$\mathbf{J} = \begin{bmatrix} \frac{\partial l_0^{\text{final}}}{\partial l_0^{\text{initial}}} & \cdots & \frac{\partial l_0^{\text{final}}}{\partial \lambda^{\text{initial}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \lambda^{\text{final}}}{\partial l_0^{\text{initial}}} & \cdots & \frac{\partial \lambda^{\text{final}}}{\partial \lambda^{\text{initial}}} \end{bmatrix} \tag{4}$$

and transporting the symmetric covariance matrix by linear error propagation simply becomes a similarity transformation

$$\mathbf{\Sigma}_{\text{final}} = \mathbf{J} \cdot \mathbf{\Sigma}_{\text{initial}} \cdot \mathbf{J}^{\text{T}} \tag{5}$$

However, because of the inhomogeneous magnetic field of ATLAS, and the resulting lack of explicit analytical functions for the propagation of the track parameters, the Jacobian cannot be calculated directly. The linear error propagation (5) is still valid, but the Jacobian must be obtained in another way. This is done in three steps; first, we find the Jacobian required for transforming the covariance matrix from the initial local track parameters to the initial global track parameters used within the STEP algorithm. Furthermore, this Jacobian is transported along with the track parameters to the destination surface. Finally, it is multiplied by the Jacobian which transforms the covariance matrix to the local track parameters at the destination surface. The resulting Jacobian (4) transports the covariance matrix from one set of local track parameters at the initial surface to another set of local track parameters at the destination surface (5). The initial and destination surfaces can be picked independently from any of the five surfaces defined within the ATLAS event data model, and positioned arbitrarily. In this paper we focus only on the second of the three steps; the transport of the Jacobian along the track.

3 Semi-analytical error propagation by using the Bugge-Myrheim method

As mentioned above, the common way of obtaining the derivatives (Jacobian) needed for the linear error propagation is to expand the parameter propagation functions to first order in a Taylor series. The lack of analytical parameter propagation functions in this case unfortunately makes this approach impossible. Another common technique is the numerical error propagation described in Section 4. This method is slow, but robust and accurate, making it useful for testing the error propagation. A third way of obtaining the Jacobian is to differentiate the recursion formulae of the numerical integration method directly. This is the essence of the Bugge-Myrheim method. For reasons of efficiency and consistency, the natural choice is to pick the same integration method as used in the STEP parameter propagation, which is the adaptive Runge-Kutta-Nyström method. The Bugge-Myrheim method, however, follows the same principles regardless of the chosen integration method, only the recursion formulae change.

The basic idea of the adaptive Runge-Kutta-Nyström method is to divide the integration interval into steps and solve each step independently in an iterative procedure. Every step becomes an initial value problem and can be solved as best suited for that particular part of the integration interval. This is especially useful when varying the step length h to make the procedure adaptive. The solution of every step is estimated by evaluating the equation of motion u'' at four different points — often referred to as *stages* — along the step. Every stage, except the first, is based on the previous stages of the step. In the end, all stages are weighted and summed to find the solution to the step.

In the parameter propagation, we find the propagated global track parameters — where Λ is the integrated change in λ, or the total energy loss — of the equation of motion

$$\boldsymbol{u} = \begin{bmatrix} x \\ y \\ z \\ \Lambda \end{bmatrix}, \qquad \boldsymbol{u}' = \frac{d\boldsymbol{u}}{ds} = \begin{bmatrix} T^x \\ T^y \\ T^z \\ \lambda \end{bmatrix} \tag{6}$$

by integrating their respective differential equations by using some recursion formulae F and G. Given the Runge-Kutta-Nyström method, one step (numbered by n) becomes

$$\begin{aligned} \boldsymbol{u}_{n+1} &= F(s_n, \boldsymbol{u}_n, \boldsymbol{u}'_n) = \boldsymbol{F}_n(\boldsymbol{u}_n, \boldsymbol{u}'_n) = \boldsymbol{u}_n + h\boldsymbol{u}'_n + \frac{h^2}{6}(\boldsymbol{u}''_1 + \boldsymbol{u}''_2 + \boldsymbol{u}''_3) \\ \boldsymbol{u}'_{n+1} &= G(s_n, \boldsymbol{u}_n, \boldsymbol{u}'_n) = \boldsymbol{G}_n(\boldsymbol{u}_n, \boldsymbol{u}'_n) = \boldsymbol{u}'_n + \frac{h}{6}(\boldsymbol{u}''_1 + 2\boldsymbol{u}''_2 + 2\boldsymbol{u}''_3 + \boldsymbol{u}''_4) \end{aligned} \tag{7}$$

To obtain the derivatives (Jacobian) of the propagated global track parameters with respect to the initial local track parameters (i denoting initial)

$$\boldsymbol{\xi}^i = \begin{bmatrix} l_0^i \\ l_1^i \\ \phi^i \\ \theta^i \\ \lambda^i \end{bmatrix} \tag{8}$$

the recursion formulae (7) have to be differentiated with respect to $\boldsymbol{\xi}^i$, giving

$$\mathbf{J}_{n+1} = \begin{bmatrix} \frac{\partial \boldsymbol{u}_{n+1}}{\partial \boldsymbol{\xi}^i} \\ \frac{\partial \boldsymbol{u}'_{n+1}}{\partial \boldsymbol{\xi}^i} \end{bmatrix} = \begin{bmatrix} \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{\xi}^i} \\ \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{\xi}^i} \end{bmatrix} = \begin{bmatrix} \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}_n} & \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}'_n} \\ \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}_n} & \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}'_n} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial \boldsymbol{u}_n}{\partial \boldsymbol{\xi}^i} \\ \frac{\partial \boldsymbol{u}'_n}{\partial \boldsymbol{\xi}^i} \end{bmatrix} = \mathbf{D}_n \cdot \mathbf{J}_n \tag{9}$$

where the derivatives $\partial \boldsymbol{u}_n / \partial \boldsymbol{\xi}^i$ and $\partial \boldsymbol{u}'_n / \partial \boldsymbol{\xi}^i$ of the 8×5 Jacobian $\mathbf{J}$ are given by the 4×5 matrices

$$\frac{\partial \boldsymbol{u}_n}{\partial \boldsymbol{\xi}^i} = \begin{bmatrix} \frac{\partial x_n}{\partial l_0^i} & \cdots & \frac{\partial x_n}{\partial \lambda^i} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Lambda_n}{\partial l_0^i} & \cdots & \frac{\partial \Lambda_n}{\partial \lambda^i} \end{bmatrix}, \quad \frac{\partial \boldsymbol{u}'_n}{\partial \boldsymbol{\xi}^i} = \begin{bmatrix} \frac{\partial T_n^x}{\partial l_0^i} & \cdots & \frac{\partial T_n^x}{\partial \lambda^i} \\ \vdots & \ddots & \vdots \\ \frac{\partial \lambda_n}{\partial l_0^i} & \cdots & \frac{\partial \lambda_n}{\partial \lambda^i} \end{bmatrix} \tag{10}$$

$\mathbf{D}_n$ is an 8×8 matrix containing the recursion formulae $\boldsymbol{F}_n$ and $\boldsymbol{G}_n$ differentiated with respect to the global track parameters

$$\mathbf{D}_n = \frac{\partial(\boldsymbol{F}_n, \boldsymbol{G}_n)}{\partial(\boldsymbol{u}_n, \boldsymbol{u}'_n)} = \begin{bmatrix} \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}_n} & \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}'_n} \\ \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}_n} & \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}'_n} \end{bmatrix} \tag{11}$$

giving the 4×4 matrices

$$\frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}_n} = \begin{bmatrix} \frac{\partial F_n^x}{\partial x_n} & \cdots & \frac{\partial F_n^x}{\partial \Lambda_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n^\Lambda}{\partial x_n} & \cdots & \frac{\partial F_n^\Lambda}{\partial \Lambda_n} \end{bmatrix}, \quad \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}'_n} = \begin{bmatrix} \frac{\partial F_n^x}{\partial T_n^x} & \cdots & \frac{\partial F_n^x}{\partial \lambda_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n^\Lambda}{\partial T_n^x} & \cdots & \frac{\partial F_n^\Lambda}{\partial \lambda_n} \end{bmatrix} \tag{12}$$

and

$$\frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}_n} = \begin{bmatrix} \frac{\partial G_n^{T^x}}{\partial x_n} & \cdots & \frac{\partial G_n^{T^x}}{\partial \Lambda_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial G_n^{\lambda}}{\partial x_n} & \cdots & \frac{\partial G_n^{\lambda}}{\partial \Lambda_n} \end{bmatrix}, \quad \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}'_n} = \begin{bmatrix} \frac{\partial G_n^{T^x}}{\partial T_n^x} & \cdots & \frac{\partial G_n^{T^x}}{\partial \lambda_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial G_n^{\lambda}}{\partial T_n^x} & \cdots & \frac{\partial G_n^{\lambda}}{\partial \lambda_n} \end{bmatrix} \tag{13}$$

By writing the recursion formulae of the derivatives as a product of $\mathbf{D}_n$ and $\mathbf{J}_n$ (9), we can differentiate the recursion formulae $\boldsymbol{F}_n$ and $\boldsymbol{G}_n$ with respect to the global track parameters $\boldsymbol{u}_n$ and $\boldsymbol{u}'_n$ instead of the initial local track parameters $\boldsymbol{\xi}^i$. This simplifies the differentiation a lot, giving

$$\begin{aligned} \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}_n} &= 1 + \frac{h^2}{6}\left(\frac{\partial \boldsymbol{u}''_1}{\partial \boldsymbol{u}_n} + \frac{\partial \boldsymbol{u}''_2}{\partial \boldsymbol{u}_n} + \frac{\partial \boldsymbol{u}''_3}{\partial \boldsymbol{u}_n}\right) \\ \frac{\partial \boldsymbol{F}_n}{\partial \boldsymbol{u}'_n} &= h + \frac{h^2}{6}\left(\frac{\partial \boldsymbol{u}''_1}{\partial \boldsymbol{u}'_n} + \frac{\partial \boldsymbol{u}''_2}{\partial \boldsymbol{u}'_n} + \frac{\partial \boldsymbol{u}''_3}{\partial \boldsymbol{u}'_n}\right) \\ \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}_n} &= \frac{h}{6}\left(\frac{\partial \boldsymbol{u}''_1}{\partial \boldsymbol{u}_n} + 2\frac{\partial \boldsymbol{u}''_2}{\partial \boldsymbol{u}_n} + 2\frac{\partial \boldsymbol{u}''_3}{\partial \boldsymbol{u}_n} + \frac{\partial \boldsymbol{u}''_4}{\partial \boldsymbol{u}_n}\right) \\ \frac{\partial \boldsymbol{G}_n}{\partial \boldsymbol{u}'_n} &= 1 + \frac{h}{6}\left(\frac{\partial \boldsymbol{u}''_1}{\partial \boldsymbol{u}'_n} + 2\frac{\partial \boldsymbol{u}''_2}{\partial \boldsymbol{u}'_n} + 2\frac{\partial \boldsymbol{u}''_3}{\partial \boldsymbol{u}'_n} + \frac{\partial \boldsymbol{u}''_4}{\partial \boldsymbol{u}'_n}\right) \end{aligned} \tag{14}$$

To calculate these derivatives explicitly, we need to differentiate the individual stages of the Runge-Kutta-Nyström method with respect to the global track parameters,

$$\mathbf{A}_k = \frac{\partial \boldsymbol{u}''_k}{\partial \boldsymbol{u}'_n}, \qquad \mathbf{C}_k = \frac{\partial \boldsymbol{u}''_k}{\partial \boldsymbol{u}_n} \tag{15}$$

where k denotes the individual stages, and $\boldsymbol{u}''_k$ is given — in a general form — by the equations of motion of the global track parameters [4]

$$\begin{aligned} x'' &= \lambda(T^y B_z - T^z B_y) \\ y'' &= \lambda(T^z B_x - T^x B_z) \\ z'' &= \lambda(T^x B_y - T^y B_x) \\ \Lambda'' &= -\frac{\lambda^3 g E}{q^2} \end{aligned} \tag{16}$$

The last equation handles the energy loss, with E being the energy and g the energy loss per unit distance. The energy loss and its gradient varies little within each recursion step, hence the values calculated in the first stage are recycled by the following stages. This lowers the computing cost considerably.

Writing the 4×4 $\mathbf{A}_k$ and $\mathbf{C}_k$ matrices in a general form, we get

$$\mathbf{A} = \begin{bmatrix} \frac{\partial x''}{\partial T^x} & \cdots & \frac{\partial x''}{\partial \lambda} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Lambda''}{\partial T^x} & \cdots & \frac{\partial \Lambda''}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} 0 & \lambda B_z & -\lambda B_y & T^y B_z - T^z B_y \\ -\lambda B_z & 0 & \lambda B_x & T^z B_x - T^x B_z \\ \lambda B_y & -\lambda B_x & 0 & T^x B_y - T^y B_x \\ 0 & 0 & 0 & \left(\frac{1}{\lambda}(3 - \frac{p^2}{E^2}) + \frac{1}{g}\frac{\partial g}{\partial \lambda}\right)\Lambda'' \end{bmatrix} \tag{17}$$

and

$$\mathbf{C} = \begin{bmatrix} \frac{\partial x''}{\partial x} & \cdots & \frac{\partial x''}{\partial \Lambda} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Lambda''}{\partial x} & \cdots & \frac{\partial \Lambda''}{\partial \Lambda} \end{bmatrix} = \begin{bmatrix} \lambda(T^y \frac{\partial B_z}{\partial x} - T^z \frac{\partial B_y}{\partial x}) & \lambda(T^y \frac{\partial B_z}{\partial y} - T^z \frac{\partial B_y}{\partial y}) & \lambda(T^y \frac{\partial B_z}{\partial z} - T^z \frac{\partial B_y}{\partial z}) & 0 \\ \lambda(T^z \frac{\partial B_x}{\partial x} - T^x \frac{\partial B_z}{\partial x}) & \lambda(T^z \frac{\partial B_x}{\partial y} - T^x \frac{\partial B_z}{\partial y}) & \lambda(T^z \frac{\partial B_x}{\partial z} - T^x \frac{\partial B_z}{\partial z}) & 0 \\ \lambda(T^x \frac{\partial B_y}{\partial x} - T^y \frac{\partial B_x}{\partial x}) & \lambda(T^x \frac{\partial B_y}{\partial y} - T^y \frac{\partial B_x}{\partial y}) & \lambda(T^x \frac{\partial B_y}{\partial z} - T^y \frac{\partial B_x}{\partial z}) & 0 \\ -\frac{\lambda^3 E}{q^2}\frac{\partial g}{\partial x} & -\frac{\lambda^3 E}{q^2}\frac{\partial g}{\partial y} & -\frac{\lambda^3 E}{q^2}\frac{\partial g}{\partial z} & 0 \end{bmatrix} \tag{18}$$

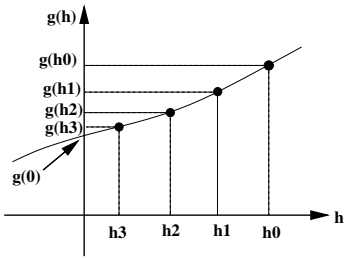


Figure 2: Schematic plot of the Ridders algorithm. The graph shows the parameterization of the symmetric derivative $g(h_i)$.

With the help of these matrices, we find the elements of $\mathbf{D}_n$, which is multiplied by $\mathbf{J}_n$ to produce the transported Jacobian $\mathbf{J}_{n+1}$ (9). This procedure is repeated for every recursion step, transforming $\mathbf{J}$ along the way.

When applied to real problems, the gradients of $\mathbf{A}$ and $\mathbf{C}$ are usually quite costly to calculate, hence it is common practice to set all of the $\partial g/\partial\lambda$, $\partial B_i/\partial x_j$ and $\partial g/\partial x_j$ gradients, i and j indicating the x, y and z components, to zero. This is, however, only correct for the material gradients $\partial g/\partial x_j$ of the blended dense volumes of the simplified ATLAS material description.

