MULTI-ANODE PHOTON-MULTIPLIER READOUT ELECTRONICS FOR THE LHCb RING IMAGING CHERENKOV DETECTORS

Nigel John Smale

St Anne's College

A thesis submitted for the degree of Doctor of
Philosophy at the University of Oxford

Michaelmas Term, 2004

I

Abstract

A readout system for the Ring Imaging CHerenkov (RICH) detectors of the LHCb experiment has been developed. Two detector technologies for the measurement of Cherenkov photons are considered, the Multi-Anode Photo-Multiplier Tube (MAPMT) and the Hybrid Photon Detector (HPD), both of which meet the RICH requirements. The properties of the MAPMT are evaluated using a controlled single-photon source; a pixel-to-pixel gain variation of ~3 and a typical signal to noise of ~20 is measured. The relative tube efficiency is found to be reduced by ~26 % due to the detailed focusing structure of the MAPMT device.

A radiation hard application-specific integrated circuit (ASIC) chip, the Beetle1.2MA0, has been developed to capture and store signals from a pair of MAPMTs. The Beetle1.2MA0 is built on the architecture of the Beetle family that was designed for silicon strip detectors, the difference being a modified front-end amplifier. The 128 input-channels of the Beetle1.2MA0 have a charge-sensitive pre-amplifier followed by a CR-RC pulse shaper, which is sampled into an analogue memory with a depth of up to 160 time slots at 40 MHz. The internal bias generator module of the Beetle has also been developed and tested. This module generates all the required voltage and currents required internally to the chip, and consists of voltage DACs, current DACs, a current-source and a voltage source. Beetle1.2MA0 evaluation chips were fabricated, and the performance when connected to a MAPMT is demonstrated in terms of pulse profile, noise, pulse spillover and overshoot. The Beetle1.2MA0 ASIC is found to be well suited to LHCb RICH readout requirements.

A demonstrator system of the RICH readout electronics has been developed, which will serve either HPD or, with some small adaptation, MAPMT applications. Multimode fibres driven by vertical cavity surface emitting lasers (VCSEL) devices are used to transmit event-blocks of data in 900ns from the on-detector (Level_0 region) to the Level_1 region, which is located off-detector in a non-radiation environment. The data words are captured into 9-Mbit Quad Data Rate (QDR) static RAM at a rate of up to 160 MHz. The scheme for Level_1 data receipt and buffering is described, and results from a full system test are presented.

Acknowledgements

Amongst the numerous people who have helped me throughout this D.Phil, I would like to mention but a few that I am particularly indebted to:

Neville Harnew for providing me with the opportunity to work with the LHCb group, for taking care of the administrative side of my D.Phil, for his scientific guidance and for his continuous support.

John Bibby for sharing his wealth of engineering knowledge, for giving up his time to help with the technical detail and for his uplifting personality.

Marco Adinolfi and Stig Topp-Jorgensen, for their help and their combined efforts to construct, maintain, operate and write endless programming code for the hardware in the Oxford LHCb RICH lab.

Johan Fopma, Colin Perry, Barney Brooks and Pete Shield, for supporting my endeavor to gain this D.Phil and for politely answering endless queries.

Everybody at the ASIC lab, Heidelberg Germany, for sharing their extensive technical and theoretical ASIC knowledge, and for their friendship.

I would like to give special thanks to Stefanie, for being my wife and giving a reason to smile every day. To Georgie for giving me the inspiration to start my DPhil. To Mia for giving me the inspiration to finish my DPhil. To my mother and father for always being there and for supporting my alternative route through life.

On a lighter note I would also like to thank Eddie for being the oracle. Niels for sharing his off road jump skills. Martin for the RTFM advice. Sven for having too many beers in his wine, and for all of his help. Daniel for writing in English. Harald for taking me to a garage so that I could get rid of 350 euros. Rainer and Sylvia for supplying an excellent place of study, with the necessary food for thought. Timo for paying my fines. Marco for making my plots blue. Stig for giving the data that I needed. Charlotte for grouping my captions and Mike, and I am sure Lois too, for enjoying my research.

Table of Contents

List of Figures

Preface

The LHCb high-energy physics experiment is under development at CERN and will measure *CP*-violation in the B-meson system with very high precision. The experiment makes use of two Ring Imaging CHerenkov (RICH) detectors for particle identification. Two choices of RICH photon detectors have been proposed; the so-called Hybrid Photon Detector (HPD) being the baseline choice and the Multi-Anode Photo-Multiplier Tube (MAPMT) being the back-up. The HPD and MAPMT have effective pixel sizes of 2.5 mm^2 and 2 mm^2 respectively, corresponding to a total of ~325k and ~440k signal channels to be captured, temporarily stored, and read out. At the time of writing the HPD was adopted as the chosen photon detector.

To reduce the readout bandwidth and the amount of data that needs to be finally stored, LHCb uses three tiers of trigger and data storage: Level_0, Level_1 and the high level (HL). The trigger system processes the event data at each storage level and makes an accept or reject decision. The Level_0 region is physically located on the LHCb detector, while Level_1 and HL are some 100 m away but still within the experiment hall. When designing the RICH instrumentation system, special design considerations have to be made for electronic components on or near the detector to ensure radiation tolerance, as they will be subjected to high particle fluxes. Therefore the RICH electronics components used for Level_0 have either been selected or designed to withstand the expected radiation level of 3 krad per year.

For the capture and storage of signals from the MAPMT, the 128-channel Beetle1.2MA0 front-end readout chip is used. The HPD readout utilizes the 1024 channel LHCBPIX1 chip, designed at CERN. Both devices are Application Specific Integrated Circuits (ASICs) with charge-sensitive input channels and 4 µs deep memories. Data capture rates are at 40 MHz in both cases.

This thesis describes the study of the MAPMT for its suitability of use as the RICH photon detector. The development of the Beetle1.2MA0 readout chip and its evaluation is fully detailed. In addition, the development of the RICH readout system that transmits, and stores, event data from the Level_0 to Level_1 regions is described. Particular detail is given to the Level_1 memory region. Although the Level_0 to Level_1 readout chain has primarily been

designed for reading out the HPD, it is also compatible with the readout of the MAPMT with minor modifications.

In this thesis the author gives a general overview of the LHCb experiment and detector specifications in Chapter 1. Chapter 2 describes the MAPMT and its properties; a detailed account of the author's evaluation of a Hamamatsu R5900-00-M64 H7546B MAPMT is given. Chapter 3 discusses the electronic requirements for reading out an MAPMT and presents arguments why the Beetle ASIC chip was selected for this purpose. The principles underlying the design of a RICH readout chip to optimize for radiation tolerance and electronic noise are also given. Chapter 4 gives an account of a core module, the bias generator, which was designed and tested by the author. This module is internally used for biasing the Beetle chip. In order to make the Beetle chip compatible with RICH specifications, the author re-designed the Beetle front-end amplifiers. This necessitated a full study of candidate front-end amplifier types and resulted in the Beetle1.2MA0 chip submission. Chapter 5 describes the amplifier design flow in schematic and layout, and highlights the successful test results of the chip. Chapter 6 describes the upstream Level_1 readout system, with particular attention given to data storage. The author's contribution was to select the Level_1 hardware and develop algorithms to capture and buffer data received from the Level_0 region. The author collaborated with the University of Cambridge in this work. Finally a summary is given in Chapter 7.

Chapter 1.

The LHCb experiment

Since 1993 physicists and engineers have been working on the circular proton-proton (pp) collider, named the Large Hadron Collider (LHC), which is situated at CERN Geneva. The LHC will make use of the existing 27 km circumference Large Electron Positron (LEP) collider tunnel. Each of the four interaction regions houses a major detector. The bunch crossings will occur at a nominal rate of 40 MHz, a centre-of-mass energy of $\sqrt{s}$ =14 TeV and a design luminosity of the order of 10^{34} cm^{-2} s^{-1}. For precision measurements of *CP*-violation and rare decays of B-mesons, a high-energy particle detector named LHCb is being built. Particle identification is an essential requirement of LHCb and is achieved with two Ring Imaging CHerenkov (RICH) detectors. This chapter briefly describes the LHCb detector, the global readout electronics and synchronisation scheme that must be adhered to, and the trigger scheme that is used to either accept or reject stored data.

1.1   The LHCb detector

The LHCb detector [TDRp][TDR03opt], situated in the existing DELPHI pit at the Intersection-8 experimental area of the LHC, will be operational around summer 2007. A side view of the 19.7 m long detector is shown in Figure 1-1. At LHC energies, B-mesons generally emerge from collisions at angles close to the beam direction. This motivates a forward detector geometry with angular coverage between 10 and 300 mrad. Referring to Figure 1-1, LHCb has a Silicon VErtex LOcator (VELO), a Pile-up Veto detector (VETO), a Trigger Tracker (TT), a 4 Tm dipole magnet, two Ring Imaging Cherenkov detectors (RICH-1 & -2), three tracking stations (T1-T3), a Scintillating Pad Detector (SPD), a Pre-Shower (PS), an Electromagnetic CALorimeter (ECAL), a Hadronic CALorimeter (HCAL) and a five-station Muon system (M1-M5).
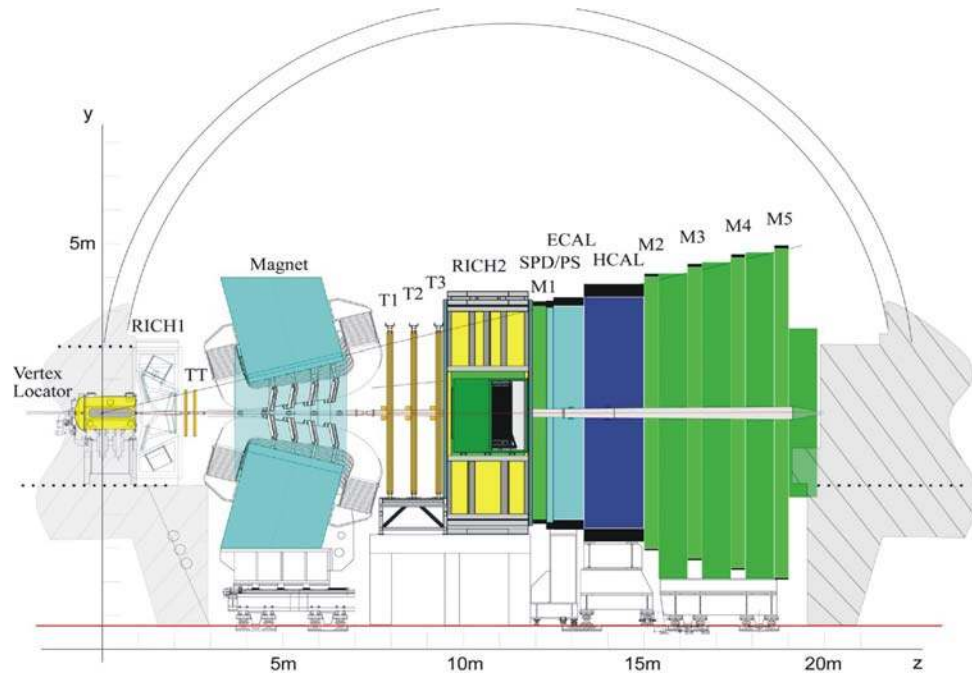
Figure 1-1 The side view layout of the LHCb spectrometer showing the VELO, the two RICH counters, the four tracking stations TT and T1-T3, the 4 Tm dipole magnet, the Scintillating Pad detector (SPD), Preshower (PS), Electro-magnetic (ECAL) and Hadronic (HCAL) Calorimeters, and the five muon stations M1-M5.

The LHCb coordinate system is such that z points along the detector axis, x points towards the centre of the LHC, and y is vertically upwards. The interaction point is at (0,0,0).

The interaction rate

For LHCb to make valid trigger decisions and reconstruct tracks with high efficiency, the detector occupancy must be kept low. A high occupancy will likely result when there is more than one pp interaction per bunch crossing. Multiple pp interactions would also cause ambiguity as to which primary vertex a specific B meson originated. Furthermore, the demands on the readout electronics would be increased and the tracking algorithms would have to deal with an increased particle density.

LHCb will operate with an average luminosity of 2 x 10^{32} cm^{-2} s^{-1} in order to have only a single pp interaction within a bunch crossing. The number of n pp interactions in each bunch-crossing will follow a Poisson distribution

$$P_n = \frac{\mu^n}{n!} e^{-\mu} \quad \text{with a mean } \mu = \frac{L\sigma}{f} \;,$$

Equ 1-1

where L is the luminosity, f is the bunch-crossing frequency and σ is the inelastic pp cross-section, which is ~80 mb at LHCb energies[1] [CHA03]. As only 2622 bunch crossings from the possible 3564 will have colliding beams, the effective bunch crossing frequency is ~30 MHz.

Interaction rate (MHz) for 80 mb cross section
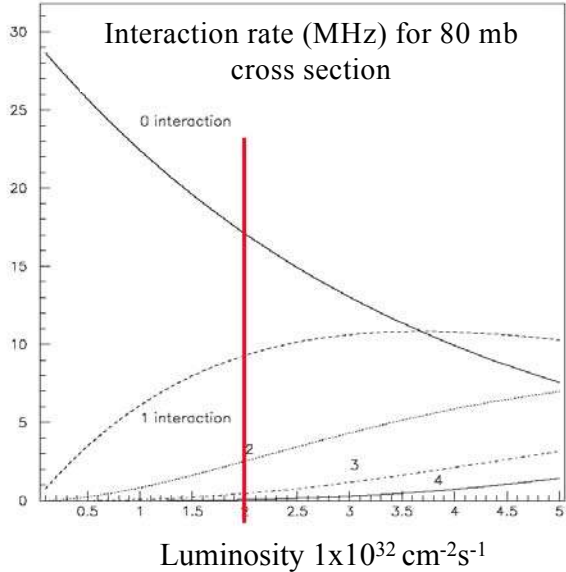
Figure 1-2 The interaction rate in MHz as a function of the luminosity for n = 1 to 4 using Equ 1-1.

At the nominal LHCb luminosity the mean μ is 0.53 per crossing, compared to the LHC design of 20.   Figure 1-2 shows the effective interaction rate as a function of the luminosity for zero to four pp interactions. The red line shows the nominal LHCb working value. As can be seen, the rate for single pp interaction is ~10 MHz.

The beam pipe

The performance requirements of LHCb demand that the position of the sensitive area of the VELO detectors is as close as possible to the beam and that there is a minimum amount of material in the detector acceptance. An increase in the amount of material leads to an increase in secondary particle interaction, which is problematic for the tracking stations and RICH detectors due to the increased occupancy that results. Secondary interactions are particularly critical for the VELO since material interactions in its vicinity make it difficult to reconstruct a

---

[1] The barn is a unit of area. 1 mb=10^{-27} cm².

reliable vertex. For this reason the ~20 m LHCb beam pipe close to the VELO is made in sections of beryllium or aluminium/beryllium. The material thickness is a compromise between the mechanical strength required and an acceptable fraction of radiation-lengths. The cost and handling constraints makes it prohibitive to use any one low radiation-length material throughout the total beam length.
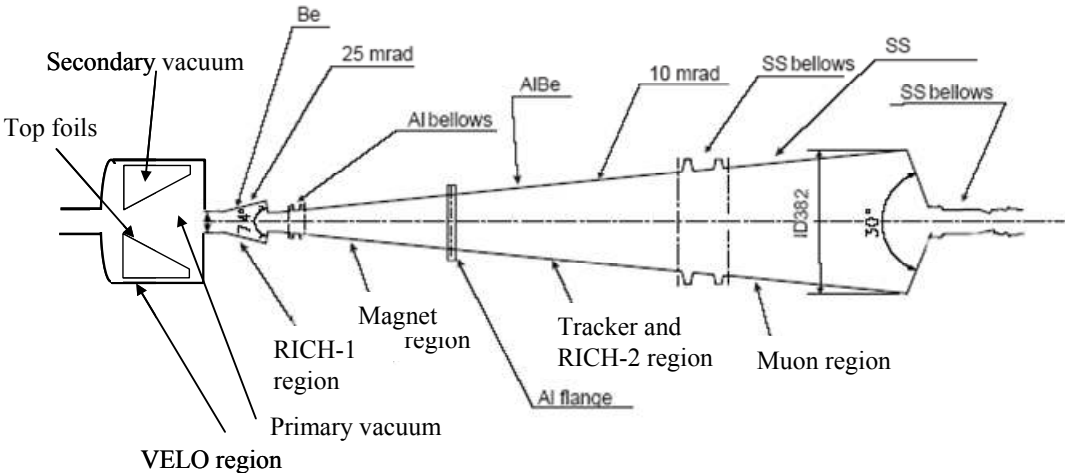
Figure 1-3 A representation of the beam pipe, including the VELO (not to scale).

Figure 1-3 is a representation of the LHCb beam pipe and shows the pipe sections and materials used. The VELO section (see also Figure 1-4) has a large primary vacuum vessel that maintains the $\sim10^{-9}$ mbar LHC vacuum. To protect the primary vacuum, the VELO detector modules are placed in a thin-walled aluminium alloy secondary container maintained at $\sim10^{-4}$ mbar. This structure also shields the detector modules within it from high frequency RF fields of the LHC beam. It also suppresses wake fields.

The RICH-1 section of the beam pipe is a 1840 mm length of 1 mm thick beryllium. The transition point from Be to AlBe, at the Al bellows, has been optimised to keep as low as possible the background induced by the beam pipe in the TT station. The final section is made from 2 to 4 mm thick stainless steel in the muon region, where secondary interactions are much less important.

The VELO

The VELO must provide precise measurements of track coordinates close to the interaction region and is the main tracking system upstream of the magnet. The VELO features a series of silicon stations placed along the beam direction that are capable of measuring the primary vertex to a resolution of ~8 μm in the x and y direction, and ~40 μm in the z direction [MOR03].

Primary vacuum vessel
Corrugated 300 μm thick
Secondary
220 μm thick silicon

a)
b)
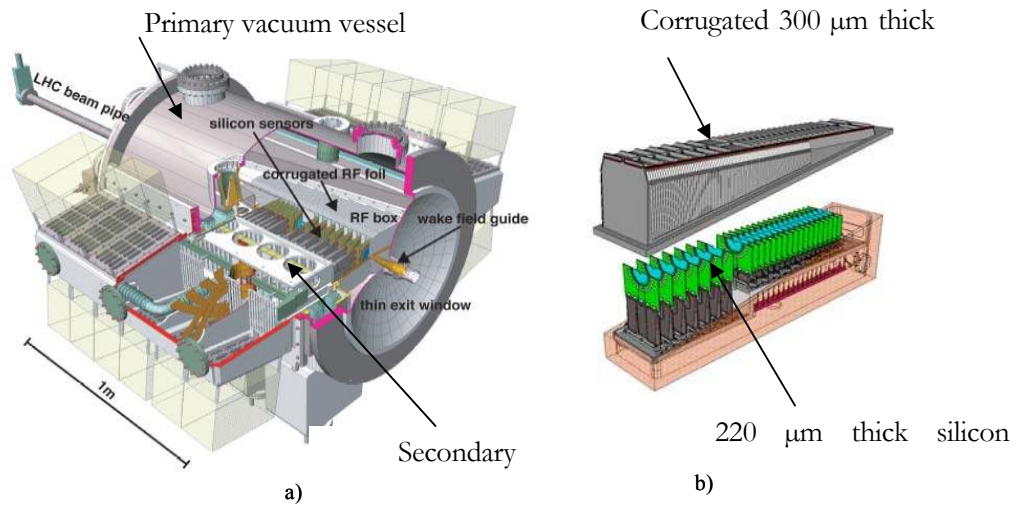
Figure 1-4 a) 3D view of the VELO showing the large outer primary vacuum vessel and the internal (cut away area) of the internal secondary vacuum that houses the detector. b) The corrugated RF shield and silicon sensors.

Figure 1-4 a) shows a 3D view of the VELO with its primary and secondary vacuum vessels, as discussed previously. Figure 1-4 b) shows a more detailed view of the secondary vacuum vessel showing the silicon sensors within the RF shield. Each sensor is semi-circular in shape with an active area between radii of 8.17 mm and 42 mm. The 220 μm thick silicon sensors have either azimuthal (R measuring) or quasi-radial (phi measuring) strips, as shown in Figure 1-5 a). To maintain a high measurement resolution for tracks close to the beam axis whilst restricting the total number of readout strips, the pitch of the strips varies over the detector plane from 40 μm to 101.6 μm for the R measuring strips. Furthermore the strips are divided into four sections to insure a uniform strip capacitance. The magnitude and variation of strip capacitance has significant influence on the readout electronics as strip capacitance directly contributes to the electronic noise.

For the phi measuring detectors the strips are divided azimuthally into an inner and outer region. For the inner region the pitch varies from 35.5 μm to 78.3 μm, and the outer region from 39.3 μm to 96.6 μm.

There are 21 stations in total, shown in Figure 1-5c. As well as covering the full LHCb forward angular acceptance, the VELO also has a partial coverage of the backward hemisphere to improve the primary vertex measurement. The Level_0 trigger, explained in section 1.2, aims to select beam crossings with only one pp interaction by reconstructing the z position of the primary interaction. For this purpose, two R measuring sensors placed upstream of the VELO stations are used, shown in blue in Figure 1-5 c). These stations are referred to as the VETO detector.
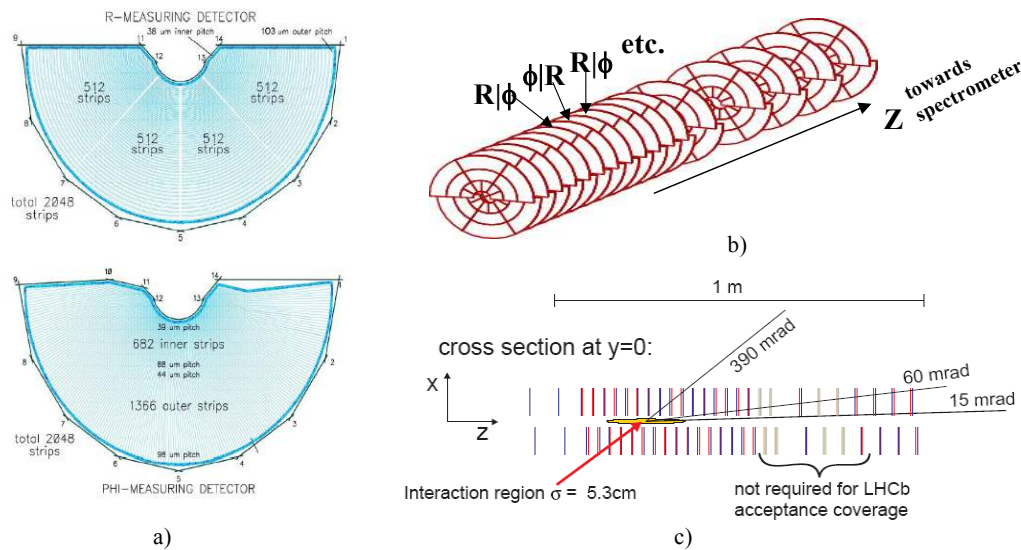
Figure 1-5 a) R and phi silicon detectors, b) R/phi arrangement per station, c) station set-up.

For reading out the VELO and VETO detectors, 16 front-end Beetle chips are used per silicon sensor. The Beetle chip development is a major work of the author and is described in chapters 3 to 5. For the VELO, the chips sample the data at a rate of 40 MHz and store them in a pipeline memory until the Level_0 trigger decision has been made. On a Level_0 trigger accept, the data are sent to a Level_1 storage unit. In the case of the VETO the data channels are sampled at 40 MHz and have their signal response discriminated on-detector. A fast OR of groups of four data channels is performed and the result contributes to the Level_0 trigger

decision-making process. Being so close to the interaction point, both the detector and readout chips are designed to be radiation hard[2] up to an expected exposed dose of the order of 10 Mrad per year.

Tracking system

The tracking system consists of the TT and T1-T3 tracking stations. The TT is the first station downstream of the VELO and is positioned in front of the magnet and just behind RICH-1. The three remaining stations are placed downstream of the magnet with equal spacing.

Together with the VELO, the TT is used in the Level_1 trigger. Large impact parameter tracks found in the VELO are extrapolated to the TT. The magnetic field in the RICH-1 region allows their momenta to be coarsely measured. The TT consists of four planes of silicon strip detectors with strip pitch of 198 μm and thickness of 500 μm, split into two pairs of planes separated by 30 cm. The first and the fourth plane have vertical silicon strips while the second and third planes have their strips rotated by stereo angles of +5° and -5°, respectively. The total active area of the combined planes is approximately 8.3 m² giving ~140k readout channels. These are also read out using Beetle chips.

Each of the three tracking stations T1-T3 consists of an Inner Tracker (IT) close to the beam and an Outer Tracker (OT) surrounding the IT. The integration of the IT and OT are shown in Figure 1-6. The type of detector technology used and the physical size of the detectors are determined by the required resolution and the expected channel occupancy. For the high track density region close to the beam, silicon strip detectors are used. The strips are of the same type used in the TT discussed previously but are either 410 μm or 320 μm thick depending on their location. The readout electronics are also the same but the IT has fewer channels, approximately 130k. The Outer Tracker consists of straw tube drift cells, which are 5 mm in diameter. The emphasis is on tracking precision in the (x,z) magnet bending plane. The stations have two planes with wires in the vertical direction and two stereo planes which are tilted by either +5° or -5°.

---

[2] Circuits that can cope with doses up to a few 100 krad can be considered radiation tolerant whereas circuits that can cope with a few Mrad are classified as radiation hard.
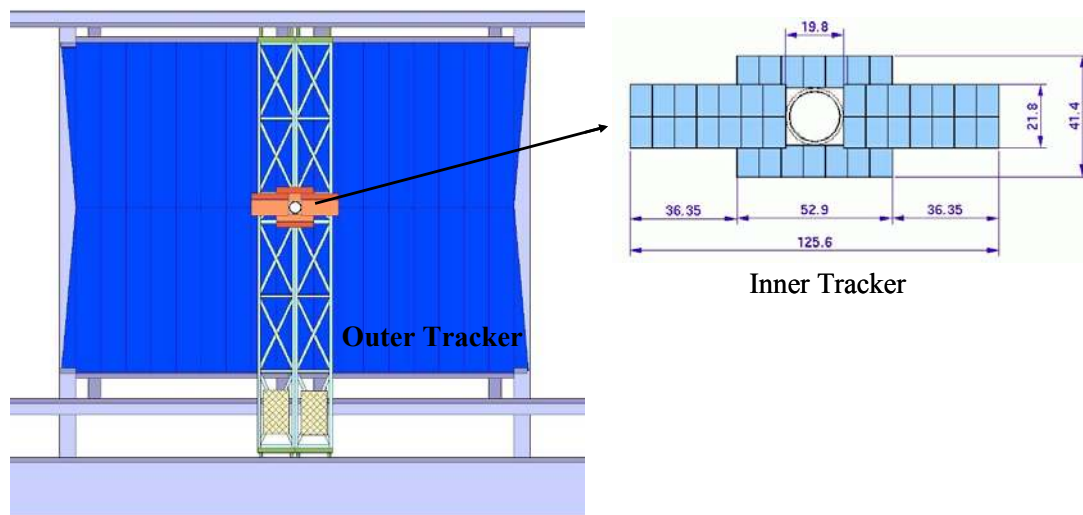
Figure 1-6 The IT and OT integration for a single station.

The occupancy for each OT cell must be below 10% to ensure efficient track reconstruction. This gives constraints on the maximum wire length and cell size. The delay in signal arrival time at the preamplifier is the sum of the transit time of the track to the chamber, the drift time and the propagation delay along the wire. Besides additional constraints on the wire length and cell size, also the choice of 'fast' drift gas will be important. It is expected that all drift signals will be measured within 50 ns. The electronics used to read out the 54k channels is the so called 'ASDBLR' chip developed for ATLAS [TDRO].

The Calorimeter

The calorimeter, which consists of the ECAL, HCAL and PS/SPD, is used to identify electrons and hadrons for the trigger and off-line analyses. The PS, which is used to provide an accurate spatial detection of photons and electrons, is constructed from 14 mm thick lead plates followed by square scintillators that are read out by photo-detectors. The SPD is placed in front of the PS and is used for background rejection. The ECAL, built from 2 mm thick lead plates and scintillator plates, provides identification and an energy measurement of photons and electrons. The HCAL is constructed from scintillator tiles that are imbedded in an iron structure and read out with photo-detectors. Its purpose is to aid in the separation of high-energy hadrons and electrons, and to provide a transverse-energy measurement of hadrons.

The muon detector

The muon detector is used for muon identification and to provide trigger information. This detector makes use of Multi Wire Proportional Chambers (MWPCs). The station M1 is especially important for the transverse-momentum measurement of the muon track used in the Level_0 muon trigger.

The magnet

The dipole magnet provides an integrated field of 4 Tm. The magnetic field is oriented vertically and has a maximum value of 1.1 T. The polarity of the field can be changed to compensate any small left-right asymmetries of the detector.

The Ring Imaging Cherenkov Detectors

Two RICH detectors, shown in Figure 1-7, are used in LHCb for the purpose of particle identification, primarily providing π/K separation in the momentum range from 1 to 100
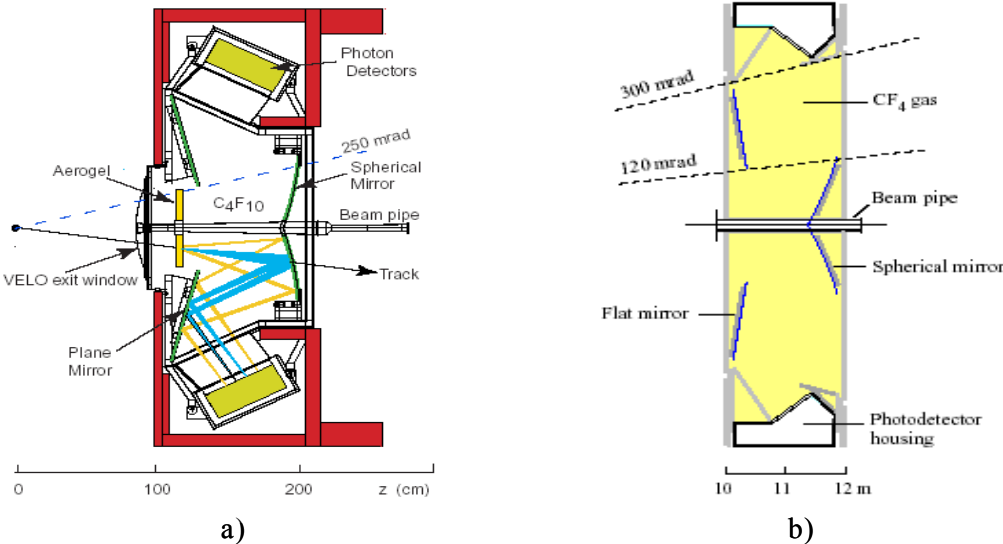
Figure 1-7 a) RICH-1 from the side [TDR03opt], b) RICH-2 from the side [TDR00].

GeV/c. The first, RICH-1, a combined gas/aerogel device positioned upstream of the magnet, has an angular coverage of 25-300 mrad. The second, RICH-2, is a gas device downstream of

the magnet, which covers 10-120 mrad. These detectors are essential in reducing backgrounds and providing an efficient flavour tag of B^0 and anti-B^0 hadrons.

Cherenkov radiation is emitted whenever a charged particle with a velocity (v) exceeds the velocity of light c in a medium, i.e. $v \geq c/n$, where n= refractive index of the medium. Light is emitted in a coherent wave-front or cone with a characteristic Cherenkov angle given by $\cos\theta = 1/(\beta n)$, where $\beta = v/c$ [YPS99]. In the LHCb RICH detectors, mirrors focus the Cherenkov photons onto photo-detection planes, which are outside the LHCb detector acceptance. Candidate photon detectors are Multi-Anode Photo-Multiplier Tubes (MAPMT) or pixel Hybrid Photon Detectors (HPD). The HPD was the baseline option and, during the time of writing, was chosen as the sole photon detector. MAPMTs were the backup solution and have been studied in depth for this thesis.

The RICH detectors are optimised to generate the maximum number of photons while maintaining the required spatial resolution. A source of imperfect resolution is the dispersion in the refractive index of the radiator. For C_4F_{10} gas the index varies from 1.0013 at long wavelength, to 1.0015 at about 200 nm, which is approximately the wavelength cutoff point of the photon detector with a UV glass entrance window. Therefore it is desirable to select wavelengths only within a small band that corresponds to the smallest change in the index. The RICH gas radiator chromatic dispersion is reduced by using small wavelengths into the UV range. However, the aerogel that is used to detect the lower momentum particles suffer mostly from Rayleigh scattering, which is dependent on $\frac{1}{\lambda^4}$, and therefore improves with larger wavelengths. Therefore the photo-detector is chosen to be sensitive to light in the visible and ultraviolet, and the Raleigh scattering is reduced with a filter, for example a glass window between the aerogel and gas interface.

The MAPMT can offer up to 200 times the gain of the HPD and is therefore less demanding on the front-end electronics due to a better signal-to-noise ratio (SNR). Conversely the HPD has far better single-photon resolution. A total photo-detector surface area of about 2.9 m^2 is instrumented in the two RICH detectors, which requires approximately 480 HPDs or 5000 MAPMTs. The upstream RICH-1 has a 5 cm-thick aerogel radiator and an 85 cm-long C_4F_{10} radiator. The expected numbers of detected photons are 7 and 30 respectively for tracks

with particle velocity, $\beta \cong 1$. The downstream RICH-2 is filled with CF_4 radiator with an approximate length of 180 cm, giving 20 expected photons.

Several sources of background can place demands on the RICH readout electronics due to an increase in channel occupancy or by introducing saturation effects. For example, in RICH-1, low momentum charged particles bent backwards by the dipole magnet can strike the photo-detectors and produce showers of photo-electrons from Cherenkov radiation in their entrance windows. Further sources of background originate from secondary interactions in the upstream detector components, Rayleigh scattering of photons in the aerogel, particles remaining from previous bunch crossings and electronic noise. An example of a typical simulated event showing hit photo-detector pixels arising from gas and aerogel rings in RICH-1, including all background processes, is shown in Figure 1-8. The figure shows the x-y projection of photons on the photo-detector plane, in cm. In this event a typical number of about 50 charged particles were simulated traversing the detector.
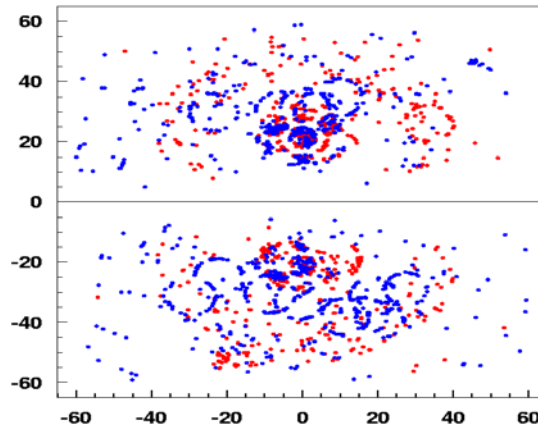
Figure 1-8 x-y projection of photons on the photo-detector plane, in cm, for RICH-1. Background sources are included.

1.2 Triggers

With an LHC bunch crossing occurring every 25ns, approximately 10^{12} B-hadron pairs will be produced per year at the LHCb luminosity. From these, approximately 10^9 will be of interest within the acceptance for storage and analysis. LHCb uses a 4-level trigger system to

make a decision on whether or not an event should be stored for off-line analysis. The trigger selects tracks originating a few millimetres away from where the pp collision took place; an average 7mm for an 80 GeV/c B-hadron. Tracks with a displaced vertex indicate that a B-hadron had been created, travelled a short way, and subsequently decayed. Each trigger level is more complex in its use of available information than the previous level. The first level, Level_0, is implemented in custom electronic hardware. Level_1 and the two High Level triggers (HLT) are computer algorithms executed on a farm of commodity processors.

To meet speed requirements, Level_0 and Level_1 triggers use coarse resolution information from selected detectors. The detectors and their trigger contributions are shown in Figure 1-9. The Level_0 trigger uses information from the Muon, ECAL and HCAL to identify high-momentum leptons, hadrons and photons. The VETO is used to count the number of primary vertices and reject events that have multiple pp interactions. The Level_0 trigger accepts events that have a single muon, electron or hadron high transverse momentum p_T
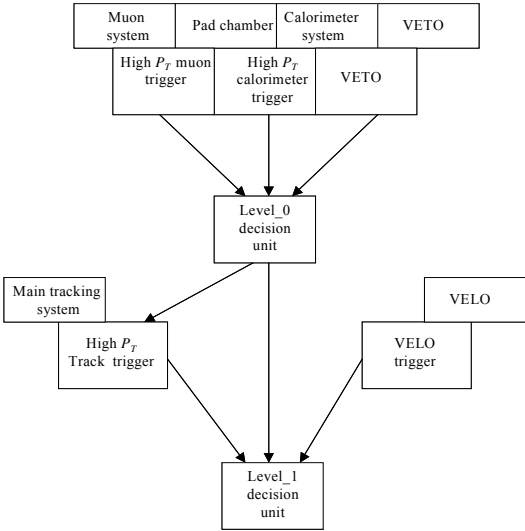
Figure 1-9 Block diagram of trigger contributions [TDR00].

cluster. If the event is accepted, then the information on the high p_T cluster which triggered the event is passed onto the Level_1 trigger. Here the cluster is used as a seed to start the Level_1 trigger analysis. All the event information from the VELO is used in the Level_1 trigger decision. The event data from the VELO are first used to reconstruct the primary vertex. From this vertex a search of secondary vertices is performed. The trigger decision is then made on an

event probability based on the number of secondary vertices found. The algorithms used are sophisticated and require a longer time to make a decision than the Level_0 trigger.

The next level of trigger, HLT-2, reconstructs the VELO information again but uses full momentum information to remove any false secondary vertices that may have been generated by low multiple-scattered momentum tracks. The final level of trigger, HLT-3, uses all the detectors, including both RICH detectors, to reconstruct specific B-hadron decay modes of interest.

To allow time for the trigger decisions to be made, event data are stored using a three-tier hardware storage system. The first tier, Level_0, stores the detector signals in pipeline memory until the Level_0 trigger accepts or rejects the data. The Level_0 trigger latency, defined as the time elapsed between a pp interaction and the time when the trigger signal is received back at the front-end electronics, is chosen to be a constant 4 μs. The data reduction for the Level_0 is about a factor of ~40. The second tier, Level_1, stores the accepted data from Level_0 in a pipeline for a maximum Level_1 trigger latency[3] of 52.4 ms and gives a further data reduction by a factor of ~25. The Level_1 trigger latency is defined as the time elapsed between the event data leaving the Level_0 until the Level_1 electronics receives back the Level_1 trigger signal. The third tier stores data in memory on a farm of commodity processors. The total reduction in data is approximately a factor of 200k. The raw event size is approximately 31 kBytes which is stored on disk at a rate of 200 Hz. For more detailed information the reader is referred to [TDRtrig].

Accommodating the latency required for the triggers has a huge impact on the Level_0 and Level_1 hardware. For the Level_0 a large memory requirement leads to a large die size for the custom-made front-end readout chips. This consequently leads to an increase in cost, an increase in complexity of the readout control logic, and introduces stringent timing issues with regards to storage, retrieval and synchronisation. The Level_1 incurs much the same problems as the Level_0 but has further demands on real estate as commercial components are employed.

---

[3] This number has yet to be frozen.

These requirements, particularly readout time and synchronisation, require an LHCb global electronics scheme. This scheme is discussed further in the following section.

1.3 LHCb Global Electronic Scheme

The LHCb experiment consists of several sub-detectors with ~one million electronic channels. Table 1-1 shows the sub-detector technology and the number of channels to be read out from each.

Sub-detector	Technology	# of Channels
Trigger Tracker	silicon-strip detectors	140 k
VELO	r and phi silicon sensors	205 k
Inner Tracker	silicon-strip detectors	130 k
Outer Tracker	straw tubes	54 k
RICH 1	HPD readout	~172 k
RICH 2	HPD readout	~268 k
SPD/PS	Lead/Scintillator with MAPMT	2 x 6 k
ECAL	Lead/Scintillator with phototube	6 k
HCAL	Iron/scintillating tile with phototube	1.5 k
Muon	MWPC	26 k

Table 1-1 LHCb sub-detectors, technology and number of readout channels [SCH02].

For LHCb to work in a seamless and homogeneous way, all sub-detectors need to be completely synchronised to the beam crossing from data capture to data readout. This is of considerable importance for the Level_0/Level_1 triggers to ensure that the correct event data are extracted. In order to have event data alignment across the sub-detectors, global specifications have been determined. These include the maximum sample time for the analogue detector signals, maximum allowable trigger rates, derandomiser buffer lengths (described later) and event readout times. Not only do these global specifications ensure event data alignment, they also lead to common solutions such as common readout chips and interface modules. An example of a common module would be the Timing Trigger Control receiver (TTCrx) described in section 1.3.1. Common solutions make the experiment more predictable and minimize the resources and manpower needed during the development and testing phase.

This section briefly describes some of the global electronic specifications and components that are important to this thesis. The reader should refer to [CHR01_L0], [CHR01_L1] and [CHR_E] for a comprehensive coverage of this subject matter.

1.3.1   General scheme and specifications

The general readout scheme for all LHCb sub-detectors is illustrated in Figure 1-10. The Level_0 region, which includes the front-end amplifiers, is physically on the detector. This is a harsh environment in terms of radiation and electronic noise. The Level_1 is in a radiation-free
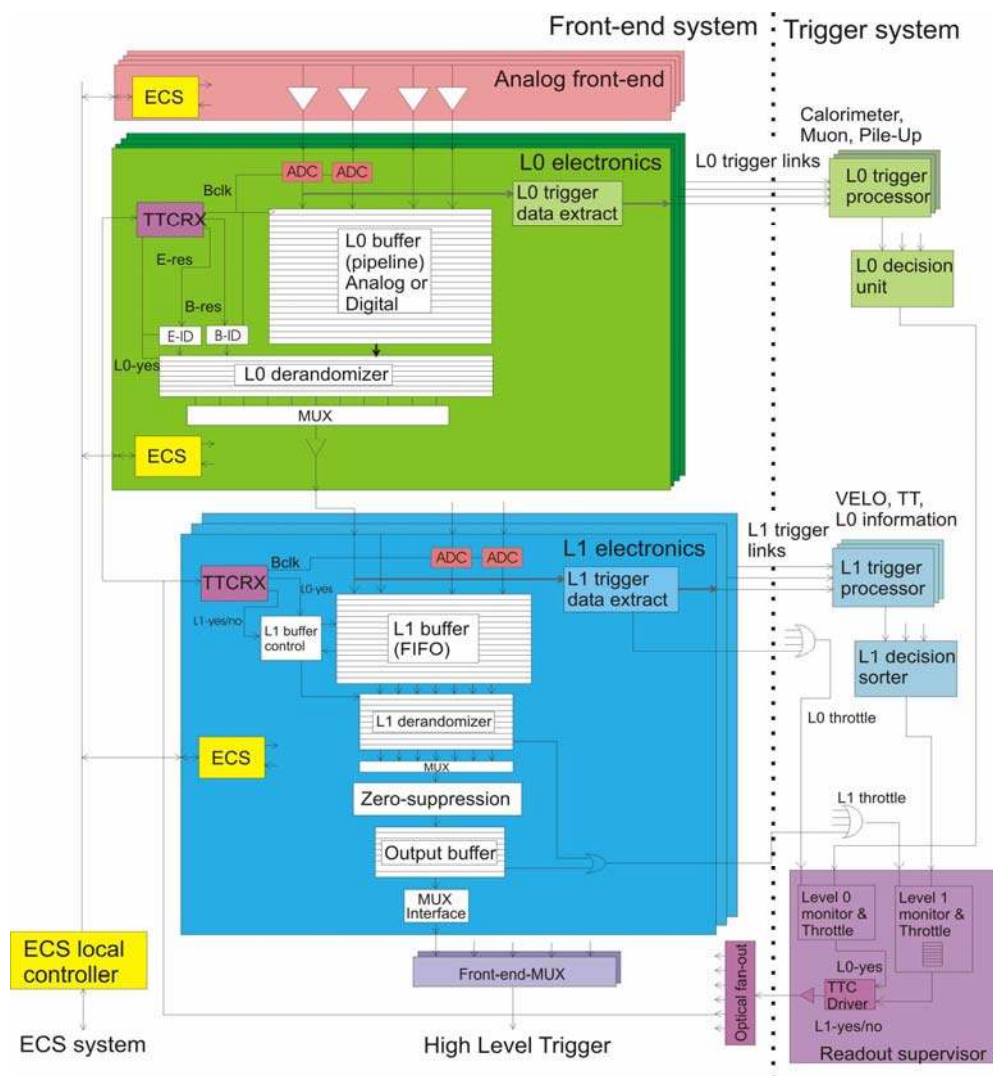
Figure 1-10 The General LHCb electronic readout scheme for Level_0 and Level_1 [TDRtrig].

environment. The TTCrx and Experimental Control System (ECS) blocks in Figure 1-10 are common to both the Level_0 and Level_1 regions. The event data are stored in the pipeline memories for the latency periods while the derandomisers are temporary storage areas for event data being read out. The following sections describe the Readout Supervisor (RS), ECS, TTCrx, Level_0 and Level_1 electronics.

Readout Supervisor

The RS monitors and controls all data read out from the entire experiment. It receives decisions from the trigger system and drives the Timing Trigger Control (TTC) system. The TTC is the name given to the system that encompasses all the sub-modules needed to distribute the timing, trigger and control signals over the experiment. The TTCrx is one such sub-module. In the case of a possible Level_0 or Level_1 buffer overflow the RS will throttle back the data-flow by vetoing Level_1 and Level_0 trigger accepts. In normal operation one RS services all sub-detectors, however, each sub-detector has its own RS to allow for stand-alone calibration and debugging. To synchronize the TTC to the LHC machine, the RS uses the LHC machine synchronization signal, which has a fixed delay with respect to the bunch crossing. To be able to compensate for this delay and any delays contributed by the TTC system, the readout supervisor is capable of adding a coarse delay (in steps of ~100 ns) up to one complete LHC machine cycle period.

Experimental Control System

The ECS is responsible for the slow control parameters such as temperature monitoring, system configuration and power supply monitoring. Transmission errors detected by the Level_0 and Level_1 electronics are also sent to the ECS for notification. The ECS interfaces to the different parts of the experiment via JTAG or I^2C. JTAG is the Joint Test Advisory Group, which in 1998 developed the JTAG testability bus specification. This can be used for communicating with or configuring logic and has the IEEE standard 1149.1 [JTAG]. I^2C is also a standard for configuring and communicating but has a different protocol [PHI95].

TTCrx

The TTCrx is an Application Specific Integrated Circuit (ASIC) designed by the CERN EP Microelectronics group and fabricated in the DMILL process (see section 3.1.1). It re-generates the 40 MHz machine bunch clock, receives, decodes and executes any individual or broadcast commands that have originated from the RS, which include the triggers and resets.

The TTCrx has been designed to deliver only one level of triggering, whereas LHCb requires both a Level_0 and a Level_1 trigger. To circumvent this problem the Level_0 trigger and the Bunch ID (BID), which is used for data tagging, are transmitted on a data channel designed for this purpose; known as 'channel-A'. However, the Level_1 trigger is distributed on a second data channel, 'channel-B', which was intended for only broadcast commands; this limits the rate at which a Level_1 trigger can be sent. The 8-bit broadcast command has a reduced bandwidth of one command every 16-clock cycles and can only deliver the two least significant bits of the BID. This requires that the 12-bit BID number is generated locally on the Level_1 electronics and checked against the 2-bits from the TTCrx. Using the TTCrx in this way effects when and how often other broadcast commands can be sent. Both the Level_0 and Level_1 electronics need to be designed around this constraint.
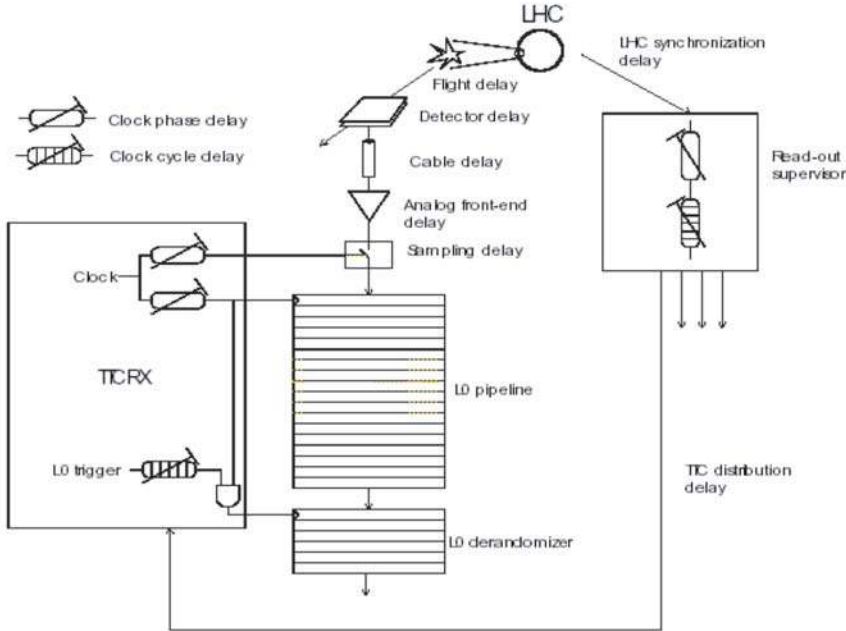
Figure 1-11 Bunch crossing synchronisation of the front end electronics [CHR01_L0]

Three clocks are offered by the TTCrx: CLK40, CLK-des1 and CLK-des-2. CLK40 is phase-aligned with the machine clock. Clocks des1 and des2 are skewing clocks that can be independently phase-adjusted up to 25 ns in 100 ps steps. This is for either aligning the Level_0 electronics to the detector, or aligning the Level_1 to Level_0 electronics. To extend the dynamic range of the fine clocks, the TTCrx offers an additional course delay of up to 15 bunch clocks on synchronous TTC signals such as the L0 trigger, resets and short broadcast commands. The total allowable local skewing is 400 ns. Figure 1-11 shows an example how the TTCrx can be utilised within the Level_0. In the case of MAPMTs, described in the next chapter, each TTCrx in the instrumentation system would need to be independently adjusted to ensure correct sampling times.