4 Numerical error propagation

To test the semi-analytical error propagation, we need an alternative way of calculating the derivatives of the Jacobian (4). The most straightforward way is by using the definition of the numerical derivative

$$f'(\xi_i) \approx \frac{f(\xi_i + h_i) - f(\xi_i)}{h_i} \tag{19}$$

where $f(\xi_i)$ propagates the local track parameters — denoted by i — from the initial surface to the target surface, while h_i is kept sufficiently small, ideally zero. By using the above definition of the derivative, we vary the initial local track parameters by a small amount h_i, one at a time. This is the key to knowing exactly how these variations translate to the final local track parameters. Registering the changes to the final parameters gives us the 25 derivatives of the Jacobian.

Though very easy and straightforward, this method is quite inaccurate. One way of increasing the accuracy is by using the symmetric derivative

$$g(h_i) \approx \frac{f(\xi_i + h_i) - f(\xi_i - h_i)}{2h_i} \tag{20}$$

which typically has a fractional error two orders of magnitude better than the original definition of the derivative [6].

To further increase the accuracy, we use a numerical method called Ridders' algorithm [6]. The essence of Ridders' algorithm is to parameterize the symmetric derivative as a function of h_i alone by calculating it for descending values of h_i, Fig. 2. This parameterization of $g(h_i)$ is used to estimate the derivative in the limit $h_i \to 0$. Since it has to be done for every derivative, this method is very time consuming and only useful for testing.

Compared to the semi-analytical error propagation even the simplest numerical error propagation is slow, needing at least five additional parameter propagations for every track, increasing the computing time accordingly.

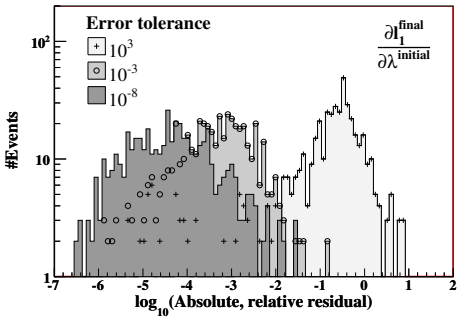


Figure 3: Logarithms of the absolute, relative residuals of the $\partial I_1^f / \partial \lambda^i$ Jacobian term in an inhomogeneous magnetic field with energy loss. The semi-analytical derivatives are calculated at three error tolerances with both gradients included, whereas the numerical derivatives are all calculated at a tolerance of 10^{-8}.

5 Validating the Jacobian in an inhomogeneous magnetic field including energy loss

To get a complete understanding of the semi-analytical error propagation, we need to study the Jacobian terms in a realistic, inhomogeneous magnetic field with energy loss. The test setup involves propagating muons — through solid Silicon in the realistic ATLAS magnetic field — in random directions, covering all azimuthal and polar angles at momenta ranging from 500 MeV to 500 GeV, starting off from an initial surface located at the interaction point of the ATLAS detector. The particles are propagated towards a target surface randomly placed and rotated in a cube with sides of 20 m centered in the detector. During this test, the derivatives required by the error propagation are calculated twice; first semi-analytically by the Bugge-Myrheim method, and then numerically by the Ridders algorithm. The numerical derivatives define the baseline for the semi-analytical terms. To assure the quality of the numerical derivatives, the STEP propagator at an error tolerance of 10^{-8} is used for calculating the symmetric derivatives of the Ridders algorithm. The error tolerance is a user specified number steering the accuracy of the propagation, a low tolerance giving a high accuracy, and vice versa. The *absolute, relative residual*

$$\frac{|\text{semi-analytical derivative} - \text{numerical derivative}|}{|\text{numerical derivative}|} \tag{21}$$

is then used to compare the 25 derivatives of the semi-analytical and numerical Jacobians.

Figure 3 shows three histograms of the logarithm of the absolute, relative residuals of the Jacobian element $\partial I_1^f / \partial \lambda^i$, f and i indicating the final and initial values. These histograms are typical of all the $\partial I_0^f / \partial \boldsymbol{\xi}^i$ and $\partial I_1^f / \partial \boldsymbol{\xi}^i$ Jacobian elements.

Figure 4 shows the effect on the residuals of two Jacobian terms by only including one type of gradient into the calculation of the semi-analytical derivatives. These terms are only sensitive to either the magnetic field gradients or the energy loss gradient. Due to the underlying single precision of the analysis program used to produce the plots (ROOT [7]), no relative difference better than approximately 10^{-7} is seen. Entries with better relative precision become identically zero and are not shown.

Figure 5 shows the mean values of residuals of a selection of Jacobian terms with and without both gradients included. All of the semi-analytical derivatives are sensitive to the gradients, especially the angular derivatives.

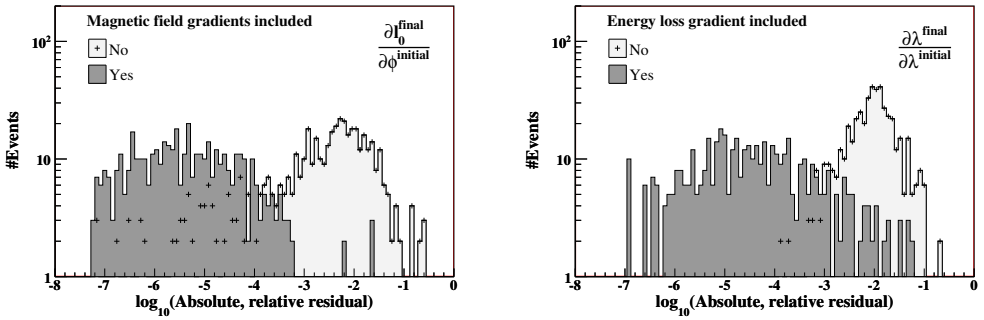


Figure 4: Logarithms of the absolute, relative residuals of two Jacobian terms in an inhomogeneous magnetic field with energy loss. The semi-analytical derivatives are calculated with, and without including the magnetic field gradients $\partial B_i/\partial x_j$ (left) and the energy loss gradient $\partial g/\partial \lambda$ (right). Both the semi-analytical and numerical derivatives are calculated at an error tolerance of 10^{-8}.

The improvements in the residuals when turning on the magnetic field gradients are presented on the left-hand side of Fig. 6, while the additional improvements by including the energy loss gradient are shown to the right. The effect of the energy loss gradient is only seen in the last column of the Jacobian, $\partial \boldsymbol{\xi}^f/\partial \lambda^i$, illustrated by the constant $\partial l_0^f/\partial \phi^i$ residual in the right-hand plot of Fig. 6, whereas the effects of the magnetic field gradients show up all over the Jacobian, except in the last row $\partial \lambda^f/\partial \boldsymbol{\xi}^i$, as illustrated by the constant $\partial \lambda^f/\partial \lambda^i$ residual in the left-hand plot of the same figure.

Figure 7 shows the additional computing time — relative to the STEP parameter propagation — spent by the semi-analytical error propagation, magnetic field and energy loss gradients. The error propagation is only done after the adaptive parameter propagation has found the optimal step length, making the nominal computing cost of the error propagation relatively stable over the whole error tolerance range. Thus, the drop in the additional computing cost of the error propagation at low error tolerances is mostly caused by the increased computing cost of the parameter propagation.

6 Verifying the propagated covariance matrix in an inhomogeneous magnetic field including energy loss

In the previous sections we have looked at the individual Jacobian elements to get a deeper understanding of the error propagation. Now, we examine the final covariance matrix produced by the linear error propagation (5). From this transformation, we see that the elements of the final covariance matrix are sums and products of many initial covariance and Jacobian terms. Evaluating the final covariance elements on an individual basis becomes prohibitively difficult, yet testing the Jacobian alone is not sufficient to guarantee the quality of the error propagation. Only a full error propagation, using a realistic initial covariance matrix, allows us to test the significance of the missing, or inaccurate Jacobian elements, and the gradients. To perform this test, we use the fact that the initial covariance matrix defines the Gaussian variances and correlations of the initial track parameters. By varying the initial track parameters according to their associated covariance matrix before propagating them to the target surface, the variation of the final track parameters should be reflected in the final covariance matrix. In short, we use the initial covariance matrix for smearing the simulated tracks and the final track parameters for statistically testing the propagated covariance matrix, Fig. 8.

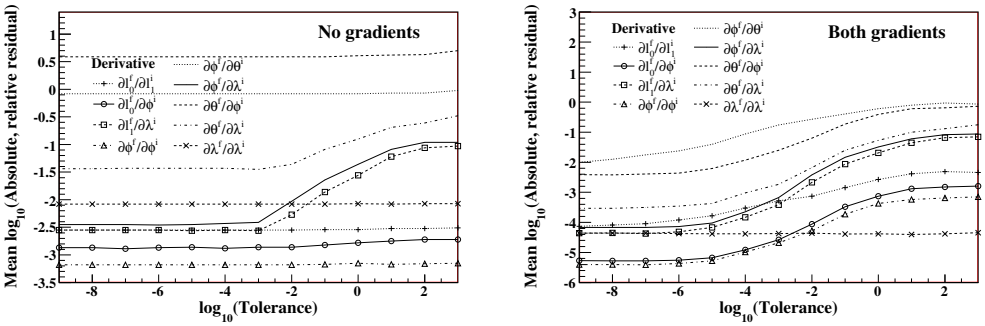


Figure 5: Mean values of the logarithms of the absolute, relative residuals in an inhomogeneous magnetic field with energy loss. The semi-analytical derivatives are calculated at different error tolerances by including no gradients (left) and both gradients (right), whereas the numerical derivatives are all found at a tolerance of 10^{-8}.

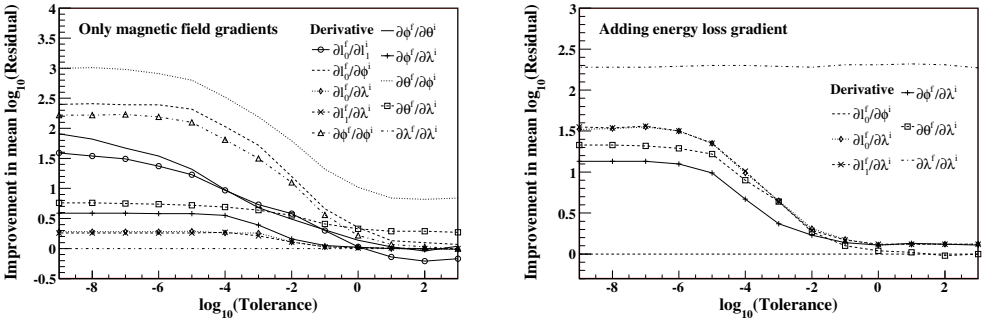


Figure 6: Improvements in the mean values of the logarithms of the absolute, relative residuals by including the magnetic field gradients (left), and the additional improvements by including the energy loss gradient (right) in an inhomogeneous magnetic field with energy loss. The semi-analytical derivatives are calculated at different error tolerances, whereas the numerical derivatives are all found at a tolerance of 10^{-8}.

6.1 Smearing the initial track parameters according to the covariance matrix

To simulate the variances and correlations of the initial local track parameters, we first decompose the initial covariance matrix into two triangular matrices by using Cholesky's method [6]

$$\mathbf{\Sigma}_{\text{initial}} = \mathbf{L} \cdot \mathbf{L}^{\text{T}} \tag{22}$$

This method is easy to use and sufficient for decomposing symmetric, positive definitive matrices such as the covariance matrix.

The elements of the initial covariance matrices are picked at random from Gaussian distributions with mean values of zero and widths of 50 μm for the positions l_0 and l_1, 1 mrad for the angles ϕ and θ, and 1% for the inverse momentum λ. These are realistic values of the resolution of the ATLAS detector, except from the 1% λ uncertainty, which is too optimistic. This is kept low due to the big amount of material — and hence large energy losses of the particles — in the test setup.

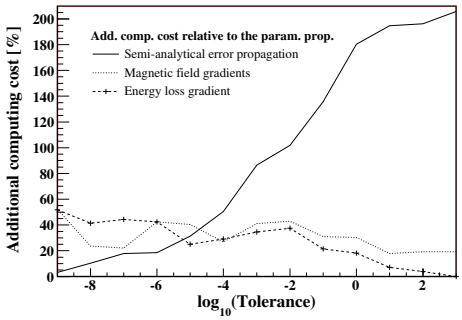


Figure 7: Additional computing time relative to the STEP parameter propagation.

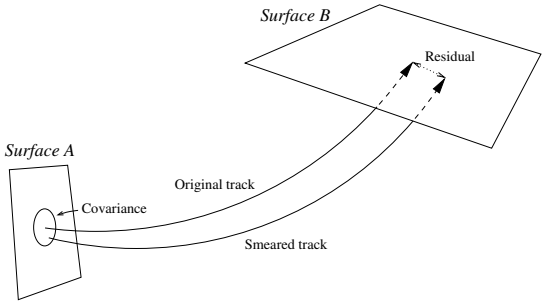


Figure 8: Testing the error propagation by smearing the initial track parameters according to the covariance matrix. The residual is normalized by the propagated covariance to produce the pulls, and multiplied by the inverse of the propagated covariance to find the chi-square. The angles and the momentum are smeared similarly to the positions shown in the figure.

After decomposing the initial covariance matrix, we use it to smear [8] the initial local track parameters $\boldsymbol{\mu}^i$ (i denoting initial)

$$\boldsymbol{\xi}^i = \boldsymbol{\mu}^i + \mathbf{L} \cdot \boldsymbol{\eta} \tag{23}$$

where $\mathbf{L}$ is the lower triangular matrix obtained through the Cholesky decomposition (22), and $\boldsymbol{\eta}$ is a vector of five independent variables picked at random from a Gaussian distribution of mean zero and variance one. Equation (23) assumes that the initial local track parameters are Gaussian distributed and smears them accordingly by using the initial covariance matrix.

6.2 Statistical validation of the semi-analytically propagated covariance matrix

In this test, muon tracks are generated — with their initial parameters smeared according to the above procedure — and propagated by using the test setup described in Section 5, replacing the Silicon with Iron. Enough tracks are generated to make the statistical uncertainties insignificant. The track parameter and error propagation is done by STEP with the magnetic field and energy loss gradients included, at an error tolerance of 10^{-8} to assure the quality of the tracks. Each track is propagated twice to produce the undisturbed $\boldsymbol{\mu}$ and smeared $\boldsymbol{\xi}$ local track parameters at the target surface. The $\boldsymbol{\xi} - \boldsymbol{\mu}$ residuals

are then statistically compared to the semi-analytically propagated covariance matrix $\mathbf{\Sigma}_{\text{final}}$ by using the normalized residuals, or *pull values*

$$\hat{\xi}_k^j = \frac{\xi_k^j - \mu_k^j}{\sqrt{\Sigma_{\text{final},k}^{jj}}} \tag{24}$$

and the *chi-square*

$$\chi_k^2 = (\boldsymbol{\xi}_k - \boldsymbol{\mu}_k)^{\mathrm{T}} \cdot \mathbf{\Sigma}_{\text{final},k}^{-1} \cdot (\boldsymbol{\xi}_k - \boldsymbol{\mu}_k) \tag{25}$$

with k indicating the simulated tracks and j the track parameters.

Since ξ_k^j is Gaussian distributed around μ_k^j, the pull values should be Gaussian distributed around zero. Moreover, if the propagated covariance $\mathbf{\Sigma}_{\text{final},k}$ is correct, the width of the pull values should be normalized to one. All of the pull values presented in Fig. 9 satisfy these requirements, showing good agreement between the semi-analytical error propagation and the simulation. The tails of the λ pull are intrinsic to the semi-analytical error propagation and arise from the information loss caused by introducing the temporary global track parameters during the error propagation.

Whereas the pull values are calculated for each parameter of the simulated track, the chi-square incorporates the whole covariance matrix and all of the track parameters. Assuming that all five track parameters are Gaussian distributed, and that these distributions obey the variances and correlations given by covariance matrix, the test chi-square distribution should be similar to the standard chi-square distribution corresponding to five degrees of freedom. By integrating the standard chi-square distribution from the test chi-square to infinity, we get the so-called *p-value*, or probability value of this test statistic. If the test chi-square distribution is correct, the p-value plot is flat. Inverting the covariance matrix by using singular value decomposition [6], we get the flat p-value plots of Fig. 10, showing good agreement between the semi-analytical error propagation and the simulation.

6.3 Estimating the impact of the gradients on the semi-analytical error propagation

The covariance matrices of the pulls of Fig. 9 are all propagated by including the $\partial g/\partial\lambda$ and $\partial B_i/\partial x_j$ gradients discussed in Section 5 into the error propagation. Including these gradients improves some elements of the Jacobian significantly. Such improvements are also seen when comparing the p-values found by only including the magnetic field gradients (right) to those found by excluding all of the gradients (left) in Fig. 10. The flat p-value plot found by only including the magnetic field gradients leaves little room for further improvement. Thus, only the magnetic field gradients — and not the energy loss gradient — are included into STEP by default. The gradients' influence on the pulls is insignificant, consequently they are not presented here.

Pull and p-values obtained by using an error tolerance of 10^{-2} for the semi-analytical error propagation — instead of the 10^{-8} used in Figs. 9 and 10 — produce similar plots, indicating little sensitivity to the error tolerance in the semi-analytical error propagation.

7 Conclusion

In this paper we have performed an extensive study of the Bugge-Myrheim method, gaining a quantitative understanding of the impact of the magnetic field and energy loss gradients on the accuracy and speed of the semi-analytical error propagation. Results show that only the magnetic field gradients have a visible effect on the covariance matrices transported by the semi-analytical error propagation in the ATLAS magnetic field, hence the energy loss gradient is left out of the error propagation by default.

The computing cost increase — relative to the parameter propagation — by adding the semi-analytical error propagation is less than 100% at medium and high accuracies. This is significantly less than

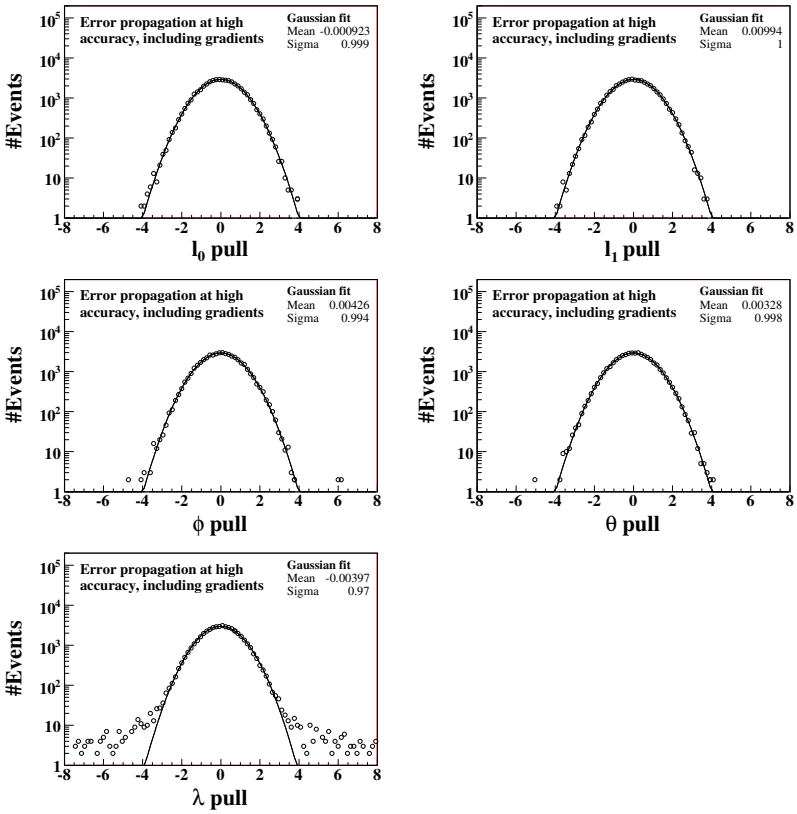


Figure 9: Pull values of the final parameters of tracks smeared by the initial covariance matrix, and their Gaussian fits (solid lines). The covariance matrices are propagated semi-analytically at an error tolerance of 10^{-8} with the gradients included.

the minimal computing cost increase of 500% seen in the numerical error propagation methods. Furthermore, the additional computing cost for including the magnetic field and energy loss gradients is around 30–40% for each type of gradient. Finally, the nominal computing cost and accuracy of the semi-analytical error propagation is relatively stable over the whole error tolerance range.

8 Acknowledgements

This work has been carried out as part of the developments of the ATLAS tracking group. We would like to thank our colleagues for their help in integrating the software and for their support in preparing this note.

References

[1] The ATLAS Collaboration, G. Aad *et al.*, JINST **3** (2008) S08003.

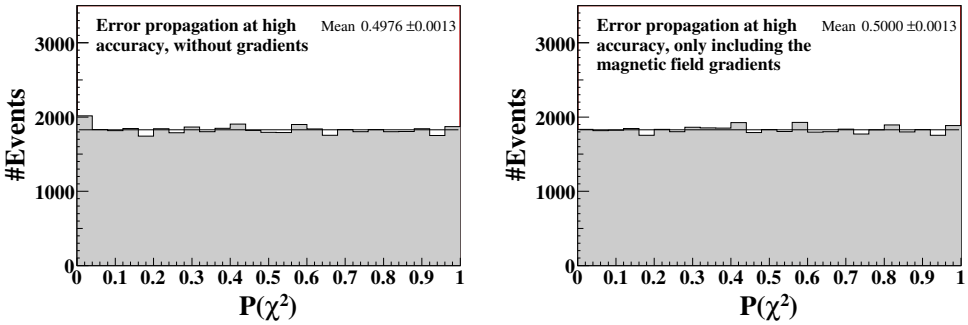


Figure 10: P-values of the final parameters of tracks smeared by the initial covariance matrix, and their linear fits. The covariance matrices are propagated semi-analytically at a tolerance of 10^{-8}, excluding all of the the gradients (left) and by only including the magnetic field gradients (right).

[2] A. Salzburger *et al.*, The ATLAS Tracking Geometry Description, ATLAS Public Note ATL-SOFT-PUB-2007-004, 2007.

[3] E. Lund *et al.*, Track parameter propagation through the application of a new adaptive Runge-Kutta-Nyström method in the ATLAS experiment, ATLAS Public Note ATL-SOFT-PUB-2009-001, submitted to Journal of Instrumentation, 2009.

[4] L. Bugge and J. Myrheim, Nucl. Inst. and Meth. **179** (1981) 365.

[5] F. Åkesson *et al.*, The ATLAS Tracking Event Data Model, ATLAS Public Note ATL-SOFT-PUB-2006-004, 2006.

[6] W. Press *et al.*, Numerical Recipes in C, (Cambridge University Press, Cambridge, 1999).

[7] R. Brun and F. Rademakers, Nucl. Inst. and Meth. in Phys. Res. A **389** (1997) 81.

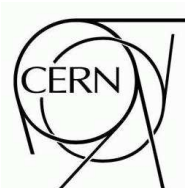
[8] Particle Data Group, C. Amsler *et al.*, Phys. Lett. B **667** (2008).

III

ATLAS NOTE

ATL-SOFT-PUB-2008-003

January 23, 2009



Treatment of energy loss and multiple scattering in the context of track parameter and covariance matrix propagation in continuous material in the ATLAS experiment

E. Lund[*,1], L. Bugge[1], E. W. Hughes[2],
D. López Mateos[2,3], A. Salzburger[4,5], A. Strandlie[1,6]

[1] *University of Oslo, Oslo, Norway*
[2] *Columbia University, New York, USA*
[3] *California Institute of Technology, Pasadena, USA*
[4] *CERN, Geneva, Switzerland*
[5] *Leopold-Franzens Universität, Innsbruck, Austria*
[6] *Gjøvik University College, Gjøvik, Norway*
[*] corresponding author (esben.lund@fys.uio.no)

Abstract

In this paper we study the energy loss, its fluctuations, and the multiple scattering of particles passing through matter, with an emphasis on muons. In addition to the well-known Bethe-Bloch and Bethe-Heitler equations describing the mean energy loss from ionization and bremsstrahlung respectively, new parameterizations of the mean energy loss of muons from the direct e^+e^- pair production and photonuclear interactions are presented along with new estimates of the most probable energy loss and its fluctuations in the ATLAS calorimeters. Moreover, a new adaptive Highland/Molière approach to finding the multiple scattering angle is taken to accomodate a wide range of scatterer thicknesses. Furthermore, tests of the muon energy loss, its fluctuations, and multiple scattering are done in the ATLAS calorimeters. The material effects described in this paper are all part of the *simultaneous track and error propagation* (STEP) algorithm of the common ATLAS tracking software.

1 Introduction

Experimental particle physics is on the verge of a new era, heralded by the Large Hadron Collider being commissioned at the European Organization for Nuclear Research — CERN — located just outside Geneva, Switzerland. The LHC accelerator will collide protons at a center of mass energy of 14 TeV, opening up a new window for particle discoveries and precision measurements of existing theories. Particle detectors are located at four beam crossings along the LHC, one of which houses the ATLAS detector [1]. This is the largest of the LHC experiments, employing a great variety of detector and magnetic field technologies to identify a wide range of particles. The complex magnetic field and high collision rate, however, make the reconstruction of particle tracks very challenging. Things are complicated further by the relatively big amount of material within the 7000-ton ATLAS detector, generating considerable disturbances to the particle tracks through material interactions such as energy loss and multiple scattering.

Track parameter and covariance matrix propagation — the so-called error propagation — is an important part of any track reconstruction algorithm. Due to the high amount of material in the ATLAS detector, energy loss and multiple scattering have to be accurately accounted for in the track reconstruction. One common way of handling these material effects is by layer-based track parameter updates. This involves making point-like updates to the track parameters and covariance matrix at predefined surfaces by using parameterizations of the detector material surrounding these surfaces. This approach is especially useful when the detector material is naturally layered, such as in the ATLAS inner detector. Another strategy involves treating the material effects continuously during the propagation of the track parameters and covariance matrix. This is the approach taken by the STEP algorithm, and it is most useful in large dense volumes with a relatively uniform mass distribution, such as those found in the ATLAS calorimeters.

Since muons are the only particles exposed to a substantial energy loss in the detector without being stopped, extra care is taken when calculating their energy loss. In addition to the well-known Bethe-Bloch [2] and Bethe-Heitler [3] equations describing the mean energy loss from ionization and bremsstrahlung respectively, new parameterizations of the mean energy loss of muons from the direct e^+e^- pair production and photonuclear interactions are presented along with new estimates of the most probable energy loss and its fluctuations in the ATLAS calorimeters. Furthermore, a new adaptive Highland/Molière [4–6] approach to finding the multiple scattering angle is taken to accomodate a wide range of scatterer thicknesses.

In Section 2 we describe the mean energy loss of particles going through matter, followed by a comparison of these values to those found by the simulation toolkit GEANT4 [7] in Section 2.4, and a short study of the mean energy loss in the ATLAS calorimeters in Section 2.5. In Section 3 and 4 we present new estimates of the most probable energy loss and energy loss fluctuations of muons in the ATLAS calorimeters. Moreover, we discuss the multiple scattering angle and contribution to the covariance matrix in Section 5. Furthermore, we perform statistical tests of the multiple scattering covariance contribution in a single volume (Section 5.3), and in the ATLAS calorimeters (Section 5.4). The additional computing cost of including the material effects is presented in Section 6. Finally, we give a brief conclusion in Section 7, followed by the properties of the STEP algorithm in Appendix A, and the derivation of the multiple scattering λ variance in Appendix B.

Much attention is given to the material interactions in the ATLAS calorimeter since this is the main source of such interactions in the ATLAS detector. Natural units ($\hbar = c = 1$) are used throughout this paper, and vectors and matrices are generally given in bold italic and bold capital letters, respectively.

2 Mean energy loss

The energy loss drains particles of energy, possibly stopping them, and in the presence of a magnetic field, the energy loss changes the path of charged particles by reducing their momenta. Energy loss is caused by ionization of the material and radiation by the incident particle, and its fluctuations — the so-called straggling — might be significant, especially at low momenta and for electrons. Straggling introduces uncertainties into the covariance matrix, which are described in Section 4. At high energies, the dominant radiative effects are bremsstrahlung, direct e^+e^- pair production and the photonuclear interactions. Only the mean energy loss from ionization and bremsstrahlung is treated here, except in the case of muons.

2.1 Mean energy loss from ionization, the Bethe-Bloch equation

The mean energy loss from ionization of the material is given by the Bethe-Bloch equation [2];

$$\left(\frac{dE}{ds}\right)_{\text{ionization}} = -Kz^2\frac{Z\rho}{A\beta^2}\left(\frac{1}{2}\ln\frac{2m_e\beta^2\gamma^2T_{\max}}{I^2} - \beta^2 - \frac{\delta}{2}\right) \tag{1}$$

where K is a constant, z is the charge of the incident particle, and A and Z are the atomic mass and number of the material. The variable I is the mean excitation energy, and ρ is the density of the material. The variable $T_{\max}$ is the maximum kinetic energy which can be imparted to a free electron in a single collision

$$T_{\max} = \frac{2m_e\beta^2\gamma^2}{1+2\gamma m_e/m+(m_e/m)^2} \tag{2}$$

with the kinematic variables β and γ given by $\beta \equiv v = p/E$ and $\gamma \equiv 1/\sqrt{1-\beta^2} = E/m$, where m is the rest mass of the incident particle and $E = \sqrt{p^2+m^2}$. The density effect δ is given by

$$\frac{\delta}{2} = \ln(28.816\frac{\sqrt{\rho Z/A}}{I}) + \ln\beta\gamma - 1/2 \tag{3}$$

for γ above 10, which corresponds to a momentum of around 1 GeV for muons. For momenta below this, the density effect becomes small and is therefore ignored. The mean excitation energy I is highly influenced by the internal structure of the atom, especially the number of electrons in the outer shell. Finding a good approximation, valid for all atoms, is difficult. Here we estimate I with the simple equation

$$I = 16Z^{0.9} \tag{4}$$

For relativistic electrons we use a slightly different Bethe-Bloch equation [8];

$$\left(\frac{dE}{ds}\right)^{\text{electrons}}_{\text{ionization}} = -K\frac{Z\rho}{A}\left(\ln\frac{2m_e}{I} + 1.5\ln\gamma - 0.975\right) \tag{5}$$

2.2 Mean energy loss from bremsstrahlung, the Bethe-Heitler equation

The mean energy loss from bremsstrahlung for relativistic particles is given by the Bethe-Heitler equation [3];

$$\left(\frac{dE}{ds}\right)_{\text{bremsstrahlung}} = -\frac{E}{X_0}\left(\frac{m_e}{m}\right)^2 \tag{6}$$

where X_0 is the radiation length of the material traversed.

2.3 Mean energy loss from the direct e^+e^- pair production and photonuclear interactions of muons

The mean energy loss from the direct e^+e^- pair production and photonuclear interactions of muons is parameterized by using the muon tables found in Ref. [9];

$$\left(\frac{dE}{ds}\right)^{E\in\langle 8\text{ GeV},1\text{ TeV}\rangle}_{\text{pair prod. + ph. i.}} = 0.5345\frac{1}{X_0} - 6.803\cdot 10^{-5}\frac{E}{X_0} - 2.278\cdot 10^{-11}\frac{E^2}{X_0} + 9.899\cdot 10^{-18}\frac{E^3}{X_0} \tag{7}$$

$$\left(\frac{dE}{ds}\right)^{E\in[1\text{ TeV},\rightarrow\rangle}_{\text{pair prod. + ph. i.}} = 2.986\frac{1}{X_0} - 9.253\cdot 10^{-5}\frac{E}{X_0} \tag{8}$$

with the energy given in MeV and the radiation length in mm. The linear parameterization above 1 TeV is applied to avoid any scaling problems at "infinite" momenta. The tables of Ref. [9] are based on an extensive study of the muon energy loss and contain details on all of the energy loss processes mentioned in this paper. The mean energy loss from the pair production and photonuclear interactions is exaggerated to compensate for the underestimated bremsstrahlung found by the Bethe-Heitler equation (6), and it is negligible for muon energies below 8 GeV.