Level_0 region.

The analogue detector signals must be amplified and stored in either a digital or analogue pipeline memory at a bunch-crossing rate of 40 MHz. The raw signals from the detector element must be fully captured within the bunch-crossing period of 25 ns to ensure event synchronisation is maintained across the detector. The length of the pipeline must be large enough to allow 4 μs to elapse between the bunch crossing and the arrival of the Level_0 trigger signal at the front-end electronics. This includes time-of-flight, cable delays, and 2 μs for the processing of the data in the Level_0 trigger hardware. On the arrival of a Level_0 accept/reject trigger signal every 25 ns, the event data should be overwritten on a reject, or moved into a derandomiser buffer for subsequent readout on an accept.

Parameter	Value
Bunch Crossing Frequency	40 MHz
Max trigger rate	1.1 MHz
Average trigger rate	1 MHz
Latency	4 μs (160 events)
Max consecutive trigger accepts	16
Derandomiser depth	16 events
Derandomiser readout time	900 ns

Table 1-2 General requirements of the Level_0 electronics.

The necessary depth of the derandomiser buffer is calculated from the average Level_0 trigger accept rate of 1 MHz and a readout speed that will maintain a maximum allowable event loss of less than 1 %. A derandomiser depth of 16 with an event-packet[4] readout speed from the derandomiser to Level_1 electronics of 900 ns was chosen.   This requirement was determined from simulations with parameters being varied [CHR_s]. The 900 ns readout time allows 36 words[5] in an event-packet to be read out at 40 MHz (36x25 ns =900 ns). Four of these words are required for transmission overheads, to be utilised by the user, such as an event number. Therefore 32-words have been allocated for event data. With these derandomiser specifications, a 10 % faster readout compared to the write of the memory is achieved, which results in a simulated 0.5 % event loss. Consecutive Level_0 accepts are allowed until the derandomiser buffer is full. To ensure against buffer overflow the RS emulates the Level_0 derandomiser and throttles the Level_0 accept rate when necessary. The TTCrx distributes the Level_0 trigger decision to the front-end electronics, along with an associated BID, which is concatenated to the event data via the available user bits. The strict synchronicity between sub-detectors for data transmitted from the Level_0 region is made less critical by the BID identification. Table 1-2 summarises some of the Level_0 specifications.

---

[4] The term 'event-packet' in this context is used to describe a block of data that has the event data plus transmission information.

[5] A word is a set of parallel bits that are treated, stored and transported as a unit. In most cases a word is 16 or 32 bits wide.

Level_1 region.

The purpose of the Level_1 electronics is to store event-packets accepted from Level_0 and to either discard them on a Level_1 trigger reject or to transport them to an event building network on an accept. The Level_1 global specifications are based on using Quad Data Rate (QDR) or Double Data Rate (DDR) Random Access Memory (RAM) for the data storage regions [CHR01_L1]. Table 1-3 gives the general specifications for the Level_1 electronics.

Event-packets, which are 36 data words in length, are written to the Level_1 buffer at an average rate of 1 MHz and a maximum rate of 1.1 MHz. The data rate is 25 ns per word. The event-packets are stored for a variable latency of up to ~52.4 ms while waiting for a Level_1 accept or reject trigger signal. The average Level_1 accept rate is 40 kHz. As the buffer size is still under consideration, provision must be made in hardware to increase this by up to a factor of 2 if necessary. The time between triggers is a compromise between slow Level_1 readout speeds to reduce bandwidth requirements, and the need to keep the memory from filling. This is achieved by issuing the reject signal at the maximum rate possible, which is limited by the TTCrx speed to 400 ns, and the accept rate at a feasibly slow rate of 20 μs. The time between an accept and a following reject signal is 900 ns and whilst not optimal, is needed to keep compatibility to electronics already built on an earlier specification.

Parameter	Value
Time between two consecutive rejects	400 ns
Time between accept and subsequent reject	900 ns
Time between accept and subsequent accept	20 μs
Buffer size[6]	58254 events (~52.4 ms)
Max buffer input rate	900 ns
Average trigger accept rate	40 kHz
Max buffer output rate (accept)	900 ns
Derandomiser buffer size	448 events

Table 1-3 General requirements of the Level_1 electronics.

---

[6] The Level_1 buffer described in chapter 6 has been designed for the predecessor buffer size of 1820 events. The decision to increase to 58254 event storage was made around August 2003.

At present the trigger signals being sent by the RS, via the TTCrx, arrive in the same order that the events are stored in the Level_1 buffer. This makes the Level_1 memory controller quite simple to implement as the address generation is always sequential. As this scheme makes use of QDR and DDR RAM that can have their memory location directly accessed, there is a proposal to send the BID with the Level_1 trigger accept signal. This will alleviate the requirement of the RS only being allowed to send the trigger signals in a sequential manner, as the Level_1 can make use of the BID to directly locate the correct event. This greatly simplifies the algorithms needed for the trigger and RS. Further to this, trigger rejects no longer need be sent, as the RS will ensure that any accepted event has been read out before the write pointer overwrites the memory contents.

Like the Level_0, the Level_1 has to interface to the TTC and ECS system. Unlike the Level_0, it also interfaces directly to the RS. The Level_1 checks the validity of the stored data and notifies the ECS system if there are data errors.

1.4 Summary

In this chapter the LHCb detector, sub-detectors, global electronic scheme and triggering system have been introduced. It gives a description of the role in which the RICH plays within LHCb and sets out the electronics readout requirements. Two photo-detector technologies have been considered for the RICH detectors; both the MAPMT and HPD detectors are sampled at 40 MHz into a 4 μs memory and read out at an average Level_0 trigger accept rate of 1 MHz into the Level_1 region. The Level_1 region stores these data while awaiting the Level_1 accept trigger at an average rate of 40 kHz. The sampling clock, experiment synchronisation, triggers and fast control are utilised via the TTC. The slow control is implemented using the ECS. Both the MAPMT and HPD detectors require specialised radiation hard ASIC chips that can capture and store the signals, while being fully compatible to the global electronics of LHCb.

The Multi-Anode Photo-Multiplier Tube

As described previously in Chapter 1, Cherenkov rings in the RICH detectors typically consist of between 5-30 photons, dependent on radiator, emitted from tracks with a particle velocity, $\beta \cong 1$. These photons are detected with wavelengths in the visible and near ultraviolet range. To reconstruct the ring for particle identification requires a photo-detector that has good quantum efficiency (~20-25 %), a wide spectral range and the required spatial detection resolution of ~2 mm^2. The Multi-Anode Photo-Multiplier Tube (MAPMT) is a suitable candidate for this application.

This chapter describes the work undertaken by the author to characterise the R5900-00-M64 H7546B MAPMT from Hamamatsu. A brief description of the principle of operation and available tube types is followed by the test scheme. Finally, the tube is characterised with respect to gain, pixel-to-pixel gain uniformity, tube efficiency and pulse-height spectrum.

2.1   Principle of operation

Figure 2-1 shows a schematic and photograph of the Hamamatsu M64 8x8 MAPMT. The principle of operation of a PMT is that photons are detected through the following processes:

a)   The photon passes through the entrance window. The window material is chosen for its spectral response and its matching to the photo-cathode material. Borosilicate glass is the most common for responses above the 300 nm wavelength range, having a thermal coefficient very close to that of the Kovar alloy used for the leads of the photo-multiplier tube [MAP97].

b)   At the photo-cathode the photon excites a photo-electron that is emitted into the vacuum of the tube (photoelectric effect). Most photo-cathodes are made of a compound semiconductor consisting of alkali metals with a low work function, which is vaporised onto the inside of the entrance window. A common photo-cathode is the bialkali, which uses a combination of two alkali metals [MAP98].

c) Photo-electrons are accelerated and electrostatically focussed onto the first dynode (electrode), where they are multiplied by means of secondary electron emission. Beyond this is arranged a series of dynodes at progressively increasing voltages e.g. +100 V, +200 V. These are coated with a good secondary emitter (BeO) or (Mg-O-Cs), which yields 2-5 electrons when struck by an electron with energy over 100 eV. Hence an electron is accelerated to the first dynode where it liberates electrons, which are accelerated to the second dynode, and so on until reaching the anode. An amplification of 10^6 is readily obtainable, with a 5 ns pulse length. Ideally, the current amplification of a photo-multiplier tube having n dynode stages and an average secondary emission ratio of S per stage will be S^n. There are a variety of dynode structures available, for example venetian-blind and circular cage. Each type exhibits different current amplification, time response, gain uniformity and secondary collection efficiency. The MAPMT has a square array of PMT structures, which allows pixallation of photon hits.
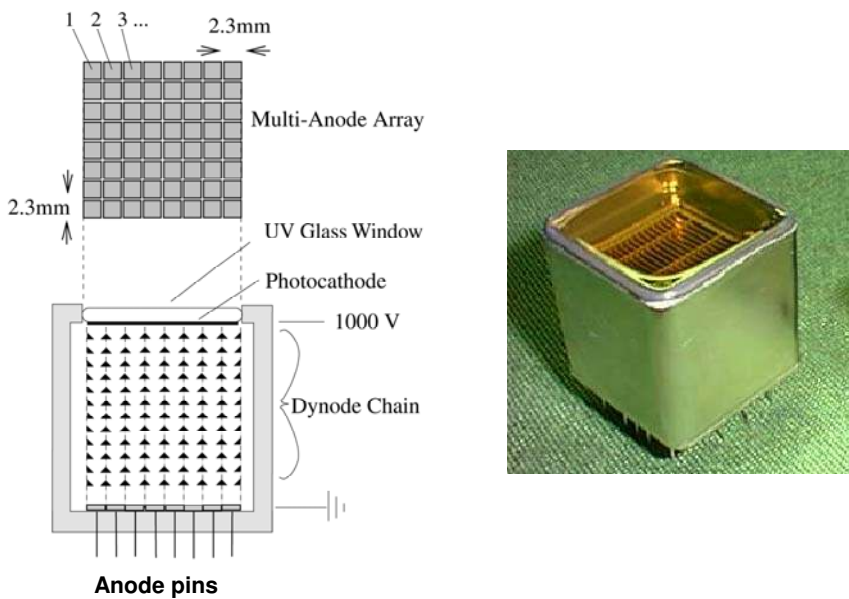
Figure 2-1 A schematic and photograph of a Hamamatsu M64 MAPMT.

2.2 The M64 MAPMT family

The photo-multiplier tubes considered for the LHCb RICH detectors are all from the Hamamatsu M64 family. Depending on the tube type, each stage has either 12 or 8 dynodes of bialkali on stainless steel [MAP97]. Photo-electrons are electrostatically focussed onto an array of 8 x 8 venetian-blind dynode stages. This gives 64 pixels each with a ~2 mm^2 detection area, where each pixel has its own anode pin. The study of the tubes first began in 1997 with an R5900-00-M64 H7546B, which is a standard Hamamatsu production tube[7]. In 1998 the tube was tested in the CERN SPS beam [ALB01]. The physical dimensions of the tube are ~ 30 mm^2 x 20 mm, front area and depth respectively, which includes a 1 mm metal flange around the circumference. The entrance window is ~ 800 μm thick borosilicate glass with an 18.1 mm^2 bialkali photo-cathode that has a peak wavelength response of 420 nm. It is a 12 dynode device.

In 1998 Hamamatsu accepted a request from LHCb for the 1 mm metal flange to be removed to improve the active area of the tube by 14 % and the borosilicate window to be replaced with a UV-glass that is transparent up to wavelengths of 180 nm, hence improving the spectral range by 50 %. This tube is the 12 dynode R7600-03-M64, with serial numbers 9Cxxxx[8]; R7600 for no rim, 03 for the window type and M64 is the number of pixels. This tube was tested in the X7 beam at the CERN SPS, [ALB02]. Inefficiencies between pixels led to a further tube, R5900-03-M64[9], with serial numbers 9Kxxxx. The modification was made to the physical layout of the 8 x 8 venetian-blind, improving on the homogeneity of the electrical field between cathode and blind. In 2002 the R7600-03-M64MOD, serial number GAxxxx, became available with an 8-dynode structure, R7600 body, UV-window (03) and with the 9K focussing. The 8-dynode chain gives a single-photon response of ~50,000 electrons at an HT of ~ -800 V. This relatively low gain makes the tube compatible with some of the silicon-sensor readout chips developed within LHCb, allowing the use of an existing readout chip rather than an in-house customised design.

---

[7] Characteristics of this tube can be found on the Hamamatsu www site [HAM_www].

[8] The 9C is common to all these tubes, the xxxx is an incremental number per tube.

[9] These were prototype tubes given by Hamamatsu and have the standard flange.

2.3 The MAPMT test set-up

The author designed and commissioned a test set-up and evaluated the Hamamatsu R5900-00-M64 MAPMT. With a bias voltage of -1000 V the gain of the tube is about $1x10^6$. Output signals typically have rise-times of 1.5 ns with 5 ns duration. For a single photo-electron with charge $1.6x10^{-19}$ C, a 2 mV signal would be expected at the load (the base is designed to match 50 Ω). The MAPMT has a quantum efficiency of ~20 %, and the gaps between pixels are 300 μm ([HAM_www] data sheet). The gain can vary between pixels by a factor of 5 [MAP97]. However this does not follow a normal distribution, and the large gain variation is dominated by a few 'bad' pixels, typically 5 % of the tube.

In order to make a reasonable assessment of the tube parameters and to ensure the suitability of the tube for the LHCb RICH application, photons from a light spot of 420 nm wavelength and 47 μm in diameter were scanned across the surface of the tube in 100 μm steps. The following subsections give details of the set-up required to deliver the light source to the tube, move the light source, and read out the 64 channels of the MAPMT.

Mechanical housing and mounting

In order to protect the MAPMT from extraneous light, a dark box housing of 0.75x0.4x0.16 m (length, width, height) has been constructed, shown in Figure 2-2. The various components are labelled in the figure:

1) An optical x, y, z stage with 200 nm spatial resolution.

2) A large fixed platform for lens mounting.

3) A blue LED.

4) A x10 objective lens.

5) A x20 objective lens.

6)   An FC-type receptacle to accept the fibre.

7)   A monomode fibre.
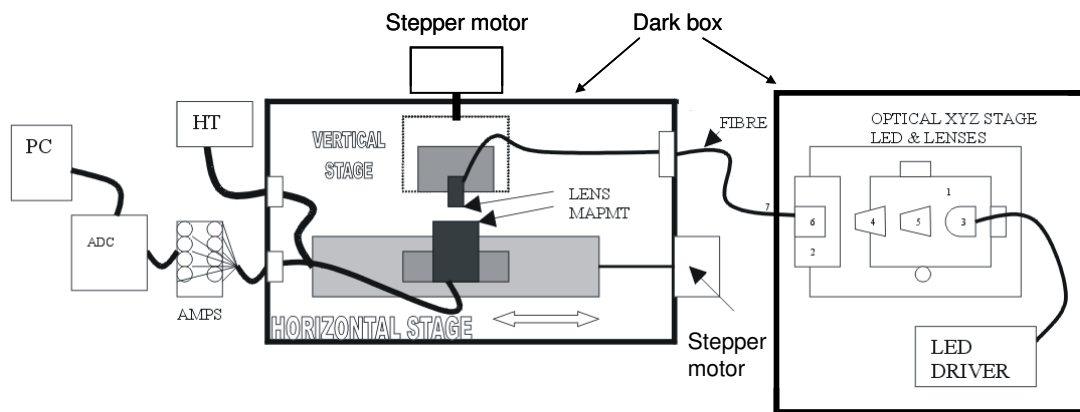
Figure 2-2 The MAPMT scanning facility.

The monomode fibre has a core diameter of 3.5 μm, a cladding diameter of 125 μm and a numerical aperture (NA) of 0.11. It has a 3 mm diameter graded index lens on its end with focal length of 25 mm. All feed-through connectors are electrically isolated from the box by nylon washers to reduce ground loops. Anodes 1 to 64 of the MAPMT are read out via individual coaxial cables with LEMO connectors, which have a common ground point on the tube. The MAPMT is supported and clamped using a silicon resin bonded fibre (SRBF) block. This is used for its insulation properties since the inner case of the MAPMT is held at bias potential. In this way the anode on the back of the entrance window is at the same potential as the case and therefore prevents leakage current between the two. Foam inserts between the MAPMT and SRBF are used to prevent damage to the tube.

Eight of the anode signals are simultaneously read out into a 16-channel NIMS amplifier, model 776, giving a gain of 100. The amplified signals are fed to a 12-channel CAMAC ADC, model 2249a, via AC coupling capacitors. The capacitors are necessary to remove the DC offsets of the amplifiers.

Light source/coupling and stages for the scan

The peak wavelength response of the MAPMT is 420 nm (blue light). It would have been advantageous to use coherent light of the type produced by a laser, suitable for coupling to monomode fibres[10]. This could be focussed to a much smaller spot than incoherent light produced by an LED. Unfortunately blue lasers are not readily available and are expensive, so a blue 470 nm LED was used. This has a maximum luminosity of 1000 milli-candela and a view angle of 15 degrees, switched at 10 kHz with pulse duration of 15 ns. The short duration was required so that the ADC gate width could be minimised to reduce capture of MAPMT dark counts and system noise.

An electronic circuit was constructed so the LED could be pulsed rapidly and is shown in Figure 2-3. The principle of operation is as follows. First the 134 pF capacitor, C_1, charges with a slow RC time constant. Once fully charged, a short square input signal is pulsed at the input of the two FETs. The FET F_2 connected to the LED anode switches 'on' and C_1 quickly discharges through the LED, causing a burst of light. Due to the delay line F_1 switches 'on' a small time after F_2, and abruptly turns the LED 'off' again. The resistor R_1 is used to limit the current through F_1 when it is switched 'on'.
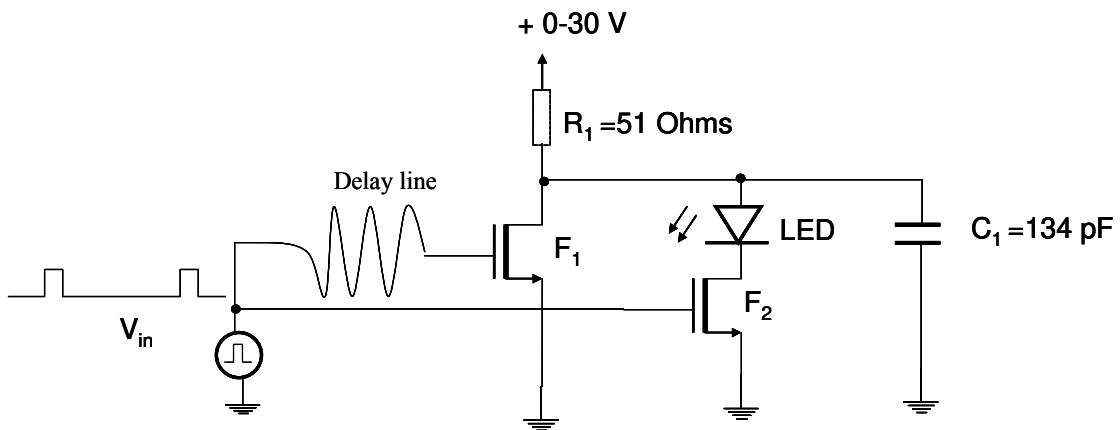
Figure 2-3 Basic FET circuit for the LED pulse.

---

[10] A mono-mode fibre is a fibre that only has one acceptable ray-path for the frequency of the light. A multi-mode fibre has a number of possible rays that light of a particular frequency may take.

To deliver the LED light pulse to the MAPMT, a monomode fibre was used for two reasons. Firstly, to restrict the pulse to a narrow light spectrum with a spot-size conforming to a Gaussian distribution, and secondly, to minimise the spot-size from the end of the fibre. This is explained in the following section.

Monomode fibre and the light spot-size

Referring to Figure 2-2, the first step in coupling an LED to a monomode fibre is to use an objective lens to form a parallel beam. This proved difficult as the LED has its own encapsulated lens and is an incoherent light source. The parallel beam is therefore focussed through a second objective lens to a spot-size that matches the fibre. A local x, y, z stage is used to align the fibre with the focussed light spot. In practice the spot was ~1 mm in diameter.

A light wave entering the fibre is either refracted into the cladding and attenuated, or is totally internally reflected at the core/cladding boundary. In this manner it travels along the length of the fibre. There are actually only certain 'allowed' incident angles, θ_{accept}, that will allow light to propagate through the fibre and these define the modes of the waveguide. The number of modes in a given waveguide depends on the optical frequency being transmitted and can be estimated from the normalised frequency or V-number, defined as $V=[2\pi\, r_{core}/\lambda]\sin\theta_{accept}$, where r_{core} is the fibre radius and λ is the wavelength of the light. For $V \leq 2.405$ the fibre is termed a monomode [HAW89]. The larger the value of the sine of the acceptance angle, defined as the numerical aperture (NA), the larger the cone of light which can be coupled into the fibre and the more the light exiting the fibre will spread out. The fibre used for this project has NA=0.11.

It is possible to deliver the light to the face of the MAPMT by placing the end of the fibre as close to the borosilicate glass window as possible. However this involves having an air gap between the fibre and MAPMT face so that the fibre can be moved across the face during a scan. Due to this air gap and the 800 μm thick window, the light spot will have spread to a diameter of ~200 μm before its incidence on the photo-cathode. Since the 300 μm gap between MAPMT pixels has to be accurately mapped, a more acceptable light spot diameter must be

constructed. Therefore a graded index lens with a confocal parameter[11] of $\sqrt{2}/0.8$ mm and a 25 mm focal length is mounted on the end of the monomode fibre.

The spot-size was measured by using a CCD camera which had a pixel size of 9 μm and was positioned at the focal length distance of 25 ± 0.5 mm from the end of the fibre, at a position where the photo-cathode surface of the MAMPT would be placed. The spot-size distribution is shown in Figure 2-4 and has an approximate upper limit on the full-width-half-maximum (FWHM) of 47 μm. The top of the distribution is missing as it overflows the ADC
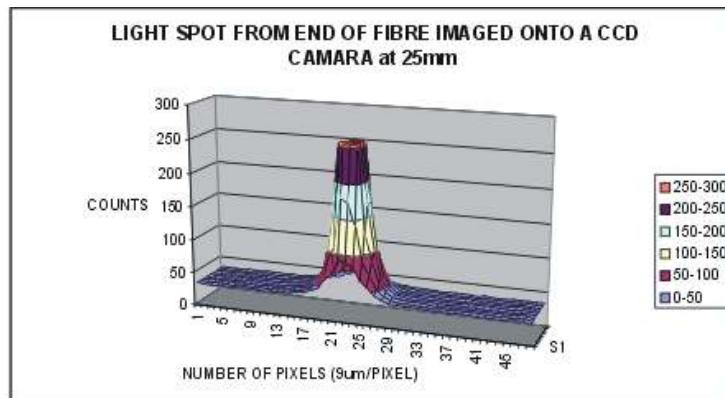
Figure 2-4 Spot-size distribution of light from the end of the fibre.

range of the camera.

To enable scanning of the light across the face of the MAPMT in x and y, two motorised stages were used, each with 1 μm resolution. One moved the MAPMT in the horizontal plane, the other moved the end of the fibre in the vertical plane. A stepper motor driver unit, interfaced to a PC, was designed and built to control the stages.

2.4   Characteristics of the M64 MAPMT

This section characterises the performance of the Hamamatsu R5900-00-M64 MAPMT when under test in the setup described in section 2.3.

---

[11] The confocal parameter is defined as the distance perpendicular to the place of the focussed spot at which the intensity falls by a factor of 2 and the spot radius increases by a factor $\sqrt{2}$.

Single-photon resolution

The single-photon resolution of the M64 was studied. Although the discrimination of one photon from two or more in the same pixel is not essential for LHCb RICH purposes, it does have some advantages [TP98].

The monomode fibre was positioned centrally on an MAPMT pixel. To select the working voltage of the LED to give single-photons, the LED voltage controlling the photon intensity, was increased from 6.1 V to 12.5 V in steps of 0.2 V. For each setting the LED was triggered 100,000 times. No photons were detected below 6.1 V. Figure 2-5 shows two typical spectra, with the LED voltage at 7.3 and 9.7V. A fit was made to these spectra to determine the
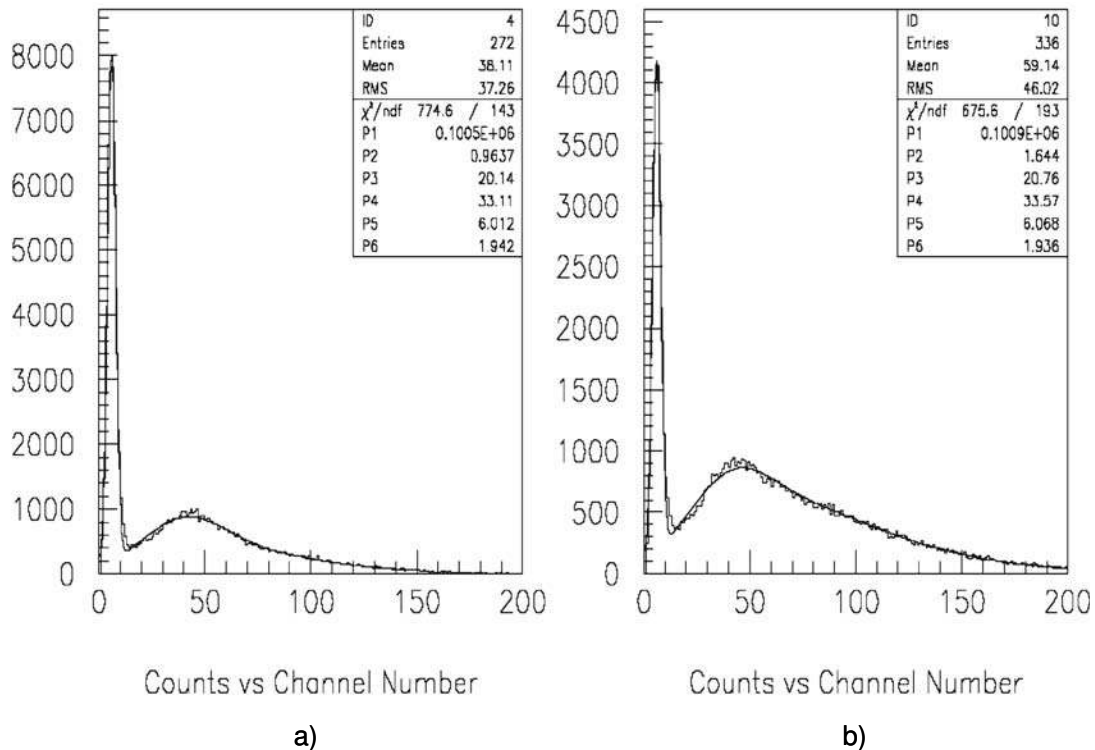
Figure 2-5 Fitted spectra in ADC counts (0.25pC/channel, bias voltage –950V). The LED voltage is a) 7.3 V and b) 9.7V. The fitted parameters are given as follows: P1=normalisation, P2 =mean number of photons detected, P3=sigma of the single photon peak, P4=position of single photon peak with respect to pedestal mean, P5=pedestal mean and P6=pedestal sigma. P3 to P6 are in units of ADC counts.

position of the single-photon peak and the mean number of photons detected. The fitting procedure uses a combination of Gaussian distributions to simulate a) the resolution of the MAPMT tube to a single-photon, and b) the noise peak. A Poisson distribution is used to

simulate the relative probability for producing 0, 1, 2 and 3 photons. For an in-depth study of fitting responses from M64 tubes the reader is referred to [RAD02]. It can be seen that good fits to the data are produced giving, in these two cases, a mean number of photons of 0.96 and 1.64, respectively. Although the multi-photon peaks are not resolved by the eye, the fits indicate that the multi-photon response of the MAPMT tube is well understood.

Relative efficiency across the tube surface

Hamamatsu quotes the photo-cathode efficiency of the MAPMT to be ~20%, however an absolute measurement of this is beyond the scope of this work. Therefore the relative efficiency across the surface of the tube was measured, as was the local detection efficiency caused by the gap between pixels. The entire area of the tube was scanned in 100 μm steps with the light source. The LED voltage was set at 9 V, and rows of eight pixels were read out in turn. For each, a pedestal run was made and this enabled a 5σ noise cut. The LED was triggered 10,000 times at each position of the light source. The signals received from the MAPMT above the noise cut were classed as hits and counted. The total counts for each position was then plotted. An example of a scan is shown in Figure 2-6.
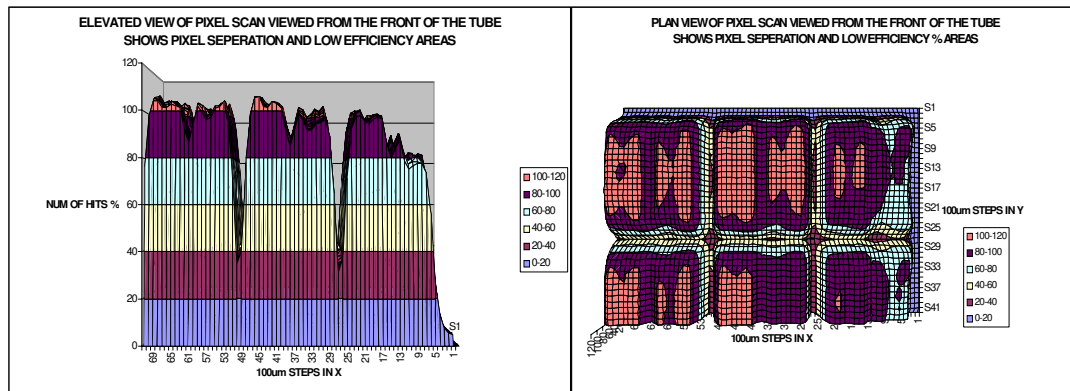
Figure 2-6 Example of 1D and 2D scans across the face of the MAPMT photocathode.

Dead areas between pixels and inefficiencies at the centre of the pixels are clearly seen in Figure 2-6. These effects are caused by the HT focusing wire geometry of the tube, explained below. A photograph looking through the MAPMT window, Figure 2-7, is of a single pixel. This has two electron entry gaps into the dynode chain. Around the pixel is a pair of focusing

wires and through the middle of the pixel is a single focussing wire, slightly off centre, which optimises the focussing properties. These areas define the inefficient regions of a pixel.
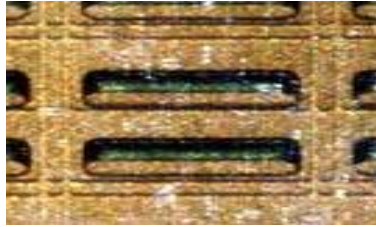
Figure 2-7 Photograph of a single MAPMT pixel.

The integrated tube efficiency, relative to the fully efficient regions of the pixels is now calculated. This is done by taking an average of the counts over the flat top of all central pixels (ignoring the dip in the centre of each pixel) and this is then taken to represent 100 % relative efficiency. The relative efficiency integrated over the surface of the tube is then found to be 74 % ± 3 %. The full efficiency of the MAPMT for a single-photon is therefore approximately $\sim 0.2 \times (0.74 \pm 0.03) = 14.8 \% \pm 0.6 \%$., assuming the photo-cathode efficiency to be 20 %.

Pixel gain uniformity

The variation of photo-electron gain at various locations of the tube (with particular interest at the edges and centres of the pixels) was studied. The bias voltage was set at –950 V and the LED voltage at 9V. Firstly, to study the gain variation within a pixel, pixel 13 (defined by the numbering scheme shown in Figure 2-1) was measured in a 5x5 grid of 300 μm steps. At each step the LED was triggered 100,000 times and the signals from the MAPMT histogrammed and the single-photon peak was fitted as in Figure 2-5. The resulting variation of gain over the pixel was found to have an RMS of 15±5 %. Secondly, to study the gain variation pixel-to-pixel, the light source was scanned over two fully efficient points on each of the 64 pixels and similar single photo-electron spectra measurements taken. A comparison of the single-photon peaks (parameter P4 of Figure 2-5) gives the relative pixel-to-pixel gain uniformity for a single-photon response. The distribution is shown in Figure 2-8. Although the gain is generally uniform, there are certain channels with relatively low gain response. As mentioned

earlier, this is a feature of the M64 tube [MAP97]. If the six pixels with low gain are ignored, the gain variation is less than 20 %.
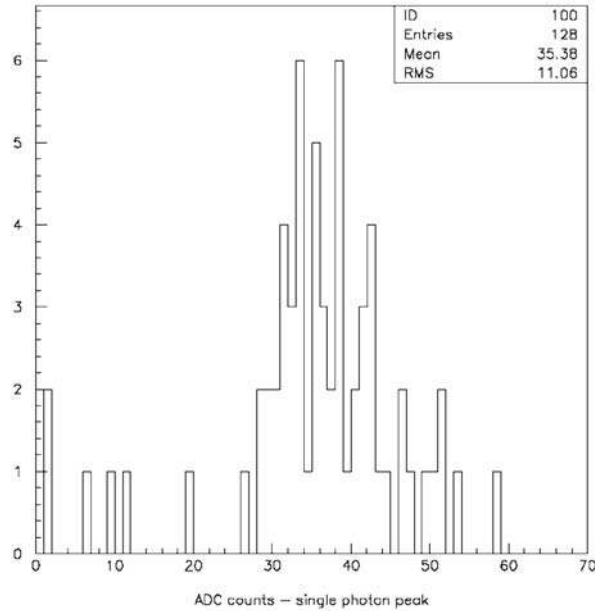
Figure 2-8 Distribution of the mean of the single photon peaks in ADC counts for 64 pixels.

The gain variation can be primarily attributed to the fabrication of the first dynode. Even very small variations in the secondary-emitter material at this dynode stage will affect the number of electrons emitted; this small variation will then be multiplied by the full stage.

Implications of gain variation for the readout electronics

Several factors will influence the SNR of the front-end electronics. These are the typical signal response of the MAPMT from a single-photon, the maximum number of photons detected in a single pixel, the variation of gain over the MAPMT, the front-end electronic noise and the dynamic output signal range of the amplifier. The specifications of SNR and the dynamic range of the input signal to which the chosen amplifier must comply are now discussed.

Considering the dynamic range requirements, the gain variation over a tube has an RMS value of ~11 ADC counts (refer to Figure 2-8). Taking plus and minus the RMS value about the mean ADC count provides a working definition of the limits over which a photon

response might be expected. Therefore the pixel gain variation over a tube can be crudely estimated to be $\frac{Mean+RMS}{Mean-RMS}\approx2$. In addition, there is a tube-to-tube gain variation which is given by [MUH00] as ~2. Taking the quadratic sum of these two values gives an overall gain variation of ~3. As the expected number of photons within a given pixel is estimated to be a maximum of 3, the total dynamic range at the input of the amplifier is, in terms of photon hits, 3x3=9. To consider this in terms of electrons at the amplifier input depends on the HV setting of the tube. To stay within the dynamic range of the amplifier, the HV has to be compatible with the amplifier's dynamic output transfer function whilst giving as large a signal as possible compared to the electronic noise. A SNR specification can be given for an average tube with a typical tube gain and gain variation; the tube-to-tube SNR variation can then be alleviated by adjustment of the tube high voltage on a tube by tube basis. The SNR of the tube studied is $\frac{33}{1.9}\approx20$ from Figure 2-5. Considering this to be a typical tube, a gain variation of 2 would imply a SNR of only 10 in the worst case. A front-end ASIC to read out the MAPMT should therefore ideally improve the SNR to ~40 for the typical case in order to give a SNR of ~20 in the worst case. Whether this SNR is possible will ultimately depend on the detailed amplifier design.

2.5 Summary

The 12-dynode R5900-00-M64 H7546B MAPMT performs to the manufacturer's specifications and is a suitable photon detector for the LHCb RICH detectors. The M64 family has improved in performance over the years. Although the statistical dynode gain variation smears the single-photon resolution, a fitting process to the signal spectra can recover the single-photon response and mean number of photons detected. Due to the focussing and gain structure of the MAPMT, the relative efficiency is reduced to $74\pm3\%$. With a single-photon gain of up to 1×10^6 electrons, the readout electronics are greatly simplified for the MAPMT. However, the amplifier must be able to accommodate the dynamic range of the input signal of 9 and give a SNR ratio of ~40 for a typical pixel response so as to ensure all signals are well above the noise.

The studies described in this chapter do not take into account the detection efficiency for the entire RICH photo-detector plane as a loss is incurred due to the packing density achievable and the absolute photo-cathode quantum efficiency. The R7600-03-M64 can be used to improve the packing density by ~14 %, since its flange is removed and its spectral range has improved by 50 % due to the new UV-glass window.

Introduction to the design of an MAPMT readout chip

This chapter outlines the important points needed to develop and select a suitable readout chip for the LHCb RICH detectors in the case where MAPMTs are used. The chip selection process and subsequently the so-called 'Beetle' ASIC readout chip and design philosophy are discussed.

3.1   Requirement

At the end of 1999 the Hybrid Photon Detector and LHCBPIX1 ASIC readout chip were chosen as the base-line detector and readout scheme, respectively, for the LHCb RICH detectors. This readout scheme was still in the early development stages and was demanding in its design requirements. Therefore it was essential that a back-up solution was available. The back-up was the MAPMT, discussed in Chapter 2. The tube had already been under study and fully characterized but there was no existing readout chip that conformed to the LHCb specifications and could be directly coupled to an MAPMT. Three options existed: 1) to design and fabricate a completely new ASIC, 2) to take an existing ASIC and match the MAPMT output with discrete or active components on an external PCB, and 3) to modify an existing ASIC to be compatible to the MAPMT. Option 1) was discounted due to the design effort and cost. Option 2) was undesirable due to the extra cost and problems associated with the manufacturing of a PCB to mount the tubes and chip along with the external components needed for input matching; this area would be very dense with track lines and components. Option 3) was considered the most favorable as the cost of the ASIC fabrication would be shared with the VELO and tracker projects, PCB complications would be removed, a minimal amount of design effort would be needed (as this would only be a modification), and support from the chip designers would be readily available.

An additional possibility exists to modify the MAPMT gain to reduce the output signal to ~20k electrons for a single-photon input, which is approximately the number of electrons that

an average silicon readout chip can tolerate. This gain reduction could be combined with options 1) and 3) to make the chip design process less demanding. Preliminary tests of modifying the MAPMT high voltage resister ladder network to give smaller gain was unsuccessful, although later in 2002 Hamamatsu would offer a tube tailored to LHCb requirements.

The following sub-sections describe some of the fabrication processes available for ASIC design, the ASICs that exist in the high energy physics community that have been considered for use with an MAPMT, and finally a description of the ASIC chip used.

3.1.1 ASIC manufacturing processes

Chip fabrication is available worldwide but becomes extremely expensive or even not available for production of low numbers, i.e. less than a few 100 wafers, as the cost of the mask-set is prohibitively high. The low number of wafers required by the high-energy physics community therefore limits the chip fabricators available. In addition, radiation tolerant layout techniques or processes must be allowed by the fabricator. For the 0.25 μm CMOS[12] deep sub-micron process the radiation tolerance is gained by special layout techniques, discussed in section 3.4. Another method of making a chip design radiation hard is to use special process techniques like TEMIC Bi-CMOS DMILL, discussed below. The following companies are available to the high-energy physics community:

Austria Mikro System International GmbH (AMS) CMOS:

AMS [AMS] provides a variety of production-proven and industry-standard process technologies. The core technologies consist of 0.35 μm and 0.8 μm digital and mixed signal CMOS and Bi-CMOS processes. The Helix chip, a predecessor to the Beetle, was fabricated in this technology. However, the increased demands on radiation tolerance now makes this technology unusable due to the limited geometry size of 0.35 μm.

---

[12] Complementary Metal-Oxide-Semiconductor.

IBM Blue Logic CMOS 6SF

This IBM [IBM_CMOS] technology is an advanced high-yield process technology featuring 0.25 μm lithography. The fine lines and high densities characterizing this silicon process can support high-performance graphics, leading-edge microprocessors, communications, and computer data-processing applications. The technology provides a P-epitaxial[13] layer on a p+ substrate feature. This gives improved latch-up protection due to the thin epitaxial layer, described in more detail in section 3.4.2. Planarized passivation and interlevel dielectrics enable manufacturing of more levels of metal, permitting higher densities and more integration of components. The IBM processes are available through Mosis, which is the US equivalent to Europractice (see below), but is considered expensive. More detail on the 0.25 μm process is given in section 3.1.3.

Europractice Alliance Network:

Europractice [EURO_p] offers a network of design houses that work together with the Europractice IC Service in order to offer a total design and fabrication solution. This includes design, low cost prototyping and production, packaging and testing. At present there are around 20 fabricators to choose from that offer specialized processes, for example high voltage and mixed processes (Bi-CMOS). A tender was placed for LHC chip development but none of the Europractice fabricators could offer both full access to the wafer during production or a competitive price.

Temic DMILL

The DMILL (Durci-Mixte sur Isolant Logico-Lineaire) process, developed in France by a consortium [HU96]. This was used for the SCTA128 chip that was considered for the LHCb RICH detectors and is therefore discussed in a little more detail. DMILL offers a radiation hard, high-speed analogue and digital process to the LHC community. The DMILL process combines bipolar and CMOS in a 0.8 μm process with a gate oxide thickness of 14 nm. It offers

---

[13] Epitaxial layer: A single crystal layer that has been deposited or grown on a crystalline substrate, both having the same structural arrangement. The chip circuitry is in the epitaxial layer while the substrate acts as a structural support.

a variety of devices: MOS FET, fast BJTs, JFETs and high value resistors with low stray capacitance. The technology employed is silicon-on-insulator (SOI) which uses a thin oxide to separate the ~70 nm active substrate from the silicon carrier substrate. The thin active substrate is a key factor in making this process radiation hard to single event upsets, as the bulk volume under the transistors does not create large numbers of electron hole pairs for an ionizing particle
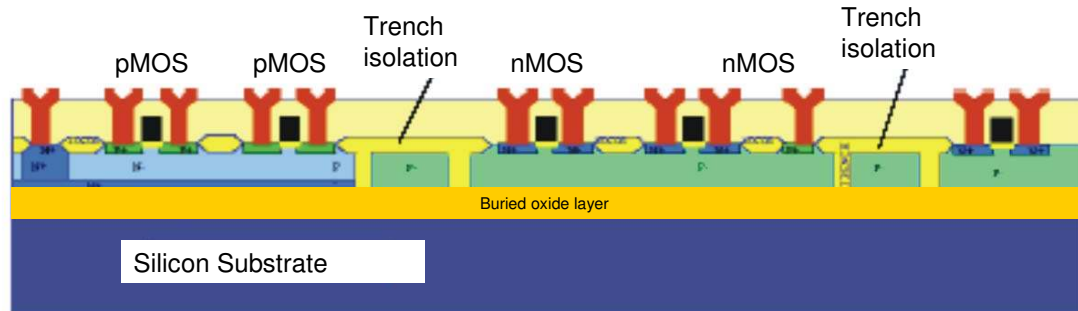
Figure 3-1 Cross sectional view of NMOS and DMOS in a DMILL process [SEX01].

passing through it. Figure 3-1 shows the cross sectional view of NMOS and PMOS transistors in the DMILL process. Steps are required to further reduce radiation effects:

- 'Trench isolation' is used to prevent parasitic paths that are created by ionizing radiation in the oxides. The trenches also reduce the volume of the bulk under the transistors.

- Trapped interface states at the oxide and diffusion boundaries are removed by the surface finishing processes.

- Etching with high frequency plasma is kept to a minimum to prevent generation of trapped interfaces during fabrication.

In comparison to the 0.25 μm process, DMILL is more suited to analogue design as it offers fast bipolar devices but requires more real estate due to the trench isolation and the fact that it is a 0.8 μm process. This makes the CMOS devices actually about four times slower than for the 0.25 μm process. Also the fabrication of identical devices (matching) is better with the 0.25 μm process as the accuracy of the lithography requires it. The price of production is comparable for both processes but the 0.25 μm has a larger market which facilitates portability of the design to other fabricators. The DMILL process does not have the dense packing properties of the 0.25 μm process and is nearing the end of its production life. Finally, with the

correct layout procedures, both DMILL and the 0.25 μm processes can be considered radiation hard to a total dose of 10 Mrad.

3.1.2   Chip selection

Several readout chips for high-energy physics experiments have been developed over the last few decades with demands for high speed performance, reliability, radiation and magnetic field tolerance, low cost, availability (both in terms of development and technology), extremely high sensitivity and low noise, all becoming ever more demanding. Generally the readout chips are similar in principle but tailored to their specific task. The majority of the readout chips have been developed for silicon detectors where low input impedance for charge collection is required, this being achieved with a charge-sensitive front-end amplifier.

For LHCb, all readout chips must conform to the LHCb global electronics scheme, outlined in section 1.3; thus the data capture rate, storage (latency) and 'time to readout' are all specified. A further consideration is the technology of chip fabrication as this not only affects the radiation-tolerance properties but also the chip availability when time comes for manufacturing. Some of the chips considered for MAPMT readout are presented in the following bullet points. They all require some form of modification for MAPMT use.

- The Beetle family was first developed for the LHCb VELO sub-detector, therefore it meets all of the LHCb global requirements including the required radiation tolerance. The Beetle in its standard form cannot accept the dynamic range of the MAPMT signal and would need customization to the front-end amplifier. A major advantage of the Beetle chip is that it allows operation in either binary or analogue mode. The Beetle is built on 0.25 μm CMOS technology Chapter 5 gives further detail of the modification requirements.

- The SCTA128 [KAP98] [KAP98n] was a backup solution for the LHCb VELO detector. This readout chip was originally designed for the LHC ATLAS experiment and then later modified for the VELO to conform to the global LHCb specifications. The SCTA128 is very similar to the Beetle chip in functionality but differs in the fabrication process; the SCTA128 is fabricated in a

DMILL 0.8 μm BiCMOS technology. Like the Beetle, the SCTA128 cannot accept the dynamic range of the MAPMT. At the time of writing, further effort on the SCTA128 for the VELO had been abandoned and the Beetle was the chosen option.

- The APVm [JON98] is designed for the LHC CMS inner tracker detector and conforms to the LHCb radiation-tolerance requirements. However to use the chip in a 40 MHz sampling manner necessitates operation in the so-called 'deconvolution' mode, which precludes the LHCb requirement to read out consecutive triggers. In the deconvolution mode the signal is passed through a Finite Impulse Response (FIR) filter which removes the slow components of the signal to reduce pile-up effects. This therefore removes some of the analogue information. The APVm was used to read out an MAPMT in a test beam during the summer of 1999 [MUH00]. The MAPMT was coupled externally to the chip by charge division using a 'tuning fork' capacitor on a ceramic fan-out. The external components caused large pulse overshoots at the amplifier output stage and cross-talk at the inputs. This was acceptable for charactering the MAPMT in the test beam but would not be suitable for the final system.

- The ALICE pixel chip [TDR00] underwent design modifications for LHCb and was made compatible to reading out pixel HPDs. The LHCb version is called the 'LHCBPIX1'. Although fully compatible to the LHCb requirements, the sensitivity of this chip is very high and would need a considerable modification to accept an MAPMT signal. Furthermore it has no analogue readout mode, which is highly advantageous for MAPMT readout due to unpredictable common mode noise and the large gain variations between channels and tubes. The 'LHCBPIX1' is a bump-bonded chip with 1024 channels, which makes the physical interconnections to an MAPMT impractical.

Following careful examination of all the possible chips, the author pursued the option to modify the Beetle to accept the relatively large input signal for MAPMT readout requirements. At that time, the Beetle was in a very early stage of development and therefore allowed the

author to join the design team to develop the chip with the RICH specifications in mind. Furthermore, the technology at the time was relatively new and would therefore still be available at manufacturing time, even if this meant transporting the design to another 0.25 μm fabricator. The modified Beetle adapted for MAPMT readout is named the BeetleMA. A version number is added that corresponds to the Beetle chip that has been modified along with a BeetleMA version number, e.g. Beetle1.2MA0.

3.1.3 The Beetle readout chip

A general description of the Beetle chip is discussed here. For more technical information the reader is referred to [BAU03] and [HD_U].

Chip name	Submission Date	Chip size [mm^2]	Description
BeetleFE1.0	May 1999	2 x 2	Front-end test chip
BeetleBG1.0	May 1999	2 x 2	Bias generator test chip
Beetle1.0	April 2000	5.5 x 6.1	Readout chip
BeetleCO1.0	April 2000	2 x 2	Comparator test chip
BeetlePA1.0	April 2000	2 x 2	Pipe-amp test chip
BeetleMA1.0	April 2000	2 x 2	Front-end test chip
Beetle1.1	March 2001	5.5 x 6.1	Readout chip
BeetleFE1.1	May 2001	2 x 2	Front-end test chip
BeetleFE1.2	May 2001	2 x 2	Front-end test chip
BeetleSR1.0	May 2001	2 x 2	SEU robust test chip
Beetle1.2	April 2002	5.1 x 6.1	Readout chip
Beetle1.2MA0	December 2002	5.2 x 6.1	Readout test chip for MAPMT
Beetle1.3	July 2003	5.4 x 6.1	Readout chip

Table 3-1 The Beetle family.

A collaboration of three institutes, Heidelberg [HD_B], Oxford [OX_B] and NIKHEF [NIK_B], undertook the design of a chip that could be used for the VELO, Inner Tracker and RICH. Heidelberg is the design centre and is where the chip database resides. The

development of the Beetle chip started in the last quarter of 1998 and the first chip submission was in May 1999. Table 3-1 lists the family members. The Beetle architecture is built on the HELIX128 [TRU00], which is a readout chip developed for the ZEUS [ZEU], HERMES [HER] and HERA-B [HER95] detectors at the HERA electron-proton accelerator at the DESY Laboratory, Hamburg, Germany.

The Beetle chips are manufactured in a 0.25 μm CMOS process that uses one polysilicon layer, three metal layers (with options for up to five), metal insulator metal capacitors (MIM), poly and diffusion resistors and a gate oxide thickness of 6.2 nm. With the appropriate layout techniques described in section 3.4, this process is radiation hard.