2.4 Total mean energy loss

Adding the mean energy loss from ionization, bremsstrahlung and, in the case of muons, the direct e^+e^- pair production and photonuclear interactions, we get the total mean energy loss per unit distance

$$g = \left(\frac{dE}{ds}\right)_{\text{ionization}} + \left(\frac{dE}{ds}\right)_{\text{bremsstrahlung}} + \left(\frac{dE}{ds}\right)^{\text{muons}}_{\text{pair production and photonuclear interactions}} \tag{9}$$

This mean energy loss causes a change in the $\lambda \equiv q/p$ of the incident particle given by [10]

$$\frac{d\lambda}{ds} = -\frac{\lambda^3 gE}{q^2} \tag{10}$$

which is used by the STEP algorithm [11] for calculating λ in the equation of motion

$$\frac{d^2\boldsymbol{r}}{ds^2} = \frac{q}{p}\left(\frac{d\boldsymbol{r}}{ds}\times\boldsymbol{B}(\boldsymbol{r})\right) = \lambda(\boldsymbol{T}\times\boldsymbol{B}(\boldsymbol{r})) \tag{11}$$

where $\boldsymbol{T} = d\boldsymbol{r}/ds$ is the normalized tangent vector to the track and $\boldsymbol{B}(\boldsymbol{r})$ is the magnetic field.

The total mean energy loss of muons found by STEP is compared to that of the tables found in Ref. [9] and the mean values calculated by GEANT4 in Fig. 1. The small bumps around 1 GeV are from the underestimation of the density effect (3) at momenta of a few GeV.

2.5 Testing the mean energy loss of muons in the ATLAS calorimeters

Figure 2 shows the mean energy loss of muons in the ATLAS electromagnetic and hadronic calorimeters as a function of the absolute pseudorapidity ($|\eta|$) for muons of momenta 10, 100 and 1000 GeV. The histograms are simulated by GEANT4 in the full detector description, while the points are found by STEP in the dense volumes of the calorimeter tracking geometry [12] by using the test setup described in Ref. [13]. All of the $|\eta|$ distributions presented in this paper are averaged over ϕ. The material in the electromagnetic calorimeter amounts to around 30 X_0, while the hadronic calorimeter adds another 100 X_0 to that. The holes in the STEP mean energy loss around $|\eta| \approx 0.8$ in Fig. 2 are probably caused by an insufficient description of the inactive material — such as the liquid Argon cryostats and the inner detector services — in the dense volumes of the calorimeter tracking geometry in these regions, illustrated in Fig. 3.

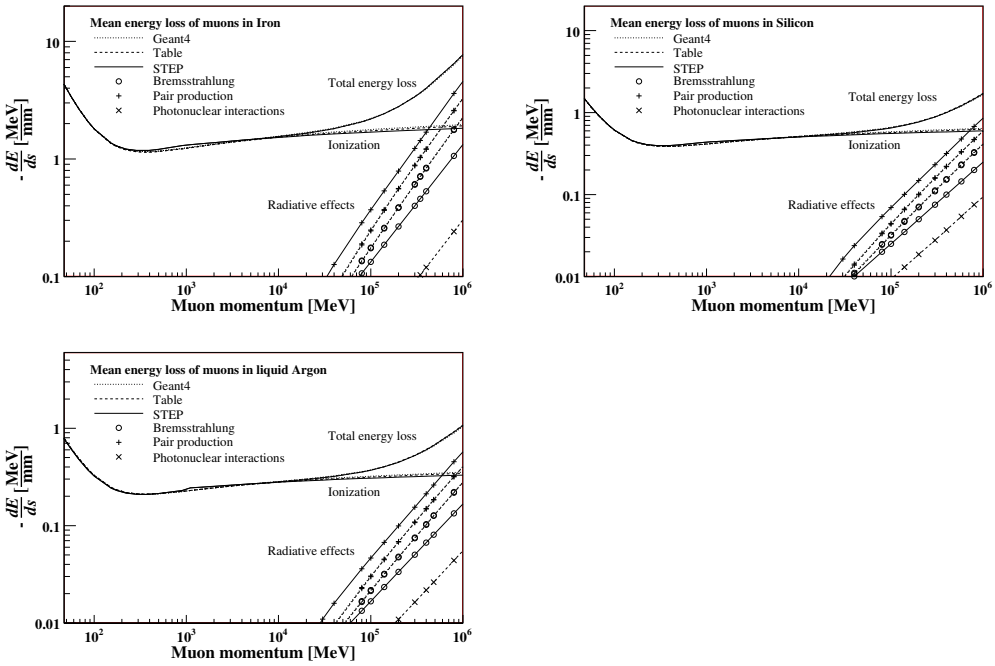


Figure 1: Mean energy loss of muons in Iron, Silicon and liquid Argon. The photonuclear interactions are not included into the GEANT4 simulation, while they are bundled with the pair production in the STEP algorithm.

3 Most probable energy loss of muons in the ATLAS calorimeters

The energy loss is Landau distributed with a relatively low most probable (peak) value and a tail going towards the higher energies, making the most probable value smaller than the mean energy loss, Fig. 4. This split between the mean and the most probable energy loss increases with the energy. A good approximation of the most probable energy loss is achieved by reducing the mean ionization and radiative energy loss by 10% and 85% respectively, Fig. 5. The mean and most probable energy loss are toggled by a property described in Appendix A, with the most probable energy loss as the default setting. Although the mean energy loss becomes three times higher than the most probable value at muon momenta of 1 TeV, the mean energy loss is still only 1.5% of the muon momentum, significantly less than the momentum resolution at this energy. At momenta below 100 GeV— where the energy loss and its fluctuations matter most — the difference between the mean and most probable energy loss is relatively small.

4 Energy loss fluctuations of muons in the ATLAS calorimeters

The energy loss of particles fluctuates considerably, especially for light particles such as electrons. These fluctuations also reduce the energy resolution of muons below 100 GeV significantly. Since the energy loss is Landau distributed, analytical calculations of the fluctuations are difficult, so here we use the following parameterization to estimate the Landau width of the energy loss fluctuations of muons going

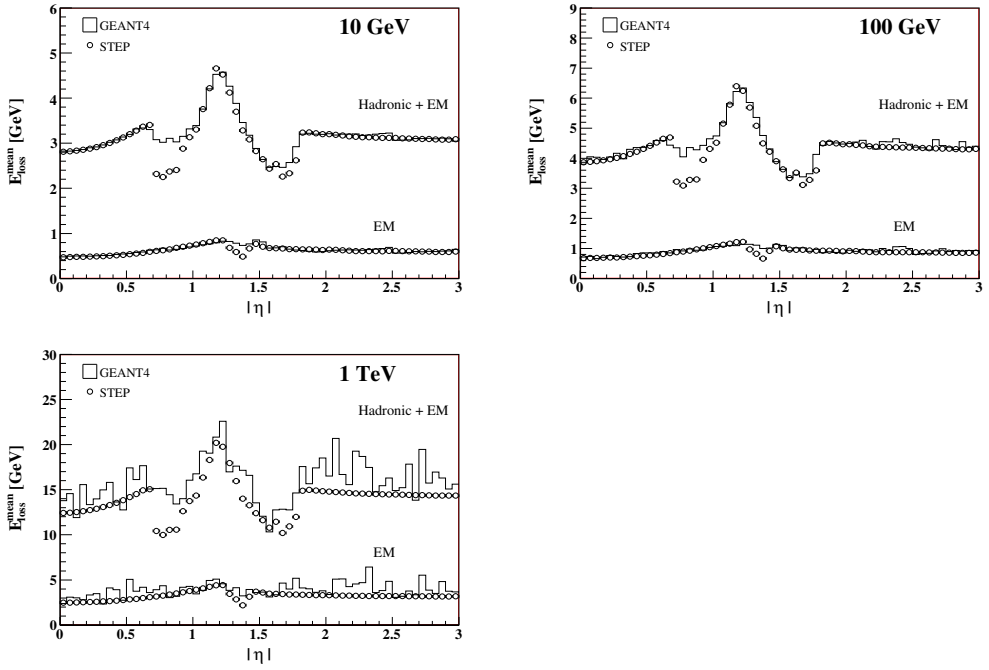


Figure 2: Mean energy loss of muons going through the ATLAS hadronic and electromagnetic calorimeters. The points correspond to the mean values calculated by the STEP algorithm in the dense volumes of the calorimeter tracking geometry, while the histograms represent the mean energy loss simulated by GEANT4 in the full detector description.

through the ATLAS calorimeters;

$$E_{\text{loss}}^{\sigma} = \sqrt{\frac{as\rho KZ}{2A\beta^2}} \tag{12}$$

where s is the traversed distance in mm and a is given by

$$a = 121 + 3.9 \cdot 10^{-3} E + 5.3 \cdot 10^{-9} E^2 \tag{13}$$

with the energy given in MeV. Figure 6 shows the Landau width of the energy loss fluctuations of muons going through the ATLAS calorimeters as given by the above parameterization and GEANT4. The Landau width can be scaled by a property described in Appendix A. Given the above Landau width, the noise contribution to the λ variance of the covariance matrix becomes

$$\text{Var}(\lambda) = \frac{(E_{\text{loss}}^{\sigma})^2}{\beta^2 p^4} \tag{14}$$

In case of large relative energy losses, the volume is sliced. This is possible because the straggling variance parameterization is proportional to the traversed distance s, hence fulfilling the convolution property; the variance of the combined straggling of all of the slices of a volume is independent of the chosen number of slices. For reasons of simplicity these slices are shared with — and described in — the

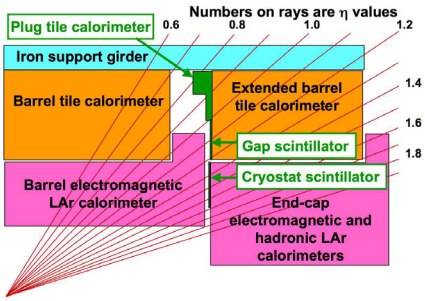


Figure 3: Schematic of the transition region between the barrel and end-cap cryostats, where additional scintillator elements are installed to provide corrections for the energy lost in the inactive material (not shown) of the gap, such as the liquid Argon cryostats and the inner detector services. (From Ref. [1]).

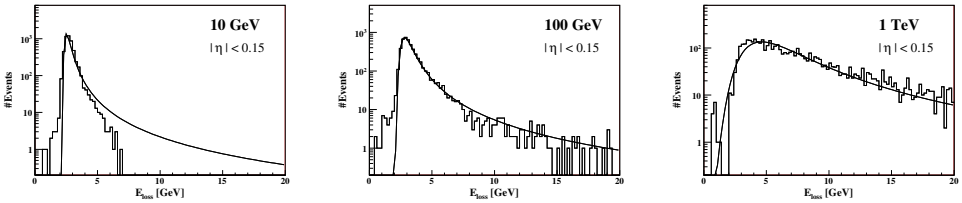


Figure 4: Energy loss of muons with $|\eta| < 0.15$ going through the ATLAS calorimeters, as found by GEANT4 in the full detector description. Fits to Landau distributions are also shown.

multiple scattering calculation, Section 5. Because the real energy loss fluctuations do not scale linearly with s, the above straggling estimate — which is a fit to the energy loss fluctuations of muons going through the whole ATLAS calorimeter — does not describe the fluctuations through the electromagnetic calorimeter very well.

5 Multiple scattering

The multiple scattering causes random angular deflections to the track path with an average projected deflection of zero. Because the mean projected deflection is zero, the multiple scattering does not affect the unbiased track estimate. It does, however, introduce correlations and a loss of resolution in the track parameters, which must be reflected in the covariance matrix $\mathbf{\Sigma}$. This is done by adding the noise contribution from the multiple scattering to the transported covariance matrix. The linear error propagation [14] then becomes

$$\mathbf{\Sigma}_{\text{final}} = \mathbf{J} \cdot \mathbf{\Sigma}_{\text{initial}} \cdot \mathbf{J}^{\text{T}} + \mathbf{\Sigma}_{\text{multiple scattering}} \tag{15}$$

where $\mathbf{J}$ is the transport Jacobian.

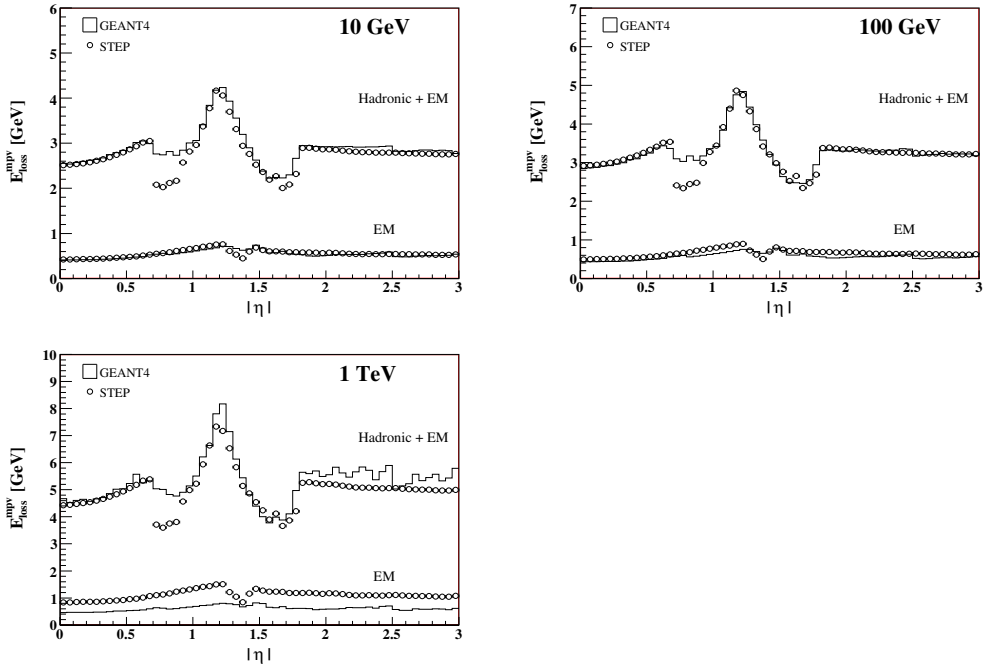


Figure 5: Most probable energy loss of muons going through the ATLAS hadronic and electromagnetic calorimeters. The points correspond to the most probable values calculated by the STEP algorithm in the dense volumes of the calorimeter tracking geometry, while the histograms represent the most probable values of the Landau fits of each $|\eta|$ bin of the GEANT4 simulations in the full detector description.

5.1 Finding the scattering angle

The most widely known model of multiple scattering is due to Molière [4, 5], giving the projected scattering angle θ_p width as

$$\sigma_M(\theta_p) = \frac{15}{\beta p}\sqrt{\frac{s}{X_0}} \tag{16}$$

where s is the traversed distance, X_0 is the radiation length of the material and the momentum is given in MeV. The scattering angle θ_s and its projection θ_p are illustrated in Fig. 7. The Molière model is based on a linear, Gaussian approach to the multiple scattering. Besides its simplicity, it also fulfills the convolution property; the variance of the combined scattering angles of all of the slices of a volume is independent of the chosen number of slices. This property is fulfilled by the Gaussian Molière model because the variance is strictly proportional to the traversed distance s. In a simulation program where the number and thickness of slices is flexible, fulfilling the convolution property becomes very important. Slicing of a volume may also be required when the energy loss becomes significant.

The popular Highland scattering model [6] is a fit to the Molière distribution of singly charged particles, assuming $\beta = 1$. The Highland model is easily generalized [15] to allow for β less than one by including $1/\beta^2$ in the logarithm, giving

$$\sigma_H(\theta_p) = \frac{13.6}{\beta p}\sqrt{\frac{s}{X_0}}\left(1 + 0.038\ln\frac{s}{\beta^2 X_0}\right) \tag{17}$$

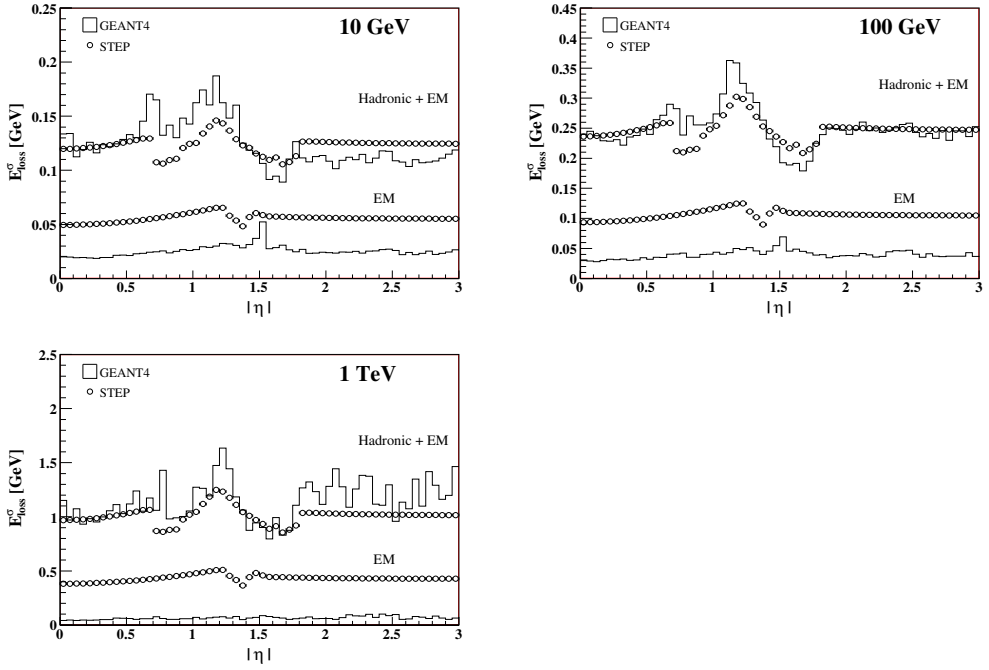


Figure 6: Energy loss fluctuations of muons going through the ATLAS hadronic and electromagnetic calorimeters. The points correspond to the parameterization applied by the STEP algorithm in the dense volumes of the calorimeter tracking geometry, while the histograms represent the width of the Landau fits of each $|\eta|$ bin of the GEANT4 simulations in the full detector description.

This expression is only better than the original Molière model as long as slicing the volume is not required since the Highland formula does not fulfill the convolution property, giving a smaller combined scattering angle variance with an increasing number of slices within a given volume.

Projected scattering angle widths of both models are shown along with the width from the GEANT4 simulation in the top left plot of Fig. 8. The GEANT4 multiple scattering uses a modified Highland formula and around 100 steps (slices) for the simulation, whereas the Molière and Highland widths are calculated for the whole volume in one go (one slice), assuming no energy loss. The Highland formula agrees well with the GEANT4 simulation for distances less than 2 X_0, corresponding to a 5% energy loss for a muon having a momentum of 1 GeV. Going through additional material, the scattering is significantly increased by the energy loss. The only parameter of the Highland formula varying substantially during the propagation is the momentum, implying a breakdown of the single layer Highland approach at a 5% energy loss regardless of X_0 and s.

The top right plot of Fig. 8 shows different applications of the Molière and Highland scattering. The first two lines show the widths found within STEP by using 10 slices, assuming a constant momentum in each slice. The Molière scattering shows good agreement with GEANT4 above 2 X_0, whereas the Highland scattering is underestimated across the whole range due to the convolution problem. The next two lines show the widths obtained by the simple simulation described in Section 5.3. The simulation slices the volume into equally thick layers before successively traversing them, while using the Molière

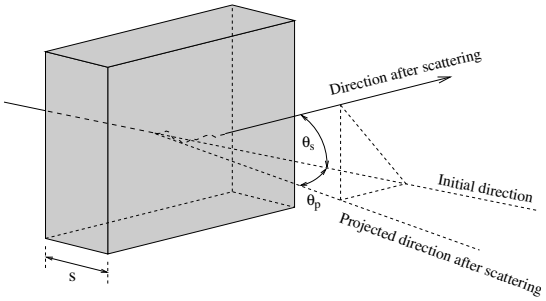


Figure 7: The multiple scattering of a particle going through a slab of material of thickness s. The particle is scattered by an angle θ_s in space, which is projected into a plane to find the projected scattering angle θ_p.

or Highland formula to calculate the scattering in each layer. Scattering angles are pulled randomly from Gaussian distributions of widths given by the formulae. The Molière simulation is in good agreement with STEP, whereas the Highland simulation is underestimated. Reducing the number of layers of the Highland simulation to the same as that used by STEP (line five in the plot) brings these into agreement, clearly demonstrating how an increased number of slices within the same volume reduces the combined Highland scattering angle variance, hence violating the convolution property.

The bottom left plot of Fig. 8 shows the final solution chosen for the projected scattering angle width in STEP. This solution draws on two lessons from the above plots; the single slice Highland formula does well below a 5% energy loss, while the multiple Molière layers perform well above. To find the width, the STEP algorithm first checks whether the energy loss is above 5%, in which case it applies multiple Molière layers, else the width is calculated by the single layer Highland formula by using the average momentum along the track. In the first case, the number of Molière layers is given by the energy loss (in percent) divided by 5% and rounded up, such that a 6% energy loss gives 2 layers and an 11% loss gives 3 layers. Testing shows that the combined scattering angle width converges at around 10 Molière layers regardless of the energy loss, which is therefore the maximum number of layers used by STEP.

Figure 9 shows four examples of projected scattering angle distributions given by the GEANT4 and Molière simulations. The Molière models' tendency to overestimate the scattering in thin layers is clearly seen. The STEP widths are included for comparison and show good agreement with the GEANT4 simulations.

5.2 Setting up the multiple scattering covariance contribution

Dividing the volume into layers normal to the track, the estimated multiple scattering covariance contribution from each layer [16] — in the curvilinear coordinate system to be defined below — by using the local track parameters $\boldsymbol{\xi} = (l_0, l_1, \phi, \theta, \lambda)$ is

$$\boldsymbol{\Sigma}_{\text{ms}}^{\text{one layer}} = \text{Var}(\theta_p) \begin{bmatrix} \frac{s^2}{3} + sd + d^2 & 0 & \frac{s}{2\sin\theta} + \frac{d}{\sin\theta} & 0 & 0 \\ 0 & \frac{s^2}{3} + sd + d^2 & 0 & -\frac{s}{2} - d & 0 \\ \frac{s}{2\sin\theta} + \frac{d}{\sin\theta} & 0 & \frac{1}{\sin^2\theta} & 0 & 0 \\ 0 & -\frac{s}{2} - d & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 3s^2(\frac{d\lambda}{ds})^2 \text{Var}(\theta_p) \end{bmatrix} \tag{18}$$

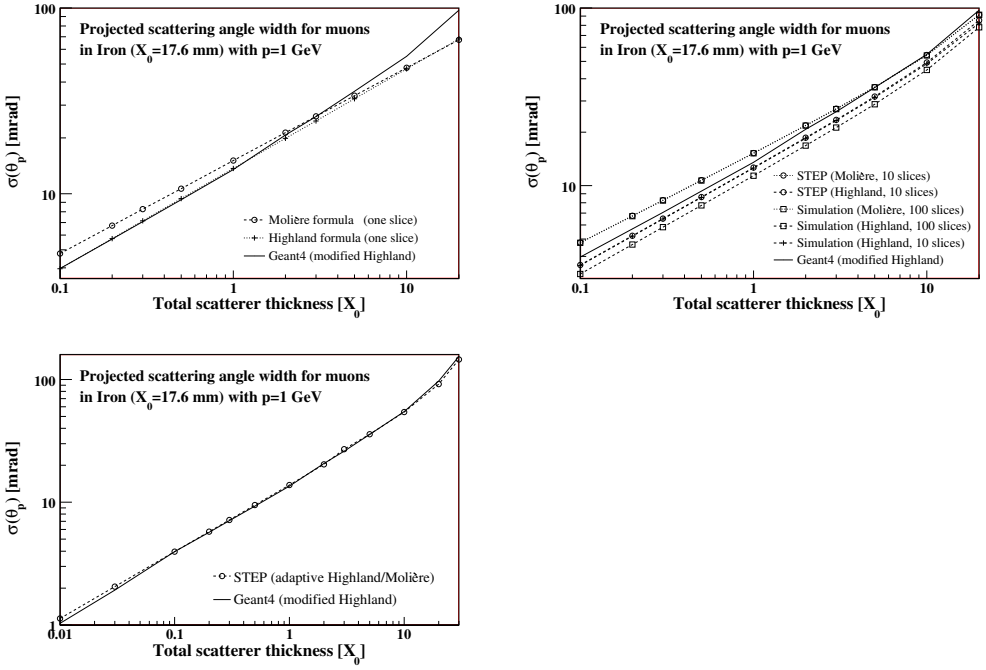


Figure 8: Projected scattering angle widths of 1 GeV muons passing through Iron.

where s is the thickness of the layer, d is the remaining distance along the track path from the end of the layer to the target surface, and $d\lambda/ds$ is the energy loss given in Eq. (10). The local track parameters $\boldsymbol{\xi}$ are defined in the ATLAS event data model [17] by the local track position at a surface (l_0, l_1) and the global track momentum (ϕ, θ, λ), while the orthonormal right-handed curvilinear coordinate system is defined by the three axes; $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{T}$. The axis $\boldsymbol{T}$ is parallel to the particle momentum, while the axes $\boldsymbol{U}$, $\boldsymbol{V}$, defined as

$$\boldsymbol{U} = \frac{\boldsymbol{Z} \times \boldsymbol{T}}{\|\boldsymbol{Z} \times \boldsymbol{T}\|} \quad \text{and} \quad \boldsymbol{V} = \boldsymbol{T} \times \boldsymbol{U} \tag{19}$$

are normal to the momentum. The vector $\boldsymbol{Z}$ is given by the global z-axis.

The multiple scattering covariance contribution is based on a linear, Gaussian approach, with all of the layers calculated in the same curvilinear system set up in the direction of the initial momentum. Therefore the θ angle in the covariance contribution is given by the initial θ throughout all of the layers. All of the elements of the covariance contribution, except the λ variance, are given in Ref. [16], and their widths can be scaled by a property described in Appendix A. The small differences between some of the terms given in the reference and those presented here, are all due to the different coordinate systems used. The λ variance, derived in Appendix B, comes from the variation in the energy loss introduced by the variation of the path length due to the multiple scattering.

The total scattering covariance contribution is the sum of the covariance contribution matrices (18) of all of the scattering layers. Before adding the total scattering contribution — given in the curvilinear system — to the transported covariance (15), it is transformed into the local coordinate system given at the target surface. To account for the variation in the energy loss caused by the combination of the variation of the path length — due to the multiple scattering — and the tilted target surface, we add two

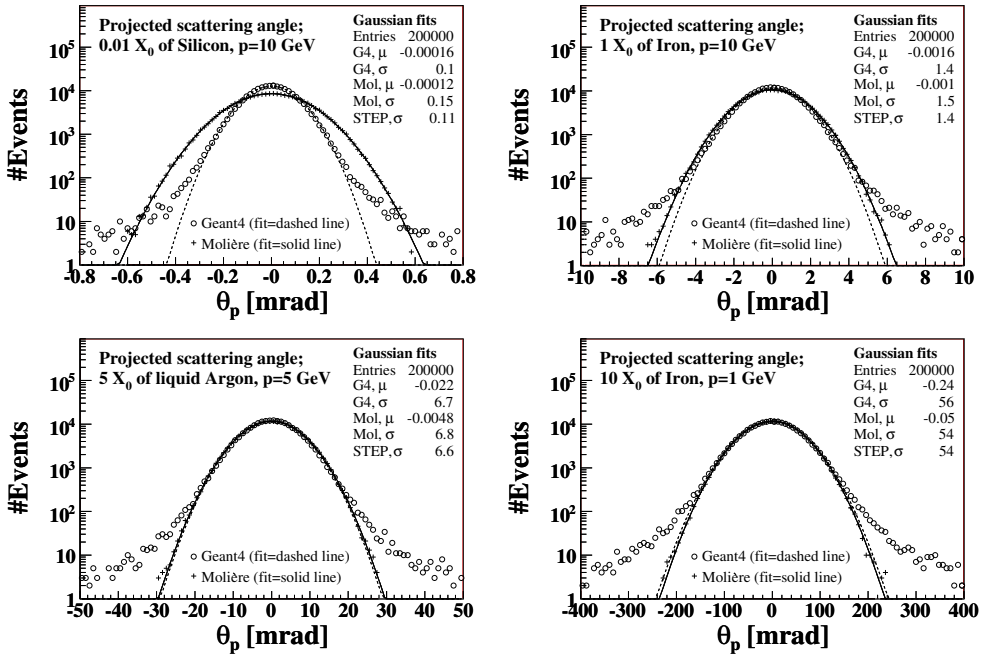


Figure 9: Projected scattering angles found by the GEANT4 and Molière simulations (100 slices), and their Gaussian fits. The STEP widths are also given for comparison.

elements to the cuvilinear to local Jacobian [18];

$$\frac{\partial \lambda^{\text{loc}}}{\partial l_0^{\text{curv}}} = -\frac{\boldsymbol{U}\cdot\boldsymbol{I}}{\boldsymbol{T}\cdot\boldsymbol{I}}\left(\frac{d\lambda}{ds}\right) \quad \text{and} \quad \frac{\partial \lambda^{\text{loc}}}{\partial l_1^{\text{curv}}} = -\frac{\boldsymbol{V}\cdot\boldsymbol{I}}{\boldsymbol{T}\cdot\boldsymbol{I}}\left(\frac{d\lambda}{ds}\right) \tag{20}$$

where $\boldsymbol{I}$ is the unit vector normal to the target plane, defining one axis of the local coordinate system at the target surface.

The multiple scattering is calculated separately after the propagation is done, making it independent of the error tolerance. The error tolerance is a user specified number steering the accuracy of the propagation [11], a low error tolerance giving a high accuracy, and vice versa. The multiple scattering contribution is only found once for every successful propagation, hence the computing cost is small, as seen in Fig. 16. It can be switched off by using one of the properties mentioned in Appendix A.

5.3 Statistical validation of the multiple scattering through a simple Monte Carlo simulation in a single volume

The stochastic multiple scattering process can be tested by using a Monte Carlo simulation, which is done by slicing a dense volume into equally thick layers normal to the track before successively traversing them while the momentum is deflected by a random spatial scattering angle θ_s at the end of each layer, Fig. 10. This angle is picked from a Gaussian distribution with a width given by the Molière formula (16) multiplied by $\sqrt{2}$ to go from the projected to the spatial scattering angle

$$\sigma(\theta_s) = \sqrt{2}\sigma_M = \frac{21.2132}{\beta p}\sqrt{\frac{s}{X_0}} \tag{21}$$

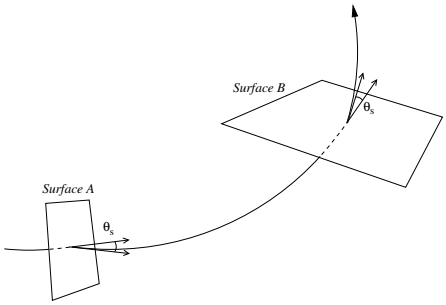


Figure 10: The multiple scattering is simulated by deflecting the track a random spatial angle θ_s at layers normal to the track.

The test setup involves propagating muons in random directions, covering all azimuthal and polar angles at momenta ranging from 500 MeV to 500 GeV, starting off from an initial surface located at the interaction point of the ATLAS detector. The muons are propagated through Iron in the realistic ATLAS magnetic field, towards a target surface randomly placed and rotated in a cube with sides of 20 m centered in the detector. All of the propagation is done by STEP at an error tolerance of 10^{-8} to assure the quality of the tracks. 100 000 muon tracks are generated to produce around 34 000 successfully simulated tracks.

In the simulation, each track is propagated twice to produce the undisturbed $\boldsymbol{\mu}$ and scattered $\boldsymbol{\xi}$ local track parameters at the target surface. The $\boldsymbol{\xi} - \boldsymbol{\mu}$ residuals are then statistically compared to the total scattering contribution $\boldsymbol{\Sigma}_{\text{ms}}$ by using the *pull values*

$$\hat{\xi}_i^j = \frac{\xi_i^j - \mu_i^j}{\sqrt{\Sigma_{\text{ms},i}^{jj}}} \tag{22}$$

and the *chi-square*

$$\chi_i^2 = (\boldsymbol{\xi}_i - \boldsymbol{\mu}_i)^{\text{T}} \cdot \boldsymbol{\Sigma}_{\text{ms},i}^{-1} \cdot (\boldsymbol{\xi}_i - \boldsymbol{\mu}_i) \tag{23}$$

with i indicating the simulated tracks and j the track parameters.