General readout scheme of the Beetle

Figure 3-2 shows a block diagram of the Beetle architecture. The Beetle is a 128 input-channel device. The physical implementation of the input pads is a fourfold staggered pad array having a pitch of ~40 μm. The pads are followed by electrostatic discharge (ESD) protection diodes. Each channel consists of a continuous sampling charge-sensitive preamplifier (CSA), the output of which is shaped by an active CR-RC shaper stage which has a rise-time of 25 ns and fall time to 30% remaining after a further 25 ns. A source follower buffers the output of the shaper. The shaper times, channel occupancy rate, overshoot and noise contribution can all be optimised in a limited way by adjusting on-chip DACs.

The buffer output can be utilised in three ways: 1) for analogue readout the signals are sampled and stored in the pipeline memory, 2) for binary readout a comparator with adjustable polarity and threshold per channel provides a binary signal which is sampled and stored in the pipeline memory, 3) for the Level_0 trigger output of the VETO detector, four adjacent binary channels can be ORed, latched and multiplexed at 80 MHz using LVDS[14]. The advantage of binary readout over analogue is that the bandwidth required from Level_0 to Level_1 is considerably reduced. Binary transmission needs only 'one bit' of information, either a '1' or a

---

[14] The ORed channels have dedicated output ports labelled 'compout' which are located on either side of the chip which use the LVDS standard, Low Voltage Differential Signals.

'0', while the analogue readout would typically require 8-bits per channel, dependent on the resolution of the ADC.

The binary/analogue choice is particularly important for the RICH MAPMT readout. Although the binary readout would have considerable financial saving due to the reduced number of fibre-optic transmission cables required between Level_0 and Level_1, for low-gain pixels, the MAPMT does not have a good separation between the signal and noise. This makes the binary threshold difficult to set. If the binary threshold is set 'low' then many noise signals would be misconstrued as genuine MAPMT photon hits. If the binary threshold is set 'high' then many of the genuine MAPMT photon hits would be discarded. The binary readout characteristics of the MAPMT are a subject of further study.

Downstream of the buffer, the sampled signal, either analogue or binary, is captured at 40 MHz and stored in a pipeline buffer (also known as a 'ring buffer'). The pipeline has a programmable latency of up to 160 sampling intervals, i.e. 4 µs at 40 MHz. The pipeline buffer is realised in FET gates which are configured as capacitors; they are referred to here as 'pipe-FETs'. These pipe-FETs reduce the silicon area required compared to MIM capacitors which must have a total area of less than 300000 $(\mu m)^2$, defined by the fabrication rules. Each pipe-FET has an associated read and write switch that allows cells to be read out or overwritten. Read and write cycles are managed by pointers which increment the pipeline buffer at the system clock speed. The read pointer follows the write pointer with a delay set by the latency register, 160 sampling intervals in the case of LHCb. On a readout request, the read pointer flags the addressed column in memory and moves on. Readout is then initiated, releasing the read switch and transferring the stored charge in the pipe-FET to a resetable charge-sensitive amplifier, the 'pipeline readout amplifier'. This initiation time takes 100 ns. Within this 100 ns an events header is added. The data from the pipeline readout amplifiers from a triggered event are multiplexed 128:4, common mode subtracted using the 'sense channel' (refer to Figure 3-2) and driven off chip within 900 ns.
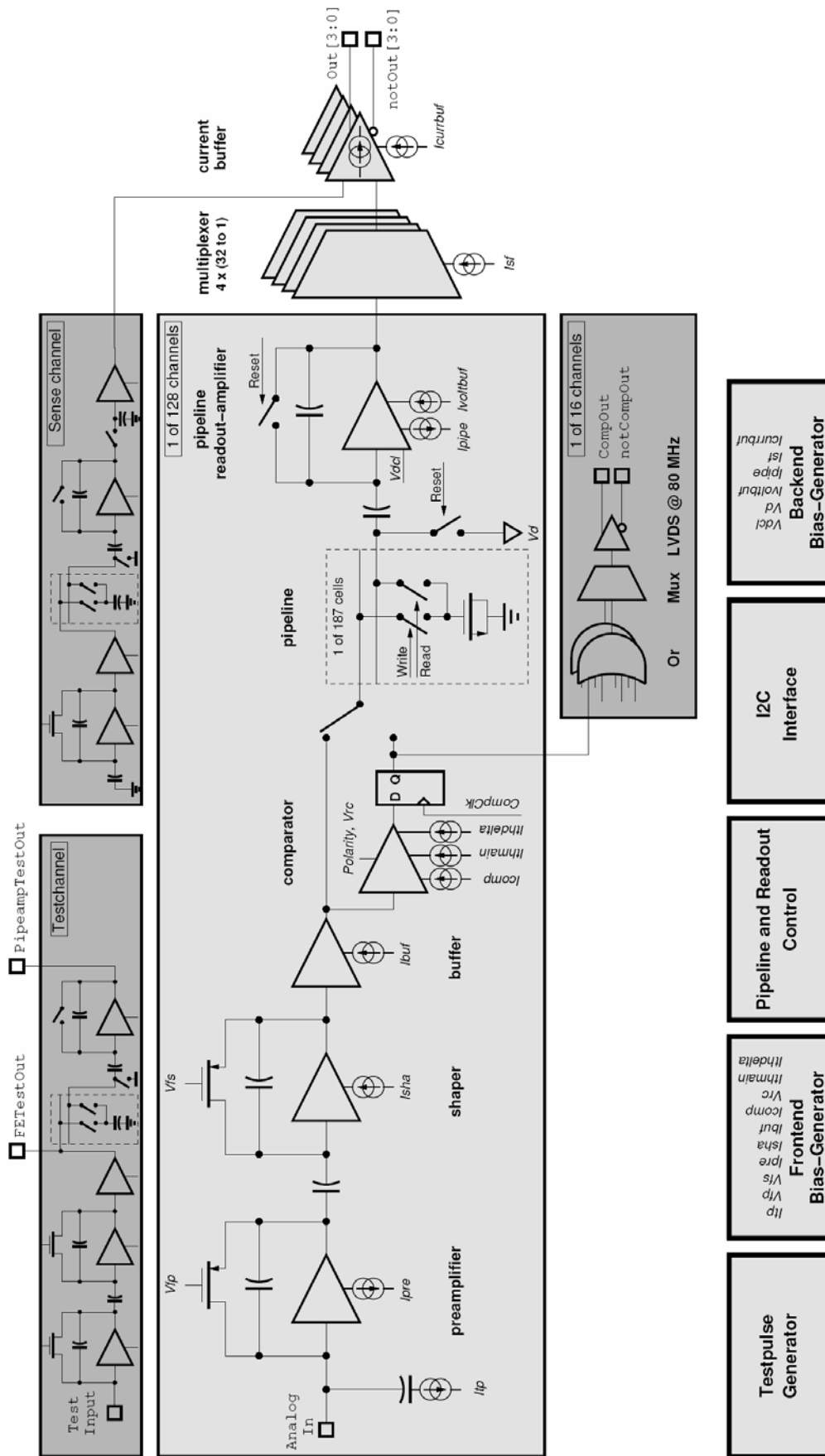
Figure 3-2 Block diagram of the Beetle ASIC.

47

3.2 MOS FET characteristics

The Beetle chip uses MOS FET technology. In order to introduce the Beetle design considerations, an overview of the theory of MOS FET devices is necessary. Characterization of MOS FETs is available in numerous text books, for example [SED98], and therefore will not be repeated here. This section is intended to be a reminder of the basic features of design with MOS FET devices.

The terminology employed throughout this thesis for signal definition is as follows: total instantaneous quantities are denoted by lower case symbols with uppercase subscripts, for example i_G and v_G. Direct-current (DC) quantities will be denoted by an uppercase symbol with an uppercase subscript, for example I_G and V_G. Finally, incremental signal quantities will be denoted by a lowercase symbol with lowercase subscript, for example i_g and v_g. This notation is illustrated in Figure 3-3.
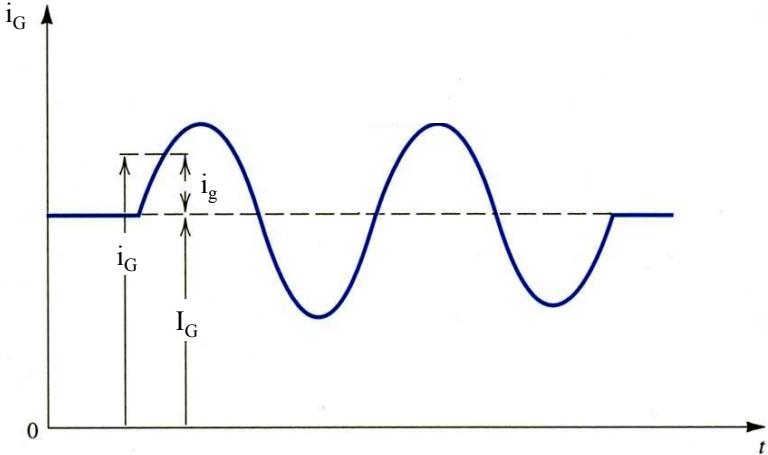
Figure 3-3 Symbol convention employed throughout this thesis.

The enhancement MOS FET is the most widely used device in integrated circuit design. n-channel devices are preferred to p-channel devices because of their higher transconductance, a result of the fact that μ_n is two to three times higher than μ_p, where μ_n and μ_p are the electron and hole mobility, respectively. Both devices are, however, utilized in CMOS technology, currently the most popular technology for the design of analogue and digital integrated circuits. In ASIC design only two parameters are at the designer's disposal, namely the aspect ratio W/L,

given in terms of the width (W) and length (L) of the gate dimensions, and the voltage between gate and source, V_{GS}. The enhancement MOS FET has two operational regions, the 'triode' and 'saturation' regions. The saturation region is usually used in amplifier designs and in the ideal case can be considered to be the region where $\frac{\Delta V_{DRAIN}}{\Delta I_{DRAIN}} = \infty$. In practice the saturation region occurs when $V_{DS} = (V_{GS} - V_t)$ and has an incremental output resistance r_o. The variables r_o, V_t and V_{DS} are defined in the following text. Figure 3-4 depicts the characteristic curves of an n-channel enhancement-type MOS FET indicating the triode and saturation operating regions and the affect that v_{DS} and i_{DS} have on the output resistance r_o.
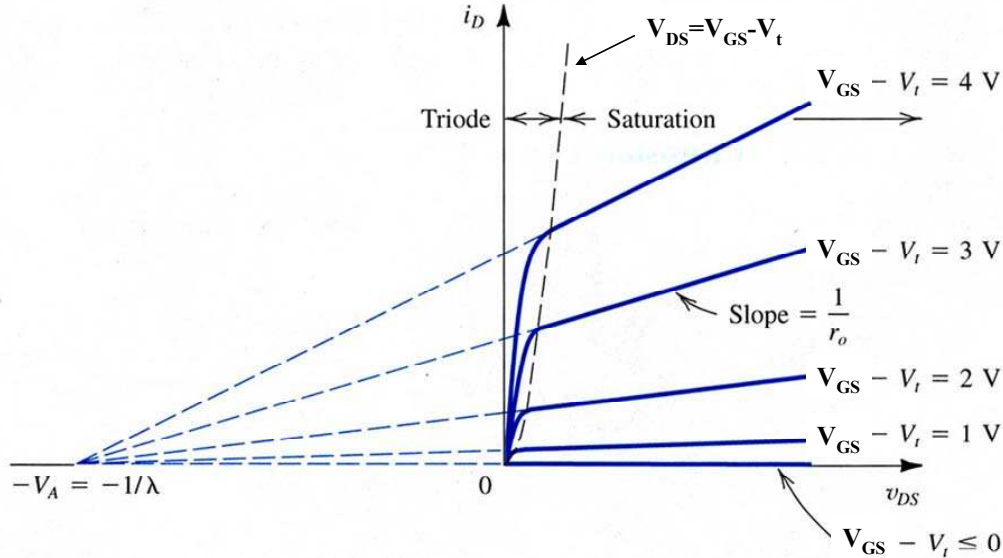
Figure 3-4 An n-channel enhancement-type MOSFET with V_{GS} and v_{DS} applied. The triode and saturation operating regions are shown. Also shown is the affects of v_{DS} on i_D in the saturation region giving the output resistance r_o [SED98].

For an NMOS FET the following equations apply:

for the triode region where $V_{GS} \geq V_t$, $V_{DS} \leq V_{GS} - V_t$

$$I_D = k_n' \left(\frac{W}{L} \right) \left[(V_{GS} - V_t) V_{DS} - \frac{1}{2} V_{DS}^2 \right],$$ Equ 3-1

and for small V_{DS}

$$R_{DS} = \frac{V_{DS}}{I_D} = \left[k_n' \left(\frac{W}{L} \right) (V_{GS} - V_t) \right]^{-1}.$$

<div align="right">Equ 3-2</div>

For the saturation region where $V_{GS} \geq V_t$, $V_{DS} \geq V_{GS}\text{-}V_t$,

$$I_D = \frac{1}{2} k_n' \left(\frac{W}{L} \right) (V_{GS} - V_t)^2 (1 + \lambda V_{DS}).$$

<div align="right">Equ 3-3</div>

In the equations above, I_D is the drain current, V_t is the 'threshold voltage' (see below), V_{DS} is the voltage between drain and source, and k_n' is the process transconductance parameter given by $\mu_n \times C_{ox}$, where C_{ox} is the capacitance between the gate and the conduction channel. The channel-modulation parameter is given by $\lambda = \frac{1}{V_A}$ where V_A is the 'Early voltage' and is an intersect point on the $-V_{DS}$ axis of the i_D vs v_{DS} plot that is extrapolated from the linear part of the curve when in saturation. Typical values of λ, normally defined for the minimum allowed L dimension used in any one technology, range from greater than 0.1 V^{-1} for short-channel devices to 0.01 V^{-1} for long channel devices [GEI90].

The value of V_{GS} at which a sufficient number of mobile electrons accumulate in the channel region to form a conduction channel defines the 'threshold voltage', V_t. The value of V_t is controlled during device fabrication and has a characteristic given by [SED98]

$$V_t = V_{t0} + \gamma \left[\sqrt{2\phi_f + |V_{sb}|} - \sqrt{2\phi_f} \right],$$

<div align="right">Equ 3-4</div>

where γ is the fabrication process parameter given by $\gamma = \sqrt{2qN_A\varepsilon_s}/C_{ox}$, N_A is the doping concentration of the p-type substrate, q is the electron charge 1.6×10^{-19} C, $\varepsilon_s = 1.04 \times 10^{-12}$ *F/cm* is the permittivity of silicon and ϕ_f is a manufacturing parameter,

typically 0.3 V. V_{to} is the threshold voltage for $V_{sb}=0$ where V_{sb} is the voltage between the FET source terminal and the bulk (also called the body).

For a PMOS device V_t and V_A are negative, and for the triode region, $V_{GS} \le V_t$ and $V_{DS} \ge V_{GS}-V_t$. For the PMOS saturation region, $V_{GS} \le V_t$ and $V_{DS} \le V_{GS}-V_t$.

Transconductance (g_m), which is the small-signal V-I transfer function for a FET device, is given by the following equations for a FET operating in the saturation region:

$$g_m = \sqrt{2k'(W/L)}\sqrt{I_D} \,, \qquad \text{Equ 3-5}$$

$$= k'(W/L)(V_{GS} - V_t), \qquad \text{Equ 3-6}$$

$$= \frac{2I_D}{V_{GS} - V_t}. \qquad \text{Equ 3-7}$$

The incremental output resistance (r_o) of a FET is the derivative $\dfrac{\partial V_D}{\partial I_D}$ for small-signals in the saturation region and is infinite in the ideal case. The value of r_o can be found from

$$r_o = \frac{|V_A|}{I_D}. \qquad \text{Equ 3-8}$$

For small-signal analysis it is usual to remove all of the DC sources and replace them with low impedance paths to a common point i.e. ground. This makes analysis of the circuit far less complex. The FET's transconductance parameter g_m and output resistance, r_o are used to represent its current and voltage characteristics. An example of a common-source amplifier and its equivalent small-signal model is given in Figure 3-5.
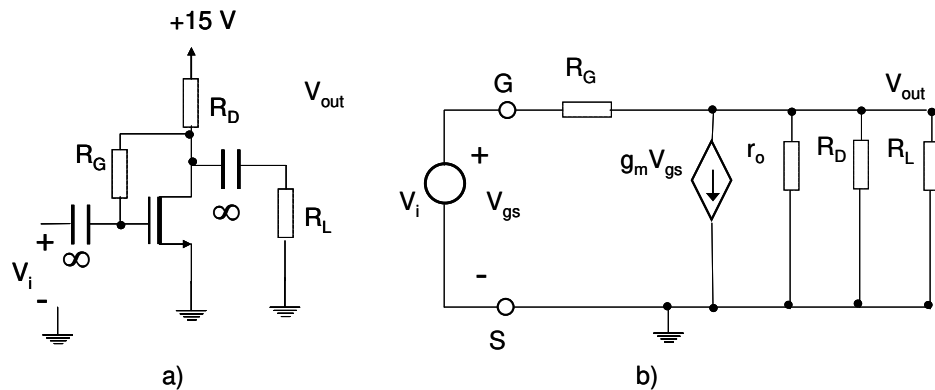
Figure 3-5 a) A common-source amplifier and b) the equivalent small-signal model.

In general the source terminal is usually tied to the bulk, as in the case of Figure 3-5. However, there are many occasions where the source will be tied to another voltage potential node of a circuit, a good example is a common-gate amplifier where the source potential is tied to the input voltage, V_{in}. As V_t determines the conduction channel thickness, which in turn determines the value of I_D, it follows that the body acts as another gate for the MOS FET. This is easily accounted for in small-signal analysis circuits by just adding another transconductance parameter g_{m_sb}, where g_m is multiplied by the body-effect parameter X, which typically lies in the region of 0.1-0.3 [SED98]. The value of X can be found from

$$X \equiv \frac{\partial V_t}{\partial V_{sb}} = \frac{\gamma}{2\sqrt{2\phi_f + V_{sb}}} \quad .$$

Equ 3-9

The current mirror configuration

In ASIC design the use of current-sources/mirrors constructed from FETs, both for FET biasing and as circuit load elements, is extensively used in CMOS design due to the small size and good matching of FET devices compared to discrete devices such as resistors. The use of a current-source as a load element is widely used as this offers a large output resistance for small signals. The large resistance consequently gives a large voltage gain. Current mirrors are useful for distributing bias currents across chips as it is much simpler to program a current bias with another current. From one initial reference current all the other required bias currents can be

generated by using a FET to mirror it. The FET can be chosen with different dimensions to increase or decrease the mirrored current.
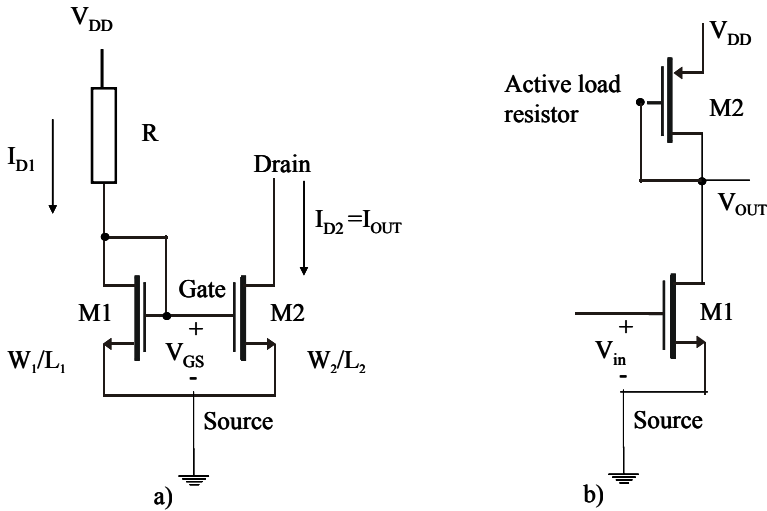
Figure 3-6 a) Current-mirror, b) common source amplifier with active R load.

Figure 3-6 a) shows a standard current mirror configuration. A current flows through M1 which is defined by the gate-source voltage, V_{GS1}. Since $V_{GS1}=V_{GS2}$, ideally the same current, or a multiple of the current in M1, flows through M2. By ignoring the $(1+\lambda V_{DS})$ term in Equ 3-3 and assuming that V_{GS} and V_t are exactly the same for both devices (which is determined by how well the devices are matched in layout), a simplified expression for the current ratio I_{D2}/I_{D1} is given as

$$\frac{I_{D2}}{I_{D1}} \approx \frac{\dfrac{K'W_2}{2L_2}(V_{GS}-V_t)^2}{\dfrac{K'W_1}{2L_1}(V_{GS}-V_t)^2} = \frac{W_2 L_1}{W_1 L_2}.$$
<div align="right">Equ 3-10</div>

This demonstrates how to adjust the W/L ratio of the two devices to achieve the desired output current $I_{D2}=I_{OUT}$. If the FETs are the same geometric size, the same drain current flows in each FET providing M2 is in the saturation region.

Figure 3-6 b) shows a common-source amplifier circuit design which has a gain of simply

$$\frac{v_{out}}{v_{in}} = g_{m1} \times R_{LOAD}.$$
<div align="right">Equ 3-11</div>

In this case R_{LOAD} is an active transistor (M2) in a gate-drain configuration. The equivalent resistance of M2 is simply $\dfrac{1}{g_{m2}}$. In this way, large values of resistance can in principle be made with little real estate. Another configuration which further increases R_{LOAD} is to disconnect the gate-drain connection of M2 and instead have a fixed reference voltage on the gate of M2. This in effect makes M2 a current-source. The value of R_{LOAD} then becomes the incremental resistance r_o, which is much larger than a typical value of $1/g_m$. The gain of the common-source amplifier is then given by

$$\frac{v_{out}}{v_{in}} = g_{m1} \times r_{o1} \| \ r_{o2} \ , \qquad\qquad \text{Equ 3-12}$$

where the $\|$ represents the parallel combination of components.

3.3 Noise and amplifiers

Most particle physics detectors deliver only a small but fast analogue charge pulse from their output; the rise-time of the MAPMT considered in this thesis is ~ 2 ns. To be able to make this signal usable it must be amplified. Due to the sheer number of channels used in particle detectors the amplifier will need to be low power. If precision is required then it will also need to be fast, analogue, and have low noise. Noise is the ultimate limit of detectability of
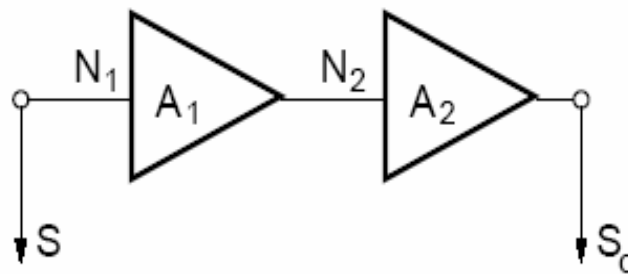
Figure 3-7 Cascade of two amplifiers.

small signals and determines the accuracy to which they can be measured. In the best case the noise performance of an entire readout system is dominated by the pre-amplifier noise. For

example consider the cascade of two amplifiers shown in Figure 3-7. A_1 and A_2 represent the gain of the amplifiers and N_1 and N_2 the noise contribution of each amplifier represented as a voltage at its input. The input SNR is simply S/N_1. Assuming uncorrelated noise, the output SNR is given by [BAU03]

$$\left(\frac{S}{N}\right)^2_{out} = \frac{(A_2 A_1 S)^2}{(A_2 A_1 N_1)^2 + (A_2 N_2)^2} = \frac{S^2}{N_1^2 + \left(\frac{N_2}{A_1}\right)^2} = \left(\frac{S}{N_1}\right)^2 \cdot \frac{1}{1 + \left(\frac{N_2}{A_1 N_1}\right)^2}$$

Equ 3-13

$$\approx \left(\frac{S}{N_1}\right)^2 \text{ in the limit } A_1 \gg 1.$$

Therefore a high gain, low noise first stage is desirable. In the case of the Beetle folded-cascode pre-amplifier described in, Chapter 5 the first stage is the input transistor.

The pre-amplifier noise can be optimised by understanding the noise sources within an amplifier and by matching the output impedance of the detector with the pre-amplifier input impedance. A low noise pre-amplifier design can be achieved if the source impedance is known and suitable layout techniques, along with correct choice of technology, is employed. This is described in detail in section 5.2.

Impedance matching between the source and the pre-amplifier can be achieved by transformer coupling, reactive tuning or reactive matching. Transformer coupling and reactive tuning do not generally lend themselves to particle physics applications due to either the high cost if realised within a monolithic device or the large area of real-estate needed if realised in discrete components. Reactive matching also requires that the source output stage has a fixed frequency response, and that the pre-amplifier input impedance matches the source while maintaining the required amplifier signal transfer function.

The way in which noise signals are characterized also depends on the load type. It is conventional to refer noise voltages to the input of an amplifier that would give the equivalent noise at the output. The advantage of this is that one can easily consider relative noise added by the amplifier to a given signal, independent of amplifier gain. For the case where the input signal is charge this would be expressed as 'total equivalent noise charge' (ENC). The ENC of a

detector readout system is defined as the ratio of the total integrated RMS noise voltage at the output to the signal amplitude due to one electron charge at the input.

The amplifiers described in this thesis are built in CMOS technology and therefore the discussions on amplifier noise sources are restricted to FET devices. The three main noise sources to be considered are thermal, shot and flicker. Thermal noise, also known as Johnson noise, exists in any piece of material that suffers thermal excitation of the atoms. The thermal movement of the charge carriers causes a fluctuation in the charge distribution and an instantaneous electric field is produced which can be measured across a conductor. The root mean squared voltage and current noise, given in many text books [HOR80], is

$$v_n = \sqrt{4kTR} \quad \left[V \middle/ \sqrt{Hz} \right] \qquad\qquad \text{Equ 3-14}$$

and

$$i_n = \sqrt{\frac{4kT}{R}} \quad \left[A \middle/ \sqrt{Hz} \right], \qquad\qquad \text{Equ 3-15}$$

where

k = Boltzmann's constant $(1.38 \times 10^{-23} \ \text{J/K})$,

T = absolute temperature [K]   K = 273 + ^{O}C,

R = resistance $[\Omega]$.

The current flowing between the drain and source terminals in a MOS FET is based on the existence of an inverse resistive channel between them, formed by the majority carries in the substrate under the appropriate control of the gate voltage. In the case where V_{DS} is less than V_{GS}-V_t (the triode region) the inverse channel current noise can be considered to be due to a homogeneous resistance and simply calculated from

$$i_{d-n} = \sqrt{4kT} \cdot \sqrt{g_o} \ , \qquad\qquad \text{Equ 3-16}$$

where g_o is the conductance (the reciprocal of R). A representation of the conduction channel in the triode region is shown in Figure 3-8 a). For the more common case where V_{DS} is large enough that the FET is operating in the saturation region, the resistance of the conduction channel must be considered in small-length Δx sections as it is no longer homogeneous. This is
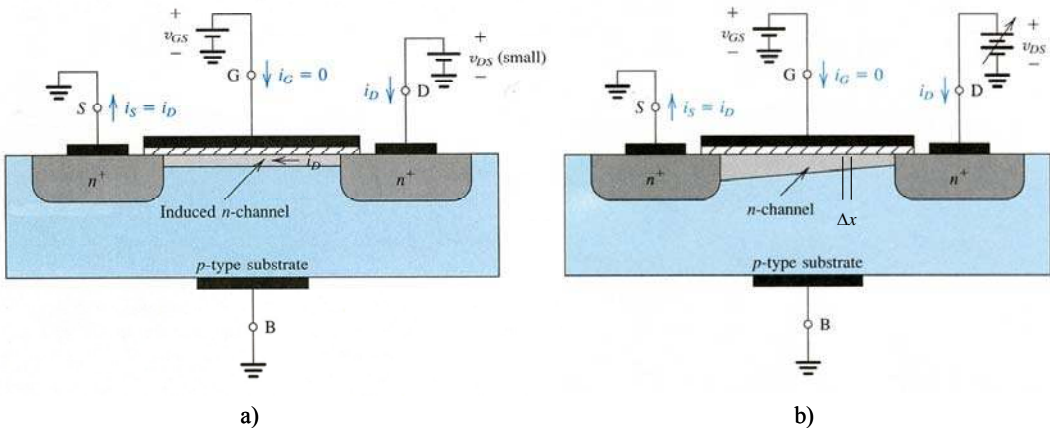
Figure 3-8 a) MOS FET in the triode region. b) MOS FET in the saturation region.

represented in Figure 3-8 b). For each Δx section the output current noise due to the noise voltage is calculated and the total drain noise is found by integrating along the entire channel. For the FET working in the saturation region, from Equ 3-2 and Equ 3-15, i_d^2 in the conduction channel is given by

$$ i_d^2 = 4kT \frac{\mu^2 W^2}{L^2 I_{DS}} \int_0^{V_{DS}} Q_n^2(V)dV , \qquad \text{Equ 3-17}$$

where μ is the electron hole mobility and $Q_n(x) = C_{ox}(V_{GS} - V_t(x) - V(x))$ is the inversion channel charge per unit area with $V_t(x)$ being the threshold voltage at position x. $V_t(x)$ is dependent on the channel potential $V(x)$. However, if the dependence of the threshold voltage V_t on the channel potential is neglected, the integral of Equ 3-17 can be carried out and is given by [CHA91]

$$i_d^2 = 4kT\mu C_{ox} \frac{W}{L} \frac{2}{3} \left[\frac{3(V_{GS} - V_t)V_{DS} - 3(V_{GS} - V_t)^2 - V_{DS}^2}{2(V_{GS} - V_t) - V_{DS}} \right] .$$ Equ 3-18

At the point of saturation where $V_{DS} = (V_{GS} - V_t)$, Equ 3-18 can be simplified to

$$i_d^2 = 4kT \frac{2}{3} \mu C_{ox} \frac{W}{L} (V_{GS} - V_t) = 4kT \frac{2}{3} g_m .$$ Equ 3-19

By experiment [CHA91] it has been found that Equ 3-19 remains valid and is a good approximation in the saturation region when $V_{DS} > (V_{GS} - V_t)$ and is the model most frequently used for Johnson noise in MOS FET devices. By writing $v_n = {i_d}/{g_m}$, Equ 3-19 becomes the equivalent v_n^2 noise equation. This is an important result as it shows that the noise voltage can be reduced with large values of g_m.

Shot noise is the effect from a current flow of discrete electric charges, not a smooth fluid-like flow. In transistors it occurs whenever current carriers cross a barrier such as a *pn-junction*. Each carrier causes a slight transient current surge as it travels across the junction, the combined effect being a random current fluctuation. Shot noise power is directly proportional to the DC current passing the junction; its effect is greatest when the junction has high internal impedance such as a reverse-biased collector-base junction. The statistical fluctuation of current per unit frequency is given by

$$i_{sn}^2 = 2qI_{DC} ,$$ Equ 3-20

where q is the electron charge and I_{DC} is the DC current passing the junction. For a MOS transistor the shot noise is associated with the leakage current of the drain and source reverse-bias diodes. Since the leakage current is generally much smaller than the drain-source current, I_{DS}, its effect is normally the least dominant noise source in FET devices.

Both Johnson and shot noise are classified as white noise, i.e. they have a flat frequency power-spectrum. Both are irreducible forms of noise, generated according to physical principles. The most expensive and carefully made resistor has the same Johnson noise as the

cheapest and poorly made. However the quality of a device will affect the 'Flicker noise'. Flicker noise, also referred to as 1/f noise, is due to random variations in diffusion processes in transistors or fluctuations in resistance in the case of resistors. As the name suggests, the power spectrum of the 1/f noise is inversely proportional to frequency, and so consists mainly of low-frequency components. It is the dominant noise source below about 1 kHz in bipolar transistors. Although the 1/f noise phenomenon has been observed in almost all kinds of devices, from homogeneous metal films and different kinds of resistors to semiconductor devices, even the sand in an hour glass, the mechanism of this noise is still little understood. For this reason the 1/f noise parameter is generally given by the MOS device fabricators and is found from empirical measurement. The general equation for the RMS voltage per unit frequency due to flicker noise is

$$v_f^2 = \frac{K_f}{C_{ox}^2 WLf} \ , \qquad\qquad \text{Equ 3-21}$$

where K_f is the flicker noise coefficient.

For the Beetle pre-amplifier discussed in section 5.2, the dominant noise source is Johnson. Although the shot noise in FET devices is usually negligible due to I_{DS} not having to cross a potential barrier, it should still be taken into account for the Beetle pre-amplifier as the input transistor is very large in physical size, leading to an increased leakage current from drain/source to the bulk. The 1/f noise is mostly below the working frequency of the pre-amplifier but, since it is possible that some MOS devices can produce flicker noise into the MHz region, good design practice should be incorporated, which generally comes down to the FET gate dimensions. This is discussed further in section 5.2.

3.4   Radiation hardened electronics

The space industry was the driving force behind the study of radiation damage to semiconductors when the first satellites, around 1960, were suffering electronic failure due to the flux of ionising particles present in the Van Allen belts. Over the following years, radiation hard electronics became applicable to space missions, nuclear power plants, military applications and research in high-energy physics. Over the next ~40 years the semiconductor radiation hard

market was large enough to sustain the use of a DMILL process that used special development steps (section 3.1.1). In the mid 90s the market became smaller, mainly due to the military share going from 90 % to 0.5 % in a 40 year span. The DMILL process therefore became more expensive, fell behind other processes in geometry size and technology, and had a large reduction in the number of fabrication plants that were available. Due to the low yield of the process, large variations of device parameters from wafer to wafer and an ever-diminishing market, in 1996 CERN formed the RD49 project [RD49www][RD49_stat] to find more suitable techniques to make radiation hard electronic circuits.

Today's electronic market is driven by digital integrated circuits for fast processing and storage which necessitates the continual demand for high levels of integration, high yields, low cost, ever faster signal speeds and low power consumption. CMOS technologies offer these requirements and, as Moore [MOO65] predicted in 1965, have tended to follow an exponential growth in the number of transistors per integrated circuit; Xilinx predicts that their 'Field Programmable Chips' will contain 2 billion 70 nm transistors by 2005. When the transistor geometries became less than ~0.3 μm, the term 'deep sub-micron' was coined to show that these transistors were now so small that special design techniques had to be used to overcome the effects of, for instance, the increased channel length modulation effects discussed in section 3.2. Fortunately for the high-energy physics community these deep sub-micron transistors had the added benefit of having inherently radiation hard gate oxides. However this was not the complete solution and special layout techniques[15] had to be invented and a solution found to solve the 'single event upset' problem, discussed in the subsequent section, that occurs in digital logic for small geometries.

For the RICH on-detector electronics the expected radiation dose (with a factor of 2 safety built in) is of the order of 3 krad/year. The radiation dose for the RICH off-detector electronics is negligible and standard commercial components can be used. However most of the ASIC designs for LHCb, including the Beetle, have been designed to cope with the radiation doses of 10 Mrad/year expected much closer to the interaction point. The following sub-sections describe the effects of radiation damage in analogue and digital transistor cells and the

---

[15] Layout techniques are down to the designer, unlike process techniques that are down to the fabricator.

techniques used to overcome these problems. The emphasis is towards the 0.25 μm CMOS process used for the Beetle chip.

Radiation damage in VLSI circuits

Radiation effects in microelectronics devices can be categorised into cumulative and single event effects (SEE). Cumulative radiation damage is a gradual effect which occurs over the lifetime of the device. Its cause is either ionization or atomic displacement and is most significant in analogue designs. The SEEs are caused by the passage of a single ionizing particle through a sensitive area of a device which leads to either non-destructive or destructive effects. Non-destructive effects cause a temporary malfunction or change of logic state in digital memories. Destructive effects generally cause a *pn*-junction to go forward biased, consequently drawing large amounts of current. The discussion on destructive effects will be restricted to MOS devices where the *pn*-junctions are only parasitics i.e. a consequence of fabrication (see later).

The way in which radiation interacts with solid material depends on the kinetic energy, mass and charge of the incident particle and the mass, atomic number and density of the target material. The incident particles are either charged or neutral. Charged particles interact mainly through Coulomb attraction or repulsion with the electronic clouds of the target atoms. The charged particles of relevance are protons, heavy ions, pions and electrons. Heavy ions, pions and protons can induce ionization, atomic excitation or displacement of an atom from its site in the lattice from collisions with the nucleus. Electrons are similar in their properties but also emit X-rays when decelerating in the target material.

Neutral particles of interest are neutrons and photons. Neutrons do not experience the Coulomb force but interact in three ways depending on their energy. Low energy neutrons tend to be absorbed by the nucleus in a nuclear reaction, the excited nucleus then emits a proton, alpha particle or photon. Another possibility is that the neutron may not be absorbed but instead is deflected in an elastic collision, but gives enough energy to the nucleus to displace it, which may lead to ionization. For the more energetic neutrons, E>100 keV, an inelastic collision can take place in which the nucleus is broken, producing heavy ions and gamma rays. These heavy ions then cause ionization. Photons, like neutrons, do not experience the Coulomb

force. The photon interaction processes are the photoelectric and Compton effects, and pair-production. The photoelectric effect is where the incident photon ionizes the target atom. As the photo-electron is emitted, an electron in an outer orbit can fall into the vacated spot and a low energy photon emitted. The Compton effect is when the photon scatters off an electron of an atom and the electron is freed from the atom. Finally the pair-production process is when the incident photon, in the presence of matter, is converted into an electron-positron pair.

It has been shown [ANE00] that both neutral and charged particles can cause ionizing and non-ionizing (atomic displacement) effects in VLSI circuits. The effects may be caused either directly by the incident particle or from secondary effects induced by the primary interaction. Ionizing and non-ionizing effects occur in different proportions depending on the type of incident particle. Neutrons tend to mainly cause non-ionizing damage, whereas photons and electrons generally cause ionizing effects.

3.4.1   Accumulated radiation effects in MOS transistors

MOS FETs are termed 'surface' devices, as their conduction mechanism is by majority carriers in a channel that only penetrates into the silicon bulk to a typical depth of 300 µm. As displacement damage leads to a reduction in minority carrier lifetimes in the silicon substrate, which MOS FETs are not dependent on, the MOS FET can be considered insensitive to this
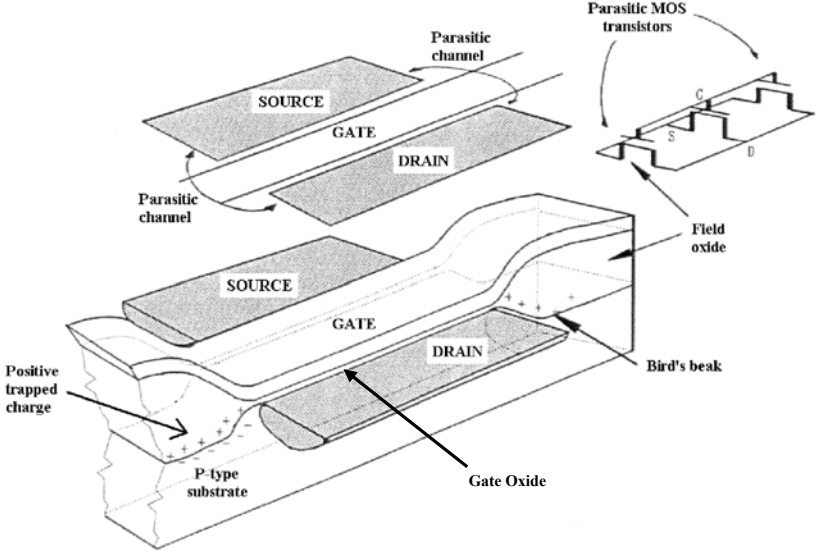
Figure 3-9 Gate and field oxides of a FET device. Also shown are the parasitic transistors between source and drain.

effect. The primary source of device degradation under irradiation is damage to the silicon dioxide insulator layers by ionization. The two oxide layers that are of most concern, the gate and field oxides, are shown in Figure 3-9 [ANE00]. Note that the point where gate oxide joins the field oxide is called the 'bird's beak'. This area, discussed later, is prone to the introduction of parasitic MOS transistors if special FET layout procedures are not followed. The effect that ionizing radiation has on the oxides has some dependency on the oxide thickness, which is determined by the technology scaling factor. In brief, for technology generations below 0.8 μm (which refers to the minimum gate lengths in the case of MOS), a constant-field scale is used where all geometrical dimensions and process parameters are scaled to increase transistor density, increase circuit speeds and decrease power consumption, while keeping the electric fields in a device unchanged. As the transconductance is dependent on the gate dimensions and C_{ox}, these reduced dimensions reduce the transconductance. To compensate this, the gate oxide thickness (t_{ox}) is reduced so as to increase C_{ox} . This reduction in gate oxide thickness turns out to be advantageous in increasing the tolerance to ionizing radiation. However, as will be shown in section 3.4.2, the scaling of the dimensions of the device is detrimental to the tolerance against single event upsets.

Ionizing dose effects in MOS FET oxides

When ionizing particles pass through a MOS FET, electron-hole pairs are generated along the track. For the gate (which can be polysilicon or metal) and the substrate, this is of little consequence as these materials offer little electrical resistance and the electron-hole pairs are quickly swept away from the region in which they were generated by local electrical fields. For the electron-hole pairs that are generated in the gate and field oxides, which are insulators, this is not the case. Figure 3-10 schematically illustrates the effects induced by ionization in a MOS device when the gate is positively biased. An ionizing particle penetrates the oxide and electron-hole pairs are generated. For pair creation in SiO_2 an energy of 17 ± 1 eV is required [LUT99]. A fraction of the electron-hole pairs generated will recombine immediately. The electron-hole pairs that do not recombine are separated by the electric field between the gate and bulk ((1) in Figure 3-10). For the case of a positive bias applied to the gate, the electrons drift towards the
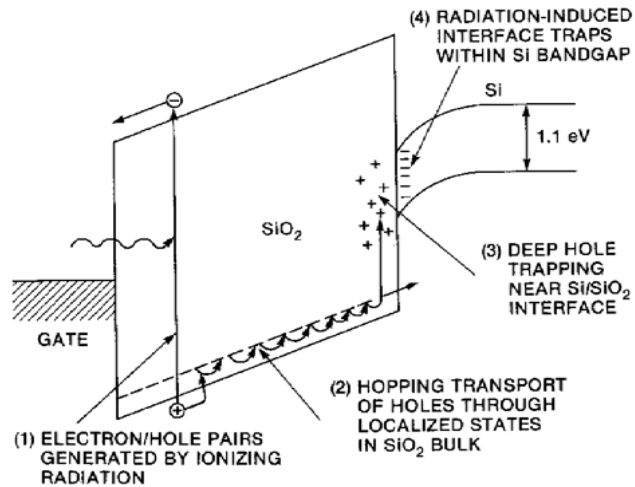
Figure 3-10 Energy band diagram of a MOS structure with a positive gate voltage applied [BAU03].

gate electrode. In SiO_2 electrons have a 10^4-10^{11} times higher mobility[16] than holes (depending on electric field and temperature) and escape out of the oxide into the gate material within the order of ps.

The remaining holes that do not initially recombine follow the electrical field in the SiO_2-layer towards the SiO_2-Si interface through a relatively slow transport mechanism (2). The transport phenomenon can last from several seconds at room temperature to several tens of thousand seconds at lower temperatures. One suitable model for this transport mechanism is continuous-time random walk (CTRW) which uses the concept of 'small polaron hopping'. Small polaron hopping is based on a strong interaction between the hole and the lattice. The transition between two nearby sites is activated by thermal fluctuations in the system. These momentarily bring the energy levels of the two sites into coincidence and the hole tunnels to the next site.

When the holes reach the SiO_2-Si interface a fraction will be captured in long-term traps (3). These trapped holes, along with the induced interface traps (4) cause a remnant negative voltage shift in the threshold voltage V_t, which is the most commonly observed form of radiation damage in MOS devices. The various processes contributing to the buildup of a negative shift in V_t are shown schematically in Figure 3-11 a) for a MOS structure with oxide

---

[16] The electron mobility in S_tO_2 is ~20 cm²V⁻¹s⁻¹ at room temperature and saturates at a velocity of around 10^7 cm/s.

thickness t_{ox} (= d_{ox}) and a positive gate voltage. A description of hole trapping and interface traps follows.

The number of holes trapped in the deep hole trapping centres (which are actually a few nano-metres away from the interface) is proportional to the number of defects in the silicon dioxide and therefore depends greatly on the control of the gate oxide quality. Holes can remain trapped from milliseconds to years and can be released by two processes. Firstly an electron can tunnel from the Si surface into the oxide and recombine, giving rise to a tunnel-effect based annealing. Secondly an electron in the oxide valence band which, having sufficient thermal energy, can jump into the oxide and recombine, this being thermal annealing. It is important to note that the tunnel-annealing probability decreases exponentially with the distance from the SiO_2-Si (or SiO_2-gate) interface.

a)   b)
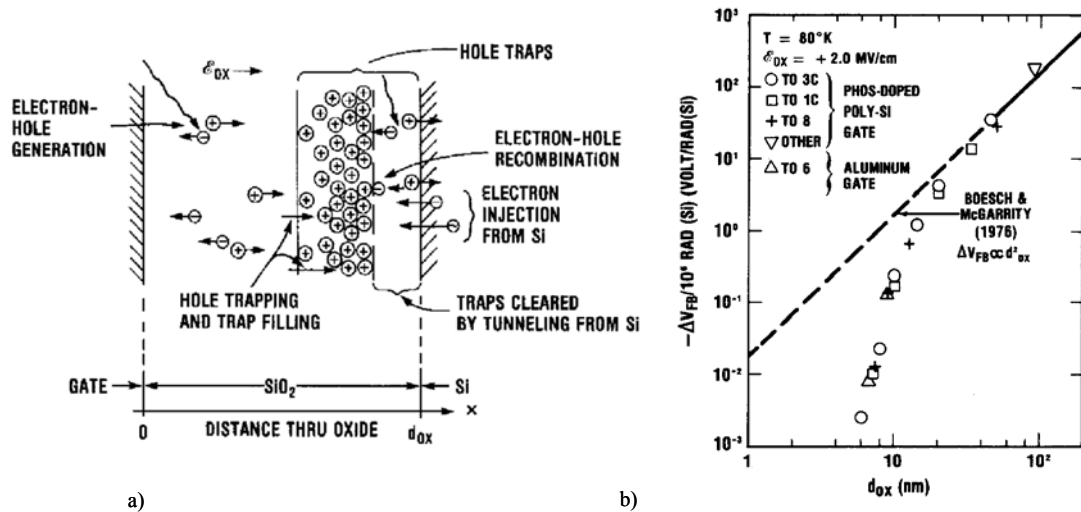
Figure 3-11 a) Hole trapping in the MOS oxide with positive gate bias applied [McL89]. b) Shift in the flat band voltageV_{FB} after an accumulated dose of 1 Mrad as a function of gate oxide thickness, d_{ox} (=t_{ox}) for various MOS structures [SAK84].

Interface traps are dangling bonds due to a deficiency in an oxygen atom at the crystalline silicon and amorphous oxide boundary. These traps have energy levels within the silicon bandgap. The number of radiation-induced interface traps depends on several factors: electric field, temperature, radiation energy dose rate and, more importantly, oxide thickness. From experiment [ANE00], traps in the upper part of the silicon bandgap are acceptors and those in

the lower part are donors. Therefore radiation-induced traps increase the absolute value of V_t for both *n* and *p* type MOS FETs.

From advances in process technolgy in cleaner and lower growth temperature oxides, the FET geomety is continously reducing, in particular a lower gate oxide thickness lowers the density of hole traps. The improved radiation hardness, due to the reduced number of traps in the gate oxide is demonstrated in Figure 3-11 b). This shows the flatband voltage (V_{FB}) after 1 Mrad irradiation as a function of t_{ox}. V_{FB} is the voltage that has to be applied to the gate eletrode to create a flat energy band within the silicon. When the voltage between source and base V_{sb} equals zero then V_{FB} is equal to V_t. Below an oxide thickness of ~12 nm the removal of trapped charge is dominated by tunnelling effects and the number of radiation-induced traps is reduced. This greatly increases the tolerance to radiation damage. For the process technology used for designs within this thesis, t_{ox} is equal to 6.2 nm and therefore can be considered to be radiation hard against trapped charge in the gate oxide.

FET layout design for thin gate oxide devices

Although the reduction in the gate oxide greatly increases the tolerance to radiation effects, there still exists the problem of the thick field oxide, shown in Figure 3-9, which is typically an order of magnitude larger than t_{ox}. Field oxide is used to define the channel width, to isolate devices from one another, and to allow interconnection layers to be routed on top of one another. For NMOS devices, as the oxide is over a *p*-type substrate, charge trapping in the bird's beak region (c.f. Figure 3-9) causes parasitic transistors, which are in parallel to the active device, to turn 'on'. The result is a leakage path directly from drain to source i.e. bypassing the gate. This problem is circumvented by the use of special layout techniques.
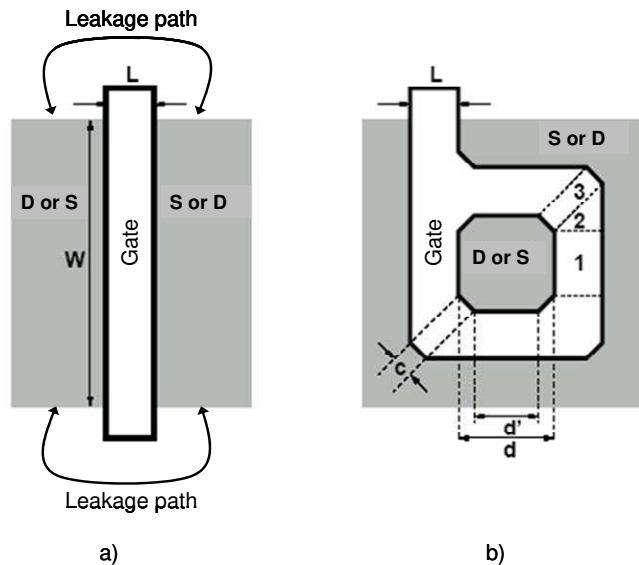
Figure 3-12 Schematic drawing a) of the standard linear FET geometry and b) an enclosed FET device.

Figure 3-12 shows two ways in which a MOS FET transistor would be in layout; the most commonly used linear method, a), and the enclosed gate (ELT) radiation hard layout method, b). Although the ELT consumes a considerable amount more area than the linear method it has only one path between drain and source, and that is under the gate region. This therefore removes any possibility of parasitic transistors being created between the drain and source regions and any leakage current around the gate is eliminated. As the linear device is completely symmetrical, the drain and source are interchangeable. However, the ELT is not a symmetric device as the two diffusion regions differ in area, and consequently capacitance and conductance. From measurement [ANE00] is has been found that, for a minimum allowed gate length L of 0.25 μm in this case, the output conductance is 20 % more when the drain is on the inside of the gate. For an L of 5 μm this increases to 75 %. Whether the drain is chosen on the inside or the outside of the gate depends on the application. For example, the current bias circuit described in section 4.2 requires a high output resistance from many of the NMOS devices used, so takes advantage of the low conductance by using the region outside of the gate as the drain.

A method to extract the effective W/L aspect ratio[17] of an ELT device has been devised by [FAC98][GIR98][RD49_stat]. The three contributions are labelled in Figure 3-12. The effective (W/L) ratio of an ELT device is described as

$$\left(\frac{W}{L}\right)_{eff} = 4.\underbrace{\frac{2\alpha}{\ln\dfrac{d'}{d'-2\alpha L_{eff}}}}_{1} + 2K.\underbrace{\frac{1-\alpha}{1.13.\ln\dfrac{1}{\alpha}}}_{2} + 3.\underbrace{\frac{\dfrac{d-d'}{2}}{L_{eff}}}_{3} , \qquad \text{Equ 3-22}$$

where

$(W/L)_{eff}$ is the effective apect ratio of the enclosed transistor,

L_{eff} is the drawn length minus the the reduction in gate length due to underdiffusion, photolithography and etching,

d, d' are the geometrical lengths shown in the ELT layout and

α, K are fitting parameters : $\alpha = 0.05$; K = 3.5 for L ≤ 0.5 μm, K = 4 for L > 0.5 μm.

The first term (1) in Equ 3-22 takes into account the linear region between corners. Term (2) describes the triangular corner segment and term (3) represents the rectangular region of the corner. This formula gives a precision of 94 % for short transistors and even better for long devices. A minimum aspect ratio (~2.3) is reached for L ≥ 7 μm so transistors with large L and small W are not possible.

3.4.2 Single event effects (SEE) in MOS transistors

In contrast to the accumulative total ionizing effects, single-event radiation effects are usually triggered by a single particle that crosses the sensitive region of a device, causing localized damage. As the technology size becomes smaller, most obviously below 0.8 μm, the sensitivity to an SEE significantly increases. For the deep-submicron technology used for the Beetle only two SEE effects are of importance, namely 'single event upset' and 'single event latch-up'.

---

[17] The effective L or W is where a reduction in the geometry due to, for example, oxide encroachment has been accounted for.

Single event upset (SEU)

Unlike the semiconductor bulk where electrons and holes generated by an ionizing particle immediately recombine, the electron-hole pairs generated in reverse bias *pn*-junctions, such as the drain/bulk, are separated by an electrical field. This can give rise to a current spike on the passage of an ionizing particle that can modify a logic cell such as a flip-flop. These effects can become static in a memory cell. For example a static RAM cell, which is made of two inverters in which the output of each is fed back to the input of the other, will have sensitive transistor drains, which if affected will cause the memory cell to capture the upset. The charge collection of this current spike has two components: a fast (~ps) collection from the drift process in the depleted region and a slow (~ns) collection from the diffusion in the bulk. The fast drift effects are magnified by an effect called 'funnelling'. An ionizing particle penetrating the junction and depleted region generates electron-hole pairs along its trajectory. The created charge distorts the local electrical field along its track and causes a nesting of funnel-shaped
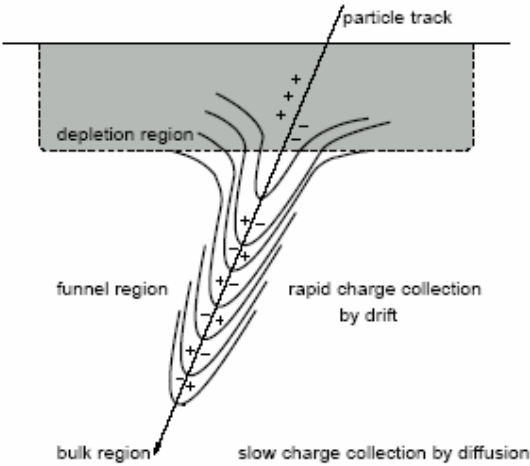
Figure 3-13 Enhanced charge collection by funnelling.

equipotential surfaces that can extend deep into the substrate bulk. This produces a large potential gradient resulting in an enhanced charge collection at the sensitive device node, demonstrated in Figure 3-13.

For each device there is a minimum charge quantity that is able to generate a SEU. The charge generated by the incident particle is directly proportional to its linear energy transfer (LET) defined as

$$LET = \frac{1}{\rho}\frac{dE}{dx} \text{ ,}$$

Equ 3-23

where ρ is the mass per unit volume of the silicon expressed in kg/m^3 and dE/dx is the mean energy transferred to the material per unit path length in eVcm^{-1}. Hence LET depends on the nature and on the energy of the incident particle and the target material. The critical LET is given by

$$LET_{crit} = \frac{E_{crit}}{\rho s_i.d} \text{ ,}$$

Equ 3-24

where E_{crit} is the minimum energy required to generate a SEU, ρs_i is the mass per unit length path of silicon and d is the sensitive depth (for example the depth of silicon under the drain region). It should be noted that the employment of LET as a single parameter can suffer limitations. For example LET yields the energy lost by the incident ionising particle which is not necessarily the same as the energy deposited in the SEU sensitive volume, which is generally referred to the stopping power. However, for very large 'd' dimensions the LET is equivalent to stopping power.

Potential candidates to cause a SEU are heavy-ions, protons, pions, or neutrons by secondary effects from inelastic interaction. Hardening integrated circuits to SEU can be achieved by technology and design methods. However, technology methods such as using SOI or the use of larger devices, generally degrade circuit performance i.e. by slowing the switching time and increasing the power consumption. The design methods make use of two forms of self-correcting digital control logic cells, described below.

Figure 3-14 a) shows a 'triple redundant cell' that would be used in, for example, a state machine. In this case the input, D, is fed to three standard flip-flops. The three flip-flop

70
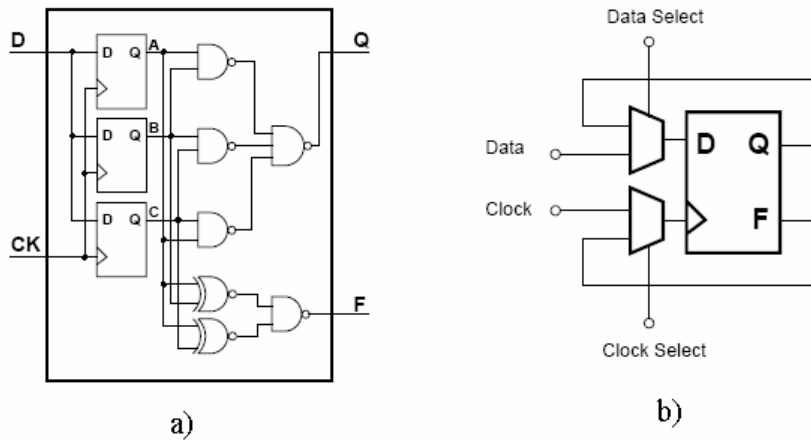
Figure 3-14 a) Triple redundant flip-flop, b) self correcting cell [BAU03].

outputs are decoded with four NAND gates so as to give an output at Q that corresponds to the majority of the logic levels at A, B and C. This is often referred to as a majority-voting scheme. In the case of a SEU causing one flip-flop to change state, the remaining two will still hold the correct logic level and will be in the majority, therefore the logic level at Q will be unaffected by the SEU. The triple redundant cell has a flag F to show that a SEU has occurred. Figure 3-14 b) shows a cell that could be used as part of a configuration register i.e. a register that is loaded from the JTAG interface. Internally it uses a triple redundant cell. In this case an analogue switch is used to select between feedback or configuration data. When in normal running mode (feedback) a SEU causes the flag to set, which is used to clock the majority vote back into the register therefore making the correction. This is normally achievable in less than 1ns, depending on technology.

Single event latch-up (SEL)

Latch-up occurs in standard CMOS technology due to the turning on of parasitic thyristors, shown in Figure 3-15. Latch-up can occur due to high temperatures, large transients on the power supply or be induced by a heavily ionizing particle. Once the parasitic thyristor is on, a short circuit develops between the power lines causing a large current on the interconnects. The device is destroyed unless this is not promptly interrupted. The mechanism of a latch-up for the schematic of Figure 3-15 is as follows. An increase in the base current of Q2 gives rise to an increase in Q2 collector current. As the collector of Q2 is connected to the

71
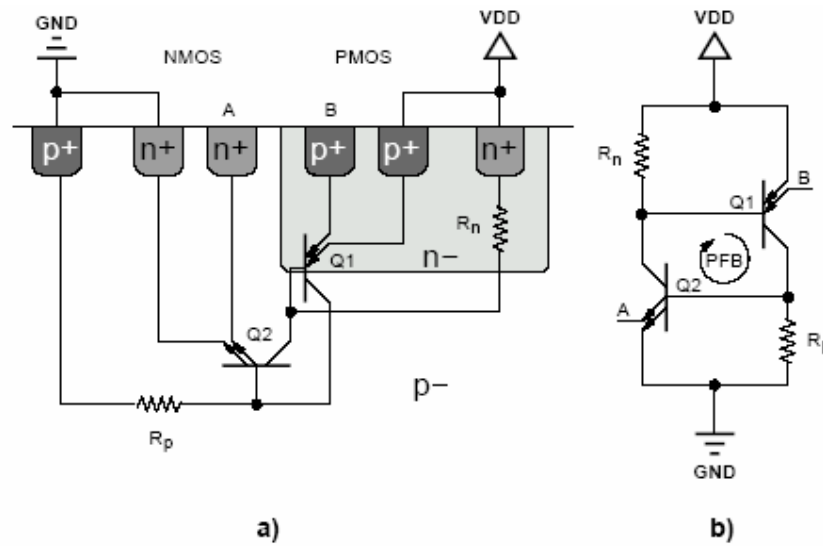
Figure 3-15 a) Cross section of a CMOS with parasitic thyristors Q1 and Q2, b) schematic representation. [BAU03].

base of Q1, the collector current of Q1 is also increased, which is positively fed back to the base of Q2 and latch-up occurs. Latch-up is only possible if the following conditions are met:

- The power supply must be able to supply enough current to keep the device latched;

- The two parasitic bipolar devices which make the thyristor must be forward biased;

- The loop gain has to be greater than one.

To prevent latch-up, the loop gain is made to be less than one by making the resistors R_n and R_p small in value, thereby reducing the voltage drop to forward bias the base-emitter junctions. These resistors are not discrete components but represent the resistance of the leakage path through the silicon. A reduction of the leakage path resistance, and therefore the potential voltage at the circuit nodes, can be achieved by flooding the circuit node areas with substrate or well contacts (discussed in the next chapter) and by the use of guard rings, shown in Figure 3-16. Guard rings surround the device and offer low impedance routes to the power rails. The other advantages guard rings have are a) the removal of leakage paths between
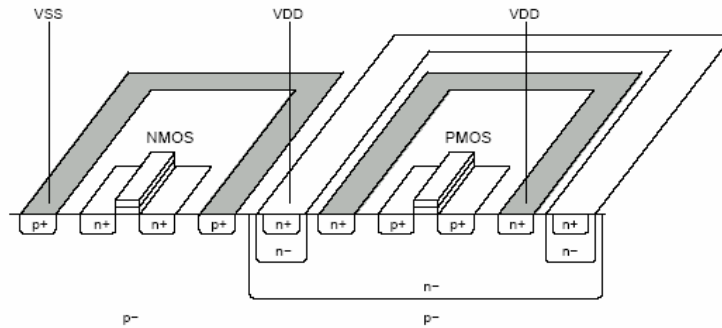
Figure 3-16 Layout principle of guard rings(grey) for NMOS and PMOS FETs [BAU03].

NMOS diffusion regions, such as the drain to the neighboring *n*-well of the PMOS, for example caused by radiation damage in the field oxide, and b) the prevention of any substrate noise injection from adjacent devices.

The following chapters make use of all these techniques for the Beetle chip design.

Beetle Bias Generator

The 0.25 μm process Beetle ASIC, introduced in section 3.1.3, needs voltage and current biasing[18] to set numerous DC operating points within the chip. The chosen bias values have an affect on the chip performance, in particular, the noise and shaping times of the front-end amplifiers. This is explained in detail in Chapter 5. To facilitate performance optimisation, the biasing needs to be adjustable.   To provide the biasing externally is cumbersome, and particularly demanding in LHCb, where thousands of closely packed chips will be used.
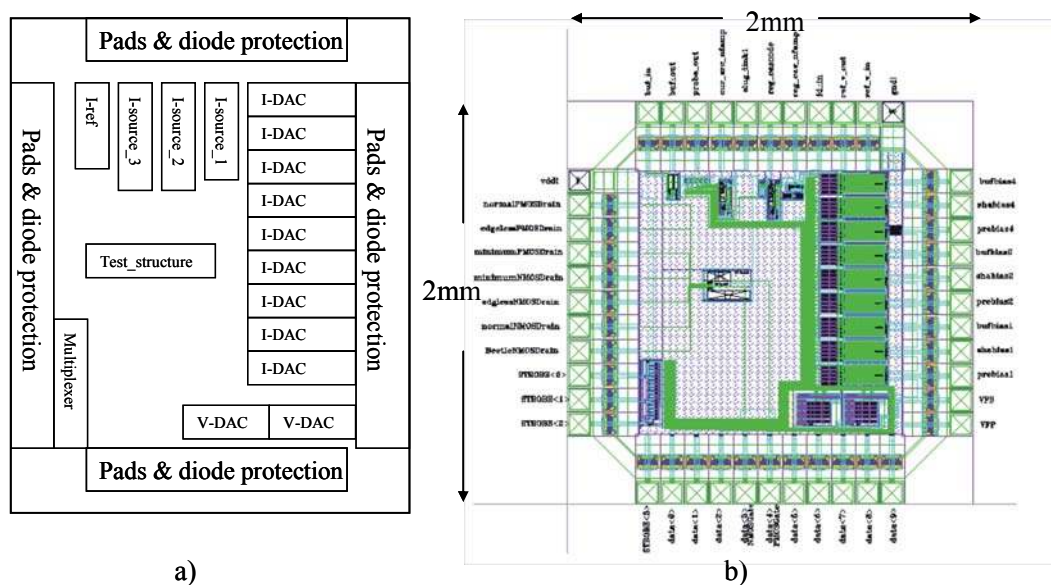
Figure 4-1 The bias generator test structure: a) is the floor plan and b) is the layout.

Therefore the bias levels are generated on-chip and are adjustable by internal voltage (V-DAC) and current (I-DAC) DACs that are controlled by the user through the I^{2}C interface.   The

---

[18] FET transistors need a fixed DC voltage applied to their gate so as to put them in, for example, the saturation mode of operation. This is termed biasing. In the case of an amplifier the small-signal response in then superimposed on the DC value.

author designed the components needed for a complete bias system and submitted the prototype designs as a test structure, in an MPW[19] run.

Figure 4-1 shows the layout and floor plan of the so-called BeetleBG1.0 prototype bias generator. The chip was designed in such a way that it could be placed next to the BeetleFE1.0 test structure, which is a set of prototype Beetle front-end amplifiers, and bonded across to give the amplifiers the necessary biasing while under test. The BeetleBG1.0 consists of two 10-bit voltage V-DACs, nine 10-bit current I-DACs, one voltage reference source, and three types of current-source. For circuit testing purposes a 16:1 multiplexer was also necessary because of the restricted number of chip I/O pads available. In this way 16 internal test points can be routed out to a single I/O at the cost of only four multiplexer addressing I/O pins. The test structure in the centre of the chip shown in Figure 4-1 was added on behalf of another project and is not related to this thesis.

The following subsections highlight layout techniques for improved matching of the bias values across the chip. Also discussed is how the components of the BeetleBG1.0 are combined together to make two circuit modules, the 'Voltage-bias module' and the 'current bias module', for generating and distributing, respectively, voltage and current to the Beetle circuit bias nodes. For these modules, the designed circuitry and chip layout down to the component level is given. For both the voltage- and current-biasing modules the measured values from the prototype designs on the test structure are compared to simulated results. The reader is referred to section 3.2, MOS FET characteristics, for general FET and current mirror principles.

4.1 Layout techniques

Ensuring that the bias points remain rigid[20] and are identical in operation for all 128 readout channels across the entire Beetle chip is of paramount importance to the signal integrity. Primary causes that affect the stability of the biasing network are the temperature and the manufacturing process variations. Variations in process parameters caused by, for example,

---

[19] MPW, Multi-Project Wafer. A wafer is shared between several independent projects to reduce cost.

[20] The set values remain constant under any various load conditions.

variation of gate-oxide thickness across the chip, oxide encroachment[21] and lateral diffusion[22] can drastically affect the performance of a component. One parameter that particularly suffers from these process variations and has a large effect on FET operation is the threshold voltage V_t (defined in section 3.2). The dominant effects of temperature are a shift in V_t and transconductance parameters, and changing resistor values. Both the process variation and temperature effects can be limited by following layout rules that are summarised in the following bullet points and given by example in the layout of the current mirror shown in Figure 4-2. In the following context a device is defined as a single component such as a FET, whereas a component is a group of devices that form a circuit.
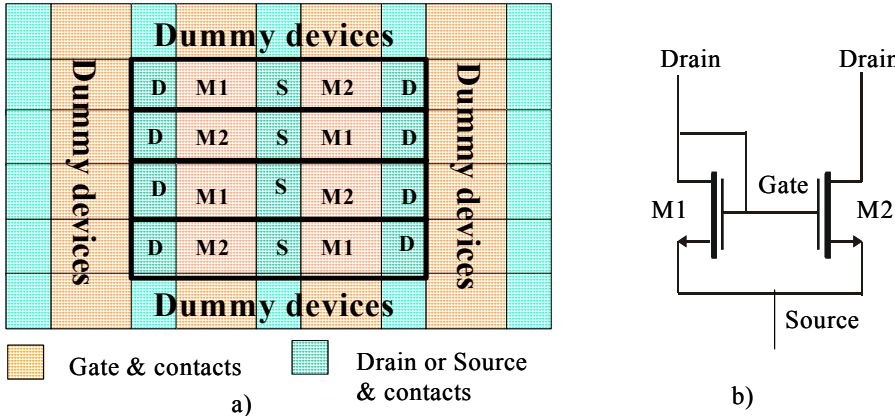
Figure 4-2 Applying the layout rules to a current-mirror component: a) Layout using W-divided FETs, where both M1 and M2 have been divided by 4, and using the centroid scheme with contact flooding and dummy devices, b) is the current-mirror schematic.

- Gate dimensions should be several times larger than the minimum allowed by the process. This reduces the effects of channel length modulation and process variations. It also improves device matching and eases the analytical modelling of the device. The gate dimensions are not explicitly shown in Figure 4-2.

- As many interconnect contacts should be used on the drain, source and gate regions as possible. Interconnects are vias with a fixed physical size that join the different metal (or poly) layers together. Increasing the number of vias used results in lower resistance, more current capability and a more distributed current load.

---

[21] Oxide encroachment is where the oxide overlaps into areas it should not.

[22] Lateral diffusion is where the *p* and *n* boundaries merge during and after fabrication, which consequently modifies the L or W of a FET device.

- Symmetrical or common-centroid layout schemes should be used. Where possible the device should be split into a number of parallel parts and laid out in a symmetric way so as to distribute process variations evenly across an entire single device. By using symmetrical or common-centroid schemes the value of, for example V_t , will be the mean V_t value of the device's smaller parts. The mean V_t value will hence represent the complete device characteristic and will be easily duplicated. For components where more than one device (resistor FET or capacitor) is used, for example the current mirror shown in Figure 4-2, a further improvement can be made by interleaving the component parts, making the device symmetrical around a central point; hence the name common-centroid. This considerably improves the matching between the two devices.

- Single FET devices should be split into n parallel devices, each with an nth of the original W. This reduces parasitics (for instance stray capacitance) associated between drain/source and substrate. An example of this scheme is shown in Figure 4-2 with M1 and M2 being sub-divided by four, and sharing drain and source connections.

- Dummy devices should surround each component. A dummy is a replica device that has no electrical purpose but does protect edge devices from over etching. The dummy device should be the same in layout as the device it is protecting.

- Guard rings should be used to isolate sensitive devices. Any precision circuit is susceptible to charge injection via the substrate from adjacent circuitry. The simplest method of reducing noise between adjacent circuits is to surround the device or component with a p+ implant (for a p-substrate) that is tied to V_{SS}. The implant removes the injected carriers and holds the substrate to a fixed reference. This has added benefits when designing circuitry for a radiation environment, see section 3.4. The guard rings are not shown in Figure 4-2.

4.2   Current bias module

The Beetle chip has several circuit nodes that need to be current biased. These current-bias nodes are mostly the DC current within an amplifier; this determines the gain, rise-time and even noise characteristics of the amplifier cell. The simplest way of setting a DC current bias is to generate a single reference DC current that can be mirrored, for example, within the input amplifier cells of each channel. The accuracy, adjustability, resolution and dynamic range of the reference current required is specified by the designer of the Beetle sub-circuit that the references serve.

In total the Beetle chip requires 11 externally adjustable current references for operation; the reader is referred to the Beetle block diagram (Figure 3-1 [BEE04]) for the bias nodes. Most

of the current references serve sensitive nodes and therefore need to be insensitive to noise and the voltage demands made of them. Furthermore they should not drift or contribute any significant noise themselves.

As discussed earlier, an effective way of distributing current biasing to many locations over a chip area is to make use of current mirrors. This has the advantage of the ease in which the magnitude of current can be changed at the design stage for each bias node i.e. by simply modifying the current mirror ratio. Furthermore, variation of parameters caused by temperature and radiation can be made to have little effect. For example, if the two FETs of the current mirror are physically close together then the same change to V_t due to temperature can be expected in both FET devices and will therefore cancel out. To satisfy the design requirements of adjustability, accuracy, resolution and dynamic range, 10-bit[23] I-DACs are used; the current magnitude at the I-DAC output being set by the user through the I^2C interface.
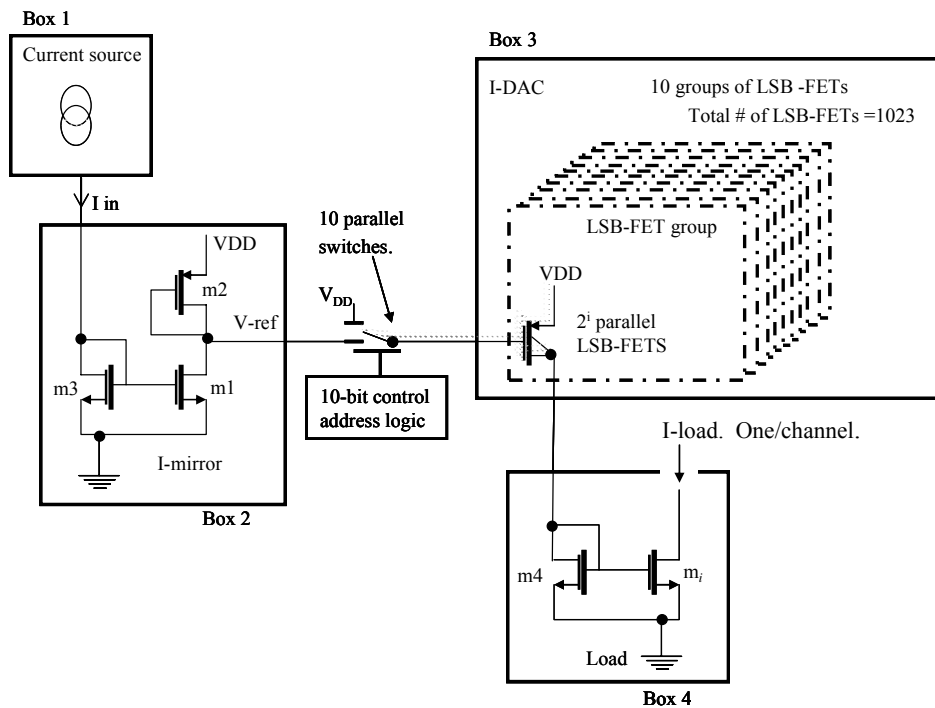
Figure 4-3 The general current-reference scheme.

---

[23] A 10-bit 2 μA resolution was more sensitive than required but allowed some insight into manufacturing process variations.

Figure 4-3 shows the general current reference and distribution scheme of a single current bias module. The principle is as follows:

- A 10-bit digital value is input via I^2C; each bit controls the operation of 10 parallel switches.

- Each switch turns "on" a circuit consisting of a group of FETs; each of these FET circuits outputs a current proportional to value given by the input control bit (i.e. a binary-weighted value), when that bit is set.

- The currents from all 10 groups of FETs are recombined to form the required output current reference, which is proportional to the input digital value.

Each group of FETs contains 2^i identical FET devices, where i runs from 0 to 9. All the FETs are connected in parallel and have common-connected drains, giving an output current which is proportional to the number of FETs within that group. Hence the 10-bit I-DAC consists of 1023 FET transistors in total, each having the exact same geometry. The basic FET building-block used in the I-DAC is termed here as the 'LSB-FET', since a single LSB-FET unit is used for the LSB setting. Hence when only the LSB of the I-DAC is selected (bit 0), then the number of LSB-FETs in the smallest group is one, and this defines the minimum possible value of bias current. If the MSB of the I-DAC is selected, then the largest group, consisting of $2^9 =$ 512 LSB-FETs, is turned 'on'. The magnitude of current from a LSB-FET is determined by its geometry and V-ref. To generate the V-ref setting, a current-source and a system of current mirrors is used. In principle a large current is generated and scaled down using a series of current mirrors to the smallest current value required for biasing. The scaling thus reduces noise and improves accuracy.

Referring to Figure 4-3, an integrated current-source (Box 1) provides the reference current for an I-mirror (Box 2) so that a single, stable V-ref (V_{GS}) can be switched to any number of the 10 groups of LSB-FETs (Box 3). When Switch #0 is closed, this corresponds to an LSB I-DAC setting, while switch #9 gives a MSB setting. When a switch is closed, m2 forms a current mirror with the LSB-FETs in the switched group. When the switch is open, the group of LSB-FETs are at V_{DD}, i.e. switched off. The load for each I-DAC is one half of a current

mirror (m4 Box 4). The other half of the current mirror is a FET (mi) that provides the bias reference for any given node of the Beetle chip. For example, the current bias I_{pre} in Figure 3-2 is distributed to each of the 128 amplifier channels, plus the test channel, with one mirror FET at each channel-node. The following sub-sections describe the components in more detail and are ordered with respect to the design flow.

4.2.1   I-DAC

The specification for the I-DAC is a 10-bit resolution with an LSB of ~2 μA. As eleven I-DACs are required, physical size is also important. A binary-weighted I-DAC was opted for, as this offers a simple DC scheme. A further advantage is that this can be implemented simply with CMOS FET transistors. Good device matching can be achieved if duplicates of the transistor used for the LSB are used to make the remaining binary-weighted outputs. In total the binary-weighted I-DAC will be composed of $2^{(10)}-1=1023$ LSB-FETs in a current-source configuration. Below, the design of a single LSB-FET device is described.

The LSB-FET

A good current-source (in this case the LSB-FET) should have a high output resistance compared to its load, implying a good compliance[24] and a low g_m. From $g_m = \dfrac{i_{out}}{v_{in}}$, a small g_m infers that the FET will be de-sensitised from the DC voltage between gate and source, V_{GS}, thereby improving the common I_D setting of duplicate devices. Furthermore, this reduces the effects of noise on the gate terminal. To obtain a low g_m, two parameters are at the designer's disposal, namely the W/L gate dimensions and V_{GS}, which determines the LSB-FET DC operating point. The L gate dimension has a direct effect on the incremental resistance of a FET device. The equation (Equ 3-8) for incremental resistance, defined in the saturation region, is $r_o = \dfrac{1}{\lambda I_D}$, where the channel-modulation parameter λ is the inverse of the Early voltage and I_D is the drain current of the FET device. Using an L gate dimension that is larger than the

---

[24] Good compliance in this case means only a small change in output current $\left(\Delta I_{OUT}\right)$ results under varying voltage load conditions.

minimum allowed will reduce the value of λ to some degree; this value is usually found empirically. Reducing λ increases r_o, where r_o in this case is the current-source output resistance. The increase in r_o, giving an improved current-source, is at a cost of real estate and needs to be optimised.

In fabrication, the W gate dimension does not suffer from the drain/source dopant atom diffusion under the gate region or λ like the L dimension, and can therefore be made closer to the minimum design rules without the consequences of the short-channel effects. In the case of the LSB-FET, W was arbitrarily chosen to have a reasonable value of 0.6 μm. With the W dimension chosen, the L dimension could now be optimised to give an I_D of 2 μA, with an appropriate V_{GS} operating point that allowed a broad range of load voltages on the LSB-FET drain terminal, an acceptable r_o and a small g_m.

Equation Equ 3-5 can be written as

$$g_m = \sqrt{K' \times 2} \times \sqrt{I_D} \times \sqrt{W/L}$$

Equ 4-1

where it can be clearly seen that g_m is proportional to $\sqrt{W/L}$, and therefore can be optimised in relation to the real estate consumed. Figure 4-4 a) is a plot of the $\sqrt{W/L}$ proportionality factor, from Equ 4-1, as a function of L with a fixed W value of 0.6 μm. Also shown is the total gate area, simply $W \times L \times (\text{the number of FET devices})$, required for all the LSB-FETs within the

a)   b)
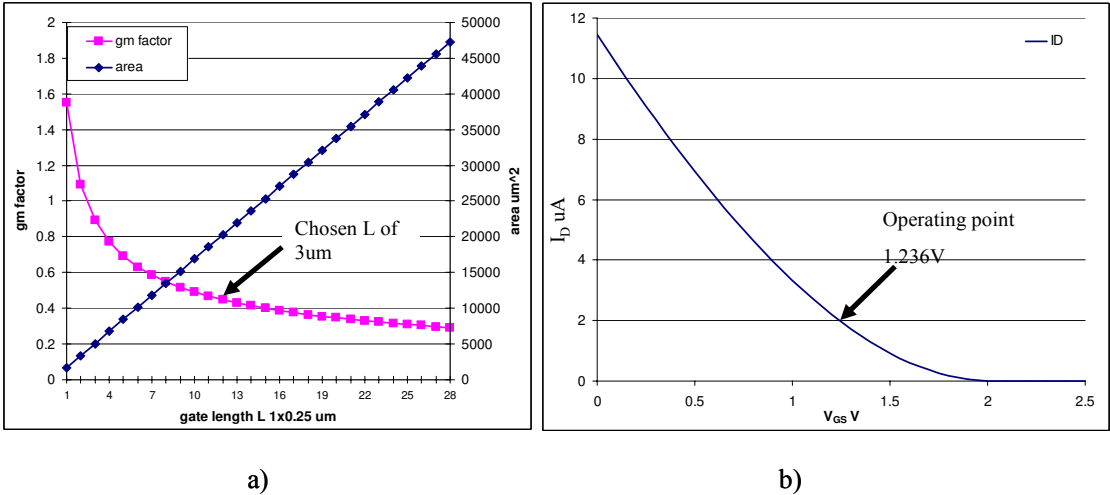
Figure 4-4 Optimisation for W/L of the LSB-FET. a) Shows the gm factor from Equ4-1 and area consumed with increasing L. b) Shows the operating point of I_D for the LSB-FET with W/L=0.6 μm/3 μm.

Beetle chip for eleven I-DACs, as a function of the gate length L. The remaining FET area required for drain- and source-terminal regions is not accounted for here; this would be a fixed area and therefore an offset. From the graph of Figure 4-4 a), the area consumed gives a linear gradient of 6800 $\mu m^2/\mu m$. From this the gate length of the LSB-FET was chosen to be 3 μm. Although an L of 1.5 μm would have saved ~ 50 % in area without much penalty in increased g_m, the longer L was beneficial in considerably reducing the short-channel effects and improved matching of the devices.

To ascertain whether the W/L aspect ratio of $\dfrac{0.6\,\mu m}{3\,\mu m}$ can give the required 2 μA with a reasonable value of V_{GS} is determined by the simplified equation of Equ 3-2 for I_D, given as

$$I_D = \frac{K'W}{2L}(V_{GS} - V_t)^2 \quad \text{for long L and } V_{DS} \leq V_{GS} - V_t \ . \qquad \text{Equ 4-2}$$

The value of K' is dependent on PMOS or NMOS technology. In order to make an NMOS FET radiation hard, the use of enclosed gate layout (described in section 3.4) is required. This has the disadvantage that it prevents the use of a small aspect ratio. Another advantage of PMOS over NMOS in this instance is that the hole mobility is ~2.25 less than the electron mobility, making g_m less for a PMOS with the same geometrical size as the NMOS. Figure 4-4 b) shows the simulated transfer curve using Equ 4-2 for a PMOS LSB-FET with an aspect ration of 0.6 μm/3 μm. Taking an operating point for I_D of 2 μA gives a V_{GS} of 1.236 V. This results in a g_m of 7 $\mu A/V$, an r_o of 47 $M\Omega$ [25] for a $0V \rightarrow V_{DS\_SAT}$ load voltage, and a total gate area for eleven I-DACs of 25,300 μm^2. The value of V_{DS_SAT}, and therefore the maximum allowable load voltage on the LSB_FET drain terminal, is found to be ~1.98 V.

Consideration needs to be given to r_o and the sensitivity of the I-DAC to V-ref when all 1023 LSB-FET devices are combined to form the maximum I-DAC current. When all 10-bits of the I-DAC are selected the effective value of r_o will be ~ $47\,M\Omega/_{1023} \approx 46\,k\Omega$. Therefore, in order to keep the generally-accepted required accuracy for a DAC of ± 0.5 LSB

---

[25] The value of λ in simulations uses empirical measurements from the fabricators.

(corresponding to an output current of 1 μA) across the full dynamic range of the I-DAC, the load resistance needs to be small so as not to cause more than 1 μA x 46 $k\Omega$ =46 mV deviation in load voltage for any I-DAC current setting. As the load of the I-DAC will be a drain-gate connected FET (see Figure 4-3, FET m4), a low ohmic load value can easily be achieved for the entire I_D range of 0.002-2.046 mA by an appropriate W/L selection of m4.

The I-DAC layout

Figure 4-5 shows a) the layout of the I-DAC, b) the floor plan and grouping scheme and c) the binary-weighted PMOS FET output stage, where n1, n2..n512 represents the number of LSB-FETs used per binary-weighted stage. The layout of the I-DAC is critical, as the monotonicity and accuracy is dependent on how well all 1023 transistors can be matched. Monotonicity is defined as the condition such that there is an increasing value of output for
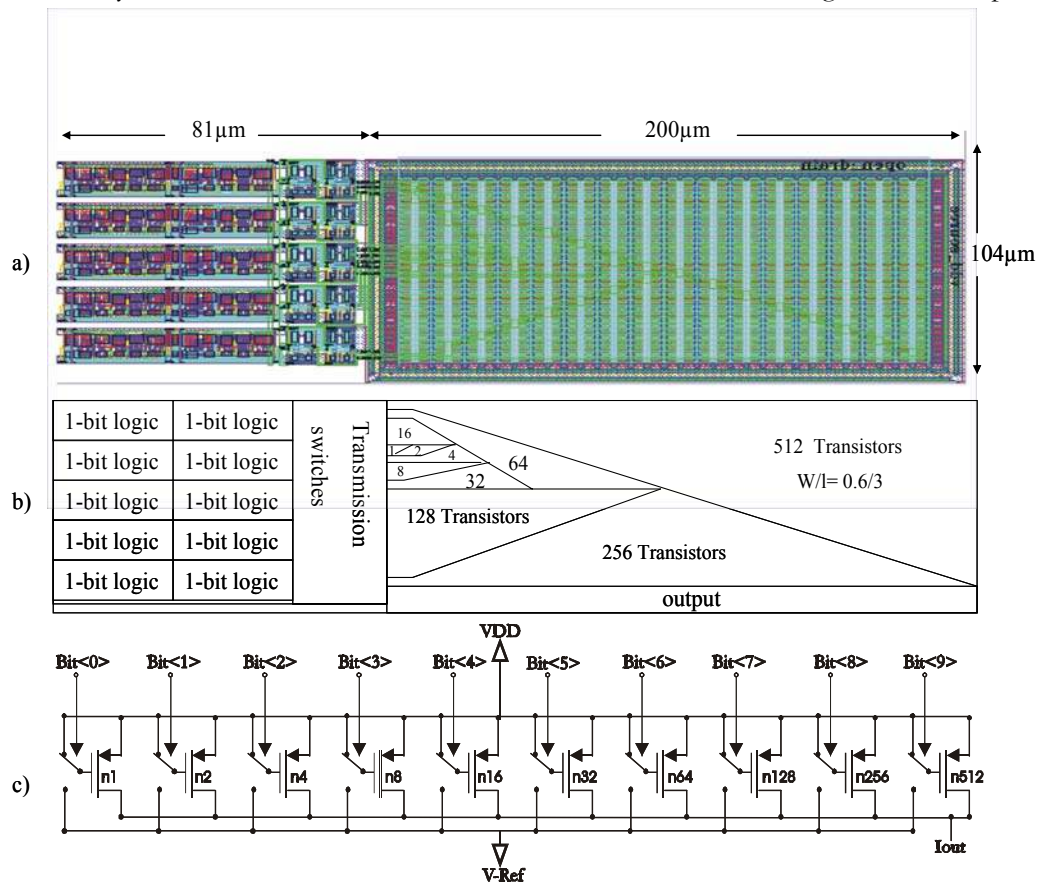
Figure 4-5 The 10-bit binary-weighted current DAC that was implemented in the BeetleBG1.0. a) The layout, b) the floor plan and c) the schematic of the binary-weighted PMOS FET output stage, where n1, n2..n512 represents the number of LSB-FETs used per binary-weighted stage.

every increasing digital input code. A non-monotonic DAC can result if the differential non-linearity error, described in the subsequent paragraph, exceeds ± 1 LSB. This leads to the requirement that the matching of the MSB-group of FETs to the LSB-FET must be better than $\frac{1}{2}_{10-1}$ or 0.195 %. To facilitate this, all of the LSB-FETs in Figure 4-5 would, preferably, all be interwoven around a common-centroid scheme described in section 4.1. However this proved difficult when routing the gate tracks. Instead the FETs are grouped together with their associated bit number, and layed out as close as practically possible to a common-centroid.

Measurement results

Four BeetleBG1.0 chips, each with 9 I-DACs, were mounted on tests boards. The input address was loaded directly to the I-DAC address bus as this was accessible via the chip I/O pins in the case of the test structure. A Hewlette Packard 34970A multipurpose data acquisition unit was used to generate the required I-DAC address and binary input, and to measure and store the I-DAC output. The unit was fully controlled by a Labview program running on a PC. Figure 4-6 a) shows the measured results of I_{OUT} as a function of the set binary input value, sampled over the 4 chips.

a)   b)
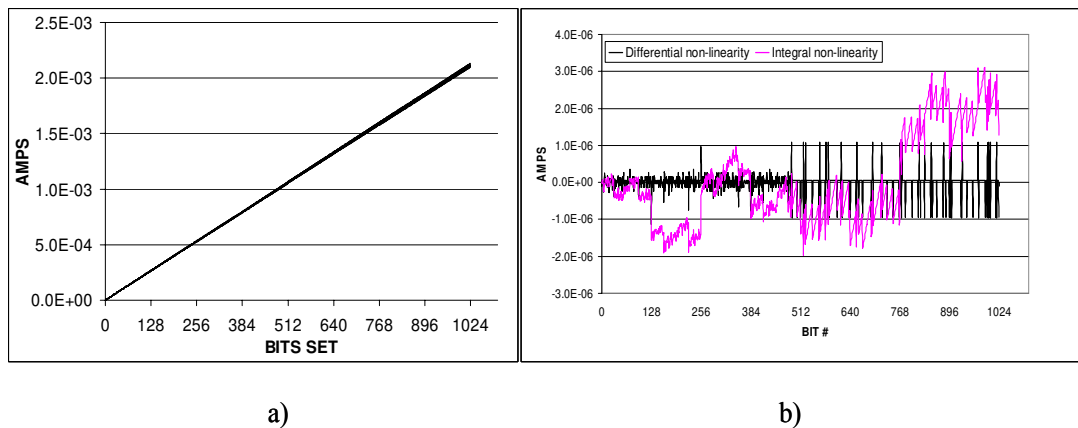
Figure 4-6 Measured results for the BeetleBG1.0 current DAC. a) Linearity plots of 36 DACs b) differential and integral non linearity for a single DAC.

For each binary input value the I-DAC measured results were histogrammed and the following results obtained. The mean dynamic range of all the I-DAC measurements shown in Figure 4-6 a) is 0.1 μA to 2.101 mA, with an LSB of 2 μA. This gives a mean gain error of

$\dfrac{\left(2^{10}-1\right) \times LSB}{DAC_{max}} \times 100 = 2.6\%$, where LSB and DAC_{max} are the measured currents. This poor

gain error is attributed to the demands of accurately setting the power supply for V_{DD} to 2.5 V, as V_{GS} of the LSB-FET is the voltage between V_{DD} and V-ref. The variation between the I-DACs is ± 1 % RMS and there are no obvious trends between the I-DACs on different chips.

Two parameters that allow a detailed view of any non-linearity that may be present in the I_{OUT} transfer function of Figure 4-6 a) are 'differential non-linearity' (DNL) and 'integral non-linearity' (INL). DNL is the measured difference ΔI_{OUT} between settings n and n-1, where n is the value set by the input bit pattern. This clearly demonstrates how well each I_{OUT} increment matched the last, and indicates whether or not the DAC is monotonic. INL shows the overall distortion from linearity by taking the difference between the ideal I_{OUT} linear curve and the measured I_{OUT}. The DAC is usually deemed acceptable if both the DNL and INL are within $\pm 0.5 LSB$. Figure 4-6 b) shows the DNL (with the LSB current subtracted) and the INL of a typical I-DAC. The measurements after the 511th binary setting suffer from rounding errors of the measurement equipment as a different resolution had to be selected. The DNL can be seen to be within ± 0.5 LSB, the I-DAC is therefore monotonic. The DNL can be seen to have some correlation with the code being set i.e. the current spikes seen at the binary input setting of 256 and 512, which indicates that there are some systematic effects, most probably due to the non-centroid scheme. The INL accentuates the mismatching of devices and is ± 1 LSB in the worst case. The dependence on the bit setting is clearly seen and indicates that the largest non-linearity effects are associated with the largest number of collective FETs i.e. the 128, 256 and 512 groups. Whilst the INL is outside the $\pm 0.5 LSB$ tolerance, a $\pm 1 LSB$ error is acceptable for the Beetle since the 10-bit sensitivity is more than necessary.

To measure I_{OUT} and r_o of the I-DAC a variable resistor was connected as the load of the I-DAC. While varying the resistance, the voltage across the resistor, V_{LOAD} and the output current of the I-DAC, I_{OUT} were measured with multi-meters. A plot of I_{OUT} as a function of V_{LOAD} for an LSB setting (2 μA output current) is shown in Figure 4-7. The ~100 nA leakage current is attributed to the summation of a small leakage from each of the 1023 LSB-FET devices. The measured LSB=2.1 μA and the measured r_o is 38 $M\Omega$ up to a load voltage of 0.5 V of the 2.5 V

V_{DD} rail, which compares well with simulation. If the offset is taken into account then the simulated value of 2 μA agrees well with that measured.
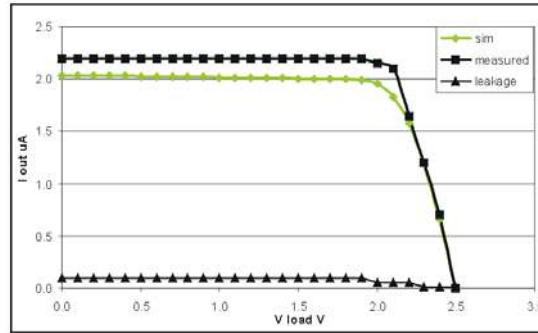
Figure 4-7 I_D vs V_{LOAD} for the LSB of the current DAC. The measured
values are: LSB=2.1μA, r_o=38 $M\Omega$ and leakage~100nA.

For a LSB setting on a typical I-DAC, the sensitivity to the V_{DD} supply voltage was measured to be an approximately linear, 1.4 μA/V, over the supply range of 1.1-3.0 V. The temperature sensitivity over a range of 25-80 °C was measured to be 1.8 nA/°C by using a hot-air gun to increase the chip temperature and a thermistor to measure it. For both the temperature and supply voltage sensitivity, the voltage and temperature gradients are multiplied by the decimal equivalent of the I-DAC input binary-bit number set. Therefore these parameters need to be kept stable to within ~ 0.5 mV and ~ 0.5 °C if the I-DAC is to maintain an accuracy of 2 μA ± 1 μA over its entire working range. The Beetle chip is less stringent in requirements than this due to the reduced resolution requirements.

4.2.2   Current-source and V-ref

The design procedure of both the current-source and V-ref is now discussed. As V-ref is fundamental in setting the correct I-DAC operating conditions the design procedure starts with V-ref.

V-ref accuracy

To determine the accuracy and stability requirement of the V-ref node shown in Figure 4-3 so as to maintain the ± 0.5 LSB I-DAC requirement and a generally acceptable requirement

that the LSB should be within 1 % of the design value of 2 μA, the I-DAC is considered when all of its binary input bits set to '1'. This consequently gives the maximum I-DAC output current, and therefore the maximum sensitivity to the setting or deviations in V-ref. To facilitate the V-ref evaluation while the I-DAC has its maximum output current set, the total number of 1023 I-DAC FETs are considered as a single FET with an aspect ratio of $\dfrac{1023x0.6\,\mu m}{3\,\mu m}$ (c.f. the aspect ratio discussed in section 4.2.1) and will be referred to as LSB-FET-m5.

For evaluating the effects of a small change in V_{GS} of LSB-FET-m5 caused by noise, the combination of the LSB-FET-m5 and drain-gate-connected m4 FET of Figure 4-3 should be considered as a parasitic common-source amplifier. The reader can refer to section 3.2 for the standard common-source amplifier configuration. In this configuration V-ref, with a superimposed small-signal v-$_{ref}$, would be the input to the common-source amplifier. The voltage at the drain of LSB-FET-m5, which is determined by the r_o resistance of the m4 FET multiplied by i_{out} , would be the output voltage. This is also v_{gs} of the m_i FET. The common-source configuration has a gain of g_{m5}/g_{m4}, where g_{m5} and g_{m4} are the transconductances of the LSB-FET-m5 and m4 FET respectively. Therefore the sensitivity of the load (Box4) to v-$_{ref}$ depends on the gain of the common-source amplifier stage as this directly modifies v_{gs} of the m_i FET. Assuming that the common-source amplifier just discussed will be designed to give a maximum gain of one i.e. an m_4 FET with a large g_m, then by extrapolating back from Equ 3-5 it is found that V-ref should not vary by more than $\pm$ 140 μV if I_D is to stay within $\pm$ 0.5 LSB over the entire I-DAC range. This does not take into account any AC decoupling that might be added to reduce noise effects at any node within the Beetle chip. From Equ 3-3 and the general requirement that the LSB I_D should be within 1 % of the design value of 2 μA, V-ref should be 1.236 V $\pm$ 10 mV. For the V-ref accuracy of $\pm$ 10 mV and the required low noise content of v-$_{ref}$ of a maximum $\pm$ 140 μV, implies that V-ref must generate little noise, be accurate and stable.

It is worth mentioning that V-ref could have been generated with a resistor divider network. To make the divider a low power circuit requires large resistors and therefore the circuit would contribute Johnson noise to V-ref. Other disadvantages are the difficulty in gaining resistor accuracy and the chip area consumed by large resistors. The implemented design achieves the required accuracy, low noise and low power by scaling down a

proportionally large current, compared to the LSB current, using a system of current mirrors. The current mirror scaling method introduces little noise compared to using discrete components and is very accurate in its scaling if the devices are matched well in layout, thus allowing a precision design.

V-ref circuit description

Again referring to Figure 4-3, the I-mirror used to set V-ref is described below in terms of both layout and circuit design. An accurate way to generate V-ref is to use a duplicate (the duplicate is m2) of the LSB-FET from the I-DAC design to form a current mirror with the I-DAC LSB-FETs. By using multiples of the 'duplicate' in parallel, the ratio of the currents can be such that a more manageable I_{D_m2} can be used, i.e. a scaled up version of the LSB current requirement. An advantage of using multiple 'duplicates' over the use of a single larger FET device is that the mean value of V_t for several devices will better match to the mean V_t value of the I-DACs. The accuracy of the scaling ratio will also be improved due to the fact that, in fabrication, devices can always be copied very precisely. The increased I_{D_m2} has a further advantage when considering noise. If m2 is considered as the active load resistor of the common-source amplifier created by m2 and m1, the increased value of I_{D_m2}, compared to the LSB of 2 μA, consequently decreases the active-load resistance (m2) given by $1/g_{m2}$. This makes the V-ref node less susceptible to noise pick-up, less of a noise generator, and also reduces the small-signal gain, g_{m1}/g_{m2}, of the common-source stage.