Since $\boldsymbol{\xi}$ are Gaussian distributed around $\boldsymbol{\mu}$ in the simulation, the pull values are Gaussian distributed around zero. Moreover, if the scattering contribution is correct, the width of the pull values should be normalized to one. The pulls presented in Fig. 11 show good agreement between the total multiple scattering covariance contribution found by STEP and the simulation. The small discrepancy in the λ pull mostly comes from shortcuts taken in the derivation of the scattering λ variance (Appendix B), and from missing elements in the curvilinear to local Jacobian applied when going from the curvilinear scattering system to the local system at the target surface.

Whereas the pull values are calculated for each parameter of the simulated track, the chi-square incorporates the whole scattering contibution and all of the track parameters. Assuming that the scattered track parameters $\boldsymbol{\xi}$ are Gaussian distributed around $\boldsymbol{\mu}$, and that these distributions obey the variances and correlations given by the scattering contribution, the test chi-square distribution should be similar to the standard chi-square distribution corresponding to five degrees of freedom. By integrating the standard chi-square distribution from the test chi-square to infinity, we get the so-called *p-value*, or probability value of this test statistic. If the test chi-square distribution is correct, the p-value plot is flat, as seen in Fig. 11.

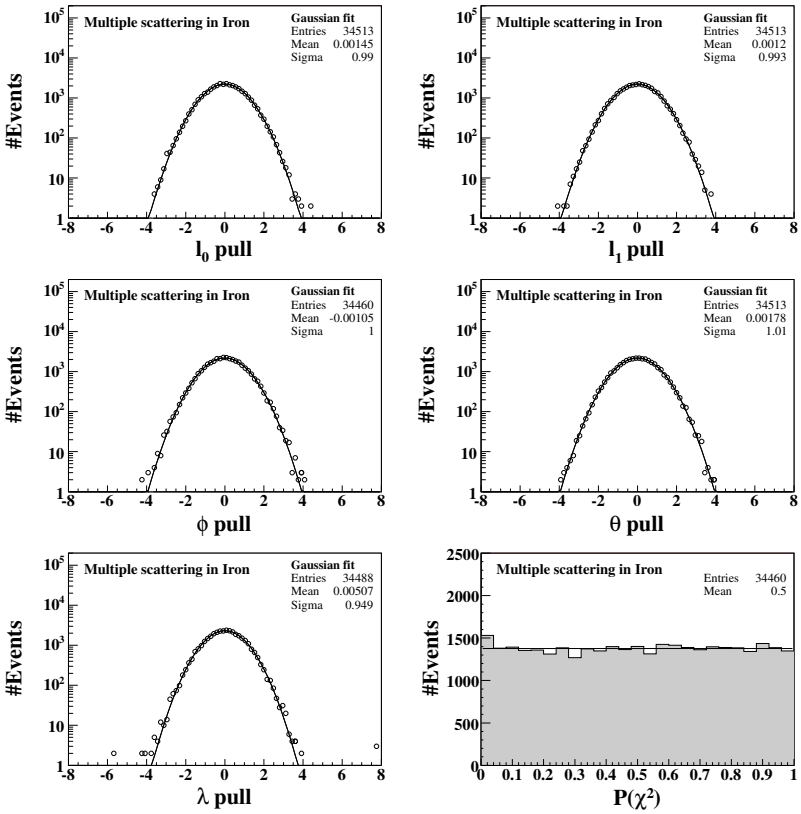


Figure 11: Pulls and p-values of simulated tracks subjected to multiple scattering in a single volume, and their fits (solid lines). The plots are made by using the multiple scattering covariance contribution found by STEP.

5.4 Statistical validation of the multiple scattering of muons in the ATLAS calorimeters

In this Section we repeat the exercise of Section 5.3 in the ATLAS calorimeters by using the test setup described in Ref. [13]. Figure 12 and 13 show the pulls and p-values of 10 GeV muons simulated by GEANT4 in the full detector description — including energy loss and multiple scattering — and statistically compared to the combined STEP multiple scattering and straggling covariance contribution. Figure 12 presents muons going through the electromagnetic calorimeters, while Fig. 13 shows muons going through both the electromagnetic and hadronic calorimeters. The l_0, l_1, ϕ and θ pulls are related to the multiple scattering, whereas the λ pull is mostly related to the straggling. Because the straggling does not scale linearly with the traversed distance it is tuned to muons going through the whole calorimeter (Fig. 13), and hence does not describe the straggling of muons only passing through the electromagnetic calorimeter very well (Fig. 12).

Figure 14 and 15 show the l_0, l_1, ϕ and θ pulls of 100 GeV and 1 TeV muons passing through both the electromagnetic and hadronic calorimeters, including all of the material interactions. However, since the muon straggling in the calorimeters is discussed in Section 4, we choose to focus on the pulls related

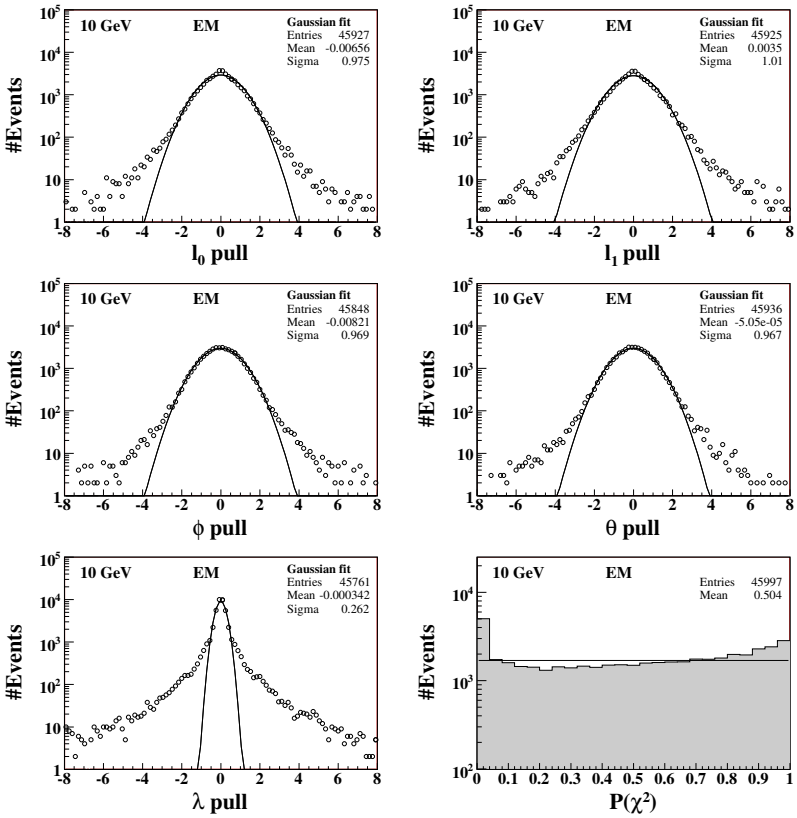


Figure 12: Pulls and p-values of 10 GeV muons subjected to multiple scattering and energy loss fluctuations in the ATLAS electromagnetic calorimeter, and their fits (solid lines). The track residuals are simulated by GEANT4 in the full detector description, while the covariance scattering and straggling contribution is found by STEP in the dense volumes of the calorimeter tracking geometry.

to the multiple scattering in these figures.

The width of the elements of the multiple scattering covariance contribution and the Landau width of the straggling are scaled by 1.5 and 4.7, respectively, in Figs. 12 – 15. Both scales can be set through the properties described in Appendix A.

6 Additional computing cost of including the material effects

Figure 16 shows the additional computing cost — relative to the STEP parameter propagation — of including the energy loss, multiple scattering and straggling, as a function of the error tolerance. The energy loss per unit distance varies little within each recursion step of the integration method, hence the value found at the beginning of each step is used throughout the step. This lowers the computing cost considerably. The computing cost is increased by around 30% by the energy loss and around 10% by the multiple scattering and straggling over the whole error tolerance range.

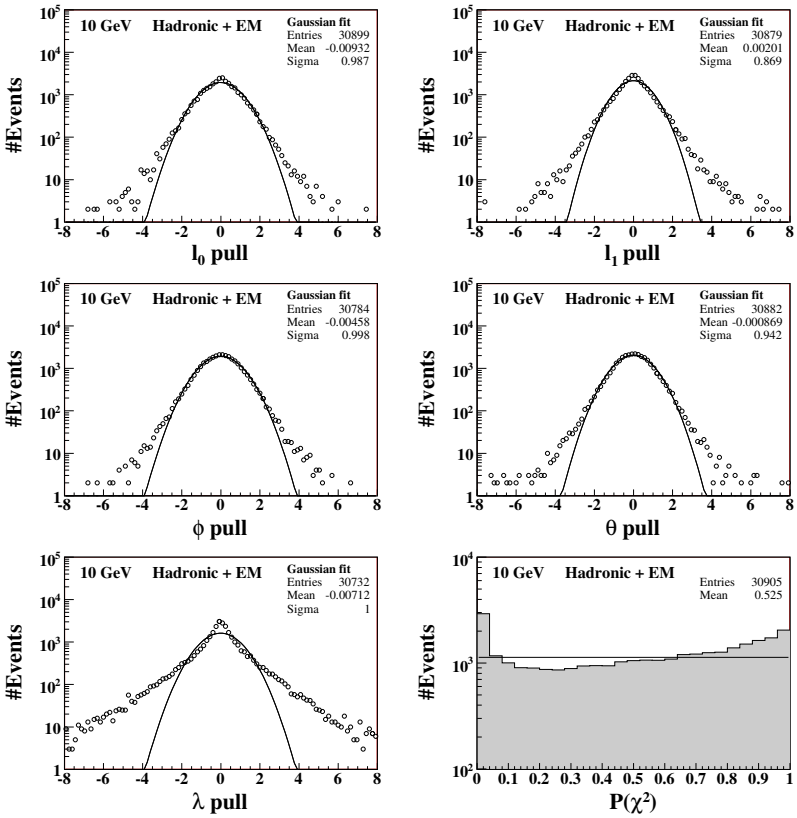


Figure 13: Pulls and p-values of 10 GeV muons subjected to multiple scattering and energy loss fluctuations in the ATLAS hadronic and electromagnetic calorimeters, and their fits (solid lines). The track residuals are simulated by GEANT4 in the full detector description, while the covariance scattering and straggling contribution is found by STEP in the dense volumes of the calorimeter tracking geometry.

7 Conclusion

In this paper some simple, yet precise, solutions to the energy loss, straggling and multiple scattering of particles in dense material have been presented, focusing on muons. Simulations performed with GEANT4 are in good agreement with these solutions. The computing cost increase — relative to the STEP parameter propagation — is around 30% by including the energy loss into the propagation, and around 10% by adding the multiple scattering and straggling contribution to the covariance matrix, over the whole error tolerance range.

All of the material interactions presented in this paper are part of the STEP algorithm, found in the TrkExtrapolation package of the ATLAS CVS repository. The properties available for steering the STEP algorithm are described in Appendix A.

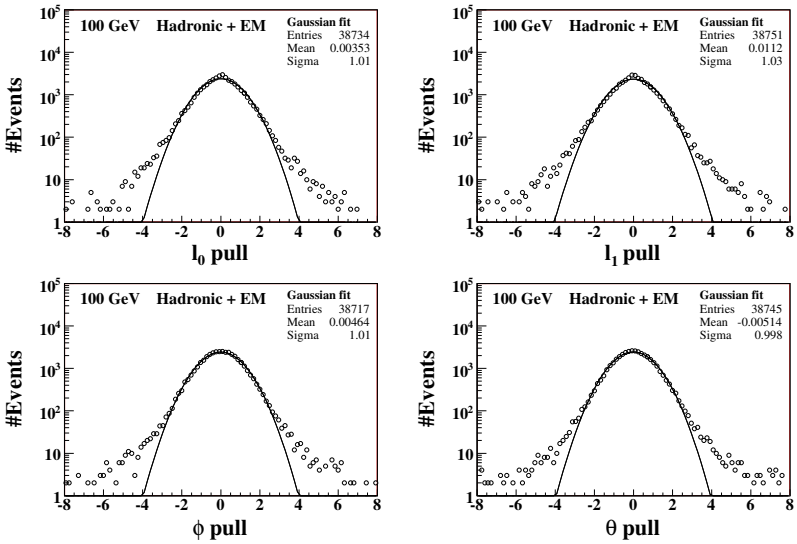


Figure 14: Pulls of 100 GeV muons subjected to multiple scattering in the ATLAS hadronic and electromagnetic calorimeters, and their Gaussian fits (solid lines). The track residuals are simulated by GEANT4 in the full detector description, while the covariance scattering contribution is found by STEP in the dense volumes of the calorimeter tracking geometry.

8 Acknowledgements

This work has been carried out as part of the developments of the ATLAS tracking group. We would like to thank our colleagues for their help in integrating the software and for their support in preparing this note.

A Properties of the STEP algorithm

The STEP algorithm has 13 properties — with the default values given in the parentheses — accessible in the jobOptions:

Tolerance (10^{-5}) Double-precision floating-point number used for controlling the accuracy of the propagation, a low error tolerance giving a high accuracy. The error tolerance is described in Ref. [11]. The default value is 10^{-5}, corresponding to an accuracy of around 10^{-7}.

EnergyLoss (true) Boolean switch for including the energy loss into the propagation.

MostProbableEnergyLoss (true) Boolean switch for using the most probable energy loss instead of the mean value.

Straggling (true) Boolean switch for adding the energy loss fluctuations to the covariance matrix. The straggling is only included in combination with the energy loss.

StragglingScale (6) Double-precision floating-point number used for scaling the Landau width of the energy loss fluctuations.

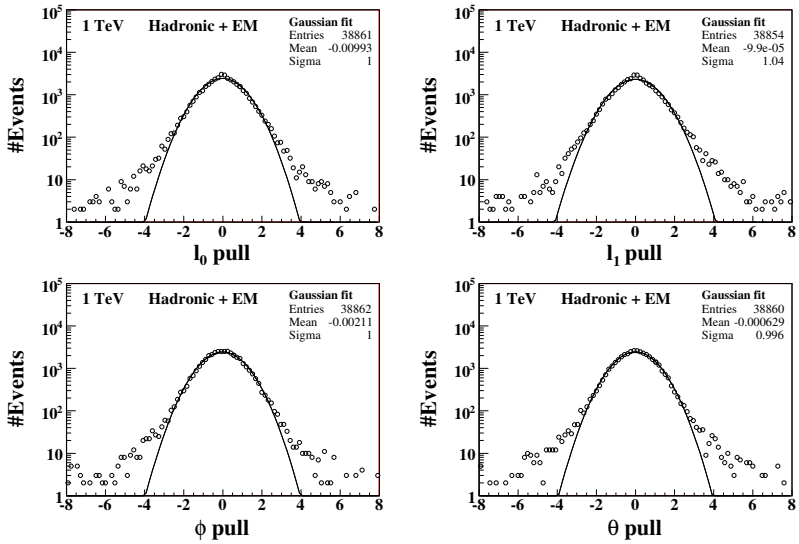


Figure 15: Pulls of 1 TeV muons subjected to multiple scattering in the ATLAS hadronic and electromagnetic calorimeters, and their Gaussian fits (solid lines). The track residuals are simulated by GEANT4 in the full detector description, while the covariance scattering contribution is found by STEP in the dense volumes of the calorimeter tracking geometry.

MultipleScattering (true) Boolean switch for adding the multiple scattering contribution to the covariance matrix.

MultipleScatteringScale (1) Double-precision floating-point number used for scaling the width of the elements of the multiple scattering covariance contribution.

MaterialEffects (true) Boolean switch for excluding the material interactions of energy loss, straggling and multiple scattering from the propagation. When false, this switch overrides the MultipleScattering, EnergyLoss and Straggling switches, setting them all to false, whereas when true, control is left to the individual switches.