Figure 4-8 shows the V-ref mirror scheme used, a) is the layout, b) is the floor plan and c) is the schematic. To have a manageable I_D of 200 μA and to keep the design to a reasonable size while maintaining good V_t matching, m2 has been built from 100 duplicate LSB-FETs. For noise considerations, m1 and m2 form the common-source amplifier with m2 being the active-load with an R_{out} of 1.4 *k*Ω (here $R_{out} = 1/g_m$ for a drain-gate connected FET and is not the incremental resistance r_o). To give the lowest g_m and largest r_o and to make use of multiple devices for improved matching, m1 composes 8 NMOS FET devices. Each have an enclosed gate, with an L of the same dimension as the L of m2 and an aspect ratio close to 4.29, which is the minimum for an enclosed gate with this L dimension.
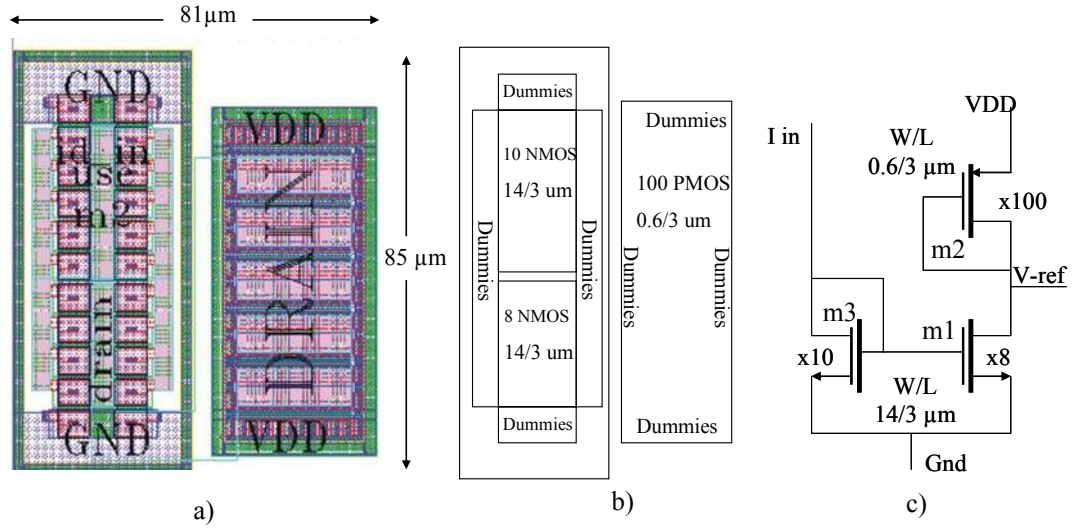
Figure 4-8 The V-ref current mirror used for the gate voltage of the current DAC. a) The layout, b) the floor plan and c) the schematic.

Current-source

The current value required to be generated by the current-source (Box1 of Figure 4-3) can simply be found by multiplying the required LSB value of 2 μA by the current mirror ratios at each stage, which is now described. While referring to Figure 4-8 c), the LSB value of 2 μA is multiplied by the first current-mirror ratio of 100:1, given by the number of m2 FETs to a single LSB-FET, this equates to 200 μA. This current flows through m1. The current mirror ratio between m1 and m3 is 10:8, and therefore m3 will have a current flow of $\frac{10}{8} \times 200\ \mu A = 250\ \mu A$. It is this value of 250 μA that the current source will generate. As the parasitic common-source amplifier gains, discussed earlier, have been made to be ≤ 1 and that R_{LOAD} of the current-source $= \frac{1}{g_{m3}} \approx 580\Omega$, which is considered to be a small R-noise generator, no special noise considerations are given to the design of this circuit.

As stated previously, an ideal current-source has a uniform output current response over the entire range of load values. To ensure this, R_{out} of the current-source needs to be large. One way in which to achieve this with a FET device is to use a small I_D, as was done for the LSB-FETs of the I-DAC (recall $r_o = \frac{1}{\lambda I_D}$). However, the current reference I_D needs to be large

compared to the LSB current of 2 μA, so that noise coupling or current fluctuations are proportionally small; therefore a small I_D approach is obviously not suitable. At the cost of circuit complexity, both a large current output and a large R_{out} can be achieved by employing 'boot-strapping' methods. Boot-strapping is a DC feedback mechanism that uses an amplifier to constrain a circuit node to a fixed voltage or current value. Three current-sources that employ boot-strapping and vary in complexity and output performance have been designed and evaluated. Table 4-1 summarises the relevant performance parameters. The regulated cascode out-performs the other designs in all aspects and is the only design reported here. The remaining two were considered as they were tune-able to the exact required current output and, although they performed as expected, this tune-ability was not required.

Type of current-source	Maximum load (delta I=1%)	Small-signal resistance (ohms)	Power consumption (mW)	Size (um²)
Regulated Cascode	1.93 V	17 M	0.859	84x23
Op-amp +Regulated Cascode Output	1.94 V	14 M	2.5	189x61
Op-Amp + Current mirror Output	1.06	4 M	2.35	164x61

Table 4-1 Three current-sources were submitted for evaluation.

Figure 4-9 shows the regulated cascode current-source, a) is the layout, b) is the floor plan and c) is the schematic. A description of operation follows. The two 2.9 $k\Omega$ resistors and transistor m5 form a simple resistor ladder to set the gate voltages of m2 and m3. The transistor m5 is used to act as a current mirror to m2. The aspect ratio of m5 and m2, both are the same
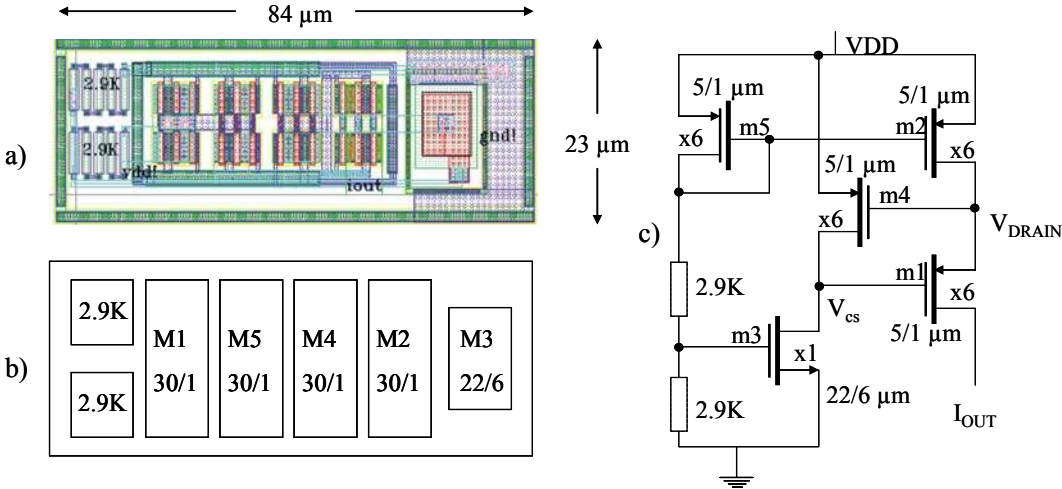
Figure 4-9 The regulated cascode. a) The layout, b) the floor plan and c) the schematic. Iout is 256 μA.

in this case, is determined by using Equ 4-2. Transistors m4 and m3 make a common-source amplifier with output, V_{cs}, connected to the gate of m1. An increased voltage ΔV on the drain of m1 caused by the load conditions gives rise to a reduced output current ΔI_{OUT}, which consequently increases V_{DRAIN} of m2 (reduces the voltage across m2). However as V_{DRAIN} is the input to the common-source amplifier, the voltage at the gate of m1 is reduced; therefore increasing I_{OUT} and consequently the original change ΔI_{OUT} is compensated by negative feedback. Hence the small-signal output resistance of m2 has been increased by bootstrapping. As the common-source amplifier stage uses m3 with a fixed V_{GS} as an active load-resistor (refer to section 3.2), the gain is expressed as, $A_v = g_{m4}(r_{o4} | | r_{o3})$. Ideally m3 and m4 would have a large r_o so that the common-source amplifier has a large voltage gain, which is a requirement for rigid boot-strapping. However, as discussed earlier, the aspect ratio is restricted for an enclosed gate device and therefore m3 has a larger I_D than desired, and this consequently reduces the value of r_{o3}. However, this has not been detrimental to the design.

If all transistors are in the saturation region and, ignoring bulk effects, the output resistance R_{out} is given by [GEI90]

$$R_{out} \cong r_{o2}\left(\frac{g_{m1} \times g_{m4}}{g_{o1}\left(g_{o4} + g_3\right)} \right),$$
Equ 4-3

where r_{o2} is the incremental output resistance of m2, g_{o1} and g_{o4} are the inverse of the incremental output resistance of m1 and m2, respectively, and $g3$ is the conductance of m3. To measure I_{OUT} and R_{OUT} a variable resistor was connected as the load of the current-source. Whilst varying the resistance, I_{OUT} and V_{LOAD} were measured with multi-meters. Figure 4-10 shows the measured and simulated results of I_{OUT} as a function of V_{LOAD}.

From simulation, I_{OUT} is a constant 250 µA and R_{out} a constant 15 $G\Omega$ up to 0.52 V of the 2.5 V V_{DD} supply rail. At this point the voltage on the drain of m1 (V_{LOAD}) supersedes the V_{GS} reference of m1, causing m1 to turn off. For this reason current sources are characterised by how close the load voltage can be taken to the supply rail; an ideal source would have I_{OUT} remain constant for an infinite V_{LOAD}. From measurement I_{OUT}=256 µA up to 0.52 V of the rail. As R_{out} is very large it is difficult to measure without employing specialised measuring
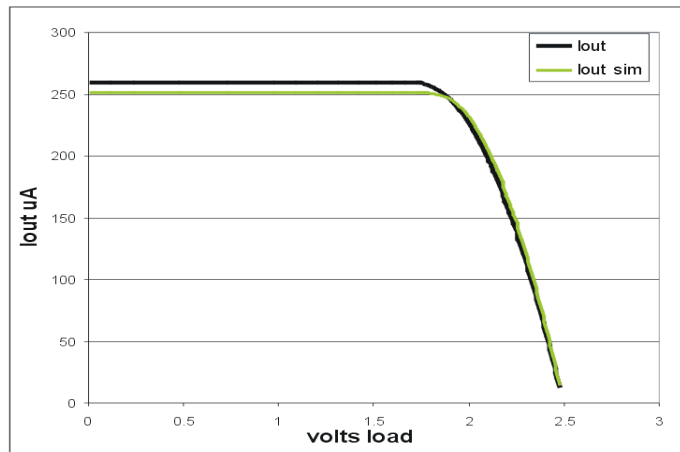
Figure 4-10 I_{OUT} from the regulated cascode current source. I_{OUT}=256 μA and is compliant to within 0.57 V of the 2.5 V supply rail. The small signal output resistance is at least 100 MΩ.

equipment. Instead, to gain a crude estimate, R_{out} was measured with a standard multi-meter and the resistance was found to be at least 100 $M\Omega$. As this translates to a $\frac{\Delta V_{OUT}}{100 M\Omega} \approx \Delta I_{OUT} \approx 20$ nA , and after scaling by the current mirrors gives an error at the LSB of ~0.008 %, the R_{out} value of 100 $M\Omega$ is already good enough for Beetle requirements. Hence an accurate R_{out} measurement was not deemed necessary.

4.2.3   Small-signal response and noise

As the current-bias generator contains circuitry that consists of parasitic amplifier stages that may add small-signal gain and cause operational instabilities if not designed correctly, a small-signal simulation was performed. The results are represented in the form of a Bode plot in Figure 4-11. The Bode plot is a plot of the gain $\frac{v_{OUT}}{v_{in}}$ and relative phase of v_{out} and v_{in} as a function of frequency. This is the common way in which the stability of a circuit is represented. For simulation purposes the output of the I-DAC is connected to an 800 Ω dummy load resistor and the I-DAC is set to deliver its maximum current of ~ 2.046 mA. An AC signal is superimposed on the V_{DD} DC supply voltage for AC analysis; $V_{DD} \equiv V_{IN}$ was set to 2.5 V with an AC sinusoidal signal content of 1 mV RMS. V_{out} is simply the voltage across R_{LOAD}.
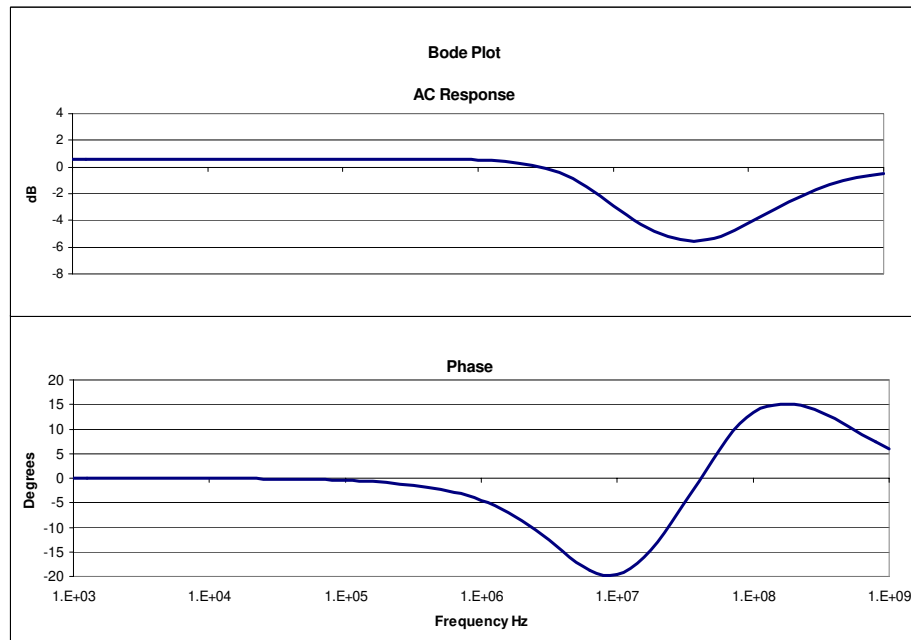
Figure 4-11 The Bode plot for the small signal analysis of the I-DAC output.

The Bode plot gives a gain of ~1 for a frequency range of 1×10^3-1×10^9 Hz and the phase has a minimum/maximum at -20° and +15°,respectively. For the sensitive front-end amplifiers the range of interest is 1-100 MHz which, from the Bode plot, can be seen to be a particularly stable operating point of the current-bias generator, as the gain drops to –6 dB. This low gain is beneficial to reducing the noise contribution from the bias generator. To ensure that the bias generator does not contribute any significant noise to the front-end amplifiers, a noise measurement is taken from simulation. The first measurement, taken at the V-ref point, gives a worst case result of $17nV \big/ \sqrt{Hz}$ for frequencies less than 1 MHz and $2nV \big/ \sqrt{Hz}$ for frequencies above 1 MHz, respectively. The 1 MHz was an arbitrary point to demonstrate the difference in noise values, where the low frequency content is attributed to flicker noise. The noise figure is approximately 20 μV RMS for the frequency range that the Beetle front-end responds to and is well within the 140 μV limit that was discussed earlier. For the I-DAC output, when connected to its load, the noise translates as $40nV \big/ \sqrt{Hz}$ and $4nV \big/ \sqrt{Hz}$ for below and above 1 MHz, respectively. The noise above 1 MHz can mostly be attributed to the noise generated by the dummy-load resistor.

93

4.3   Voltage-bias module

The Beetle chip requires five externally adjustable voltage references; refer to Figure 3-2 for V_{fs}, V_{fp}, V_d, V_{dcl} and V_{rc}. As with the current-bias nodes, the specifications are given by the designer of the sub-circuit that the bias will service. The voltage-reference nodes all have high input impedance[26] and a low voltage precision requirement. These voltage references can therefore be generated with a simple V-DAC design, one for each node. The physical size of the V-DAC's and power consumption should be optimised, and the dynamic range should be large, preferably from the ground reference up to the V_{DD} supply rail (rail-to-rail).

There are many approaches to designing a V-DAC. Most applications require high precision or fast switching; the Beetle chip does not require either of these. The simplest approach is to have a resistor ladder of 2^n resistors from rail-to-rail, where 'n' is the V-DAC resolution in number of bits. This removes the requirement of a voltage reference[27] and it is easy to minimise the DNL as only one node is switched to the output for a given binary input i.e. it is not a superposition of switched currents or voltages. The disadvantage of the resistor ladder is the balance between power and chip area, as large resistors (2^n of them) are required for low power consumption. Further chip area is consumed for the 2^n switches that are required.

An alternative method of using resistor ratios, the one implemented in this case, uses the *R-2R* ladder network. This has a full rail-to-rail dynamic range and requires only 2n+2 resistors and n switches. The main disadvantage of the *R-2R* is the degraded DNL caused by the resistance of the binary switches. Also there are stringent matching requirements between the R2R resistor network used for the MSB and LSB settings in order to make the V-DAC monotonic (see below, and refer to section 4.2 for discussion of DNL and monotonicity).

---

[26] The V_{dlc} and V_d circuit nodes do have current demands, therefore these nodes have a buffer at their inputs.

[27] Band-gap diodes are usually used as voltage references but were not available within the fabrication process at this time.
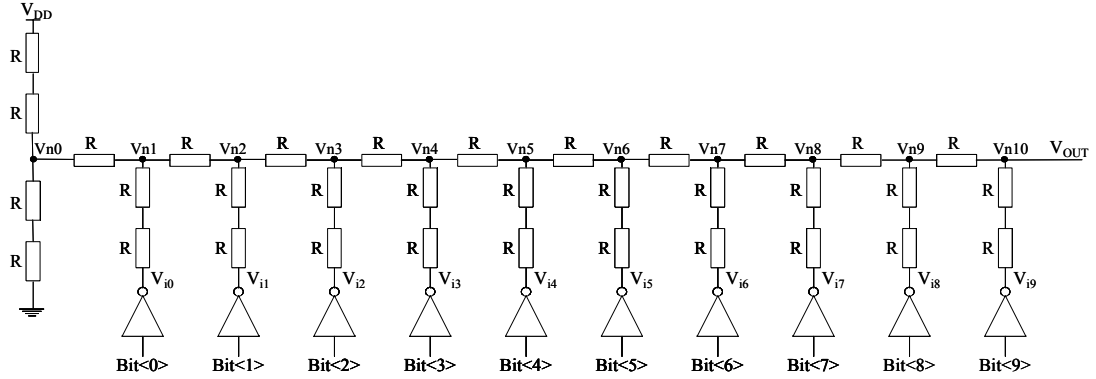
Figure 4-12 Schematic circuit diagram of a 10-bit R-2R ladder. For clarity the digital control logic is not shown. The implemented value of R for the BeetleBG1.0 chip is 3 kΩ.

The circuit diagram of a 10-bit *R-2R* ladder V-DAC is shown in Figure 4-12. The *R-2R* configuration consists of a network of resistors that alternate in value by R and 2R. For ASIC device matching purposes only multiples of R are used in this case i.e. the 2R value is made from two resistors of value R. Each node voltage is related to V_{ref}, in this case V_{DD}, by a binary-weighted relationship caused by the voltage division of the ladder network. The digital inverters pull $V_{i(n)}$ to either ground potential or V_{DD}. The circuit can be analysed by replacing the circuitry at the nodes $V_{n(0-10)}$ with a Thevenin-equivalent circuit. For example, the resistor chain between V_{DD} and ground is equivalent to connecting a voltage source of $V_{DD}/2$ with an output resistance of R to node V_{n0}. The Thevenin circuits for the remaining nodes will depend on whether $V_{i(n)}$ is 0 or V_{DD}. If V_{i0} is 0 volts, then an equivalent voltage source of $V_{DD}/4$ can be connected to the node V_{n1}. For V_{i0} equal to V_{DD}, then a voltage source of $\frac{3}{4}V_{DD}$ is used. The Thevenin-equivalent source output impedance is always R for any node. A general equation can be given for V_{OUT} that is a function of the binary input as

$$V_{OUT} = \frac{V_{DD}}{2^n}\left(\frac{1}{V_{DD}}\sum_{j=0}^{n-1}2^j V_{ij}\right) + \frac{V_{DD}}{2^{n+1}},$$ Equ 4-4

where n is the resolution i.e. 10 bits, j=0 represents the LSB, j=9 represents the MSB and V_{ij} is the voltage at the inverter node associated with the jth bit. V_{ij} is 0V and 2.5 V for a logic 0 and logic 1 input, respectively. The last term of Equ 4-4 accounts for the 0.5 LSB voltage offset

95

which is inherent to this design due to the V_{OUT} node being the output of a potential divider resistor network formed by the last two resistors in the ladder.

The power specification for the Beetle chip was originally defined at 2 mW/channel. With 60 % of this required for the front-end amplifiers, an approximate total power of 4 mW was allowable for the V-DAC. This was divided between 2+2 mW for both the analogue and digital parts. The maximum current demanded from an *R-2R* is calculated to be V_{DD}/R from the Thevenin-equivalent circuit. Therefore this gives a reasonable R value of 3 $k\Omega$ for a V_{DD} of 2.5 V. Larger values of R would be advantageous for reducing the power further and improving the DNL, but would require more real estate. The noise voltage output for this circuit is $\sqrt{(4kTR_{OUT}\Delta f)}$, where R_{OUT} is the output resistance, simply R in this case. Therefore there is a further advantage in the use of small ohmic values, although for this design, analogue noise is not a demanding parameter.

To consider the effects of R on the DNL, the impedance of the inverter switch must be known, as this resistance is in series with the 2R leg of the *R-2R* circuit and causes an imbalance in the ladder network. The DC resistance of the inverter gates can be found from

$$R_{FET} \approx \frac{L}{K'W(V_{GS} - V_t)} \quad , \text{where } V_{DS} \leq V_{GS} - V_t \; , \qquad \text{Equ 4-5}$$

which gives an approximate resistance of the gate channel when the device is operating in the ohmic region. The W/L aspect ratio of the inverter gate is taken from the standard 0.25 μm process library. The resistance of the inverter switch for both the *p* and *n* type FET is approximately 30 Ω and is therefore 1 % of R. If the layout allows it, then some compensation of the switch resistance can be made either by selecting an appropriate inter-track width and length to give a compensation of approximately 30 Ω/2, or to add compensating resistors in the single R legs of the ladder. When selecting the inter track aspect ratio, care should be taken of electromigration[28] which has a significant effect on tracks with small widths.

---

[28] Electromigration is movement of atoms with the current flow that deteriorates the tracks in analogy to wind erosion.
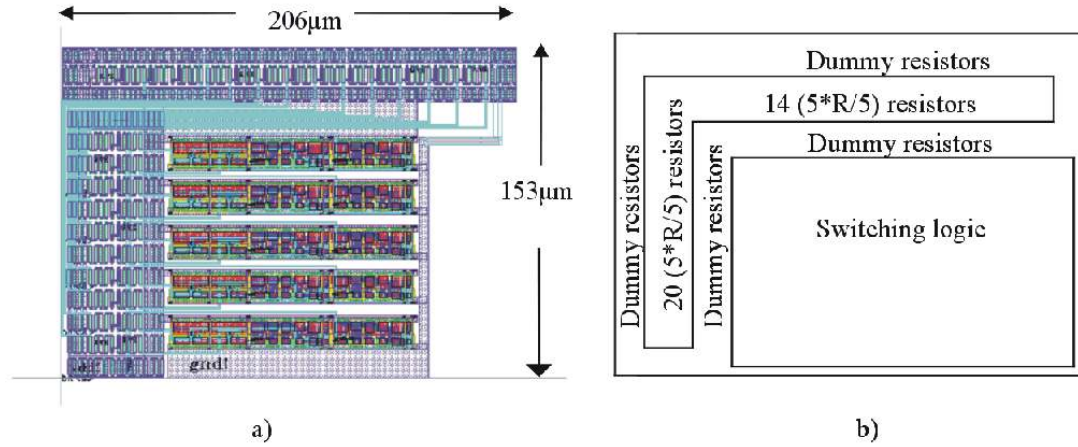
Figure 4-13 The BeetleBG1.0 *R-2R* 10-bit voltage DAC, a) layout and b) floor plan.

Figure 4-13 shows a) the V-DAC layout and b) the floor plan. The resistors are implemented as 2.9 $k\Omega$ N$^+$ diffusion OP types, where OP is a fabricators definition of this resistor. These can be made within a 10 % core resistor tolerance and within a 50 % end-cap tolerance. The simplified equation for calculating the resistance of the drawn N$^+$ diffusion OP resistor is

$$R = 2 \times \left(\frac{R_{END}}{W} \right) + \left(R_{SOP} \times \frac{L}{W} \right) , \qquad \text{Equ 4-6}$$

where R$_{END}$= $20\,\Omega \times 0.46\,\mu$m , the 0.46 µm is fixed in this design and is the distance between the contact and OP border, and R$_{SOP}$= $63\Omega/sq$ for the N$^+$ diffusion resistor. R$_{SOP}$ is the sheet resistance and is extensively used in ASIC design for calculating resistor values. The sheet resistance is a measure of the characteristics of a large, uniform sheet or film of material with a thickness that is considered to be negligible. The sheet resistance is specified in terms of ohms per square (sq) of surface area. If a rectangular sheet of material of length L and width W has measured resistance R between the two opposite ends, then the sheet resistance of the material is given by $R_{sq} = R \dfrac{W}{L}$.

As with the I-DAC, the largest non-linearity will occur when switching from the decimal input 511 to 512, i.e. the transition where the MSB is switched from off to on, and all other bits

are turned from on to off. For the V-DAC to remain monotonic this transition should produce a non-linearity of less than one LSB. In the case of the R-2R ladder this linearity is determined by the matching of the R and 2R resistors associated with the MSB. Therefore, all *R* and *2R* resistor ratios should be better than $\frac{2R_i}{R_j} - 2 \le 2^{-(n-1)} = \pm 0.195\,\%$ for a 10-bit V-DAC, where i and j are labels. The limit that can be achieved is ultimately down to the process variations but by using the special layout techniques, described in section 4.1, the matching of devices can be drastically improved. This design was implemented with each R resistor being composed of 5 sub-resistors in series, each with W=1 μm and L=9 μm.

To measure V_{OUT} as a function of the binary number set at the V-DAC input, a multi-meter with an input resistance of 10 *M*Ω was connected as the load to each V-DAC in turn. For the simulation, a 10 *M*Ω load resistor was assumed. Figure 4-14 a) shows a graph of V_{OUT}

a)   b)
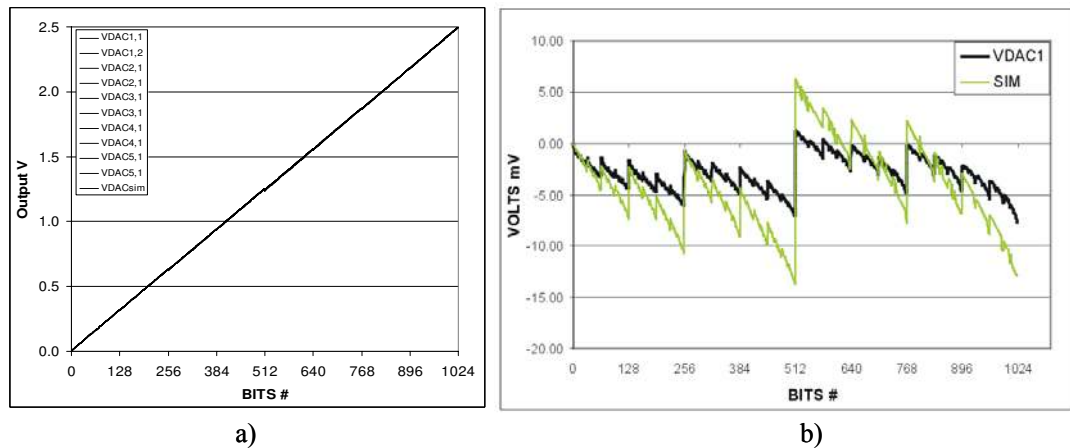
Figure 4-14 Results and simulation from the voltage DAC, a) Linearity of 10 measured DACs and simulation, b) simulation of DNL without compensating resistors, and measured DNL of a single V-DAC.

as a function of the input binary number, measured for a sample of ten V-DACs. The simulation is also included, although it is not visible as it matches well with that measured. The mean dynamic range of all the measured V-DACs is 1.57 mV to 2.498 V; the LSB is 2.43 mV and the variation between the V-DACs is ± 1 % RMS. The gain error (defined in the I-DAC section) is 0.5 %. Figure 4-14 b) shows the measured V-DAC DNL using the layout of Figure 4-13, compared to simulation of a V-DAC without compensating switch resistors. The DNL is measured to be ± 1.4 LSB. Although this does not meet the general requirements for a 10-bit

V-DAC it is more than adequate for the Beetle chip requirements, which needs a precision of only 8-bits. Like the I-DAC, the 10-bit V-DAC was submitted to evaluate V-DAC process variations and will have the last two LSBs removed for the full chip submission.

For compliance[29] measurements, a variable resistive load was connected to the output of a V-DAC and varied, while measuring ΔV_{OUT} of an LSB setting with a multi-meter. The value of V_{OUT} when $R_{LOAD} = \infty$ is taken as 100 % compliant. Figure 4-15 shows V_{OUT} to be compliant to better than 99 % for loads greater than 70 $k\Omega$, and compares well with simulation. From the value of load resistance that gives $V_{OUT\_max}/2$, the DC output resistance ($\equiv R_{LOAD}$) of this V-DAC is found to be 2.7 $k\Omega$. The voltage offset in Figure 4-15 is due to the last term in Equ 4-4 and is calculated to be 1.22 mV (0.5 LSB). This agrees well with the measured value of 1.55 mV.
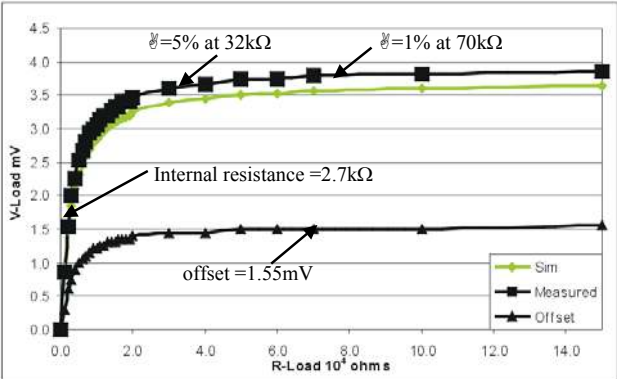
Figure 4-15 Vout vs R load for the voltage DAC with 1 LSB set.

To measure the circuit sensitivity to temperature and voltage-supply variations, a multi-meter was again used as the load resistance. For a 2.4 mV LSB setting on a the V-DAC, the sensitivity to the V_{DD} supply voltage was measured to be approximately linear, 1.65 mV/V, over the supply range of 0.7-3.0 V. When all bits of the binary input number are set to '1' the sensitivity to the supply voltage is 0.969V/V. This is easily understood as the voltage V-DAC is very dependent on the inverters being 0V or V_{DD}.   By using a hot-air gun to increase the chip temperature and a thermistor to measure it, the chip temperature sensitivity over a range of 25-

---

[29] Good compliance in this case means only a small ΔV_{OUT} results under varying resistive load conditions.

80 °C was measured to be 1.8 mV/°C at the V-DAC output, for any binary input setting. This figure is surprisingly high, as variation to resistors due to temperature should be common across the V-DAC and cancel out. Therefore the effects are attributed mainly to temperature effects in the inverters.

Although the V-DAC worked within specifications the author learned of several disadvantages with working with N^+ diffusion resistors, outlined below, and therefore changed all resistors to a non-silicided[30] polysilicon type for the Beetle chip submission. The N^+ diffusion creates a reversed-biased diode and therefore has a reverse leakage. Under irradiation this diode could also go forward biased. These resistors exhibit an undesirable non-linear relationship between voltage and current, most evident in larger devices. N^+ diffusion sheet resistance is $63\,\Omega/_{sq}$ compared to $210\,\Omega/_{sq}$ for non-silicided polysilicon, and therefore the N^+ diffusion resistors consume more chip area for the same resistance given by a polysilicon resistor. Although the absolute accuracy of the N^+ diffusion resistors is better than the polysilicon, this does not have great importance when the circuit is designed to work on good matching and ratio techniques. Also it should be noted that polysilicon has an increased radiation tolerance due to the energy bands present at the grain interface and its amorphous structure [BAU03].

4.4   Summary

A fully integrated bias generator that supplies all of the necessary current and voltage biasing for the Beetle chip has been designed, fabricated and tested.

The 10-bit I-DACs use a binary-weighted current-source scheme with 1023 PMOS transistors. The linear PMOS transistors, with an aspect ratio of 0.6μm/3μm, were chosen because of the restricted aspect ratio available with edgeless devices. The necessary voltage references are generated with current mirror scaling circuits. The size of each I-DAC is 281x104 μm^2. The mean dynamic range of the 36 I-DACs measured is from 0.1 μA to 2.101 mA (± 1 % RMS) with an LSB of 2 μA and a gain error of 2.6 %. These values are very dependent on the

---

[30] Silicides and/or refractory metals are alloyed with the poly to reduce resistance.

supply voltage and temperature. The differential and integrated non-linearity has been measured to be ± 0.5 LSB and ± 1 LSB, respectively. The output resistance for an LSB setting is 38 $M\Omega$ for a load voltage of 0-2 V. The noise level is negligible. This I-DAC is suitable for the Beetle chip requirements.

The 10-bit V-DACs are of the *R-2R* type utilising 3 $k\Omega$ N$^+$ diffusion OP resistors. The mean dynamic range measured for ten V-DACs is from 1.57 mV to 2.498 V with an RMS of ± 1 % and a gain error of 0.5 %. The mean LSB is 2.43 mV and is better than 99% compliant with a load resistance > 70 $k\Omega$. The differential non-linearity is ± 1.4 LSB making the V-DACs non-monotonic, although this is not important for the final Beetle design. The size of each V-DAC is 206x153 μm^2. The V-DAC has a worse-case sensitivity to the supply rail setting of 0.969 V/V. The effects of temperature are easily managed. This V-DAC is suitable for Beetle chip use but benefited in performance when the resistors were changed from N$^+$ diffusion to non-silicided polysilicon for the Beetle fabrication submission.

The BeetleMA ASIC

In the context of this thesis, one of the objectives was to modify the Beetle ASIC so as to make it compatible with the MAPMT. The need for a modification is primarily due to the large MAPMT output signal compared to that of a silicon strip detector, which the Beetle was designed for. The modification required a replacement of the Beetle front-end, consisting of a charge sensitive amplifier (CSA), a CR-RC shaper and a buffer, shown in Figure 5-1. The modified/replaced front-end was renamed the BeetleMA.
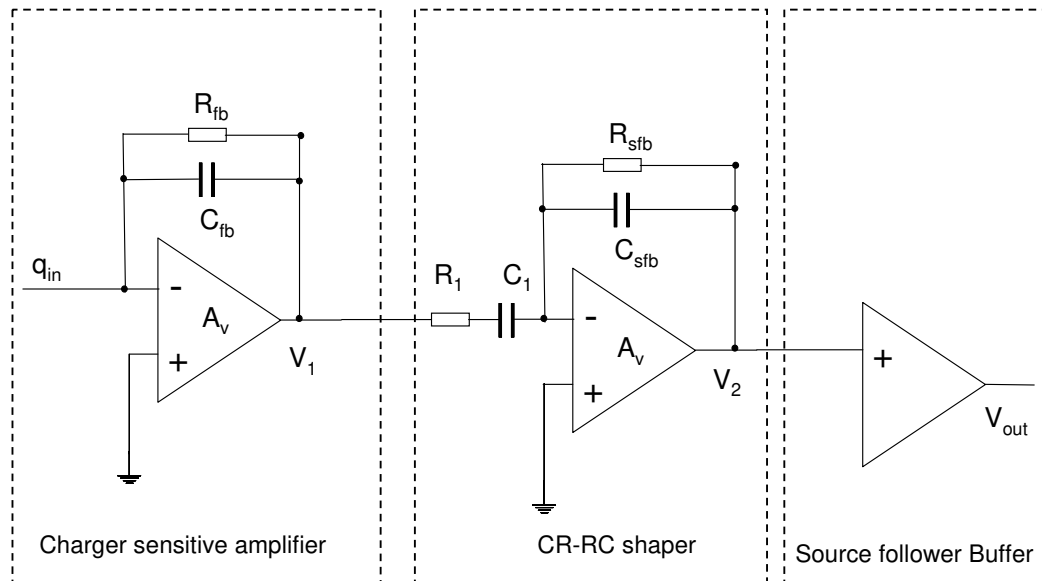
Figure 5-1 The Beetle front-end amplifier consisting of CSA, shaper amplifier and buffer.

5.1   Front-end amplifier selection

Amplifier selection requires considerable knowledge of the input load (the source) that will be connected to it, in this case the MAPMT. The MAPMT characteristics were described in Chapter 2 with the emphasis on photon detection efficiency, while here the emphasis is on electronic characteristics.

102

For the tests described in Chapter 2, the MAPMT output impedance was 50 Ω by the addition of a load resister of that value in parallel to the large output impedance of the tube. A voltage amplifier was then utilised as this was suitable for the job at hand and was easy to implement into the test system. However, ideally the MAPMT should be considered a current-source as it has an output impedance of $G\Omega s$. The output current of an MAPMT is dependent on the high voltage bias but must stay within limits; too high a voltage the tube electrically breaks down, and too low a voltage the tubes SNR becomes poor [EIS03]. For the R7600-03-M64 MAPMT used for the BeetleMA studies, a high voltage of -800 V gives a typical output signal of 300 k electrons (ke$^-$). Taking into account the gain variations between MAPMT tubes and pixels and the variation in number of photon hits within a channel, the MAPMT electronics must be optimised to operate with an input dynamic range of 0.3-2.7 Me$^-$. The dark current can be neglected as it is the order of pA.

Parameter	Average or expected value
Average single-photon response	300,000 e$^-$ @ -800 V
MAPMT capacitance without the base	1.5 pF
Pulse rise-time	2 ns
Pulse fall time	3 ns
Pulse duration	5 ns
Gain variation tube-to-tube	2
Gain variation pixel-to-pixel	2
Total gain variation	3
Range of number of input photon	1-3
Required signal to noise	40
Dynamic range	0.3-2.7 Me$^-$
Max load capacitance with PCB	10 pF
Possible photon saturation	yes
Typical channel occupancy	1 % & 10 % depending on location

Table 5-1 MAPMT characteristics.

The MAPMT does not suffer from internal electronic cross-talk [MUH00]. However, cross-talk may occur at impedance points on the PCB or between channels on the same substrate of the chip if correct physical layout and design of both the PCB and chip are not employed. For example, the large signals from the MAPMT may produce a shift in the bulk voltage which will have an effect across the entire chip. Another consideration is that charged particles traversing the MAPMT can produce Cherenkov photons in the quartz window; the number of photons produced is highly dependent on the angle of the particle with respect to the tube axis. For normal incidence tracks 5 to 10 photons are produced, for angles around 45 degrees up to 30 of the 64 channels can be activated [MUH00]. This puts greater demands on the channel occupancy and necessitates fast recovery from front-end saturation. This is discussed in the subsequent sections. Table 5-1 summarises the MAPMT characteristics.

As the Beetle chip was primarily designed for a silicon strip detector, a general comparison is now made between the MAPMT output characteristics and that of the silicon strip detector used with the Beetle. Table 5-2 shows the relevant characteristics of each device. It is clear that the main difference between the two types of detector is the magnitude of the response to a single-photon or MIP input; the MAPMT typically delivers 13 times more electrons. The SNR

Parameter	Approx MAPMT value	Approx Silicon value
Input response photon/MIP	300,000 e$^-$	22,000 e$^-$
Load capacitance/resistance	1.5 pF/~10 G ohms	10-30 pF/~10 ohms
Rise-time	2 ns	~3 ns
Required signal to noise	40	20
Dynamic range	2.7 Me$^-$	0.22 Me$^-$
Typical channel occupancy	1 % to 10 %	~3 %

Table 5-2 Comparison of MAPMT output characteristics to that of a silicon detector.

requirement of 40 in Table 5-2, discussed in Chapter 2, is based on the noise measurements of the Beetle chip and the MAPMT gain variations. This gives a SNR requirement for the MAPMT of twice that of a silicon detector. However, as the MAPMT has a larger output signal
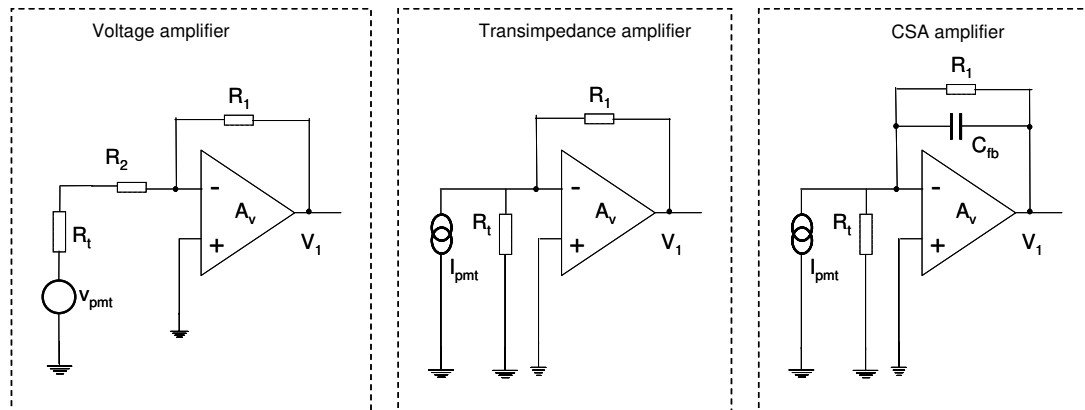
Figure 5-2 Three possible amplifier types.

and a smaller capacitance with respect to a silicon detector, the SNR is intrinsically improved for the Beetle CSA MAPMT connection. This is described further in section 5.2. The other major difference between the two dectectors is the output resistance; the MAPMT has a very large output resistance while the silicon detector is very low. Guided by the characteristics of Table 5-2, a consideration of different amplifier types for use with the MAPMT can now be made.

Three basic amplifier configurations are considered, and these are shown in Figure 5-2:

- Voltage amplifier. For the voltage amplifier, the MAPMT output source is shown as a Thevenin-equivalent voltage source. The voltage amplifier has a gain of $-R_1/(R_2+R_t)$, where R_t is the output resistance of the MAPMT. Therefore, in order to have a practical gain value, R_1 would have to be made unreasonably large. The effective value of R_t can be reduced by having a parallel resistor which loads the output of the MAPMT, as was described in chapter 2. However as the input of the amplifier at the inverting terminal sees a noise impedance of R_2 in parallel with R_1, this is not the most noise-quiet amplifier configurations. The same would apply in a non-inverting configuration.

- Transimpedance amplifier. With this type of amplifier the MAPMT load is better represented as a Norton current-source equivalent circuit. The transfer function, in this case v_{out}/i_{in}, is simply $-R_1$. This amplifier suffers the same noise problems

105

as the voltage amplifier. The transimpedance amplifier can easily become unstable if its input has inductance and capacitance, which in this case could be the bonding wire inductance and gate capacitance of the amplifier's input transistor. For both the voltage and transimpedance amplifier, the downstream Beetle shaper stage would need considerable modification, or need to be removed completely. The removal would degrade the noise performance considerably as the shaper stage is a band-pass filter.

- CSA: Again with this type of amplifier, the MAPMT load is better represented in the Norton current-source equivalent circuit. The CSA actively integrates the detector charge (q) on the amplifier feedback capacitor C_{fb} and the transfer function, in this case v_{out}/q_{in} , is approximately $-1/C_{fb}$. To remove the charge from C_{fb} a parallel resistor, R_1, is chosen so as to give the required CR discharge time-constant. This is the quietest noise configuration as a suitably chosen C_{fb} offers a low impedance point at the frequencies of interest. A noise evaluation of this configuration is given in the following section.

To minimise the changes to a working chip, it is preferable to modify the existing Beetle front-end as little as possible. Significant modifications can lead to design or layout errors and discontinuities with further Beetle design iterations, which may be required for 'bug' fixes within the Beetle Family. Since the CSA offers the best noise configuration, the Beetle CSA was modified to accept the MAPMT as a source. So that no other part of the Beetle chip required modification, it was also important that the output of the buffer was ensured to be compatible to the dynamic range of the downstream analogue pipeline memory.

5.2 Beetle front-end amplifier characteristics

Before a description of the Beetle modifications can begin, a good understanding of the operation of the front-end amplifier is required. This section discusses the CSA, the shaper and buffer stages. A more detailed description of the shaper amplifier is given in section 5.3.

The Beetle core amplifiers

Figure 5-3 shows the three front-end core amplifier stages used for the Beetle1.2 ASIC. The CSA, shaper and buffer cores, Figure 5-3 (a), (b) and (c) respectively, were designed by E.Sexauer [SEX01] and later improved by S.Lochner [LOC03]. The dimensions of the transistor aspect ratios (the gate widths are the upper numbers) are given in μm. The CSA and shaper cores are of a folded cascode type. The buffer stage is simply a source follower.
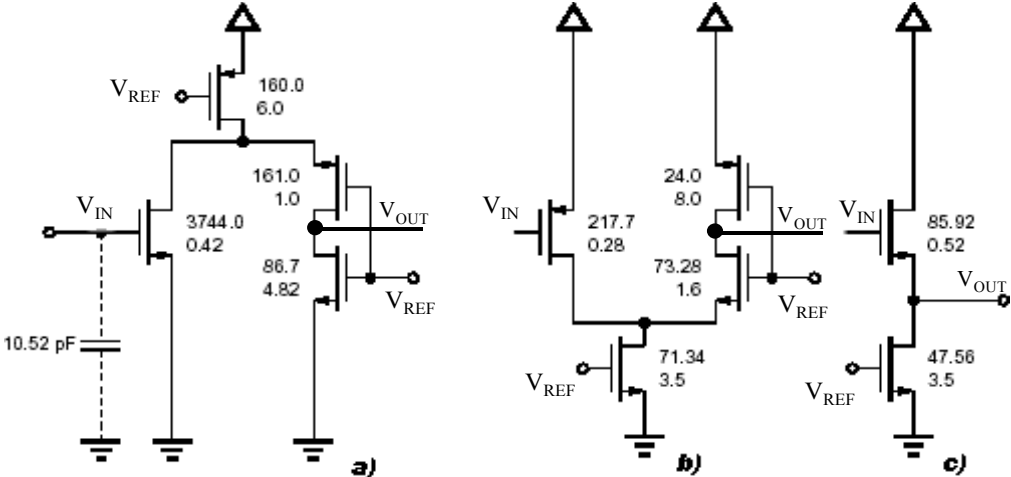
Figure 5-3 The core amplifiers of the Beetle1.3 chip. Figure a) shows the charge sensitive amplifier, b) the shaper and c) the buffer.

The CSA requires a very large open loop gain for good performance and a very large g_m for low noise, which implies in CMOS a large input FET aspect ratio. Consequently this amplifier is large in layout, the Beetle input transistor alone is $8,500 \left(\mu m \right)^2$. This large input device is complex in design and layout due to:

- The number of parasitics that have to be considered;

- The increase in gate, source and drain capacitance making high speeds and stability more demanding;

- The variation of fabrication process which is augmented over large areas.

If suitable, it is very desirable to use the existing CSA core of the Beetle chip and modify only the feedback network. As the shaper and buffer stages were designed around the CSA amplifier, the same also holds true for those components.

The circuits of Figure 5-3 a) and b) are both of the folded cascode type, the difference being that a) is an NMOS design while b) is a PMOS. The NMOS is used for the CSA due to the faster speed gained by the improved mobility over PMOS; the shaper uses PMOS so that a voltage shift can be introduced into the cascade so as to match the buffer DC bias requirements. The folded cascode configuration is composed of a common-source and folded common-gate amplifier, the reason for this is now described.

Figure 5-4 a) shows the common-source amplifier. The gain of this configuration is $g_m \times R$, where R is the output resistance of the current-source. With carefully designed components the gain can be reasonably large (~1000). The disadvantage of this configuration is
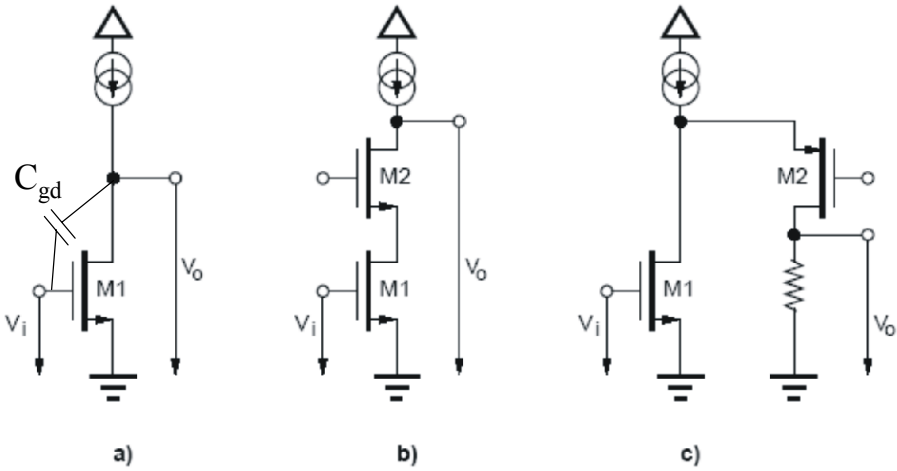
Figure 5-4 Stages of the NMOS folded cascode: a) the source follower, b) the linear cascode and c) the folded cascode [BAU03].

that the parasitic capacitance between gate and drain, C_{gd}, is multiplied by the gain, which is the so-called Miller effect. The application of the Miller's theorem is to replace C_{gd} with a Miller capacitor, C_m that is equal to $C_{gd}(1+g_mR_1)$ between the gate and ground. Therefore this circuit is

108

can be considered to be a low pass filter with a phase shift of up to 90° and a corner frequency of $f_c = \dfrac{1}{2\pi C_m R}$.

The Miller capacitance can be minimised by reducing the gain of the input amplifier and using a cascade of further amplifiers to recover the larger gain downstream. However this may lead to possible instabilities and an increase in noise. Another way in which to reduce the Miller effect is to add a second FET in series with M1, as shown in Figure 5-4 b). This circuit is referred to as a linear cascode. In this case FET M2 is operating as a common-gate amplifier and the following equations can be applied:

Input resistance

$$R_{in} = \frac{1}{\left(g_{m2} + g_{m2\_bs}\right)}\left(1 + \frac{r_{o(m)}}{r_{o2}}\right), \qquad \text{Equ 5-1}$$

Voltage gain

$$A_v = \left(g_{m2} + g_{m2\_bs}\right)\left(r_{o(m)} \| r_{o2}\right), \qquad \text{Equ 5-2}$$

Output resistance

$$R_{out} = \left(r_{o(m)} \| r_{o2}\right), \qquad \text{Equ 5-3}$$

where g_{m2_bs} represents the added transconductance due to the source of M2 not being connected to the bulk, and r_{o2} and $r_{o(m)}$ are the incremental output resistances of M2 and the current-source respectively. For simplicity, FET M1 is assumed to be a perfect voltage-controlled current-source so that it need not be accounted for in these equations. From Equ 5-1 to Equ 5-3, it can be concluded that v_{ds} of M1 will be considerably reduced due to the low input resistance of a common-gate amplifier and that v_o will be reasonably large due to the large output resistance and the common-gate amplifier's open loop voltage gain, A_v, which is about 20 % greater than that of the common-source. In this configuration, the Miller capacitance has been virtually eliminated from the feedback path.

For the low noise and large open-loop gain requirements of the Beetle chip, the input transistor requires a high drain current to achieve the high g_m requirement, refer to section 3.2. This large current flows through the entire chain in the case of the linear cascode and reduces the output resistance and therefore the overall gain. Figure 5-4 c) shows a circuit referred to as the folded cascode, which overcomes this problem by folding the common-gate. This requires an additional load resistor but allows the common-gate (M2) to operate at a lower current, therefore increasing the output resistance. For the folded cascode CSA and shaper cores in Figure 5-3, this additional load is an active device.

CSA noise evaluation

The noise of a circuit can be evaluated by replacing each component within the circuit with an ideal noiseless component and a dedicated current or voltage noise source. For resistors the Thevenin or Norton principle is used to covert from one to the other. For FET devices, the channel noise is represented as a current-source in parallel with the channel, given by Equ 3-19. By taking the inverse of g_m in Equ 3-19, the channel noise can be represented by a voltage source on the gate of the device. Figure 5-5 shows the different representations; the one chosen is dependent on the circuitry to be evaluated.
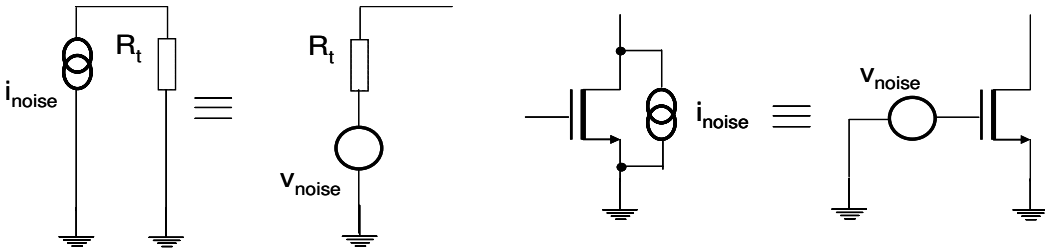
Figure 5-5 Norton and Thevenin equivalents for a resistor (left) and channel noise (right).

In the normal case of describing the noise of an amplifier, the noise performance is characterized by a series voltage noise, e_n^2 in units of $\left[V^2 \middle/ Hz \right]$, and a parallel current noise source, i_n^2 in units of $\left[A^2 \middle/ Hz \right]$.

110

The first step in evaluating the CSA noise performance is to replace the core of the CSA shown in Figure 5-3 a) with its equivalent noise component. As stated previously, the predominant noise source of the CSA is the input transistor. The noise figures e_n^2 and i_n^2, previously discussed in section 3.3, are given by

$$e_n^2 = \frac{K_f}{C_{ox}^2 WLf} + \frac{8}{3} kT \frac{1}{g_m} \ ,$$

Equ 5-4

$$i_n^2 = 2qI_D \ ,$$

Equ 5-5

respectively. The first term in the expression for e_n^2 represents the $1/f$ noise of the input transistor and the last term represents the input transistor channel thermal noise. The expression for i_n^2 represents only the shot noise. Both terms of e_n^2 are dependent on the gate dimensions, the second term is also dependent on the drain current which can be increased to reduce the noise, which increases the g_m of the device. The increase in drain current is achieved by making the aspect ratio of the input transistor large. As the CSA core of Figure 5-3 a) was specifically designed around the requirement for reading out silicon strip detectors, the W and L dimensions (3744 and 0.42 µm) and drain current were chosen through an optimisation process that considered the thermal noise, the $1/f$ noise, the low impedance capacitive load and shot noise of a silicon detector, and the power consumption of the input transistor.

When extending the design to an MAPMT, the components, other than the MAPMT source, are unchanged from that used for the Beetle CSA. Figure 5-6 shows the representation of the CSA configuration with its input connected to an MAPMT. All of the equivalent noise sources have been added. In the Beetle design, the feedback resistor R_{fb} has been implemented with a FET, however it is depicted here as a standard resistor. The feedback resistor is in parallel with a feedback capacitor C_{fb} (400 fF). As the input device has large gate dimensions, the capacitance is not negligible (10.52 pF) and shown here as C_g. The MAPMT is a current noise source in parallel with its output resistance, R_t, which is the normal representation for a device that has a very large output resistance. The MAPMT current noise contribution, i_{pmt}^2, is

111
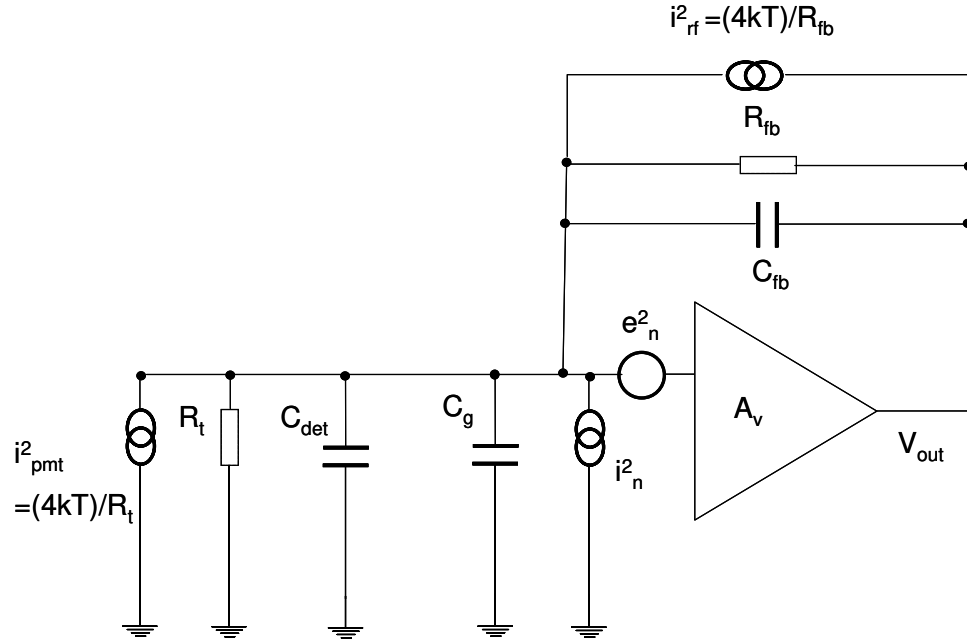
Figure 5-6 The noiseless CSA circuit when connected to an MAPMT.

just the thermal noise from R_t since the leakage current is negligible. We will see later that the MAPMT thermal noise contribution is also negligible. The MAPMT output capacitance of 1.5 pF is relatively small compared to that of a silicon strip detector (10-30 pF). However, to make allowance for the capacitance added through PCB design, the MAPMT output capacitance is taken as 10 pF, represented as C_{det}.

With the source and feedback impedances known, along with associated noise sources, an overall noise spectral density, e^2_{ni}, referred to the input of the pre-amplifier, can be found. This is achieved from the superposition of all of the noise sources added in quadrature, under the assumption that there is no correlation between them. For the Beetle CSA pre-amplifier, e^2_{ni} is given by

$$e^2_{ni} = e^2_n + i^2_{pmt}\left(R_t \left\|Xc_{det}\right\|Xc_g\left\|Xc_{fb}\right\|R_{fb}\right)^2 + i^2_n\left(R_t\left\|Xc_{det}\right\|Xc_g\left\|Xc_{fb}\right\|R_{fb}\right)^2 \qquad \text{Equ 5-6}$$

$$+ i^2_{rf}\left(R_t\left\|Xc_{det}\right\|Xc_g\left\|Xc_{fb}\right\|R_{fb}\right)^2 ,$$

where $X_c = \dfrac{1}{2\pi f C}$ and the labels are self explanatory. The current noises of Figure 5-6 are multiplied by the parallel load impedance so as to convert them to a voltage equivalent. The noise currents i^2_{rf} and i^2_{pmt} are simply due to thermal resistance calculated using the MAPMT output resistance ($\sim 10\,G\Omega$) and the FET channel resistance used for the feedback resistor ($\sim 10\,M\Omega$), respectively. As the resistance of both of these devices is large and the i^2_n shot noise term is very small, the overall current noise contribution is negligible. Furthermore, for the CSA operating frequency of $\sim$100 MHz the impedance seen by the current noise sources is also very low, $\sim 80\,\Omega$. Therefore the current noise contribution in Equ 5-6 can be disregarded and the predominant noise source is simply e^2_n, given by Equ 5-4. It can therefore be concluded that using an MAPMT as a source will not degrade the noise performance of the CSA compared to when using a silicon detector. The MAPMT does not introduce any more noise than the silicon detector, and with the PCB load capacitance, offers about the same load impedance. The ENC of Beetle chip $= 497e^- + 48.3e^- / pF$.

5.3   BeetleMA front-end design

The gain function of any amplifier is given by

$$A_f = \frac{v_{out}}{v_{in}} = \frac{A_v}{1 + A_v \beta} \ , \qquad\qquad \text{Equ 5-7}$$

where A_v is the open loop voltage gain and β is the fraction of output signal fed back to the input. For the Beetle CSA, A_v is $3.8.10^3$ and if the large parallel resistances R_t and R_{fb} are ignored β is seen from the feedback network to be

$$\beta = -\frac{C_{fb}}{C_{fb} + C_{det} + C_g}. \qquad\qquad \text{Equ 5-8}$$

Combining Equ 5-7 and Equ 5-8, the gain function of the CSA is

$$v_{OUT} = -\frac{q_{in}}{C_{fb} + \dfrac{C_{det} + C_g + C_{fb}}{A_v}} \quad , \qquad \text{Equ 5-9}$$

where $q_{in} = C_{in} \times v_{in}$. $C_{in} = C_{fb} + C_{det} + C_g$ and q_{in} is the charge from the MAPMT. From inspection, this shows that the gain of the Beetle CSA can be reduced merely by increasing C_{fb}. Also, the charge collection on C_{fb} will improve by increasing its physical size i.e. the larger A_v or C_{fb} in equation Equ 5-9 is, the more charge will be collected on C_{fb}. This leads to an improved noise performance in Equ 5-6. Another benefit is the reduction of any cross-talk due to the reduced impedance point at the inverting node of the amplifier.

From Table 5-2, the number of electrons generated from an MAPMT compared to that of a silicon detector is ~13 times more. To make the CSA feedback capacitor large enough so
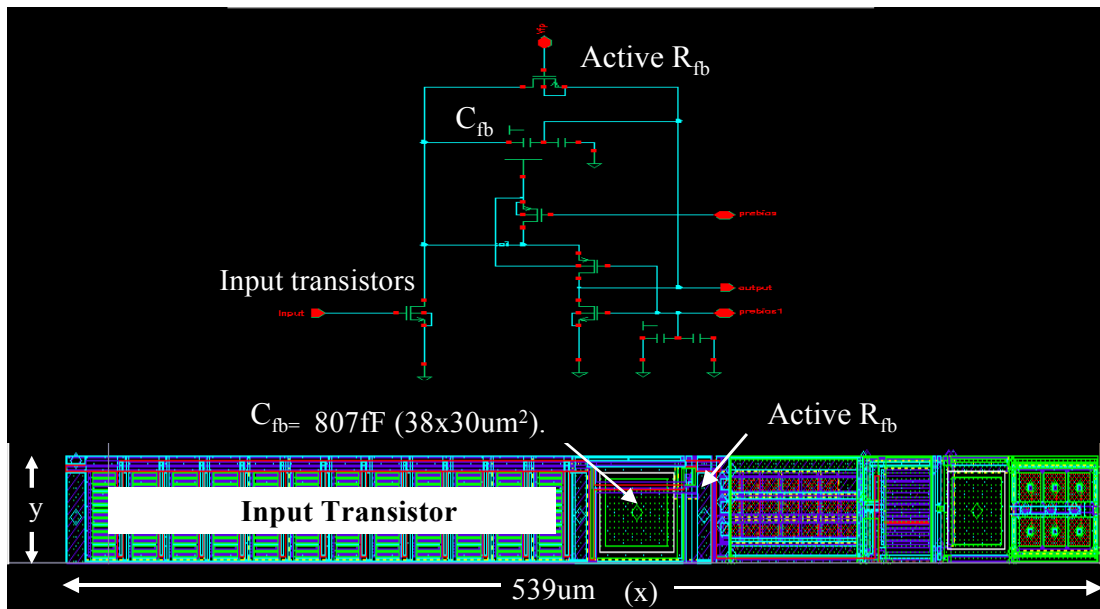
Figure 5-7 The Cadence schematic and physical layout of the modified CSA

114

that the CSA output voltage remains compatible to the existing downstream shaper stage, requires that C_{fb} be made 13 times larger than the present size used in the Beetle ASIC. Unfortunately such a large capacitor is not feasible within the existing layout of the chip, hence the capacitor has simply been made as large as the layout allows. Figure 5-7 shows the schematic and layout of the CSA, where the feedback capacitor C_{fb} has been made as large as feasibly possible by an increase in the x dimension. The y dimension could not be changed. The resulting capacitance is 807 fF, which is approximately twice as large as the standard Beetle C_{fb} capacitor. From simulation using HSPICE running under Analogue Artist, v_{out} of the modified CSA remained linear up to the required 3 Me⁻ at the input and stayed within the allowable output dynamic range. From the approximation $q/C_{fb} = v_{out}$ and the expected 300 ke⁻ for a single-photon input pulse from an MAPMT, the voltage output from the modified CSA is 6.5 times larger than that of a MIP from a silicon detector, and a non-modified CSA. This increase in v_{out} gives an immediate improvement of the SNR of 6.5 compared to the Beetle CSA. The remaining gain-reduction factor necessary, 6.5, is added at the shaper (see the next section).

To study the effects of increasing C_{fb} in the time-domain, the response of the amplifier to a δ-function current pulse on the input is evaluated. In order to evaluate the time-domain response the frequency domain has first to be considered. For this, the loaded input and output
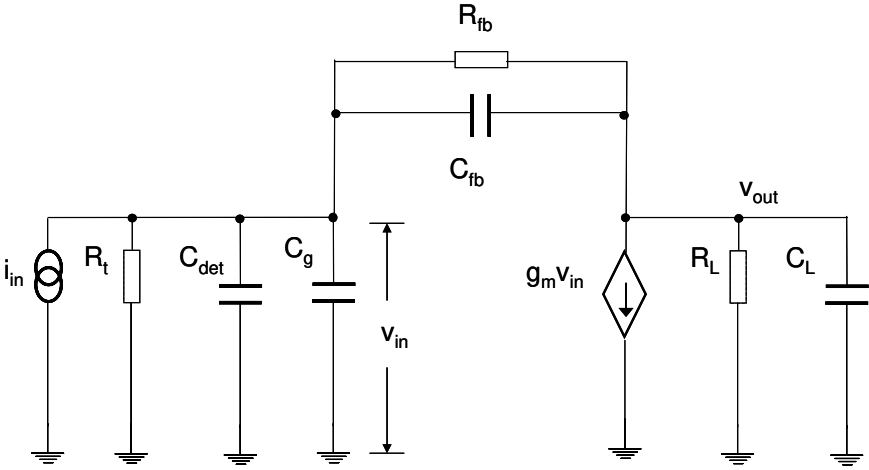
Figure 5-8 Small-signal model of the CSA.

115

impedance of the CSA has to be evaluated. As the CSA output stage is a common-gate amplifier which has large output impedance, the amplifier gain stage will be described as a voltage-controlled current-source output i.e. $g_m.v_{in}$. Figure 5-8 shows the small-signal model used for evaluating the transfer function. R_L in parallel with C_L represents the resistive and capacitive output load of the common-gate amplifier. The transfer function in the frequency domain is then given by

$$\frac{v_{out}(s)}{i_{in}(s)} = -\frac{g_m}{\dfrac{g_m}{R_{fb}} + sg_m C_{fb} + s^2 C_{in}(C_{fb} + C_L)} \quad , \qquad \text{Equ 5-10}$$

where the s term is used to represent $j\omega$ for use with Laplace transforms. In Equ 5-10, R_t and R_L are assumed to be very much greater than R_{fb} , and $C_{in}=C_{det}+C_g+C_{fb}$. Equ 5-10 contains two poles, i.e. the roots of the polynomial expression in the denominator of the transfer function. The two poles are

$$\rho_1 = \frac{1}{2\pi\tau_1} = \frac{1}{2\pi R_{fb} C_{fb}} \qquad \text{Equ 5-11}$$

and

$$\rho_2 = \frac{1}{2\pi\tau_2} = \frac{g_m C_{fb}}{2\pi C_{in}(C_{fb} + C_L)} = GBW\frac{C_{fb}}{C_{in}} \quad , \qquad \text{Equ 5-12}$$

where GBW is the gain-bandwidth product of the CSA [SEX01] . The first pole is produced by the feedback time constant $R_{fb}C_{fb}$ and determines the continuous discharge time of the CSA. The second pole is a result of C_{fb} and defines the rise-time of the voltage output signal of the amplifier. The rise-time is given for the transition time between 10% and 90% as

$$t_{rise} = 2.2\tau_2 = 2.2\frac{C_{in}}{2\pi.GBW.C_{fb}} \quad . \qquad \text{Equ 5-13}$$

116

From Equ 5-13 it is clear that minimising the input capacitance and increasing the feedback capacitance will reduce t_{rise}. The v_{out} response of the CSA in the time-domain can now be obtained by taking the inverse Laplace transformation of Equ 5-10 with a δ-function input, and is given by

$$v_{out(t)} = -\frac{q\tau_1}{C_{fb}(\tau_1 - \tau_2)}(e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}}) \ . \qquad \text{Equ 5-14}$$

The first exponential term is the rising edge, the second term is the falling edge; τ_2 is much smaller than τ_1. Therefore $v_{out}(t)$ has a rising step with a slowly decaying tail.

Figure 5-9 a) shows the simulation result of the BeetleMA CSA output response to a 300 ke⁻ input signal from an 'MAPMT-like' pulse. This asymmetric Gaussian MAPMT-like signal has a rise-time of 2 ns and falls in 3 ns and is generated by the differentiation of a step voltage across a capacitor. Both τ_1 and τ_2 are indicated in Figure 5-9 . The rise-time of the output, t_{rise}

a)   b)
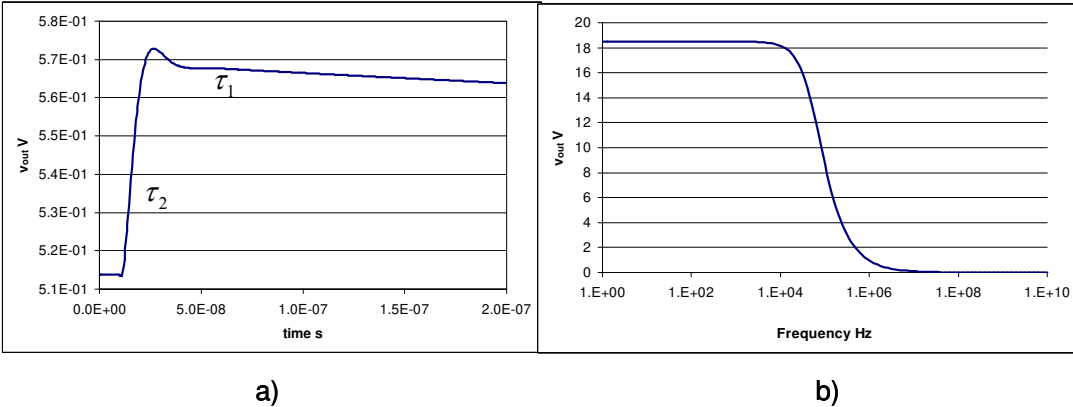
Figure 5-9 a) The simulated CSA output response to a 300 ke⁻ input signal from an MAPMT-like pulse. b) The simulated CSA output response to a 5 μA ac frequency sweep.

, is 12 ns and the decay is of the order of μs (note that in the actual BeetleMA chip, the decay time can be adjusted so that the output returns to baseline after 300 ns by varying R_{fb} via the I²C chip interface). Figure 5-9 b) shows the CSA output response to a 5 μA RMS AC frequency

117

sweep input. The flat response seen between 1 Hz and ~2 kHz is due to the DC response of R_{fb}, as C_{fb} is effectively open circuit at these low frequencies. As expected the output is of a low-pass filter type.

The CSA feedback resistor R_{fb} enables continuous operation by steadily discharging the feedback capacitor causing the amplifier output to return to the baseline. If a new charge pulse should arrive on the input while the CSA output is still returning to baseline, the corresponding voltage step would superimpose on the previous one. If the amplitude of v_{out} is sampled directly at the CSA output, its value would be dependent on history. This history could be removed if the return time to the baseline was fast enough to ensure decay before the next sample was taken. For LHCb this would require a return to baseline within 25 ns, less the signal peaking time. To achieve this, R_{fb} would have to be made small, which would consequently increase the parallel noise content. Furthermore, in the case of the CSA being used with a silicon detector, the pulse amplitude would be reduced to an unacceptable level due to the fast discharge time. A more effective way to remove the history content is to add a following stage that consists of a high-pass filter (CR filter) on the input. In this way the slowly decaying tail is removed and only fast-rising pulses are sampled. To maintain the SNR from the CSA, the CR filter should also be bandwidth limited with a low-pass filter (RC filter). These requirement lead to a CR-RC band-pass filter or otherwise referred to as the Shaper.

Shaper

The task of the pulse shaper is to convert the voltage step at the output of the CSA into a time-limited pulse, with a peak height that is proportional to the charge delivered by the detector. The CR-RC shaping method is a widely used technique. The frequency spectrum is confined by filtering low frequencies by differentiation (CR) and high frequencies by integration (RC). The design boundaries are the SNR performance, the time constraints of the shaped pulse, the dynamic range, the power consumption and the physical space requirements. For LHCb, the maximum allowable pulse peaking time of 25 ns and the constraint that a maximum 30 % of the pulse should remain after a following 25 ns is defined by the global electronic

scheme, described in section 1.3. Since the pulse characteristics are defined, this constrains the SNR optimisation.

As the modifications made to the BeetleMA CSA results in a voltage amplitude 6.5 times higher than the Beetle CSA, the gain of the shaper stage needs to be reduced by approximately 6.5 so its output remains within the dynamic range of the downstream buffer and subsequent pipeline memory. This reduction in shaper gain is ideally achieved by making modifications to only the peripheral components, R_{sfb}, C_{sfb} and C_1 of Figure 5-1. As R_1 is the output impedance of the CSA, it is not accessible for modification, however it is one of the adjustable variables via the I²C chip interface.
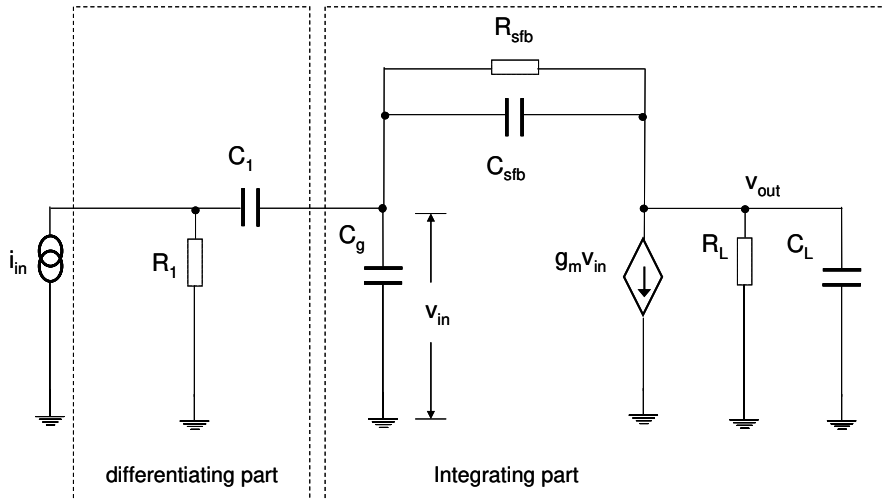
Figure 5-10 Small-signal model of the Shaper.

The evaluation of the shaper can be crudely achieved by separating the differentiating part from the integrating part as shown in the small-signal model of Figure 5-10. R_1 and i_{in} are the output impedance and current from the previous CSA stage. The small-signal model for the integrating part is the same as that of the CSA shown in Figure 5-8, with the exception that the detector capacitance is not present in this case. Therefore equation Equ 5-9 is valid by writing $q_{in} = i_{in} \times R_1 \times C_1$ and making the appropriate substitutions, namely,

119

$$v_{OUT} = -\frac{i_{in} \times R_1 \times C_1}{C_{sfb} + \dfrac{C_g + C_{sfb}}{A_v}} \quad . \qquad \text{Equ 5-15}$$

The gain of the shaper can be reduced by increasing C_{sfb} and decreasing C_1, both of which are beneficial to circuit performance. The increase in C_{sfb} reduces the noise and improves the charge collection, as was the case for the CSA. The decrease in C_1 makes the output less responsive to the slow decaying tail of the CSA output. However these adjustments will affect the time constants of the circuit.

The transfer function in the frequency domain of the equivalent circuit of Figure 5-10, as an approximation, is given as [FAL98]

$$\frac{v_{out(s)}}{i_{in(s)}} \approx -\frac{s g_m C_1 R_1}{s^2\left(C_{sfb}^2 - C_{in}C_{out}\right) + s\left(\dfrac{2C_{sfb} - C_{in} - C_{out}}{R_{sfb}} - g_m C_{sfb} - \dfrac{C_{in}}{R_L}\right) + \dfrac{1}{\dfrac{1}{R_{sfb}}\left(\dfrac{1}{R_{out}} - g_m\right)}} \quad ,$$

$$\text{Equ 5-16}$$

where $C_{in} = C_1 + C_g + C_{sfb}$, $C_{out} = C_L + C_{sfb}$ and $R_{out} = R_{sfb} \parallel R_L$. From inspection of Equ 5-16 it can be seen that this circuit has one zero and two poles, where the zero appears in the numerator of the transfer function. To keep the same bandpass characteristics as the unmodified system, i.e. the same pole and zero separation after the capacitors C_{sfb} and C_1 have been modified, the feedback resistors R_{sfb} and R_1 must be adjustable.

As with the CSA, the maximum physical size that C_{sfb} can be made is governed by the space available on the Beetle chip, without modifying any other part of the chip layout. Therefore, the simple approach was taken of first maximising the size of C_{sfb} in the x dimension and then decreasing the size of C_1 until the desired gain was obtained. To maintain the rise and fall times of the shaper unit, R_{sfb} had to be physically modified to reduce its resistance, even though it is externally adjustable via the I^2C chip interface. This was to bring it into the necessary working range.

The final schematic and layout of the shaper unit is shown in Figure 5-11. The modifications changed C_1 from 700 fF to 190 fF, C_{sfb} from 48 fF to 197 fF, and the adjustable range of R_{sfb} was halved.
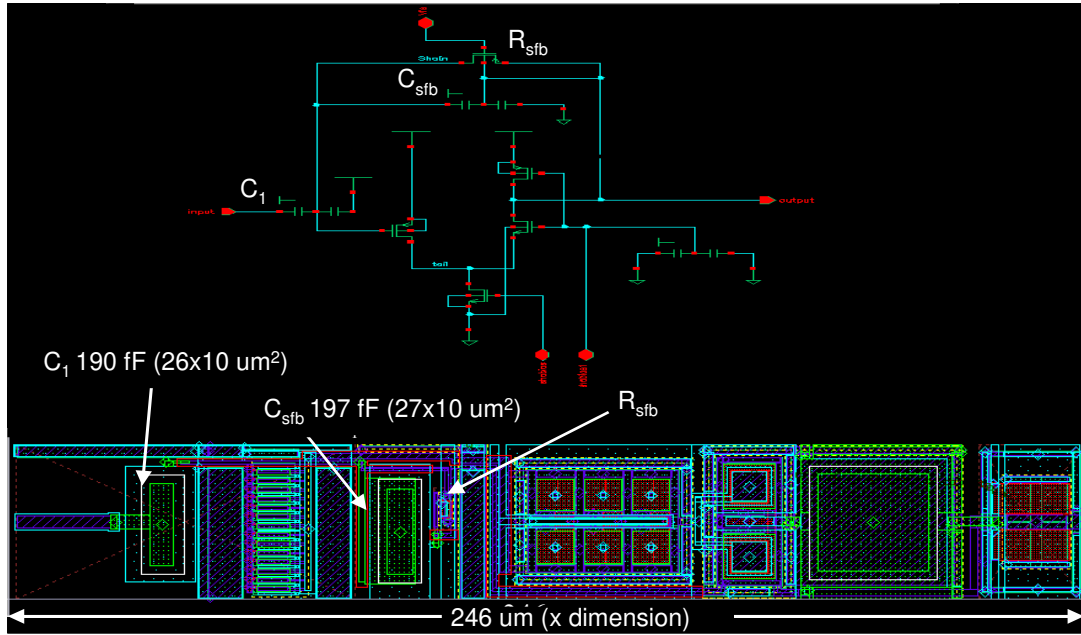
Figure 5-11 The schematic and physical layout of the modified shaper amplifier.

Figure 5-12 a) shows results of simulation of the BeetleMA shaper output response to a 300 ke$^-$ input signal, input at the CSA, from an MAPMT-like pulse. The rise and fall times are 7 ns and 17 ns, respectively. Both the rise and fall times can be adjusted by modifying R_1 and R_{sfb} using the I^2C chip interface. Figure 5-12 a) shows the fastest adjustable response time. Figure 5-12 b) shows the shaper output response to a 5 μA RMS AC frequency sweep at the CSA input. As expected it has a band-pass response with a FWHM of ~50 MHz.

From simulation of the CSA, shaper and buffer, it was concluded that the output buffer did not require any modifications; the output response to injected MAPMT-like signals has the correct magnitude, has the necessary rise and fall time, has the required output dynamic range of 10x300 ke$^-$ and recovers quickly from saturation.

121

a)   b)
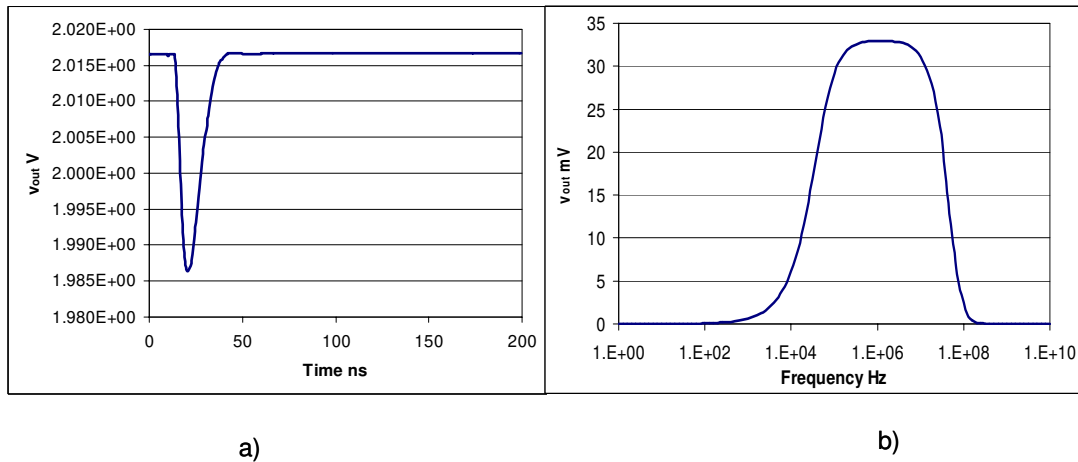
Figure 5-12 a) The simulated shaper output response to a 300 ke⁻ input signal from an MAPMT like pulse at the CSA input. b) The simulated shaper output response to a 5 μA ac frequency sweep at the CSA input.

As the front-end performed satisfactorily in simulation, a prototype chip was submitted in an MPW run for fabrication. The chip is named Beetle1.2MA0 and is discussed, along with the measured results, in the following section.

5.4   Beetle1.2MA0 submission

At the end of September 2002 the Beetle1.2MA0 chip, shown in Figure 5-13, was submitted for fabrication. Other than the front-end amplifiers, this chip had the same architecture as the Beetle1.2 [BAU03]. In total, four different amplifiers were used for testing purposes: 1) 'FBRmod' is a modified Beetle1.2 amplifier for debugging purposes, 2) '1.2T' is a standard Beetle1.2 amplifier for cross-reference purposes, 3) '1.2Div' uses a series capacitor for charge division, and 4) '1.2Att' has the modifications made to the Beetle that have been described earlier in this chapter. The layout plan in Figure 5-13 shows how these amplifiers are grouped together and distributed among the 128 available input channels of the Beetle1.2MA0. With the exception of the FBRmod, The amplifiers are grouped in multiples of at least 3 so that cross-talk measurements could be studied. To gain direct measurement access to the front-end amplifier outputs, as well as through the pipeline readout, probe points were added. The 1.2Att was allocated 64 channels so it could be fully connected to all outputs of an MAPMT. Only the

122
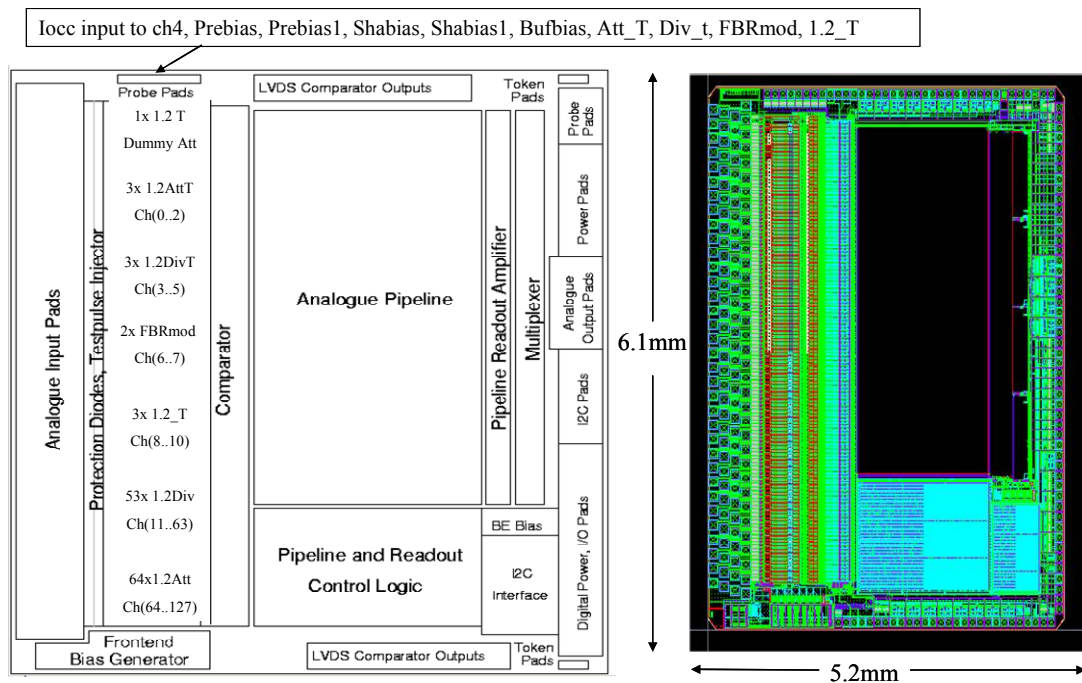
Figure 5-13 The Beetle1.2MA0, a) the floor plan and b) the layout.

results from the 1.2Att amplifier will be presented in this chapter, as this amplifier configuration was the primary reason for the chip submission.

5.4.1   Measurement test set-up

For testing the Beetle1.2MA0 and MAPMT, an electrically-sealed containment box was constructed to reduce external noise sources, shown in Figure 5-14. The size of the box is $60 \times 40 \times 2.8$ cm^3; the only common ground point for the electronics is a star-point connection in the corner. The MAPMT is fixed firmly to the inside of the box where a window has been located to allow an external fibre-optical light source to be directed at the face of the tube. An external dark box protects the MAPMT from extraneous light.   A motherboard from Heidelberg University provides ports and CLC400 amplifiers to read out the pipeline of a Beetle1.2MA0 chip, this being housed on an interchangeable daughter-board.   All power supplies are external to the box.
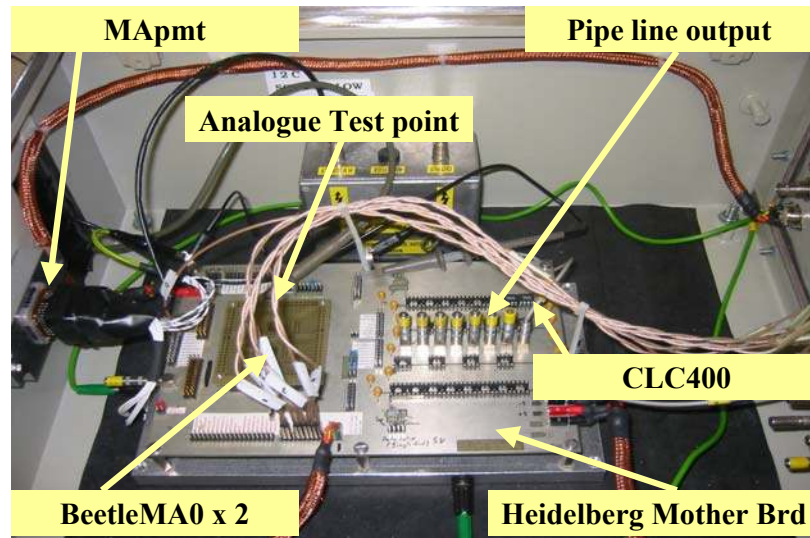
Figure 5-14 Test set-up

The external light source and fibre-optical system is a modified version of the arrangement used in Chapter 2; however the latter version gave some photon timing jitter. For testing an amplifier that is part of a pipeline chip such as the Beetle1.2MA0, the signal sampling point must be fixed and therefore timing jitter would randomize the amplitude reading. In order to improve the photon jitter in the laboratory setup, the mono-mode fibre was replaced with a multi-mode and the storage capacitor C_1 of Figure 2-3 was removed. This improved the LED-to-fibre coupling, therefore reducing the amount of light output required from the LED. Consequently the charge injected into the LED is reduced, allowing a faster LED switching. However, this was at a cost of a larger light spot diameter.

To make the jitter measurement, one output channel of the MAPMT was connected to a 50 Ω resistor and the remaining channels grounded. An oscilloscope with a sample rate of one Giga sample per second (GS/s) measured and stored the MAPMT voltage response across the resistor to single-photons. The oscilloscope trigger was synchronised to the pulse used for the LED light source. From each captured response, the time at which the signal voltage peak occurred was plotted. The LED was triggered at 100 kHz. Figure 5-15 shows the results from the modified light source. The RMS jitter was found to be ±2.5 ns which is an improvement of ~9 ns on the previous system. The photon timing jitter for the LHCb RICH experiment will be

considerably less than ±2.5 ns, and will therefore further reduce the randomizing of the amplitude measurement. The LHC timing jitter is determined by the bunch crossing and transit time, which is of the order of 175 ps RMS [CHR01_L0] and the inherent MAPMT signal jitter is of the order of a few ps.
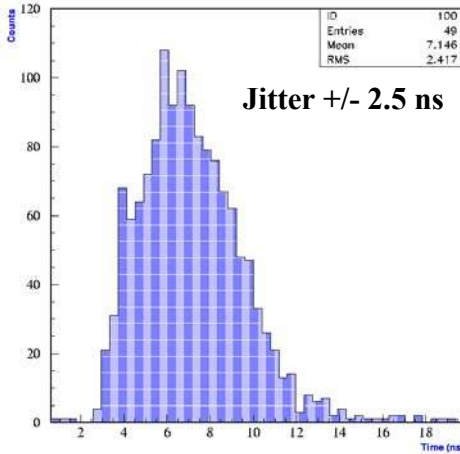
Figure 5-15 The measured time jitter from the light source system.

In order to evaluate the characteristics of the front-end amplifier, three modes of testing have been incorporated. Firstly it is advantageous to measure the voltage signal directly at the output of the amplifier. To do this the motherboard was modified so that it was possible to read out the analogue outputs directly from probe pads. Secondly, to remove complications of MAPMT pixel gain variations and to allow the front-end output dynamic range to be measured in a controlled way, a test-pulse circuit was used. The test-pulse circuit applies a voltage step pulse across a capacitor to generate an asymmetric semi-Gaussian MAPMT-like signal. The capacitor value is chosen to deliver a controlled amount of charge to the Beetle1.2MA0 input. A 1.5 pF capacitor gives 300 ke$^-$/2.6 V. The Beetle1.2MA0 chip does have internal test-pulse generators but these are upstream of the input pads and protection diodes and therefore externally generated test-pulses emulate the MAPMT more effectively.

125

Figure 5-16 shows the schematic for the test-pulse injection circuit for a single channel. The yellow shaded box area is the front-end amplifier within the Beetle1.2MA0 ASIC. The CLC400 amplifier and the resistor attenuation circuitry, along with the capacitor used for test charge generation, are peripheral to the Beetle1.2MA0 and are mounted on the Heidelberg motherboard. The TDS oscilloscope is fully controlled via Labview. Input load capacitance, not shown in Figure 5-16, can be added to the motherboard for testing purposes. For testing with the MAPMT connected to the input, the test-pulse circuitry is simply removed. It should be noted that measurements taken showed that the analogue output signal suffered a 7 % attenuation due to measurement devices used (i.e. the CLC400 non-inverting amplifier of Figure 5-16). The front-end amplifier could also be read out through the chip pipeline and differential output ports, shown in the photograph of Figure 5-14.
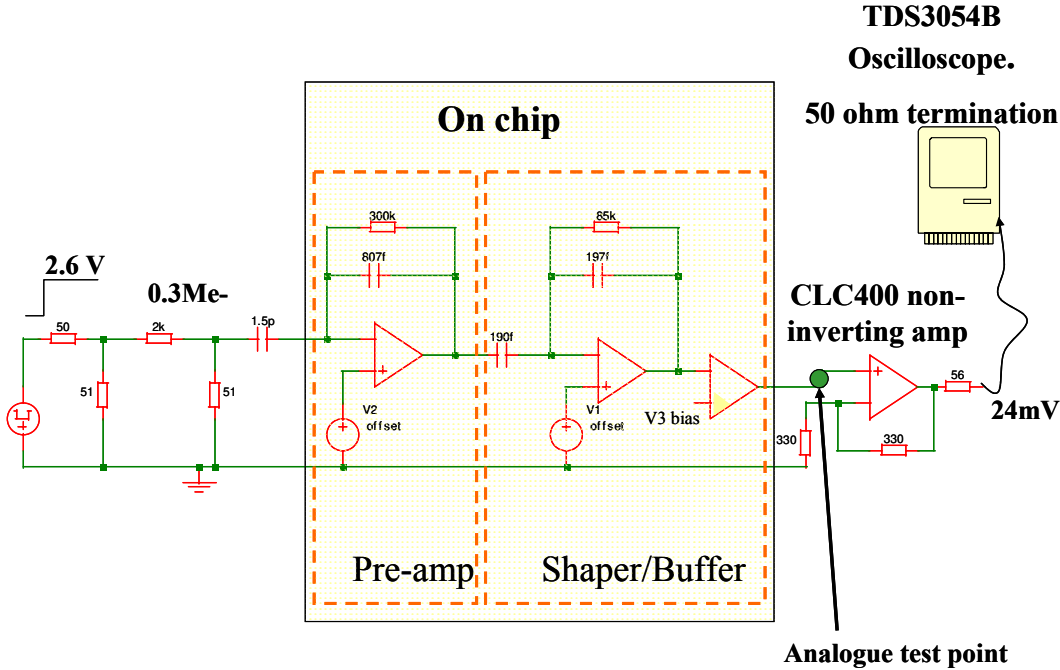
Figure 5-16 The test circuit for a single channel readout from the analogue probe points.

The standard LHC mode of operation of the Beetle1.2MA0 is to multiplex 128 input channels to four output channels. However, for these tests, the alternative readout mode of 128 input to one output channel was used for simplicity.

5.4.2   Beetle1.2MA0 results

The ASIC under test was the Beetle1.2MA0, Chip 42 from wafer D4KKGZKT. All measurements were taken with a 10 pF load at the input of the 1.2Att amplifier.

Test system authentication

To authenticate the Beetle1.2MA0 test system, the output of the amplifier was measured at the probe point for an input test-pulse compared against simulation. Figure 5-17 shows the results. There is a discrepancy between the simulated, blue trace, and measurement. This can be understood in terms of the 7 % attenuation in the readout electronics, mentioned previously,
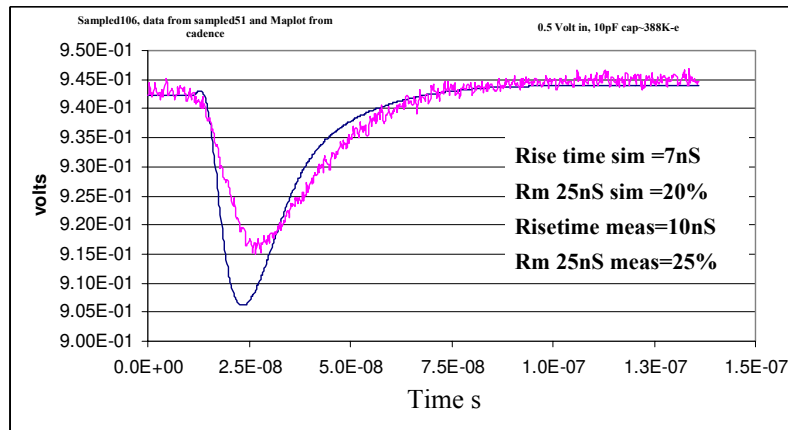
Figure 5-17 Comparison of simulated (blue trace) and measured (purple trace) analogue output response from an external test pulse injection.

and the 10 pF input load capacitance which was not accounted for in the simulation. The measured rise-time is 10 ns and the fraction of the pulse remaining 25 ns after the peak is 25 %, which is acceptable for LHCb operation. The same measured analogue output from the probe point was then compared to the signal at the output of the pipeline, shown in Figure 5-18 a). The measurement from the pipeline, blue trace, is the result of a scan over the pulse profile

127

(pulse-height scan). Here the Beetle1.2MA0 analogue sampling point was fixed, and 10,000 samples taken from the output of the pipeline, the input pulse was then repeatedly delayed by 6 ns and samples retaken. Excellent agreement is seen.

a)   b)
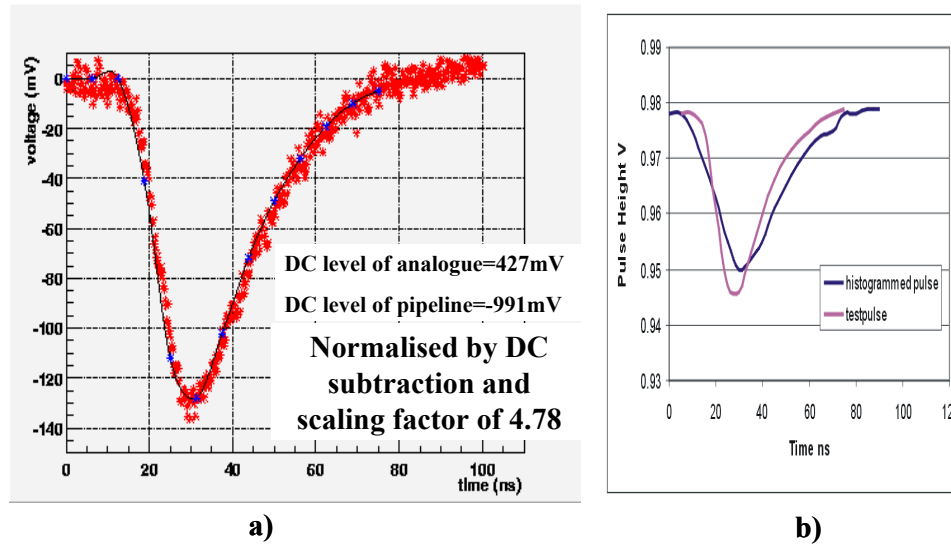
Figure 5-18 a) A pulse height scan from the pipeline, blue trace, compared to a measurement taken at the analogue probe point for an external injected pulse. b) A pulse height scan of a single photon response from and MAPMT (blue trace) compared to the measured response from a test pulse (purple trace).

An MAPMT was then connected up to the Beetle1.2MA0 input to study the response to a 'typical' photon signal. The MAPMT bias voltage was set at -800 V. An average MAPMT pulse was measured by triggering the LED light source and sampling at the pipeline output port 20,000 times at each 6 ns interval. The resulting pulse profile is compared to the test-pulse in Figure 5-18 b. The test-pulse is seen to emulate the signal from the MAPMT to a reasonable level and there can be confidence that the output of the pipeline readout truly represents the input signal.

Results from test-pulse injection

The purpose of these measurements was to characterise the amplifier with a known and stable input pulse. This also allowed an insight into the effects of coupling an MAPMT to the

input. In addition, measurements such as dynamic range can only be reliably done with a controlled source.