IncludeBgradients (true) Boolean switch for including the magnetic field gradients $\partial B_i/\partial x_j$ into the error propagation. The magnetic field gradients are described in Ref. [14].

IncludeGgradient (false) Boolean switch for including the energy loss gradient $\partial g/\partial \lambda$ into the error propagation. The energy loss gradient is described in Ref. [14].

MomentumCutOff (50 MeV) Double-precision floating-point number giving the smallest acceptable momentum of a propagated particle. If the energy loss is switched on and the momentum goes below this value, the propagation is stopped. The Bethe-Bloch calculation breaks down at low momenta, hence the default cut-off is set to 50 MeV.

MaxPath (100 000 mm) Double-precision floating-point number defining the maximum allowed path length of the propagation, given in mm.

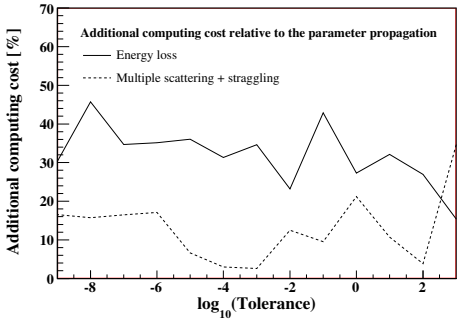


Figure 16: Additional computing cost — relative to the STEP parameter propagation — of including the energy loss, multiple scattering and straggling.

MaxSteps (10 000) Double-precision floating-point number giving the maximum allowed number of steps during the propagation.

B Multiple scattering λ variance

The λ variance comes from the variation in the energy loss caused by the variation of the path length due to the multiple scattering. The new path s' and the spatial scattering angle θ_s are illustrated in Fig. 17.

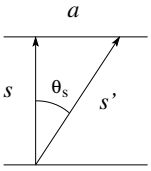


Figure 17: The changed trajectory s' caused by the multiple scattering of a particle going through a layer.

To calculate the λ variance we start by observing that

$$d\lambda = \frac{d\lambda}{ds} ds \quad \text{with} \quad ds = s' - s \tag{24}$$

and

$$s'^2 = s^2 + a^2 \approx s^2 + (s\theta_s)^2 = s^2(1 + (\theta_s)^2) \tag{25}$$

$$s' = s\sqrt{1 + (\theta_s)^2} \approx s(1 + \frac{1}{2}(\theta_s)^2) = s + \frac{1}{2}s(\theta_s)^2 \tag{26}$$

giving

$$ds = s' - s = \frac{1}{2}s(\theta_s)^2 \quad \text{and} \quad d\lambda = \frac{d\lambda}{ds}\frac{1}{2}s(\theta_s)^2 \tag{27}$$

The λ variance then becomes

$$\text{Var}(\lambda) = \langle (d\lambda)^2 \rangle = \frac{1}{4}s^2 \left(\frac{d\lambda}{ds}\right)^2 \langle (\theta_s)^4 \rangle \tag{28}$$

which contains the fourth central moment of θ_s, defined as

$$\langle(\theta_s)^4\rangle = \langle(\theta_s-\mu)^4\rangle = \frac{1}{\sqrt{2\pi\sigma_s^2}}\int_{-\infty}^{\infty}(\theta_s-\mu)^4 e^{-\frac{(\theta_s-\mu)^2}{2\sigma_s^2}}d\theta_s \tag{29}$$

Noting that the mean angular deflection in a plane caused by the multiple scattering is zero ($\mu = 0$), and that the integral is symmetric, we get

$$\langle(\theta_s)^4\rangle = \frac{2}{\sqrt{2\pi\sigma_s^2}}\int_{0}^{\infty}\theta_s^4 e^{-\frac{\theta_s^2}{2\sigma_s^2}}d\theta_s \tag{30}$$

Solving this integral gives

$$\langle(\theta_s)^4\rangle = \frac{2}{\sqrt{2\pi\sigma_s^2}}\frac{1}{2}\left(\frac{1}{2\sigma_s^2}\right)^{-\frac{5}{2}}\Gamma\left(\frac{5}{2}\right) \tag{31}$$

Putting in $\Gamma(5/2) = 3\pi/4$ and contracting, we get

$$\langle(\theta_s)^4\rangle = 3\sigma_s^4 \tag{32}$$

Going through the same procedure to find the second central moment of θ_s, also known as the θ_s variance, we reassuringly find that

$$\langle(\theta_s)^2\rangle = \mathrm{Var}(\theta_s) = \sigma_s^2 \tag{33}$$

From this, and the fact that $\mathrm{Var}(\theta_s) = 2\mathrm{Var}(\theta_p)$, where θ_p is the projected scattering angle, we conclude that

$$\langle(\theta_s)^4\rangle = 3\sigma_s^4 = 3(\sigma_s^2)^2 = 3(\mathrm{Var}(\theta_s))^2 = 3(2\mathrm{Var}(\theta_p))^2 = 12(\mathrm{Var}(\theta_p))^2 \tag{34}$$

The λ variance (28) finally becomes

$$\mathrm{Var}(\lambda) = \frac{1}{4}s^2\left(\frac{d\lambda}{ds}\right)^2\langle(\theta_s)^4\rangle = 3s^2\left(\frac{d\lambda}{ds}\right)^2(\mathrm{Var}(\theta_p))^2 \tag{35}$$

which is the expression found in the covariance contribution (18).

References

[1] The ATLAS Collaboration, G. Aad *et al.*, JINST **3** (2008) S08003.

[2] Particle Data Group, C. Amsler *et al.*, Phys. Lett. B **667** (2008).

[3] H. A. Bethe and W. Heitler, Proc. R. Soc. London A **146** (1934) 83.

[4] G. Molière, Z. Naturforsch. **3A** (1948) 78.

[5] H. A. Bethe, Phys. Rev. **89** (1953) 1256.

[6] V. Highland, Nucl. Instr. Meth. **129** (1975) 497.

[7] S. Agostinelli *et al.*, Nucl. Inst. and Meth. A **506** (2003) 250.

[8] D. Stampfer *et al.*, Comp. Phys. Comm. **79** (1994) 157.

[9] D. E. Groom *et al.*, Atomic Data and Nuclear Data Tables **78** (2001) 183–356.

[10] L. Bugge and J. Myrheim, Nucl. Inst. and Meth. **179** (1981) 365.

[11] E. Lund *et al.*, Track parameter propagation through the application of a new adaptive Runge-Kutta-Nyström method in the ATLAS experiment, ATLAS Public Note ATL-SOFT-PUB-2009-001, submitted to Journal of Instrumentation, 2009.

[12] A. Salzburger *et al.*, The ATLAS Tracking Geometry Description, ATLAS Public Note ATL-SOFT-PUB-2007-004, 2007.

[13] D. López Mateos *et al.*, A Parameterization of the Energy Loss of Muons in the ATLAS Tracking Geometry, ATLAS Public Note ATL-MUON-PUB-2008-002, 2008.

[14] E. Lund *et al.*, Transport of covariance matrices in the inhomogeneous magnetic field of the ATLAS experiment by the application of a semi-analytical method, ATLAS Public Note ATL-SOFT-PUB-2009-002, submitted to Journal of Instrumentation, 2009.

[15] R. Frühwirth and M. Regler, Nucl. Inst. Meth. A **456** (2001) 369.

[16] H. Eichinger and M. Regler, Review of track fitting methods in counter experiments, CERN Yellow Report 81-06, 1981.

[17] F. Åkesson *et al.*, The ATLAS Tracking Event Data Model, ATLAS Public Note ATL-SOFT-PUB-2006-004, 2006.

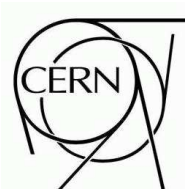
[18] A. Strandlie and W. Wittek, Nucl. Inst. and Meth. in Phys. Res. A **566** (2006) 687.

IV

ATLAS NOTE

ATL-SOFT-PUB-2008-005

January 23, 2009



Refitting of combined inner detector and muon spectrometer tracks from Monte Carlo samples by using the Kalman fitter and the STEP algorithm in the ATLAS experiment

E. Lund[*,1], L. Bugge[1], T. G. Cornelissen[2], M. Elsing[2], I. Gavrilenko[2,3], W. Liebig[4], D. López Mateos[5,6], A. Salzburger[2,7], A. Strandlie[1,8], S. Todorova[9]

[1] *University of Oslo, Oslo, Norway*
[2] *CERN, Geneva, Switzerland*
[3] *P. N. Lebedev Institute of Physics, Moscow, Russia*
[4] *NIKHEF, Amsterdam, The Netherlands*
[5] *Columbia University, New York, USA*
[6] *California Institute of Technology, Pasadena, USA*
[7] *DESY, Hamburg, Germany*
[8] *Gjøvik University College, Gjøvik, Norway*
[9] *Tufts University, Boston, USA*
[*] corresponding author (esben.lund@fys.uio.no)

ATL-SOFT-PUB-2008-005
23 January 2009

Abstract

In this paper we refit combined muon tracks using the Kalman fitter and the *simultaneous track and error propagation* (STEP) algorithm of the ATLAS tracking software. The muon tracks are simulated by GEANT4 in the full detector description, reconstructed by MUID, and refitted by the Kalman fitter in the ATLAS TrackingGeometry. The relative transverse momentum resolution of the refitted tracks is compared to the resolution of the refits done by the global χ^2 track fitter, along with the resolution found by the MUID and STACO muon combination algorithms. Reconstructed invariant masses are compared in a similar way.

1 Introduction

Experimental particle physics is on the verge of a new era, heralded by the Large Hadron Collider being commissioned at the European Organization for Nuclear Research — CERN — located just outside Geneva, Switzerland. The LHC accelerator will collide protons at a center of mass energy of 14 TeV, opening up a new window for particle discoveries and precision measurements of existing theories. Particle detectors are located at four beam crossings along the LHC, one of which houses the ATLAS detector [1]. This is the largest of the LHC experiments, employing a great variety of detector and magnetic field technologies to identify a wide range of particles. The complex magnetic field and high collision rate, however, make the reconstruction of particle tracks very challenging. Things are complicated further by the relatively big amount of material within ATLAS, generating considerable disturbances to the particle tracks through material interactions such as energy loss and multiple scattering.

Track reconstruction is mostly done in the inner detector and muon spectrometer, while the calorimeters match energy deposits to the reconstructed tracks and to tracks that go undetected in the inner detector, such as photons and neutrons. In addition to the standalone inner detector and muon spectrometer tracks, the track reconstruction matches these tracks to construct the combined muon tracks. To guarantee the quality of the reconstructed tracks, competing algorithms exist for performing individual tasks in the track reconstruction process. For the combined muon reconstruction, two of the most prominent algorithms are MUID [2] and STACO [3].

As the detector calibration and material description are better understood, old data might be improved through reconstruction with an upgraded material description, which is handled by the so-called TrackingGeometry [4] within the new ATLAS tracking realm [5]. Running the complete reconstruction chain is quite computing costly. However, it might be sufficient to only refit the existing tracks using the new detector information in some cases, thereby reducing the computing cost significantly. The ATLAS reconstruction mainly uses two mathematically equivalent, yet different approaches to track refitting; the global χ^2 track fit [6] and the Kalman filter [7], implemented in the ATLAS global χ^2 track fitter [8] and Kalman fitter, respectively. The global χ^2 fitter handles the material interactions in the final track fit, whereas the Kalman fitter incorporates the material interactions at points along the track or continuously during the track propagation. Hence, material interactions can be integrated in two ways; as point-like corrections to the track trajectory in detectors that have a discrete material distribution — such as the Silicon layers of the ATLAS inner detector — or by incorporating the material interactions as continuous corrections during the propagation process itself. The latter method is carried out by the newly developed STEP propagation algorithm [9–11] and is well suited to the transport of track parameters through the dense material of the ATLAS detector. In this note, we focus on the effect of using STEP to propagate through the muon spectrometer and calorimeters, and we investigate the quality of the refitted muon tracks in such cases. Specifically, we present the relative transverse momentum (p_T) resolution of the Kalman and global χ^2 track refitting algorithms, along with the resolution of the STACO and MUID combined muons. The combined muon tracks reconstructed by STACO and MUID are quite similar, but only MUID performs a track fit, hence we choose to restrict the track refitting by the Kalman and global χ^2 track fitters to the MUID combined muons. Reconstructed invariant masses are compared in a similar way.

In this paper, we start out by describing two of the most prominent ATLAS muon combination algorithms; MUID and STACO, along with the Kalman and global χ^2 track fitters, in Sections 2 and 3. Furthermore, we introduce the Monte Carlo samples and quality selection cuts in Section 4, followed by the relative p_T resolution and invariant mass reconstruction in Section 5. Finally, we present the conclusion in Section 6.

Natural units ($\hbar = c = 1$) are used throughout this paper, and all results are produced with ATLAS offline software release 14.4.0, unless otherwise stated.

2 Reconstruction of the combined inner detector and muon spectrometer tracks

Generally speaking, the ATLAS detector can be divided into three parts; the inner detector, the calorimeters and the muon spectrometer. The reconstruction of the interaction point, secondary vertices and most tracks is done by the inner detector located in the center of the detector. Covering the inner detector, we find the calorimeters stopping most particles — except muons — while measuring their energy deposits. Outside of the calorimeters we find the second tracking device of ATLAS; the muon spectrometer.

In this paper we focus on muons coming from the interaction point with sufficient momentum to penetrate the calorimeters and pass through the muon spectrometer, allowing a combined reconstruction by using information from both tracking devices; the inner detector and muon spectrometer. This covers most muons coming from Z, W and H, and excludes low-p_T muons stopped in the calorimeters, or muon spectrometer, and muons created through decay processes in the inner detector or calorimeters. The inner detector and muon spectrometer are complementary in the sense that the inner detector has a good p_T resolution for muon tracks with a p_T up to 50 GeV, while the muon spectrometer covers the remaining p_T range, giving the combined muons a good p_T resolution over the whole momentum range.

The ATLAS software contains several algorithms for reconstructing combined muons. Here we concentrate on two of the most established algorithms; MUID [2] and STACO [3]. Both algorithms pair inner detector (ID) and muon spectrometer (MS) tracks by using the match chi-square to identify the combined muons;

$$\chi^2_{\text{match}} = (\boldsymbol{\xi}_{\text{MS}} - \boldsymbol{\xi}_{\text{ID}})^{\text{T}} (\boldsymbol{\Sigma}_{\text{MS}} + \boldsymbol{\Sigma}_{\text{ID}})^{-1} (\boldsymbol{\xi}_{\text{MS}} - \boldsymbol{\xi}_{\text{ID}}) \tag{1}$$

where $\boldsymbol{\xi}$ are the five local track parameters defined in the ATLAS event data model [12] — expressed at the point of the closest approach to the beam line — and $\boldsymbol{\Sigma}$ is the track parameter covariance matrix at the same point. The match chi-square requires the propagation of the muon spectrometer track through the calorimeters, taking material interactions into account. The above track matching decides which pairs of inner detector and muon spectrometer tracks to combine.

Upon matching the inner detector and muon spectrometer tracks, STACO does a statistical combination to produce the combined track parameters at the point of the closest approach to the beam line;

$$\boldsymbol{\xi}^{\text{STACO}}_{\text{combined}} = (\boldsymbol{\Sigma}^{-1}_{\text{ID}} + \boldsymbol{\Sigma}^{-1}_{\text{MS}})^{-1} (\boldsymbol{\Sigma}^{-1}_{\text{ID}} \boldsymbol{\xi}_{\text{ID}} + \boldsymbol{\Sigma}^{-1}_{\text{MS}} \boldsymbol{\xi}_{\text{MS}}) \tag{2}$$

MUID, on the other hand, fits the combined track, starting from the inner track fit and adding the points from the muon spectrometer track.

As of today, MUID is the only combined muon reconstruction algorithm — discussed in this note — which incorporates the calorimeter energy loss measurements into the track fitting. This feature has, however, been replaced by a parameterized calorimeter energy loss to better compare MUID to the other algorithms presented here.

3 The Kalman and global χ^2 track fitters

The Kalman filtering approach [7] to track fitting is a progressive method in the sense that the track measurements are added to the track one at a time, Fig. 1. This linear dependency on the number of track measurements lends itself well to the refitting of combined muon tracks, which often contain close to 100 measurements. Mathematically, the Kalman fitter performs a χ^2 minimization, equivalently to the global χ^2 fitter [8]. This method, however, combines all measurements into one big calculation to minimize the χ^2. Both methods have their strengths and weaknesses; the Kalman fitter is quick but might be steered off course by bad measurements, especially early in the fitting process. The global χ^2 fit is

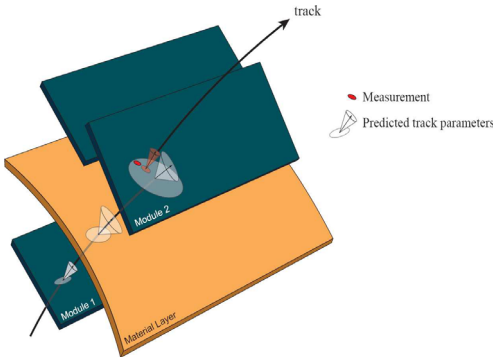


Figure 1: Illustration of a typical Kalman filtering step. The track parameters associated to module 1 are propagated through a material layer onto module 2 to find the predicted track parameters at this module. In the Kalman filter formalism, the weighted mean between the prediction and the associated measurement form the updated track parameters, which are the starting point of the next filtering step. (From Ref. [13]).

less sensitive to bad measurements — often called outliers — but more computing costly, particularly for tracks containing many measurements.

The prediction step is an important part of the Kalman filtering, which is why a dedicated propagator, STEP [9–11], has been developed with this in mind. In this note, the STEP algorithm — encapsulated within the ATLAS extrapolator tool [13] — propagates the track parameters and the associated covariance matrix through the dense volumes of the muon spectrometer and calorimeter tracking geometry [4], taking the material interactions continuously into account. These interactions introduce noise into the filtering process which has to be well understood in order to produce a good fit. In the inner detector, where the detector material is mostly layered, the Kalman fitter uses another ATLAS propagator — the RungeKuttaPropagator, originally part of the xKalman package [14] — for the prediction step, along with point-like updates to the track parameters from the material interactions at these layers. The outermost part of the inner detector (the TRT) can also be incorporated into the track propagation as dense volumes, which is not done here.

4 Monte Carlo samples and quality selection cuts

The primary goal of this paper is to find out how the Kalman fitter — employing the STEP algorithm — matches up to the other methods of finding the combined muon track parameters, hence the source of the muons is of less importance. Here we use Monte Carlo samples of $J/\psi \to \mu^+\mu^-$, $Z \to \mu^+\mu^-$, $H(600\text{ GeV}) \to ZZ \to 4l$ and single muons ($p = 1$ TeV and $p_T = 10, 100, 1000$ GeV) — simulated by GEANT4 [15] in the full detector description — to cover a wide range of muon momenta. The H and the $p = 1$ TeV single muons are produced with ATLAS software release 12.0.6, while the rest of the samples are simulated with release 13.0.40. All of the samples are simulated in a misaligned detector description, which is mirrored in the reconstruction software to compensate for the misalignment. This correction might, however, be less than perfect, giving reconstructed tracks of worse resolution than those of an ideal detector description.

A consistent set of quality selection cuts for the reconstructed tracks has been used throughout this note; requiring at least seven hits in the pixel and SCT layers of the inner detector, and demanding that

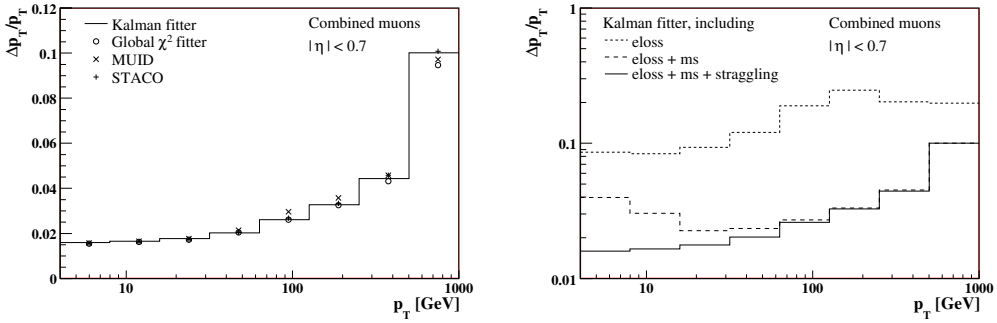


Figure 2: Relative transverse momentum resolution as a function of p_T in the barrel region for different methods (left) and configurations of the STEP algorithm within the Kalman fitter (right). The Kalman and global χ^2 refitting is applied to the MUID tracks. The straggling refers to the energy loss fluctuations.

the transverse and longitudinal impact parameters at the point of closest approach to the primary vertex must fulfill $|d_0| < 2$ mm and $|z_0 \sin\theta| < 10$ mm respectively, where θ is the polar angle of the track.

5 Relative transverse momentum resolution and invariant mass reconstruction

Since we are only doing track refitting — and not track finding — the impact on particle discovery related parameters, such as track finding efficiencies, fake rates, and exclusion and discovery limits, is minimal. Improvements to the reconstructed invariant mass resolutions and other precision measurements might, however, be seen. To evaluate the quality of the track fits, we study the relative p_T resolution, defined as the width of the Gaussian fit restricted to 3 sigma of the peak;

$$\frac{\Delta p_T}{p_T} = \frac{1/p_T^{\text{reco}} - 1/p_T^{\text{true}}}{1/p_T^{\text{true}}} = \frac{p_T^{\text{true}} - p_T^{\text{reco}}}{p_T^{\text{reco}}} \tag{3}$$

The transverse momentum p_T is used for invariant mass reconstruction and for estimating the missing transverse energy, among other things. Reconstructing the p_T is particularly challenging, making the p_T resolution the most common yardstick for evaluating the quality of a track fitting method. The accuracy of the direction measurements are, on the contrary, typically much better than that required by any physics analysis.

Figure 2 shows the relative p_T resolution of the combined muons from all of the datasets as a function of p_T in the barrel region ($|\eta| < 0.7$), while Fig. 3 shows the relative p_T resolution as a function of $|\eta|$, averaged over ϕ, for three transverse momenta. All of the methods are in good agreement with each other, producing reasonable results.

Figure 4 shows the reconstructed invariant masses of J/ψ, Z and $H(600\text{ GeV})$ from the combined muons of their respective Monte Carlo samples. The invariant masses are reconstructed from the two (or four) most energetic muons of the event regardless of their summed charge, only requiring all of the muons to be located in the barrel region. The Higgs mass is reconstructed without any Z-mass constraint. The reconstructed invariant masses of the respective methods of Fig. 4 are quite similar because the quality of the invariant mass reconstruction relies heavily on the momentum resolution of the combined muons, which is quite similar for all of the muon combination methods discussed here. The non-Gaussian tails are partially due to radiative decays, but mostly due to muons poorly measured in certain regions of

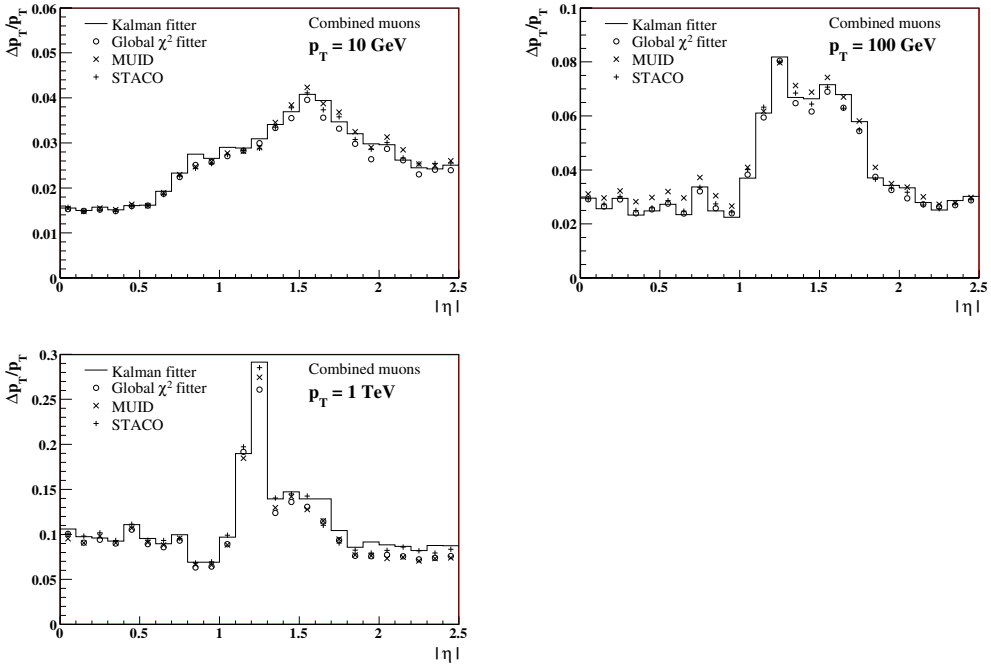


Figure 3: Relative transverse momentum resolution as a function of $|\eta|$, averaged over ϕ, for different methods, at $p_T = 10, 100$, and 1000 GeV, within the $|\eta| < 2.5$ acceptance of the inner detector. The Kalman and global χ^2 refitting is applied to the MUID tracks.

the muon spectrometer, while the lower reconstructed masses — with respect to the true masses — are mostly caused by radiative energy loss during the initial decay and radiative energy loss of the individual muons. The radiative energy loss of the muons is particularly hard to reconstruct due to its non-Gaussian, Landau distribution. The asymmetric Landau distribution has a tail towards high energy loss, which grows with increasing muon momenta. In the combined muon reconstruction, it is common practice to transport the muon track parameters through the calorimeters assuming the most probable energy loss of the Landau distribution — to optimize the momentum resolution — thereby underestimating the true energy loss and initial momenta of the muons in many cases. These underestimated momenta are then passed on to the reconstruction of the invariant masses, giving lower than expected masses. Since the radiative energy loss of muons rises rapidly with increasing momenta, the negative shift in the reconstructed mass is particularly big in the H sample of Fig. 4. In some cases, the reconstructed mass can be improved through various combinatorial techniques, such as mass and charge constraints, none of which are applied here.

6 Conclusion

In this paper we have shown that the p_T resolution of combined muons refitted with the Kalman fitter — employing the STEP algorithm — and the invariant masses reconstructed from these muons, are on par with the p_T resolution of the combined muon refits, and reconstructed invariant masses, found by the

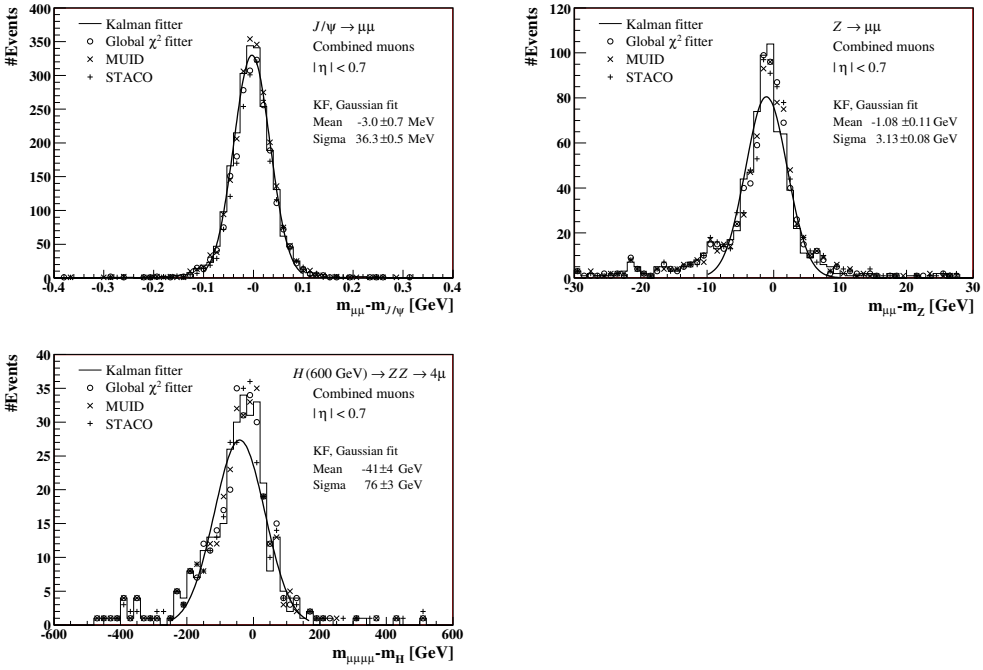


Figure 4: Reconstructed invariant masses of $J/\psi \to \mu\mu$, $Z \to \mu\mu$ and $H(600\,\text{GeV}) \to 4\mu$ from the combined muons of their respective Monte Carlo samples. The invariant masses are reconstructed from the two (or four) most energetic muons of the event regardless of their summed charge, only requiring all of the muons to be located in the barrel region. The Gaussian fit is done on the invariant masses reconstructed from the combined muons of the Kalman fitter.

global χ^2 fitter and the results obtained by the STACO and MUID muon combination algorithms.

The Kalman fitter's main strengths lie in its straightforward way of including the material interactions, and the fact that the computing time scales linearly with the number of track measurements, which is particularly useful when refitting tracks with many measurements attached, such as the combined muons. These factors cause the Kalman fitter to be around two times faster than the global χ^2 fitter when refitting the combined muons of the Monte Carlo samples presented here. The timing does, however, rely on many things, such as the ATLAS software release version and material description of the detector.

The refitting of combined muons with the Kalman fitter is partly based on new non-optimized software — such as the dense volume description of the ATLAS calorimeters in the tracking geometry and the energy loss fluctuations of the STEP algorithm — hence there is still room for improvements.

7 Acknowledgements

This work has been carried out as part of the developments of the ATLAS tracking group. We would like to thank our colleagues for their help in integrating the software and for their support in preparing this note.

References

[1] The ATLAS Collaboration, G. Aad *et al.*, JINST **3** (2008) S08003.

[2] Th. Lagouri *et al.*, IEEE Trans. Nucl. Sci. **51** (2004) 3030–3033.

[3] S. Hassini *et al.*, NIM **A572** (2007) 77–79.

[4] A. Salzburger *et al.*, The ATLAS Tracking Geometry Description, ATLAS Public Note ATL-SOFT-PUB-2007-004, 2007.

[5] T. Cornelissen *et al.*, Concepts, Design and Implemenentation of the ATLAS New Tracking (NEWT), ATLAS Public Note ATL-SOFT-PUB-2007-007, 2007.

[6] R. Frühwirth *et al.*, Data analysis techniques for high-energy physics, (Cambridge University Press, Cambridge, 2000).

[7] R. Frühwirth, Nucl. Inst. and Meth. A **262** (1987) 444.

[8] T. Cornelissen *et al.*, J. Phys.: Conf. Ser. **119** (2008).

[9] E. Lund *et al.*, Track parameter propagation through the application of a new adaptive Runge-Kutta-Nyström method in the ATLAS experiment, ATLAS Public Note ATL-SOFT-PUB-2009-001, submitted to Journal of Instrumentation, 2009.

[10] E. Lund *et al.*, Transport of covariance matrices in the inhomogeneous magnetic field of the ATLAS experiment by the application of a semi-analytical method, ATLAS Public Note ATL-SOFT-PUB-2009-002, submitted to Journal of Instrumentation, 2009.

[11] E. Lund *et al.*, Treatment of energy loss and multiple scattering in the context of track parameter and covariance matrix propagation in continuous material in the ATLAS experiment, ATLAS Public Note ATL-SOFT-PUB-2008-003, 2008.

[12] F. Åkesson *et al.*, The ATLAS Tracking Event Data Model, ATLAS Public Note ATL-SOFT-PUB-2006-004, 2006.

[13] A. Salzburger, The ATLAS Track Extrapolation Package, ATLAS Public Note ATL-SOFT-PUB-2007-005, 2007.

[14] I. Gavrilenko, Description of Global Pattern Recognition Program (XKALMAN), ATLAS Note ATL-INDET-97-165, 1997.

[15] S. Agostinelli *et al.*, Nucl. Inst. and Meth. A **506** (2003) 250.