To cater for the considerable signal spread for a single-photon response from an MAPMT, the SNR has been specified at 40, explained in chapter 2.4. The measured signal and noise from the analogue probe points were measured separately with and without the chip system clock running, shown in Figure 5-19 a) and b) respectively. To make these measurements, an input charge of ~300 ke$^-$ was injected on every other trigger signal given to the oscilloscope, which consequently gave rise to two peaks in each plot, one corresponding to signal, the other to noise. Figure 5-19 c) shows the signal and noise from the pipeline readout port, where the system clock has to be running to invoke readout. Straightforward Gaussian fits were performed on the signal and noise peaks. The SNR is then found by dividing the signal peak by the σ of the noise signal. The SNR is 38, 19 and 22 for the three cases of analogue without and with clocks, and pipeline, respectively.
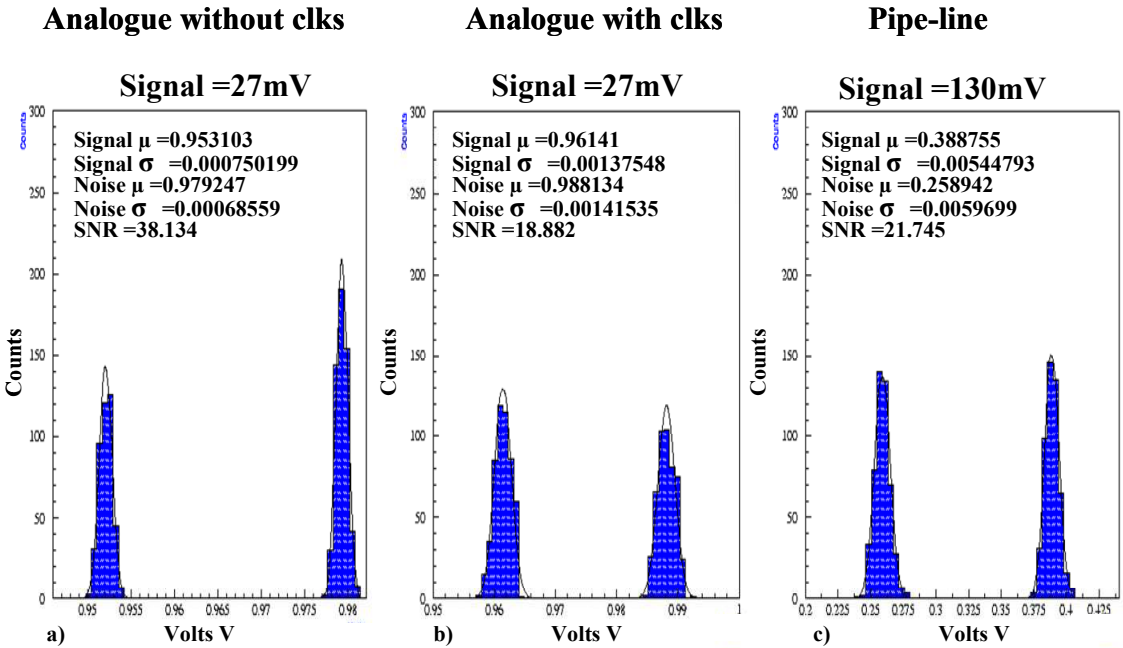
Figure 5-19 Signal and noise measurements from the analogue probe point, a) with and b) without the chip clock running and c) for the pipeline where the clocks have to be running to invoke readout.

129

From the SNR measurements it is clear that the front-end amplifier suffers considerably from cross-talk of the clock system. At the time of these measurements this was understood to be a problem arising from the system 'clock tree' of the Beetle1.2 digital logic and was under repair; the fabrication of Beetle1.3 solved this fault. As this problem is understood, allowance should be made for the rather poor SNR obtained when the clocks are running. It should be noted that reading through the pipeline does not degrade the SNR compared to that read out from the analogue probe points with the system clocks running.

With the system clocks turned off, the dynamic range of the front-end amplifier was measured at the probe point. Figure 5-20 shows the front-end voltage response to a charge injection at the input of ~0.722 to 2.7 Me$^-$ (note that a typical single-photon charge input is ~300 ke$^-$), in multiples of 0.722 Me$^-$. The inset shows the peak value from each response
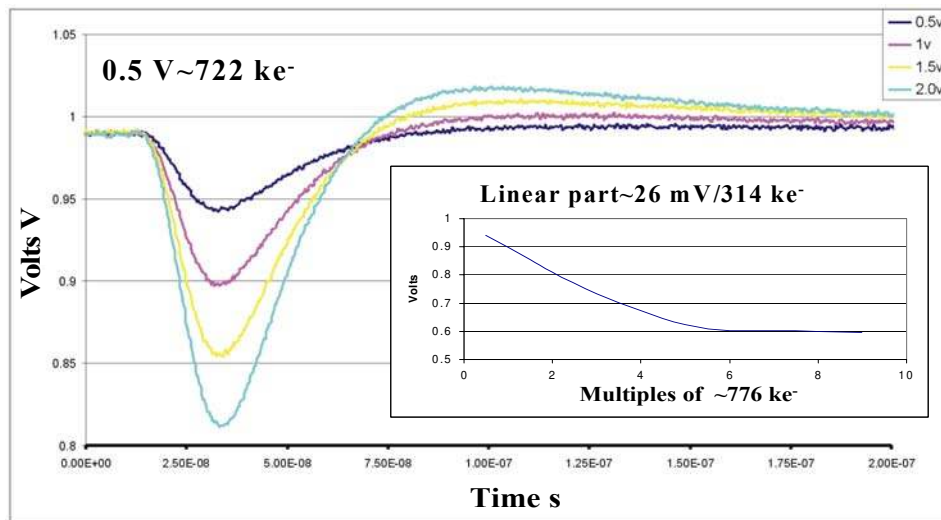
Figure 5-20 The pulse shape and linearity of the front-end amplifier at the probe point. The colours correspond to increased voltage steps across the charge injection capacitor (0.5 V=722 ke$^-$). The inset shows the linearity of the data.

plotted against charge injected. In this case the charge injected was continued up to 6.9 Me$^-$ in order to show linearity and the point of saturation; more will be discussed on saturation effects later. From the inset it can be seen that there is linearity up to an input charge of ~2.8 Me$^-$; 2.7 Me$^-$ was the specification for the dynamic range given in Table 5-1. Taking into account the 7 % readout loss, the gradient gives an output of ~26 mV per 314 ke$^-$.

The four pulse profiles in Figure 5-20 show overshoots exist that are approximately 10 % of the pulse-height, and recover to < 10 % of a single-photon pulse of 300 ke⁻ within 250 ns. The overshoot is caused by an integrator (the CSA) being coupled to a differentiator/integrator stage (the shaper). The output of the shaper responds to the slow recovery of the integrator; the following section on channel occupancy explains this in greater detail. The overshoot can be removed by tuning the front-end time constants but this is at the expense of reduced channel occupancy capabilities i.e. a longer pulse fall time. There are three mitigating factors to the overshoot problem: 1) As the maximum average channel occupancy is 10 %, or one channel hit every 250 ns, then on average the amplifier will have recovered before a second pulse arrives, 2) the average channel hit will be only 300 ke⁻, then the average loss of signal would be only 10 % of a typical single-photon response, and finally 3) any signal loss on subsequent beam crossings can be recovered offline by suitable computer algorithms, but is not a favoured solution.

Another consideration is that of saturation. To measure this, an input charge was injected from 0.776-8.5 Me⁻ in steps of 0.776 Me⁻, or in terms of photons from 2 to 28 in steps of 2, giving a maximum of ~3 times the dynamic range requirement. From Figure 5-21 it can be seen
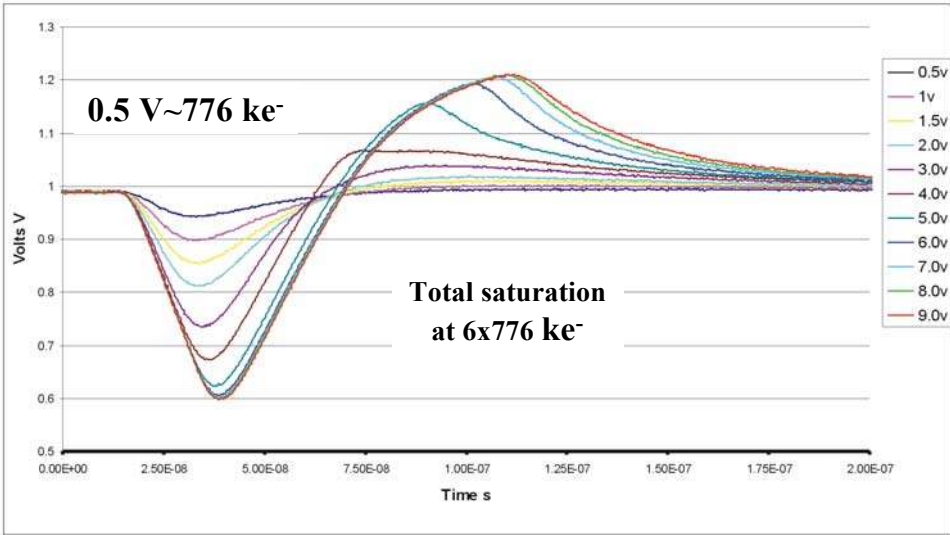
Figure 5-21 Front-end saturation effects at the probe point.

that the peak response saturates at 4.7 Me⁻ but the overshoot takes on a new shape. This is caused by saturation of the feedback resistors on the CSA and shaper amplifiers, which are implemented as FET devices. Although the overshoot goes into saturation, it does still return to zero within approximately 250 ns.

Next the output from the pipeline was studied in detail. Figure 5-22 shows all 128-amplifier channels read out from the pipeline, and clocked out in turn in 25 ns steps on one output port. Note that the readout has been inverted to comply with the DAQ system in use. A test charge of 0.776-7 Me⁻ in steps of 0.776 Me⁻ is injected at the channel 1 input; 776 ke⁻/0.5 V. The output starts with a 16 bit header that gives, among other things, the pipeline column number (the column number is the memory location within the 186 deep pipeline, i.e. the
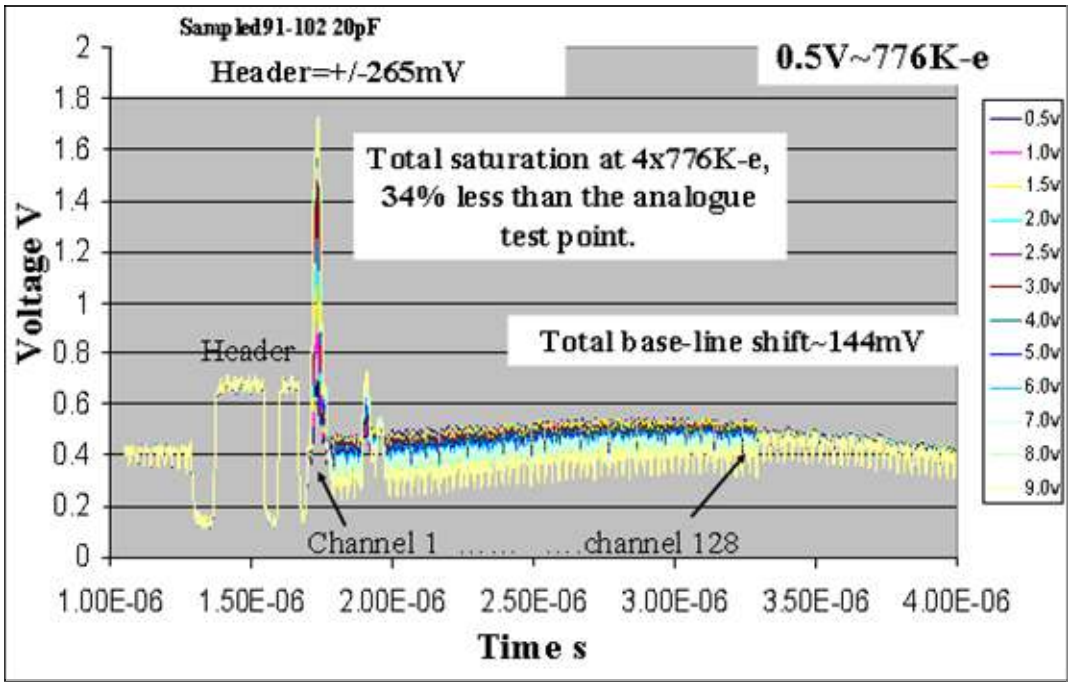
Figure 5-22 Readout of 128 channels clocked through the pipeline. The data are on channel 1.

pipeline is 128 input channels by 186 columns). There then follows the 128 channels, with channel 1 coming first. Clearly seen, from the voltage spikes on the header and data channels, are the effects of the system clock being 'on', as discussed earlier. A curvature of the base-line

can also be seen. This is due to a further feature of the Beetle1.2 architecture which has been remedied in the Beetle1.3. Data are being read out on channel '1'. Noise signals are seen on channels 8, 9 and 10, as these channels use standard Beetle1.2 amplifiers, which are much more sensitive to input noise and cross-talk.

Figure 5-23 shows a zoom of Figure 5-22 showing the channel 1 data for the full range of input test charge. The inset shows the peak value from each sample measured against charge injected. Saturation occurs at 3.1 Me⁻, which is equivalent to a 10 photon hit. The pipeline response saturates at an input signal 34 % less than from the analogue probe points; this is
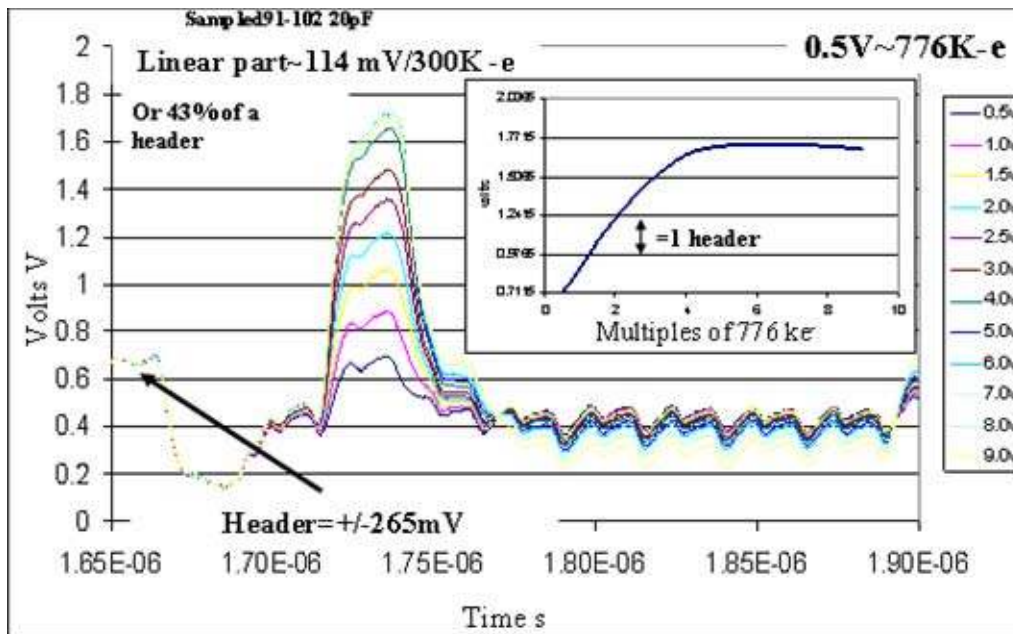
Figure 5-23 Zoom of data in channel 1 read out from the pipeline. The insert shows the linearity of the data.

attributed to the pipeline readout amplifier. Close inspection of the saturation effects on channel 1 shows that there are no detrimental effects. The linearity of the pipeline is 114 mV per 300 ke⁻, approximately 4.5 times that of the analogue probe point.

The amplitude of the header signal is used to define the output voltage range of the chip. It is fixed internally in the Beetle family of chips and therefore makes a good benchmark when comparing the output characteristics from alternative electronic systems used to read out the

Beetle chips. In this case the pipeline output has a dynamic range of 4 headers, where 1 header equates to the response from 0.7 Me$^-$ at the input.

Results from MAPMT readout

The test-pulse input circuitry of Figure 5-16 was removed and selected channels of the MAPMT were coupled to the input of the Beetlee1.2MA0, the remaining unconnected channels of the MAPMT were all grounded. Figure 5-24 shows a typical response of the MAPMT from a single-photon hit with the HT bias at –882 V, measured at the front-end amplifier probe point. The output signal has a rise-time of 12 ns and 33 % of the signal remains after 25 ns.

Pk=58 mV

Rt$_{10-90}$=12 nS

Ft25 ns=33%

Ch1 10.0mVΩ   M 20.0ns A Ch4 ∫ 1.00 V
                     ⊓→▼ 202.400ns
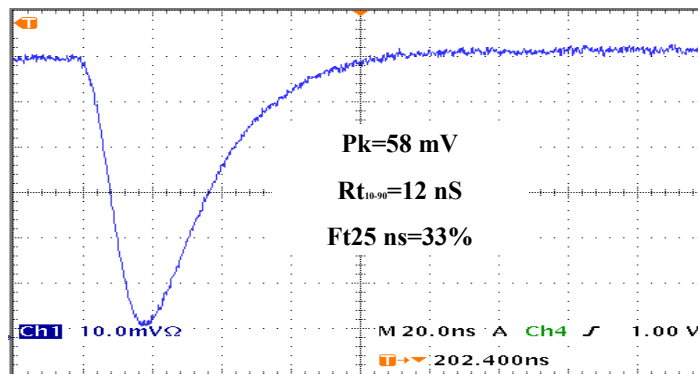
Figure 5-24 A typical response to a single photon with the MAPMT bias at –882 V. The oscilloscope time base is set to 20 ns per division and the voltage scale is 10 mV per division.

To ensure the conditions to give a single-photon response, the intensity of the LED was varied by increasing the voltage across it, as in Figure 2-3. Figure 5-25 shows the measured response from the pipeline for two LED voltage settings, a) 15 V (50,000 events) and b) 25 V (30,000). Two Gaussian distributions were fitted to the signal and noise distributions. The ratio of the number of events under the two fitted Gaussians is 6.3 % and 14 % for Figure 5-25 a) and b), respectively. These statistically low numbers of signal events gives confidence that light leaving the end of the multi-mode fibre consists mainly of single-photons. The noise cut adopted was ~ 5σ in all cases.

134

Signal μ = 0.678884
Signal σ = 0.095676
Noise μ = 0.476847
Noise σ = 0.00697458
SNR = 28.96

Signal μ = 0.680212
Signal σ = 0.102166
Noise μ = 0.476189
Noise σ = 0.00695115
SNR = 29.35

a)   b)
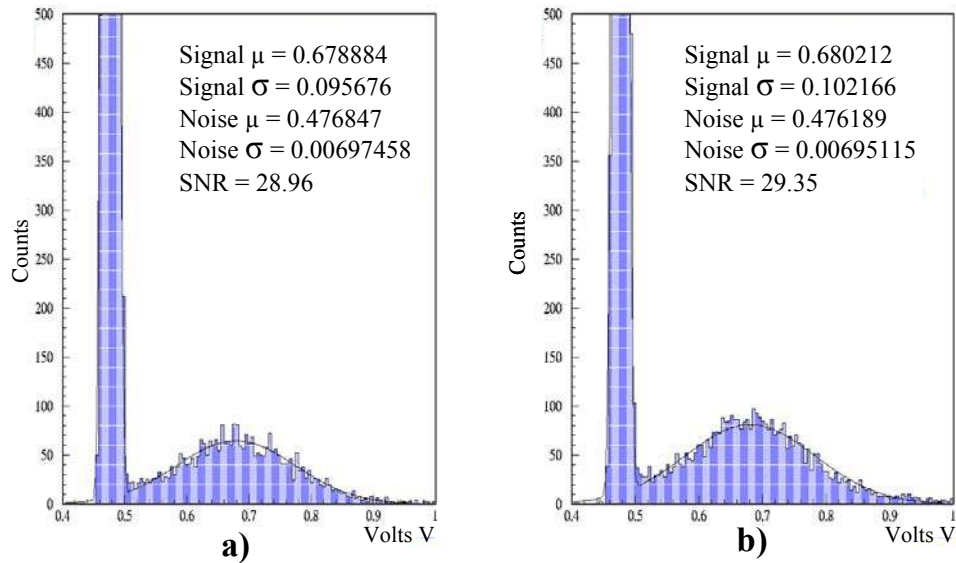
Figure 5-25 Pipeline response to an intensity-increasing light source on the MAPMT.

By adjusting the high voltage bias, the tube gain gives further freedom to adjust for a single-photon response. Figure 5-26 shows the response to bias settings of –750 to –900 Volts in steps of –50 V, a), b), c) and d), respectively. A setting of –850 V, is judged to give the best operating performance. This setting gives 142 mV per photon, therefore being compatible to the dynamic range of the Beetle1.2MA0 chip, a SNR of 21, and has low MAPMT dark counts. From measurement of single-photon response taken at the analogue probe points with the system clock on and off, the SNR can be expected to be improved by at least a factor of 1.5 when cross-talk due to the clock has been eradicated. Note that the HV setting of -900 V achieves the SNR requirement, although at the expense of reduced dynamic range.

Signal μ = 0.482971
Signal σ = 0.0217785
Noise μ = 0.444546
Noise σ = 0.00654792
SNR = 5.86

HT = -750 V

Signal μ = 0.523898
Signal σ = 0.0451165
Noise μ = 0.445295
Noise σ = 0.00652272
SNR = 12.05

HT = -800 V

Signal μ = 0.587598
Signal σ = 0.0770761
Noise μ = 0.445046
Noise σ = 0.00661191
SNR = 21.56

HT = -850 V

Signal μ = 0.694787
Signal σ = 0.111599
Noise μ = 0.445363
Noise σ = 0.00657979
SNR = 37.90

HT = -900 V
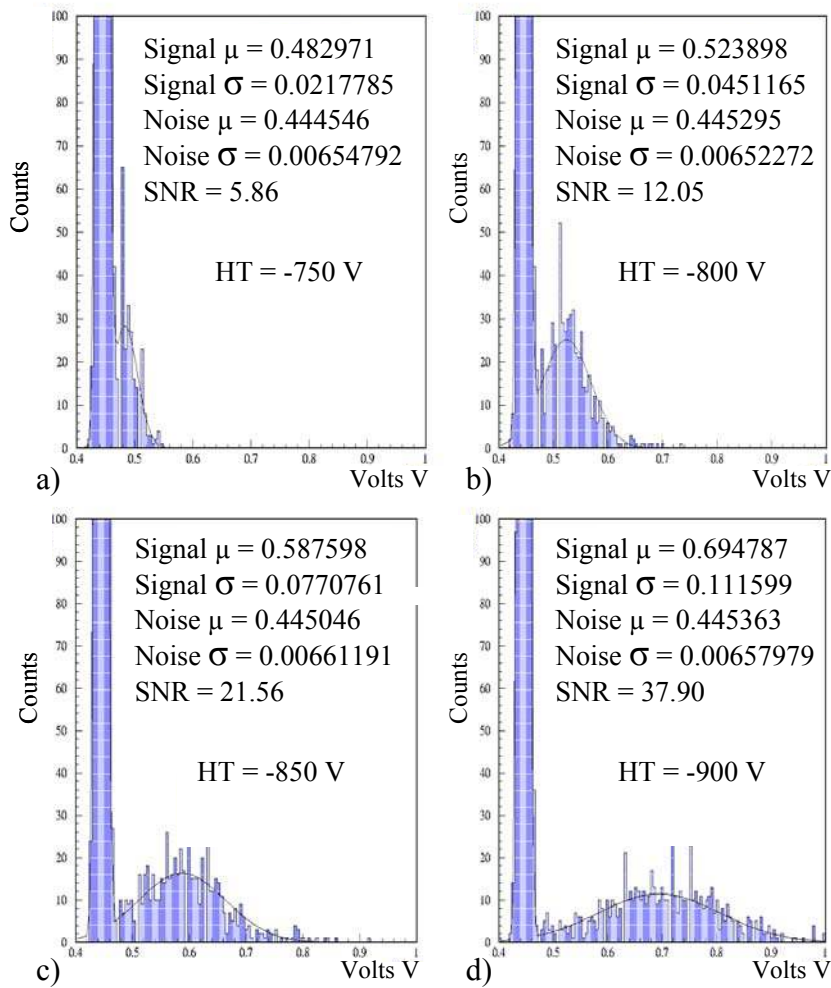
Figure 5-26 Photon response for –750 to –900 V high voltage bias settings.

Spill-over

'Spill-over' occurs when a fraction of an analogue signal pulse recorded in a given bunch crossing still remains at the next 25 ns sample time. An example of a 30 % spill over measured from the Beetle1.2MA0 is demonstrated in Figure 5-17. The amplitude of the remaining signal is sampled and stored in the pipeline and is read out on a Level_0 accept trigger (section 1.2). As there is no prior knowledge of the previous event, the sampled remainder can be construed as a valid signal; this leads to the term 'ghost hit'. The fraction of signal remaining can be tuned

between 0 and 50 % by adjustment of the CR time constants of the front-end amplifier. To remove the remainder completely, the output signal from the shaper amplifier has to be fully contained within 25 ns, this only being achievable with fast rise and fall times. An increase in rise/fall time means higher bandwidth, noise, an increase of signal overshoot and a smaller pulse plateau in which to set the sampling point. The Beetle1.2MA0 was optimised around a 30 % remainder, however ongoing studies are suggesting that this may lead to an unacceptable number of ghost hits when the chip is used in the binary mode.
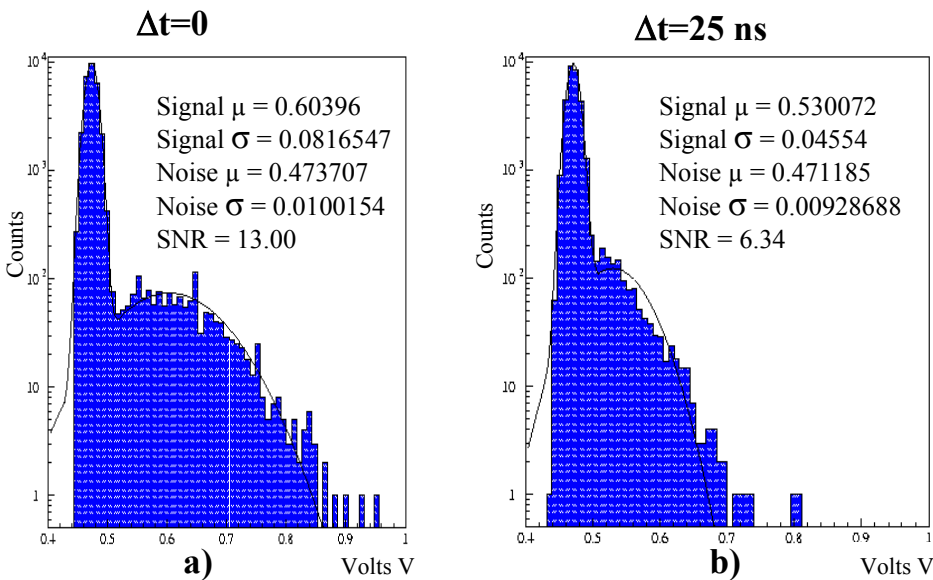
Figure 5-27 Measurement from the pipeline at time a) t=0 and b) 25 ns later.

The studies here quantify the spill-over value for the Beetle1.2MA0. The light source was synchronised to the chip sampling time so that the peak of a photon response would be captured into the pipeline. Figure 5-27 a) shows the measured response from the MAPMT for 30,000 events. Figure 5-27 b) shows the measured response captured 25 ns later into the next pipeline address column. Taking the mean signal value from Figure 5-27 b) gives ~50 % of the mean signal value of Figure 5-27 a). A 50 % remainder was chosen to give a good signal and noise separation of the spill-over measurement for first analysis. By using these measurements it was possible to determine if the plot of Figure 5-27 b) could be obtained mathematically by just scaling the pedestal-subtracted histogram of Figure 5-27 a) by 50 %. Figure 5-28 [SOM03]

137

shows the pedestal-subtraction spectrum of the peak of a photon response, which has a more typical CR time constant of ~30 % signal remaining, in blue. The red histogram is the corresponding spill-over measured 25 ns after the photon peak response. The green trace is the result of scaling the blue trace by 30 % after a pedestal cut of 3σ has been applied. It is obvious that this scaling method is an effective technique for representing the pulse spectrum of the spill-over fraction.
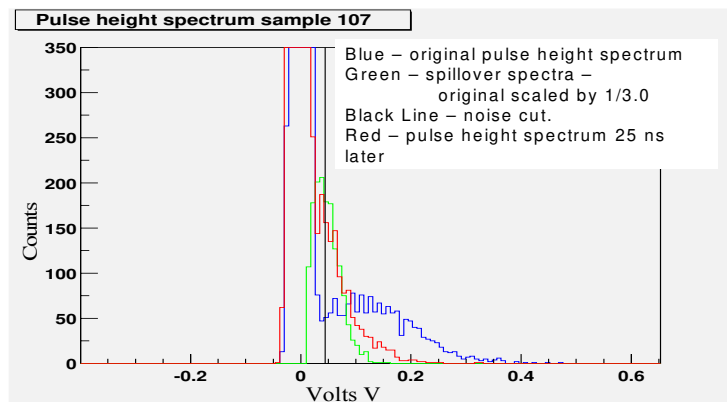
Figure 5-28 Spill-over scaling plot. The histograms are described in the

It is clear that the spill-over fraction can be reduced by simply increasing the noise threshold cut, the black line in Figure 5-28. However this will also cause a loss of genuine single-photon signal. To demonstrate this, Figure 5-29 shows the efficiency of primary signal as a function of
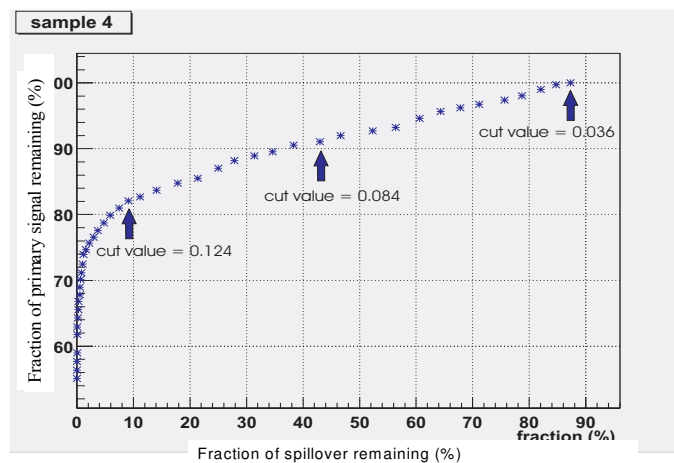
Figure 5-29 Signal efficiency verses the fraction of spill-over remaining, for different noise cuts (in Volts) **[SOM03]**.

138

spill-over detection efficiency as the noise cut is varied. To obtain this plot, the original measured pulse height spectrum and spill over data of Figure 5-28 was used. For each noise threshold position, of 0.036 V to 0.216 V in steps of 4 mV, the fraction of signal above the threshold of both primary and spill-over signal were plotted against each other. It can be seen that a threshold cut which retains only 10 % of the spill-over, also will result in a ~20 % loss of photon signal.

To investigate the effects of spill-over into later bunch (≥ 50 ns) crossings, six columns of the pipeline representing t=0 ns to t=125 ns , in steps of 25 ns, were measured as a function of the input pulse amplitude. Multiples of 300 ke$^-$ were injected at the input using the test-pulse generator, corresponding to approximately a 140 mV pulse amplitude at the output. Figure 5-30 shows the amplitude of the sampled pulse remainder from sampling times up to 125 ns, as a function of the sampled amplitude at t=0. As expected, 30 % of the sampled pulse remains in the next time bin for amplitudes $\leq$ 280 mV (two photons). This increases to $\approx$ 50 % for a 4-photon pulse, which can be understood from observation of the pulse profiles of Figure 5-20. It can be seen that the sampled amplitude fraction after two or more crossings is negligible. The
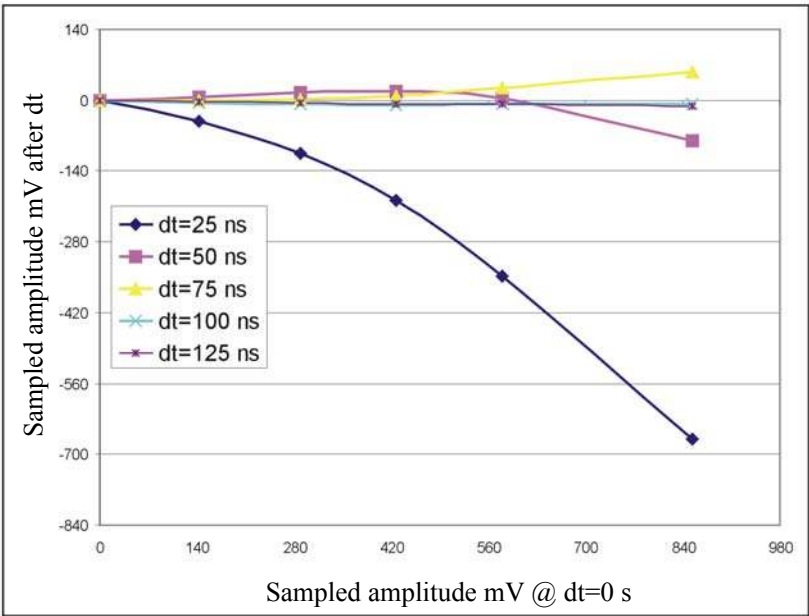
Figure 5-30 Pulse amplitude versus pulse remainder for Δt = 25-125 ns.

139

Beetle1.2MA0 does offer some external fine-tuning of this remainder, but at the expense of an increase in overshoot. Pulse overshoot is observed in Figure 5-30 where the sampled voltage goes negative for the times $\Delta t \geq 25$ ns.

The optimal CR time constant, and threshold hold setting, for operating the BeetleMA in the binary mode requires further studies. Another constraint on the optimal CR time is the effects of channel occupancy, discussed in the following section.

Channel occupancy

Channel occupancy is expected to be a maximum of 10 % for the RICH detectors, in the inner most region of RICH-1. For the electronics this means that on average there will be a ~300 ke$^-$ input signal at 250 ns intervals. To take into account fluctuations in the channel occupancy and MAPMT gain variation, the Beetle1.2MA0 is designed to accept a minimum ~900 ke$^-$ at a rate of 125 ns without being degraded by pile-up effects. Consideration must also be given to the effects of charged particles traversing the MAPMT, causing large numbers (~10) of photons to be produced in the quartz window.
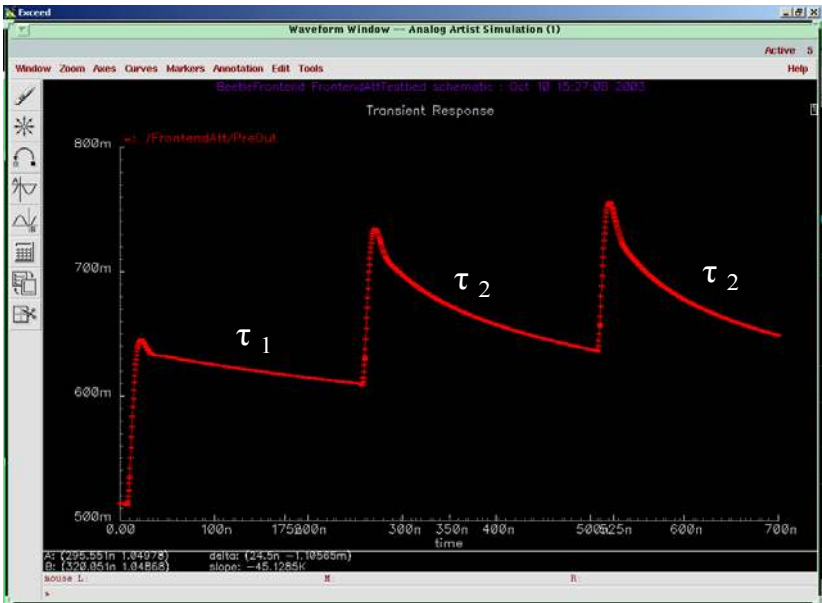
Figure 5-31 Simulated v_{OUT} (mV) of the CSA for a regularised 10 % channel occupancy.

For the Beetle1.2MA0, the occupancy is limited by the time constant of the feedback capacitor and the active feedback resistor of the CSA, C_{fb} and R_{fb} in Figure 5-7. As R_{fb} is a FET device and is accessible via the I^{2}C interface, the resistance value can be readily chosen within limits. The source connection of R_{fb} is connected to v_{OUT}, and therefore V_{GS} of the FET is dependent on v_{OUT}. As R_{fb} is a PMOS device, then the larger the small-signal amplitude v_{out} is, the more R_{fb} is turned on and the smaller the value of its resistance becomes. This gives the CSA the ability to recover quickly from a large input charge. By simulating a 10 % uniformly-spaced channel occupancy, the change in the time constant can be seen in Figure 5-31 which shows the output of the CSA, before the shaper. The first output voltage pulse has a decay time constant τ_1 followed by pulses all with a consistent decay time constant τ_2; this is seen by the difference in their gradients. In this case v_{OUT} settles down to a ~250 ns recovery rate for a 10 % occupancy.

Ideally for large channel occupancy rates, the resistance of R_{fb} would be small. The lower limit of R_{fb} is governed by the effects that the CSA recovery rate has on the shaper stage. This is because the shaper stage is a differentiator and integrator, as shown in Figure 5-10. Figure 5-32
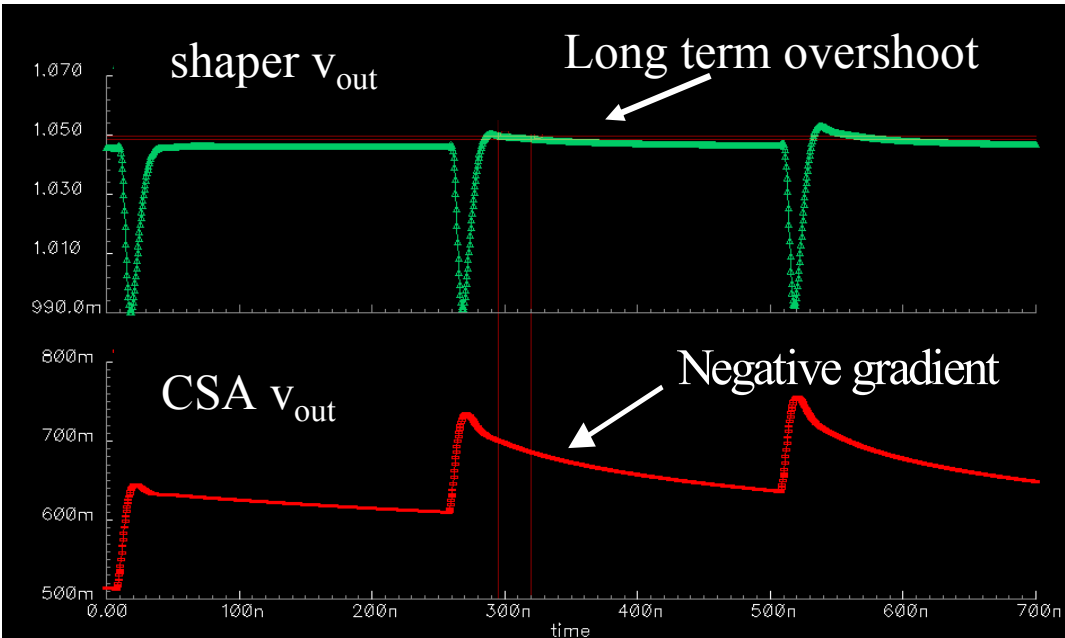
Figure 5-32 Simulation of the output voltage from the CSA and shaper.

shows, on the same time base, the CSA output from Figure 5-31 and the output from the shaper. It is observed that the shaper output has a long-term overshoot representing the negative gradient of the CSA due to the shaper differentiator. This is not accumulative in the sense of pile-up, however a baseline shift will occur depending on the occupancy and average pulse-height. The base-line will then find an average value, which in this case in a few mV greater than the shaper DC offset voltage of 1.046 V.

To study the extreme case, Figure 5-33 shows v_{OUT} at the CSA and shaper for single-photons received at 100 % occupancy. There is no degradation in pulse amplitude from the shaper, but a shift in the baseline of 30% of a single-photon response. Although the shaper output suffers from overshoot, this can, in principle, be tuned out on a per-chip basis for channels that have regular and known occupancy rates. Simulations have also been performed that show no degradation in v_{out} for 5 photon signals being received at 50 % channel occupancy. In conclusion the performance of the Beetle1.2MA0 is not degraded under high occupancy rates.
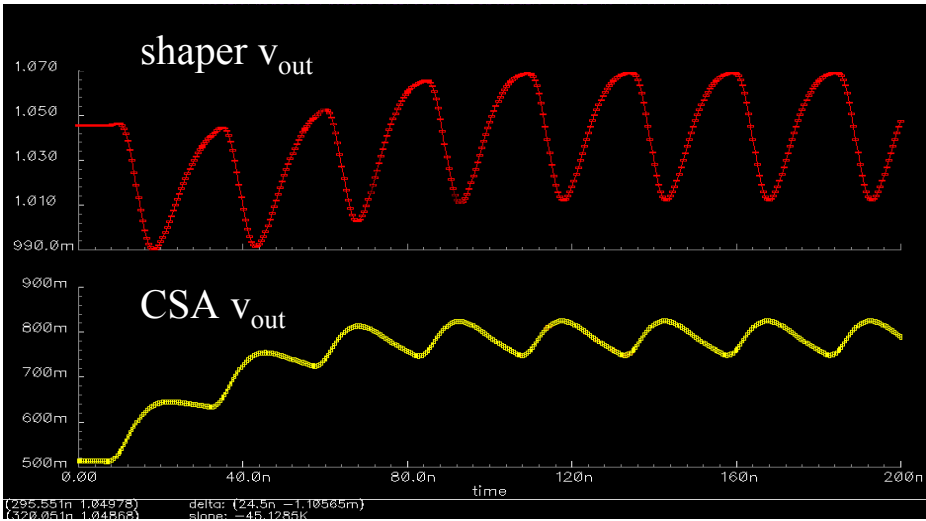
Figure 5-33 v_{out} of the pre-amp and shaper for the extreme case of 100 % occupancy.

142

5.4.3 Beetle1.2MA0 FE simulations

Secondary effects such as channel cross-talk and time jitter were studied only in simulation. The simulation test bed is shown in Figure 5-34. Although not obvious from the schematic, there are three input channels, charge can be independently injected into any one of
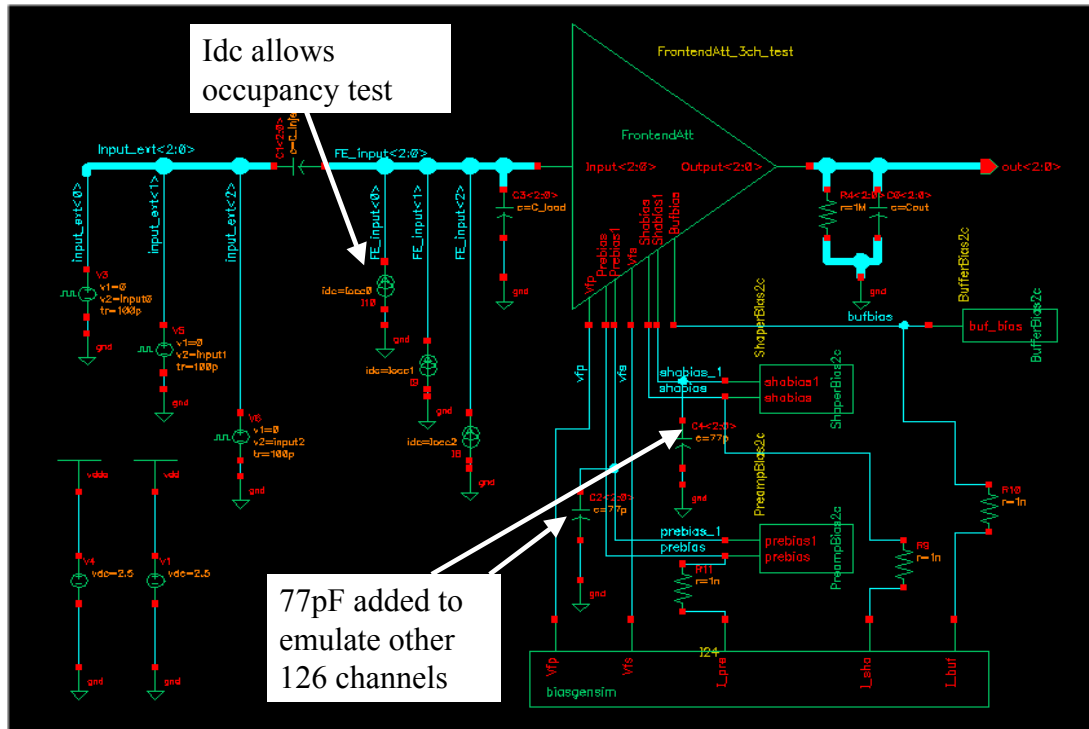
Figure 5-34 The simulation test bed.

the three. All components related to the front-end amplifier such as bias generators, current mirrors and current references have been added into the test bed. The associated parasitic capacitances, from the components and layout, were all included. To emulate the capacitance that would be present on the Beetle1.2MA0 due to the remaining 125 input channels, two 77 pF capacitors have been added in the simulation.

To evaluate the cross-talk, the signal on the centre channel (channel 2) was simulated for all 64 combinations of 0, 0.3, 1.2 and 2.7 Me⁻ on the three inputs; this represents 0, 1, 4 and 9 photons respectively. The results are shown in Figure 5-35 where the cross-talk can be seen as a
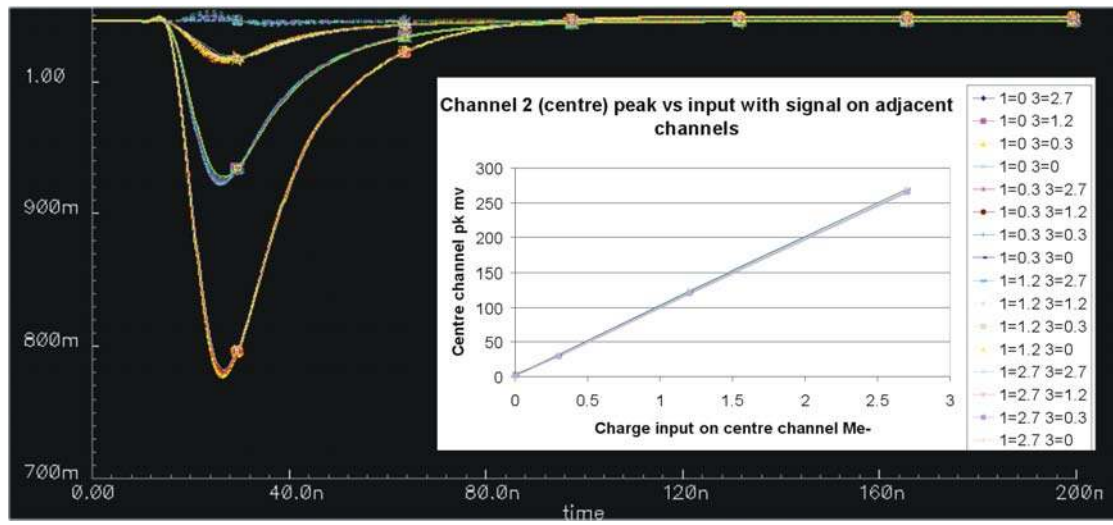
Figure 5-35 Cross talk evaluation on the centre channel (channel 2).

very slight smearing of the traces. The inset shows the measured peak signal values of channel 2, for each input combination. The curves are labelled as a function of the number of electrons injected on the adjacent channels, i.e. channel 1 and 3. The measured peak signal values are used to give the linearity of channel 2 and, although barely visible, the aberrations due to signals on the adjacent channels. As would be expected, the cross-talk distortion becomes increasingly worse the larger the injected charge on the adjacent channels becomes. The linearity is 30 mV/0.3 Me$^-$ with an RMS error of ±1.6 mV. For the worst case, when both adjacent channels have a 9 photon response, while a single-photon response is observed on channel 2, the cross-talk is 26 %. This would only happen in the case of a charged particle directly traversing the photo cathode. A more realistic case is when adjacent channels are responding to only one photon, this results in less than a 1% cross-talk seen on channel 2.

Another effect of cross-talk is jitter or time walk at the sampling point. Figure 5-36 a) shows the simulated response to one photon on channel 2 with the adjacent channels having charge injected between 0.3-2.7 Me$^-$, 16 combinations in total. The result is that the peak of the signal slightly moves in time. For a binary system, which is an option with the Beetle1.2MA0, the time walk affects where to set the threshold; the higher the threshold setting the larger the time walk. To evaluate this, the time was found at which each signal pulse in Figure 5-36 a)

144
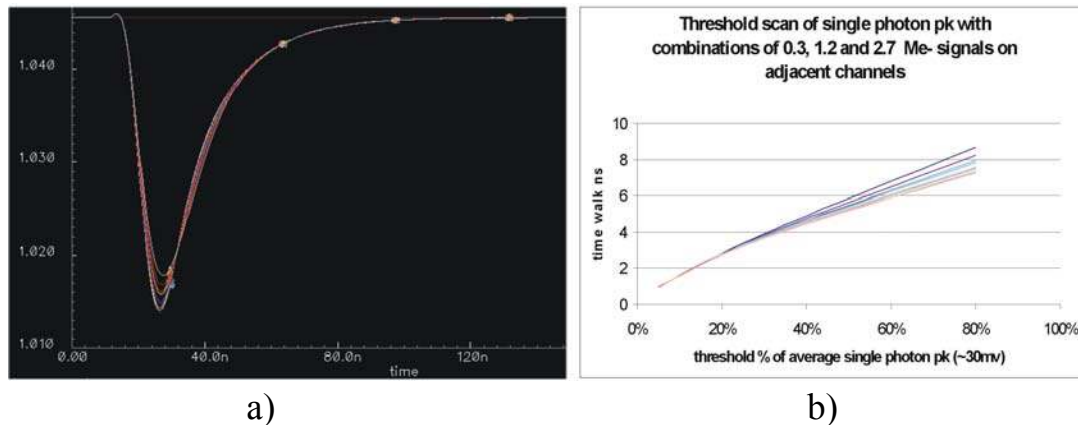
Figure 5-36 Time walk on a single photon response due to cross talk. a) The output pulse shape, b) the time walk.

crossed 5, 30 and 80 % of 30 mV, which is the amplitude for a single-photon response. Figure 5-36 b) shows the results of the time walk as a function of the crossing point, for a number each of input combinations. The curve colour scheme is the same as that used for Figure 5-35. The plot in Figure 5-36 b) shows a worse-case spread of ~2ns for a threshold setting of 80 %. Realistically the average jitter would be less than 1 ns.

5.5   Conclusions

Consideration has been given to the most appropriate front-end amplifier for capturing photon signals from an MAPMT. A study was undertaken to find the best approach for incorporating the new front-end into the architecture of the exiting Beetle ASIC by optimizing the gain performance of the amplifier. The modifications have improved the performance of the front-end amplifier in relation to SNR, a reduction in the input impedance, and a faster rise and fall time from the output.

The test-setup for evaluating the BeetleMA has been described and the measurement results given. The output has a rise-time of 10 ns and ~30 % of the signal remaining after a further 25 ns after the peak, dependent on the value of the feedback resistance. With a test-pulse to represent the MAPMT signal, the output of both the front-end and pipeline readout remained linear up to 10x300 ke$^-$. With the MAPMT coupled to the input of the BeetleMA the single-photon response was 142 mV, with a SNR of 21. Using measurements from the pipeline

145

readout the spill-over of signal from one time bin to the next was evaluated. Simulations were made of the occupancy and crosstalk effects and it was found that the BeetleMA has, in general a less than 1 % crosstalk and can operate up to an occupancy of 100 %.

From the measured and simulated results it is concluded that the BeetleMA is a suitable chip for reading out the MAPMT in an analogue mode.

The RICH demonstrator readout system

In November 2003, the final decision to adopt the Hybrid Photon Detector (HPD) as the RICH photon detector was taken. The HPD was chosen ahead of the Multi-Anode Photo-Multiplier Tube (MAPMT), which had been the alternative option. To advance the HPD development, the author was heavily involved in the commissioning and testing of the proof-of-principle HPD readout chain, and solely developed the precursor algorithms for the Level_1 region. In this regard, the demonstrator system described in this chapter was used to evaluate in the laboratory the complete readout scheme, from a single HPD readout chip to DAQ[31]. As the development of the proof-of-principle readout chain started before the final detector choice had been made, the scheme was also made compatible with reading out the MAPMT with some minor hardware modification.

In August 2003 the complete LHCb readout scheme was modified as part of a detector re-optimisation programme [TDR03opt]. This programme was necessary in order to reduce the amount of material (radiation lengths) in the detector and improve the efficiency of the Level_1 trigger; the latter improvement led to a considerable increase in the data storage requirements at Level_1. Consequently the Level_1 RICH detector electronics had to undergo a significant re-design, which changed the architecture, modularity and components to be used. Although this chapter describes the Level_1 electronics designed for the predecessor of the present system (i.e. pre-August 2003), the work presented here demonstrates the proof-of-principle on which the current system is based.

The outline of this chapter is as follows. First an overview of the full demonstrator readout chain is given, followed by an explanation of the major differences between the present and predecessor (demonstrator) systems. Next the major components of the demonstrator

---

[31] **D**ata **Ac**Quisition (DAQ). The process of acquiring data for storage. For the demonstrator system a PC was used.

system are described in detail, followed by a presentation of results and conclusions. The reader should also refer to the description of the global electronics readout scheme presented in section 1.3, as well as the system specifications.

6.1   Overview of the RICH electronics demonstrator system

The first stage of the electronics readout system is the Pixel HPD, and a description is given here for completeness. The CERN/DEP-developed HPD [TP98], shown in Figure 6-1, has active elements which comprise a photo-cathode, electrostatic imaging system and an encapsulated 62.5x500 μm^2 pixellated silicon detector. The silicon detector is bump-bonded to a 'LHCBPIX1' binary readout ASIC [LEB99], also encapsulated within the HPD vacuum envelope. The LHCBPIX1 chip OR's groups of eight pixels, resulting in a 500x500 μm^2 logical pixel size. The electrostatic imaging system of the HPD introduces a magnification factor of five, therefore giving an effective pixillation[32] of 2.5x2.5 mm^2 at the photo-cathode. Although a full-speed HPD was not available for the system tests reported here, a 40 MHz LHCBPIX1 readout chip was used to generate the input data.
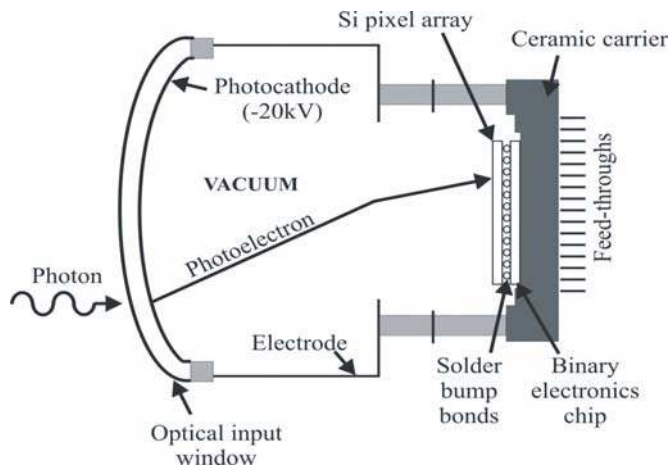
Figure 6-1 Schematic representation of the Pixel HPD.

---

[32] The effective pixel size was designed to meet the photon spatial measurement resolution required.

Figure 6-2 shows the modular conception of the demonstrator readout system. There are two basic electronics prototypes, the 'on-detector' Level_0 board and the 'off-detector' Level_1 board, connected via a 100m optical link.
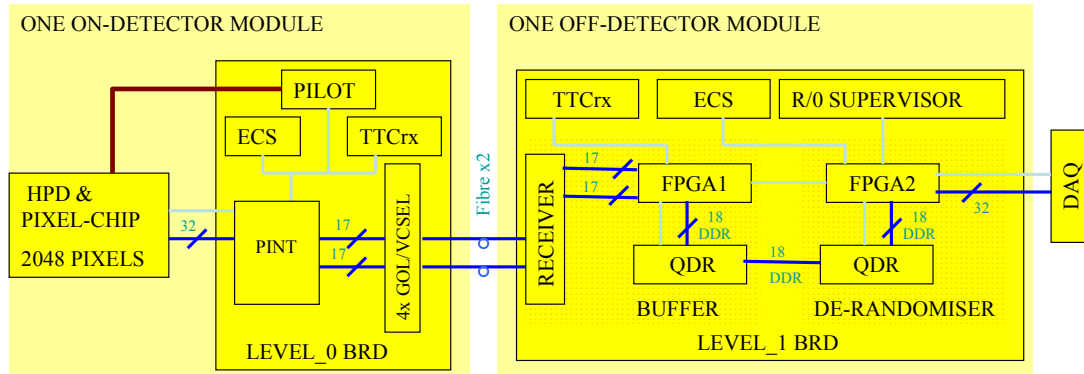
Figure 6-2 Block diagram of the demonstrator-system concept for HPD readout. The blue lines show the dataflow, and the grey lines the control.

Central to the Level_0 board is the Pixel INTerface (PINT) chip. The PINT provides the synchronisation and communication for the CERN-developed TTCrx, PILOT, and GOL chips, the latter used for optical data transmission [GOL01]. It also communicates to the Experimental Control System (ECS) using the JTAG protocol. A Spartan XLINX FPGA is used to implement the PINT logic (although an antifuse ACTEL AX1000 FPGA device will be used in the final system). The TTCrx provides clock, trigger, reset and fast control distribution; the PILOT ASIC is used for voltage-biasing of the LHCBPIX1 chip. The GOL serialises data from the PINT and drives Vertical Cavity Surface Emitting Laser (VCSEL) devices [VCSEL]. The VCSELs optically transmit at 800Mbit/s over 100 m of multimode fibre to the Level_1 board. The transmission speed from Level_0 to Level_1 is 900 ns, which is compatible with the global electronics specification (described in section 1.3.1).

The Level_1 module has two regions, shown in Figure 6-2, namely the buffer and derandomiser. The buffer region is able to receive 2 fibres from a single Level_0 board. The data are received by a HDMP-1034 optical de-serialiser component; one component services one pair of fibres. Under the control of FPGA-1, the event-blocks are checked for correct bunch crossing identification (BID) and transmission. Any subsequent errors found in the

149

event-block or in the FPGA-1 built-in-logic-observer (BILBO) are appended to the side of the event-block as an extra column; the ECS is also notified of an error occurrence via FPGA-2. The event-block, now 18x36 bits, is sent to the buffer region QDR[33] SRAM at a DDR rate of 40 MHz and stored there for the duration of the Level_1 trigger latency. The ECS interface and TTCrx are used in the same way as for the Level_0 board but further demands are made of the TTCrx with the full use of the channel-B (see the TTCrx section in 1.3.1) to receive Level_1 triggers and user commands. On a Level_1 trigger accept, the event-blocks are transferred to the de-randomiser QDR SRAMs at a DDR rate for temporary storage. Under the control of FPGA-2, data are then released from the derandomiser QDR, at a 40 MHz DDR rate, to FPGA-2 where they are formatted into 32 bit-wide words so as to be read out by the DAQ. FPGA-2 can perform different functions on the event-block such as zero-suppression, further error checking and some local error correction. All communication between the ECS and the Readout Supervisor (RS) is achieved through FPGA-2.

6.2   Consequences of detector re-optimisation

The re-optimisation of the trigger [TDRtrig] resulted in a major change to the Level_1 electronics scheme. The Level_1 trigger reject capability was improved by introducing a small magnetic field (~100 kGcm) in the region between the VELO and RICH-1. This enables a coarse momentum measurement at Level_1, but involves a longer processing time, hence increased latency. As a result, the Level_1 global electronic specifications (c.f. section 1.3.1) are now based on using QDR and DDR RAM which can have any part of their memory directly accessed in any order. This contrasts with the earlier specifications which were based on standard FIFO memory which does not have direct memory access. In addition, the Level_1 latency was increased from 2 to 52 ms. This increase has a significant effect on the RICH Level_1 electronics as the already significant buffer size of ~40 Gbits had to be increased by a factor of 26. Although the Level_1 demonstration scheme described in this chapter was designed for the FIFO memory-type specifications, the memory used was QDR with direct

---

[33] DDR (double data rate) is when data are transferred on both the rising and falling clock edges. QDR (quad data rate) is when the device can transfer data into and out of the device at the same time with both input and output operating at a DDR.

memory access. The main advantage of this memory is that the QDR can read and write to the memory within one LHCb clock period of 25 ns, thus removing any complex timing issues related to priority between read or write commands. Due to the fact that QDR memory has been used, the Level_1 demonstrator architecture can meet all the specifications of the new re-optimised scheme. However, incorporating an increased number of QDRs to meet the increased buffer size would make this design, in its detail, cost prohibitive.

In addition to the differences in the Level_1 architecture described above, the demonstrator system also differs in a number of significant aspects from the new baseline system:

- The prototype Level_0 board uses a non radiation hard Xilinx FPGA as this offers multiple re-programming capabilities. This is in contrast to the ACTEL anti-fuse device to be used in the final system which is expensive and can be programmed only once.

- The PILOT ASIC chip used for external biasing of the LHCBPIX1 chip was not available, so commercial off the shelf (COT) DACs have been used.

- The physical size of the hardware is considerably larger for testing purposes.

- The optical transmission is Glink rather than Gigabit Ethanet which will be used in the final system. This means the speed of optical transmission in the demonstrator system is 800Mbits/s as opposed to 1.6Gbits/s.

- The Level_1 receives only two fibres. It will receive 48 fibres in the final system.

- The FPGA multiplexes data into the QDR at a DDR of 40 MHz. This allows easy use of a four-burst QDR device (see later for the description) when connected to only one HPD.

- The Level_1 to DAQ-PC uses a readily available and simple CERN-developed optical Gbit Link which is based on an S-Link specification, also a CERN development [BRA00].

- The ECS interface protocol is not adopted. Instead a PC and a PM3705 JTAG interface from JTAG Technologies is used.

6.3 The readout system demonstrator

In this section the components of the demonstrator system are described in detail. Figure 6-3 shows the demonstrator readout scheme and data flow for a single HPD. The Level_0 incorporates the PINT, HPD, GOLs and VCSELs, and the Level_1 incorporates the fibre-optic receivers (HDMP-1034), FPGAs, QDRs and S-link. Both hardware regions use the TTCrx for clock and trigger distribution. For the demonstrator system the TTCrx is controlled locally with the CERN interface modules TTCvi and TTCvx [TTCstat] and a PC, these are not described within this thesis.
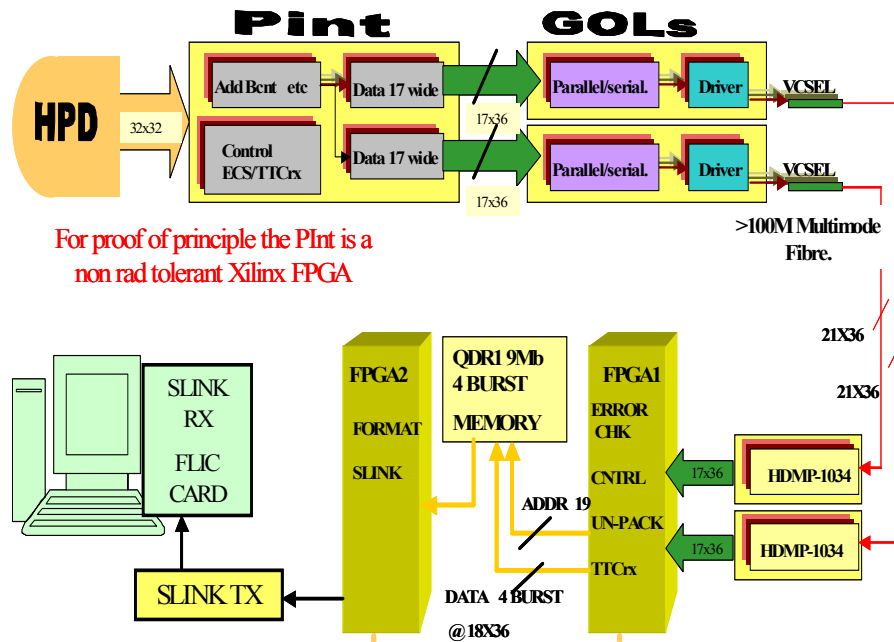
Figure 6-3 Demonstrator readout scheme.

152

6.3.1 Level_0

Figure 6-4 shows the demonstrator Level_0 board. The board size is 28x14 cm^2 and along the front edge are two pin connectors to receive data from a single HPD. The TTCrx plugs into the back of the board, not shown here. Components a) are the power-supply regulators and associated circuitry, b) the GOLs and VSCELs, c) the DACs for LHCBPIX1 biasing and d) is the Xilinx FPGA. These units, together with the data transmission protocol, are now explained.
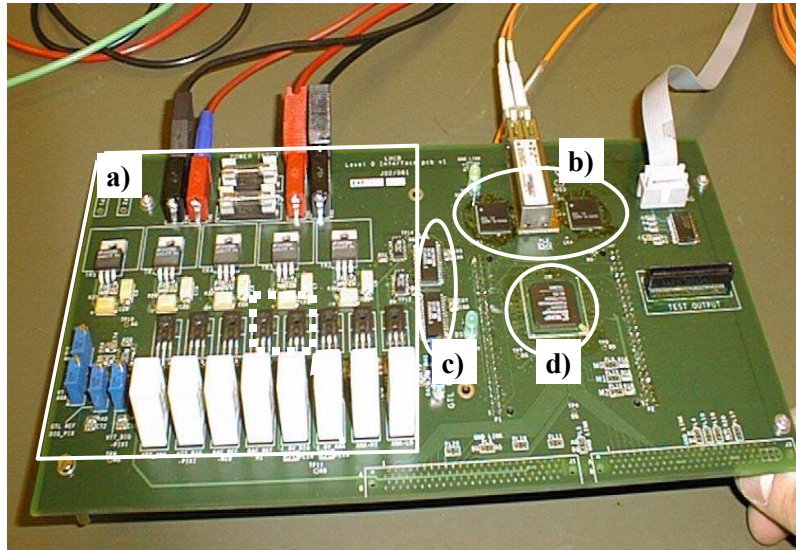
Figure 6-4 The Level_0 demonstrator board.

The PINT Xilinx FPGA chip

Figure 6-5 shows the block diagram of the PINT and its supporting blocks. The PINT translates all incoming signals to the LHCBPIX1 chip I/O standard of GTL and outgoing signals to CMOS by internal I/O configuration banks. It generates all of the required LHCBPIX1 test-pulse signals. Different PINT operation modes can be selected through the JTAG interface. The two external DACs needed for setting the LHCBPIX1 bias voltages and currents are configured from the PINT and its JTAG interface. In order to use the JTAG
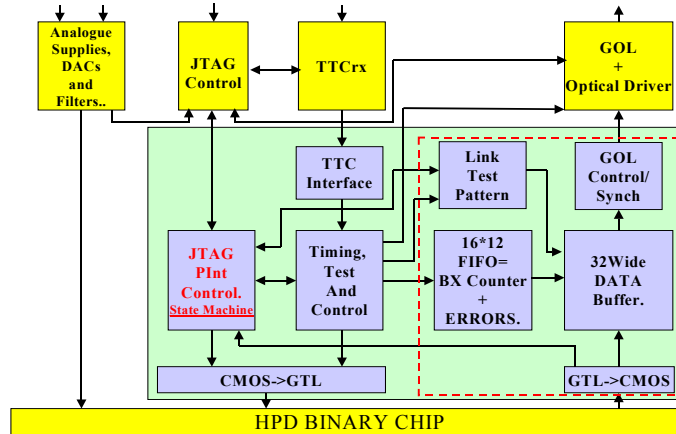
Figure 6-5 Functional block diagram of the PINT.

protocol within the PINT, a JTAG test access port (TAP) has been written in VHDL[34] and imbedded within the core logic. The TAP controller is a 16-state finite state machine (FSM) that responds to the control sequences supplied from the PM3705. The TTCrx bunch-crossing clock of 40 MHz is used to synchronise the PINT with the LHCBPIX1 and GOL chips.

As mentioned previously, for the final system, the PINT algorithms will be ported into an anti-fuse radiation-tolerant AX 1000 FPGA 896 BGA[35] from the ACTEL Axcelerator family. This FPGA uses a 0.15 μm process, has seven-layers of metal, ~1M system gates and ~4 thousand flip-flops. As the AX 1000 is an anti-fuse device, the registers to hold the configuration are themselves configured by electrically building an internal link between the gate of a flip-flop and a control voltage. Once configured the device is not re-programmable. These anti-fuse nodes are not susceptible to the radiation levels found in the LHCb RICH detectors. The flip-flops used for the core logic must have different voltage stimuli applied to their gates and are therefore not anti-fuse devices. To protect the core logic against SEU, triple redundant logic is used, described in section 3.4.2. In the demonstrator, a Spartan XLINX FPGA replaces the ACTEL, for the reasons already discussed, but the triple logic philosophy is fully adopted.

[34] **V**ery high speed integrated circuit **H**ardware **D**escription **L**anguage.

[35] Ball grid array package type.

154

Data format and transmission protocol

Data arriving from the LHCBPIX1 chip are of binary format i.e. a binary '1' for a hit pixel. On a Level_0 trigger accept, 32x32 pixels are read out into the PINT chip where the data are formatted into event-blocks. The size and format of these event-blocks are determined by the GOL protocol, discussed in the next subsection. Figure 6-6 shows a representation of the event-block transmitted down a single fibre after having being formatted by the PINT chip. A discussion of the format follows.
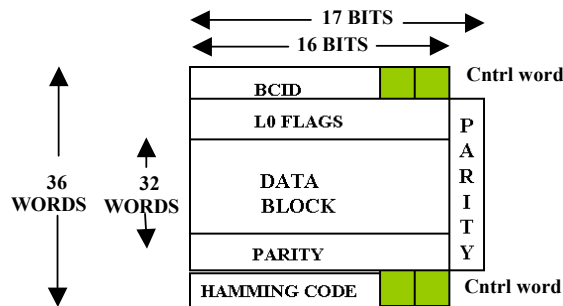
Figure 6-6 Event-block for one fibre after PINT formatting.

The PINT splits the 32-column-wide incoming data block from the LHCBPIX1 chip into two blocks of 16x32 bits per HPD. Each 16-bit wide data block then has transmission information added, bringing the block size to 17x36, and this is the event-block in this context. In the Glink protocol, the first and last words of the event are control (Cntrl) words. The data width within a Cntrl word is restricted to 14 bits. The event block is constructed as follows:

- A header is added as the 1st event-block word (first Cntrl word), which comprises the bunch crossing identification (BID).

- A following event word is added that contains any error conditions that the PINT chip may have identified (L0 FLAGS). These are either internal PINT logic errors, or errors due to the event data from the LHCBPIX1 having an unexpected format.

155

- A block has each row parity-checked and the result is added as a 17th row parity bit.

- A column parity check is added as a 35th event-block word.

- A Hamming-code is added as a 36th event-block word (last Cntl word).

This brings the total data transmission from PINT to GOLs for every Level_0 trigger accept to 17x36x2 bits per HPD (17 bit wide word, 36 words deep data event and 2 channels for each HPD).

Ideally each Cntl word that is transmitted from the Level_0 should have within it, a data word that is predictable and recognisable by the Level_1 logic for synchronicity checks. The 12-bit BID from channel A of the TTCrx[36] is common to both Level_0 and Level_1 regions and therefore is a good option. By checking that the BID arriving at the Level_1 is a correct control word, then a misalignment of event-block data, or a loss of clock, or a wrong event-block sent, can all be detected and action taken. On receiving the BID, the Level_1 logic checks that no event words are missing between the start and end control words by incrementing a counter and checking that 36 words have been received.

The GOL and HDMP-1034

The GOL [GOL01] chip is a radiation hard, SEU tolerant, multi-protocol high-speed serial transmitter developed at CERN. One such protocol is suitable for use with the Hewlett Packard HDMP-1034 de-serialiser receiver chip.

The GOL transmitting and HDMP-1034 receiving devices are used in the '17-bit Glink' protocol mode, the 17th bit, known as a flag bit, being used as a parity flag. The event-block is transferred serially, a row at a time, plus transmission protocol. The transmission rate is 20 bits in 25 ns (800 Mbits/s), 16 of which are data and the remaining 4 are overhead bits for encoding.

---

[36] The BCID is provided on every Level_0 trigger accept.

In hardware the GOL and HDMP-1034 have various configuration pins; 16 user data pins, a flag pin, two pins for marking words as either control, data or idle words and a force_ff0 pin. With the force_ff0 pin set to '1', idle words consisting of ff0 are sent in periods when no event-blocks are sent so that the transmitting and receiving devices can remain locked to each other (the clock for the receiver is recovered from transmission data).

Optical fibre transmission

The serialised data from the GOLs are converted to optical signals using VCSEL devices. VCSELs emit light perpendicularly to their p-n junctions. High output luminosity, focusing, and large spectral width allows for easy coupling to multimode fibres. The wavelengths of these devices are generally 650 nm, 850 nm or 1300 nm [HP] and the output power is typically 5 mW. VCSEL arrays can be easily incorporated into single ICs, which allow for a much more compact multiple fibre package. The VCSELs have been proven to be very robust in terms of radiation and magnetic-field tolerance. At the receiving end the optical signal is converted back to an electrical signal using the Stratos M2R-25-4-1-TL pin-diode receiver and amplifier packages. Two GOLs and one VCSEL package are used for transmission over two multi-mode fibres.

Transmission errors

Example of transmission errors can be a missing clock, an inversion of a data bit, or a loss of event-block synchronisation. These errors can be detected and a correction mechanism used for repair. The detection and repair methods used will depend on the number of errors that need to be corrected per second, the time taken to detect and repair, and the overhead in logic density to implement such a mechanism. One method is to perform a reset of all the electronics, but this should be kept to a minimum of occurrences, as several hundreds of thousands of events will be lost during the reset period. Another option, which is used in this case, is to make full use of the allowable readout time from Level_0 to Level_1 i.e. 900 ns per event [CHR01_L0]. The data block from the LHCBPIX1 chip is 32 words deep and is transferred off the LHCBPIX1 chip in 800 ns. This allows four extra words that contain error-checking information to be added to the data block in the available 100 ns (i.e. 900-800 ns). These words can be used to implement error

detection and correction codes that can then be used for diagnostic purposes further downstream.

Each data block has its row and columns evaluated with a parity and Hamming scheme. The results of both are transmitted, along with the data block, to the Level_1 receiver. The parity and Hamming-codes of a data block received at the Level_1 are evaluated with the same two schemes, and the results compared to those sent with the data block. A discrepancy between the two results indicates a transmission error. The parity and Hamming schemes are now described.

For parity, the bits of a row or column are taken through a tree of eXclusive-OR (XOR) gates so as to generate a parity bit '1' only if there are an odd number of ones in the input. By taking the XOR of both row and column, a parity grid is generated. This parity grid is compared at the receiving end where it is translated into an error grid, thus allowing the correction of single errors. Double errors are identifiable but not necessarily correctable; this depends on the positions of errors within a block. Triple errors are not recognisable or correctable. The parity generation for columns requires storage of a running parity over 32 clock cycles, which can also suffer from SEUs. An error in the parity flags results in false error detection, which is impossible to distinguish or correct. Studies have shown that a Hamming-code is more robust in terms of reducing these false errors, whilst not so good at being able to correct real errors, while the opposite is true for parity checking. Therefore the use of both Hamming and Parity is used.

The Hamming-code uses a given polynomial of the Nth power, N+1 bits, by which the data stream is modulo-2 divided. There is no carry operation between places, i.e. each place is computed separately, which makes the division simpler. The quotient is disregarded and the remainder of N bits is appended to the data stream. When the data with the appended remainder are divided by the same polynomial, the remainder should be zero if no errors occurred. The Hamming-code can be applied to the full event-block, as it is the last word to be appended. One suitable polynomial for the Level_0 to Level_1 transmission is $x^{11}+x^{10}+x^4+x^3+x+1$, (110000011011), which gives an 11-bit remainder [DAM02]. Hardware

158

for Hamming coding is simple and compact. For full checking of an event-block, an 11-bit shift register and five 'OR' gates are required. The PINT and FPGA-1 of Level_1 both use exactly the same Hamming-code algorithms to ensure compatibility.

6.3.2   Level_1

In the LHCb experiment, the Level_1 electronics will be situated in the counting room in a non-radiation and non-magnetic field region, about 100 m away from the Level_0 area. The counting room can be considered as an electronically 'friendly' environment and therefore standard COTs components can be used. This has the advantage of availability, ease of maintenance, and cost effectiveness. There are a broad range of products, and the use of SRAM FPGA devices is allowed. Error checking, error correction and self-test algorithms are built into the Level_1 electronics to ensure that synchronisation is not lost and corrupt data are not being transmitted either to, or from, the Level_1 region.
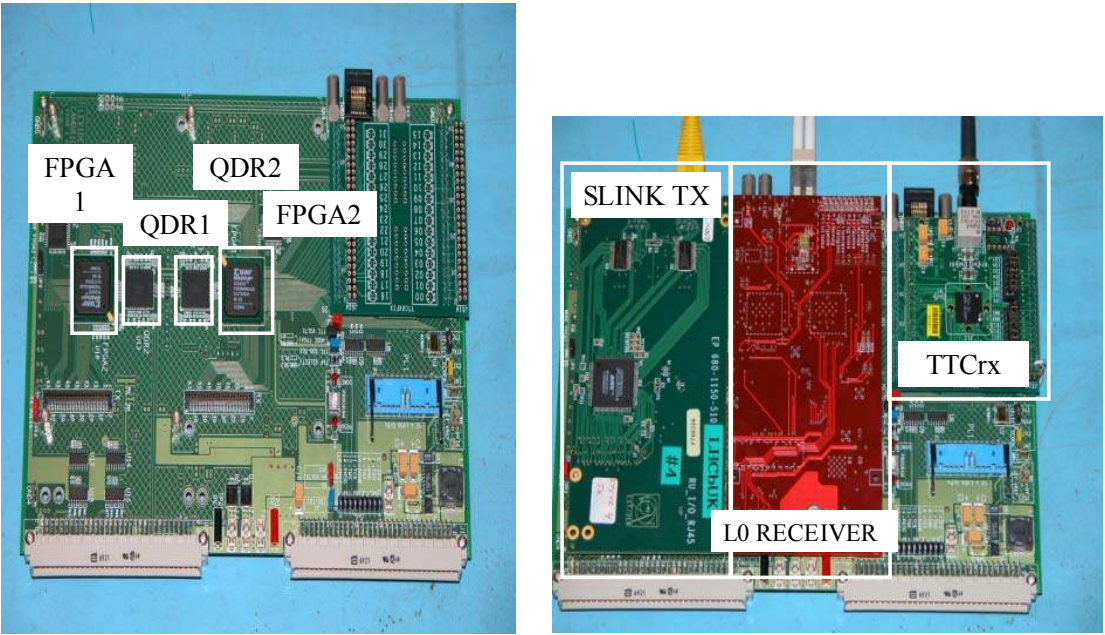
a)   b)

Figure 6-7 Level_1 board a) not loaded with interface cards, b) loaded.

Figure 6-7 a) shows the demonstrator Level_1 board. The board size is 23x17 cm^2. The board can connect into a VME crate for power or use external supplies. The FPGAs are close to the QDR memories to optimise speed performance. Figure 6-7 b) shows the Level_0 fibre-optic receiver, TTCrx and S-Link interface boards that plug onto the face of the Level_1 board. A detailed description of the Level_1 and plug-in boards is now described. The TTCrx plug-in board has already been described in section 1.3.1.

Plug-in interface boards

The Level_1 receiver accepts two multimode optical fibres into a M2R-25-4-1-TL device, which contains two pin-diodes and amplifiers. The data are de-serialised using two Hewlett Packard HDMP1034 devices, one per channel. Figure 6-8 shows a block diagram of a single data channel. The synchronisation between the Level_1 receiver board and the Level_1 mother board is maintained by the signals RxReady, RxCntl and RxData. The timing relationships are
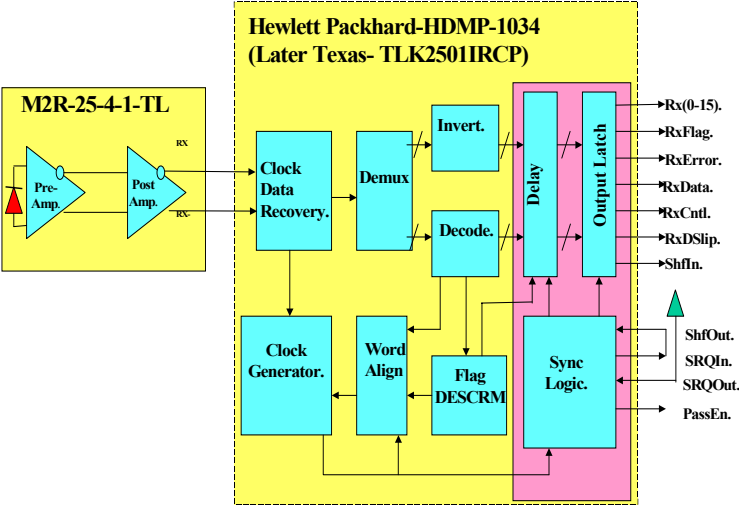
Figure 6-8 Block diagram of the Level_1 receiver board.

shown in Figure 6-9. RxReady is a flag from the de-serialiser and gives a logic '1' when the device has tuned and phase-locked into the clock recovered from transmission data. The

160

RxData flag is a logic '1' for event data transmission or a logic '0' for idle-word data transmission. RxCntl is the flag used to mark control words in the event data block.
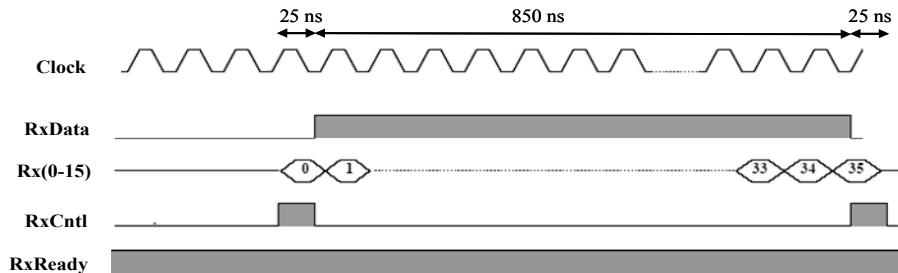
Figure 6-9 Timing relationship between RxData, RxCntl and RxReady.

The Slink TX card, seen in Figure 6-7b), is part of a readily available CERN-developed optical Gbit Link, which is based on an S-Link specification. The S-Link is a CERN specification for an easy-to-use FIFO-like data link that connects front-end to readout at any stage in a data-flow environment [BRA00]. It is used here to send the event-blocks over an optical fibre to a CERN-developed PCI receiver card named 'FLIC' where the data can then be written to disk on a PC. The readout speed of this system is a maximum of 20 MHz. FPGA-2 formats and controls the data-flow for the S-link card.

QDR SRAM

The MT54V512H18EF QDR SRAMs were provided by Micron. A single package is a 13x15 mm^2 BGA with a 1 mm pitch. This device has a total memory bank of 9 Mbits given in blocks of 128k words and can therefore store up to ~7k events from one HPD pending the Level_1 trigger decision. The QDR architecture is shown in Figure 6-10. The key features and maximum operating rates of the device are as follows:

- 9-Mbit Quad Data Rate Static RAM (upgradeable to 64-Mb).

- Manufacturers: Cypress, IDT, Micron and NEC.

- Separate independent read and write data ports support concurrent operations.

161

- 4-word burst for reducing addressing frequency.

- 167 MHz clock frequency (333 MHz data rate).

- Upgradeable clock frequency to 250 MHz (500 MHz data rate).
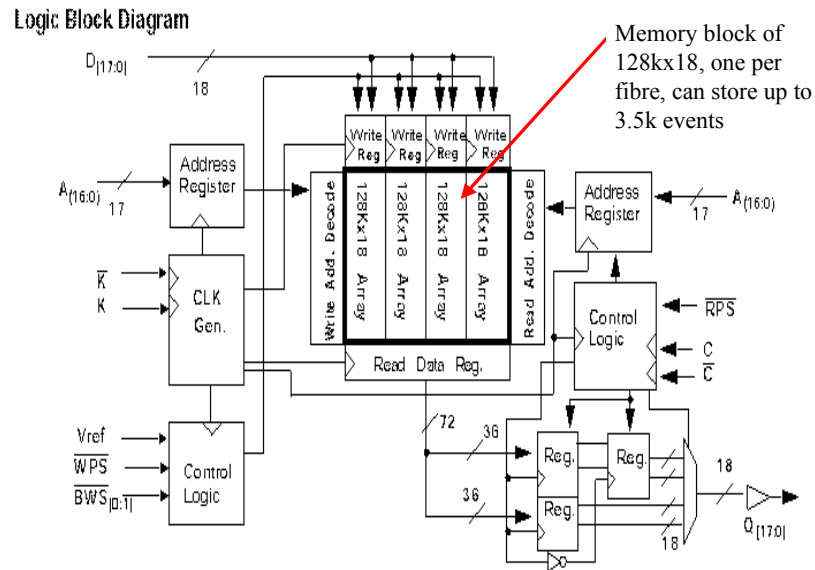
Figure 6-10 Block diagram of QDR SRAM.

Data can be read in and read out of the QDR SRAM on the same clock edge in bursts of four data words. The four-burst devices require only one write address to be generated to store four 18-bit words. The same is also true for reading from memory i.e. one read address releases a burst of four data words. This therefore means that for the demonstrator scheme, addresses are generated at a rate of 20 MHz while the data are being transferred at 80 MHz i.e. 40 MHz DDR. This allows a data word from each of the two receivers to be stored in 25 ns.

Figure 6-11 shows an example timing diagram for storing and retrieving four data words to the arbitrary memory location of address 3. On the rising edge of the k-clock, the memory address on the address bus (SA) and the write pointer signal ($\overline{wps}$), are latched into the chip. On the following rising edge of the k-clock the first of the received data words are clocked into the memory location. The three remaining words are written to memory on the following 3 k-

162
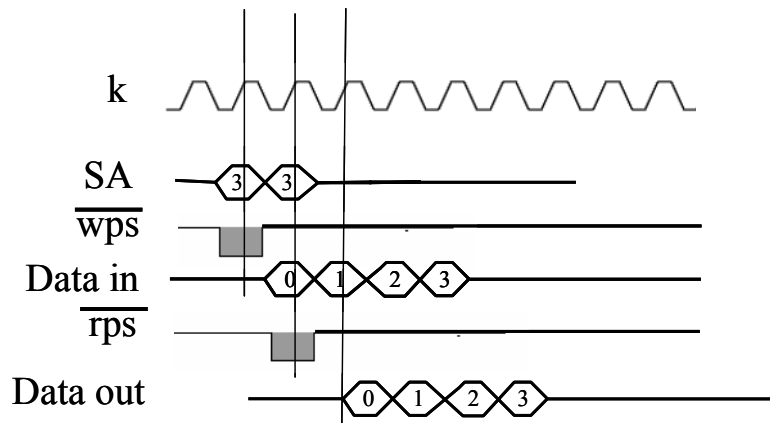
Figure 6-11 Timing diagram for writing and reading data to memory address 3 of the QDR. Read and write occur simultaneously.

clock edges. For a read sequence, the memory location to be read from, again on the SA bus, and the read control signal $\overline{rps}$ are latched into the chip on the rising edge of the k-clock. The four words in memory are then clocked out on each of the k-clock edges starting with the rising edge.

As the data from the receivers are in 17-bit wide words and the QDR accepts 18-bit words, the remaining 1x36 bits of the memory are used for error flagging and data validation in the Level_1 buffer stage. The 1x36 bit error word is appended onto the side of an event.

6.3.3   Level_1 buffer

FPGA-1 is a low-cost Xilinx Spartan-II XC2S200 FG456 FPGA and is used as the buffer controller chip. The device offers 284 input/output pins (I/Os) and access frequencies of 200 MHz, internal clock speeds of 333 MHz, 1176 control logic blocks[37] (CLB) and 5292 logic cells. The chip has four imbedded Delay Lock Loops (DLL) that can be used for clock management, for example clock multiplication, division, and phase shifting. The Spartan-II is programmable directly by JTAG or PROM.

---

[37] CLBs are imbedded blocks containing a number of logic functions such as AND and OR, along with multiplexers, shift registers and flip-flops.

In order to take full advantage of the QDR four-burst memory access, four input channels will be multiplexed into one QDR, instead of two channels as is the case for the demonstrator. To multiplex four input channels requires the QDR write and read speed to be 160 MHz i.e. 80 MHz DDR, compared to 40 MHz DDR in the case of the demonstrator. For this reason the FPGA-1 algorithm and QDR interface were designed to operate in two modes: the four-channel and demonstration mode. The demonstration mode simply selects the internal system clock to run at half speed of that of the four-channel mode and disables two input channels; both modes of operation have been tested in hardware. Below describes the four-channel 80 MHz DDR mode of operation.

Figure 6-12 shows the general architectural scheme with the support blocks for the FPGA-1 chip. Four fibre-optic input channels are multiplexed to one QDR. All four of the available DLLs are used in a cascaded way within the clock block. The first DLL is used to distribute the 40 MHz input clock within the chip. The output of this DLL is the input to the next DLL in the chain, the output of which is a doubling of the 40 MHz clock. This 80 MHz
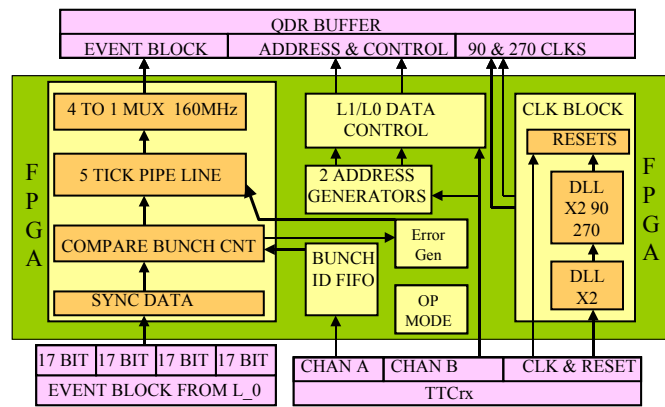
Figure 6-12 Block diagram for the FPGA-1 algorithm and general interface scheme.

clock is required for the QDR k-clocks. A further DLL is used to introduce phase shifts to the 80 MHz clocks of 90° and 270°. The 90° phase shift is used for clocking the event data out to the QDR while the 270° shift is for clocking the $\overline{wps}$ and write pointer signal ($\overline{rps}$) off chip. The final DLL is used to double the 80 MHz clock for the 4:1 multiplexer stage.

The timing for moving data blocks through the Spartan chip is critical. For example, a data block cannot always be routed directly from the chip input pins to the chip output pins. Instead it is routed through internal CLBs and flip-flops that can add delays of up to several nano-seconds. In this case each bit of the data-block can suffer different time delays and therefore cause a misalignment. Two ways in which to keep the data aligned and prevent metastable[38] states is to use the flip-flops that are available within the Spartan I/O blocks and to use pipelining techniques. These two methods are now described.

- Each of the I/O pins around the Spartan chip have some locally built in logic including a flip-flop. By assigning, within the VHDL code, a clocked one-bit register to each of the incoming, or outgoing data-bus lines, the I/O flip-flops can be forced into use. These input registers are shown as the 'sync data' block in Figure 6-12, the output registers and I/O pads are not shown. This has the effect of re-synchronising input data to the internal clock regardless of the physical location of the I/O pin used.

- Pipelining is a method of reducing the size of synchronous logic cells, the size of which is determined by the automatic physical device placement tool that is used to layout the internal architecture of the chip. The size of a synchronous logic cell is defined by how many CLBs are utilised between two synchronous registers i.e. the input register to a group of CLBs to the output register from that group of CLBs. By placing several one-bit deep clocked registers with the same width of the data-bus in the data path, the number of CLB blocks between registers can be reduced. The size of the synchronous cell should be such that the time for the data to pass through the CLBs to the next register is less than the sum of delays and set-up and hold times within the cell. For Figure 6-12 the five pipeline registers are depicted as one block in the diagram.

_____

[38] When the setup and hold time of a flip-flop is violated causing the output to be unstable.

Referring to Figure 6-12, while the event data are passing through the pipeline, the FIFO-stored BID from the Level_1 TTCrx is compared to the BID control word of the event data. The 1-bit result, along with any other error flags generated by FPGA-1, is formatted into a 1x36 bit word by the 'Error_Gen' block and is serially fed back into the pipeline to be appended as the 18th bit word of the event-block. FPGA-1 carries out many internal error checks. In particular it checks the event-block for size, parity and Hamming-code. The Op-mode block is configurable by JTAG and puts FPGA-1 into several different operational modes. Most of the modes are for self testing by generating pseudo event blocks.

The QDR read and write addresses are generated by two independent counters under the control of the L1/L0-Data-Control unit. This unit also generates control signals for the QDR memory chip. The control logic is built around two state machines, a write machine shown in Figure 6-13 and a read machine, which is very similar in operation but receives different hand-shaking signals. The state machines ensure the correct timing of the QDR signals and address generation. These are implemented using the so-called 'one hot' method, which uses one flip-flop output for each state. By taking the appropriate logic function of these state outputs ensures that only one state will ever be active at any one time. A state is represented by the circles in Figure 6-13, the movement between each state is determined by the blue boxes and the actions to be taken at each state are shown next to the states. If there is a choice to move to one of two available states the priority is given by the number given in the smaller circles, 1 being the highest priority. The 'one hot' method has the advantage of ease of design and clarity of circuit operation as the inputs to the state flip-flops directly describe the condition in which that state can be entered. Further to this, it is compact in size for designs with less than ~9 states as the decoding for each state is still quite small. Above 9 states, the decoding logic becomes disproportionate to the logic that is being implemented. The operation of the write state machine is now described.

Package List
LIBRARY ieee;
USE ieee.std_logic_1164.all;
USE ieee.std_logic_arith.all;

GET_DATA='0'

idle 2

1

GET_DATA='1'
AND
K='1'

Dout_cld<=DATA4;
HOLD_read_cld <= '0';
continue <= continue;

WR_D4 2

1

Aout_cld<=A;
WPSbar_cld<='0';
HOLD_read_cld<='1';
continue <= '0';

ADDRESS

continue = '1'

Aout_cld <= A;
WPSbar_cld <= '0';
HOLD_read_cld <= '1';
Dout_cld <= DATA3;
continue <= '1';

address2

Dout_cld<=DATA3;

WR_D3

Dout_cld<=DATA1;
incrADD_cld<='1';
HOLD_read_cld<='0';

WR_D1

GET_DATA = '1'

1

2

K'EVENT AND K='0'

WR_D2

Dout_cld<=DATA2;
incrADD_cld<='0';

Figure 6-13 The write state machine.

For the state machine to make a next state decision and have enough time to multiplex four fibres, the state machine master clock operates at 160 MHz. The state machine sits in the idle state until there is logic '1' of both the k-clock and GET_DATA, which is the logic AND of the RxCntl and RxReady signals (refer to Figure 6-9). It then moves to the address state. This state releases the $\overline{wps}$ signal and the data from the write counter. The following state, WR_D1, sets the multiplexer to output the event data received on the first of four fibres. The state machine is held in the WR_D1 state until the machine is synchronised to the rising edge of the k-clock at which point the state machine moves through each of the remaining states, multiplexing the remaining three event-blocks as it goes. Every time the write machine is activated, the wrap-around write counter is incremented by 1. The state 'address2' and 'continue' line are used to prevent the state machine going back to the idle state in the case where there has been a consecutive Level_0 trigger and incoming data are continuous between received events.

167

As stated earlier, the read state machine is similar in operation, but on receipt of a Level_1 trigger the read machine is activated and cycles 36 times consecutively. In the case of a Level_1 trigger reject the state machine multiplexes the event-blocks out but does not release the $\overline{rps}$ signal to the QDR. Logic has been incorporated into the design to ensure the read and write pointer cannot overtake each other.
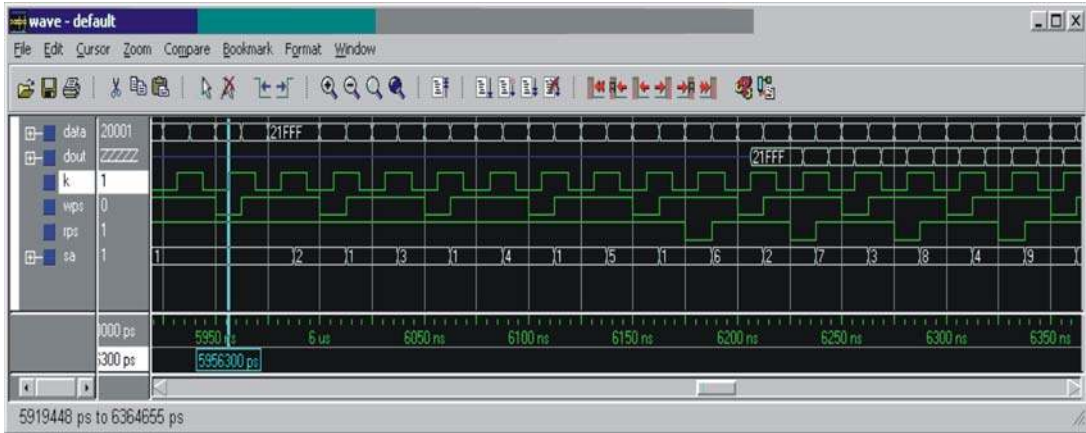
Figure 6-14 Simulation of data to (labelled data) and from (labelled dout) the QDR chip.

Figure 6-14 shows a simulation of test data flow to and from the QDR for the demonstrator system. Here event data arrive on two fibres and the internal clocks are running at half speed, i.e. the k-clock is 40 MHz. The FPGA-1 chip has been set to a test mode where after receiving the first internally generated event word, an automated self-generated Level_1 trigger accept is asserted 9 k-clocks later. Shown in the figure is the cursor which is positioned at the beginning of a write sequence. With the rising edge of k-clock and $\overline{wps}$ set to '0' the write address is taken from the input port SA, the address is 1 in this case. On the following rising edge of the k-clock the first data word from the event-block (fibre #1) is stored to memory. This is an arbitrary header word, 21FFF. On the next falling edge of the k-clock the header word, also 21FFF, from the second event-block (fibre #2) is stored. The remaining 35 words in each of the two event data blocks are then stored on every k-clock edge with $\overline{wps}$ ='0' every time a new write memory address is required. After the automated Level_1 trigger accept

168

has been asserted, $\overline{rps}$ is set to '0' and the read address is taken from SA. In this case the address is 1 so as to read out the last event-block stored. The first BID is read out on the following k-clock rising edge, the second on the next falling edge. The remaining data words from the two event-blocks are then read out on each k-clock transition. With the $\overline{rps}$ and read address on SA being asserted every other rising k-clock edge, the device can read and write event-blocks at the same time.

6.3.4   Delay Lock Loop (DLL)

The Level_1 FPGA-1 takes advantage of the Xilinx internal DLLs to multiply the TTCrx 40 MHz output clock signal by a factor of 2 and 4, for use internally and externally to the FPGA. The DLL also adjusts the phase of the internally generated 80 MHz clocks by 90° and 270°. To ensure that the TTCrx output clock signal is compatible with the DLL input requirements, and to find how the DLL responds to a corrupted input clock, the following studies were undertaken to investigate:

- Whether the lock signal[39] responds to missing input clock pulses.

- How a DLL x2 output responds to a missing input clock pulse.

- Whether the DLL can tolerate jitter from a TTCrx output clock signal.

Two types of jitter were considered. Cycle-to-cycle jitter is measured from one clock edge to the next, whereas the long-term jitter of a clock is measured with respect to some stable clock of the same frequency. Jitter measurements were obtained by using a Lecroy oscilloscope with a 2Gs/s sampling rate. The Lecroy uses its internal clock for a timing reference so as to log the input signal voltage level at each sample. On the transition of a user set voltage level the oscilloscope uses the logged voltage levels to extrapolate a time of transition. In this way the

---

[39] The lock signal is given from the DLLs when synchronisation is achieved.

oscilloscope is able to make timing measurements between rising edges of an input signal to within a 10 ps accuracy.

The DLL within the Xilinx Spartan FPGA can tolerate a maximum cycle-to-cycle jitter of +/-300 ps at its input. The DLL tolerance to the long-term jitter was not stated by the manufacturer. The TTCrx used for the demonstrator to generate the DLL input is a Version-3 chip and was found by measuring its output clock signal, to have a cycle-to-cycle jitter of +/-290 ps and to have a long-term output clock jitter (or drift) of 400 ps. The histogram of the cycle-to-cycle and long-term jitter measured at the TTCrx output is shown in Figure 6-15 a) and b), respectively. The problem of the two discrete jitter peaks seen in the cycle-to-cycle jitter distribution is caused by the internal chip cross-talk when having data traffic on both channel A and B of the TTCrx at the same time. This problem has been eliminated in the new version of

a)   b)
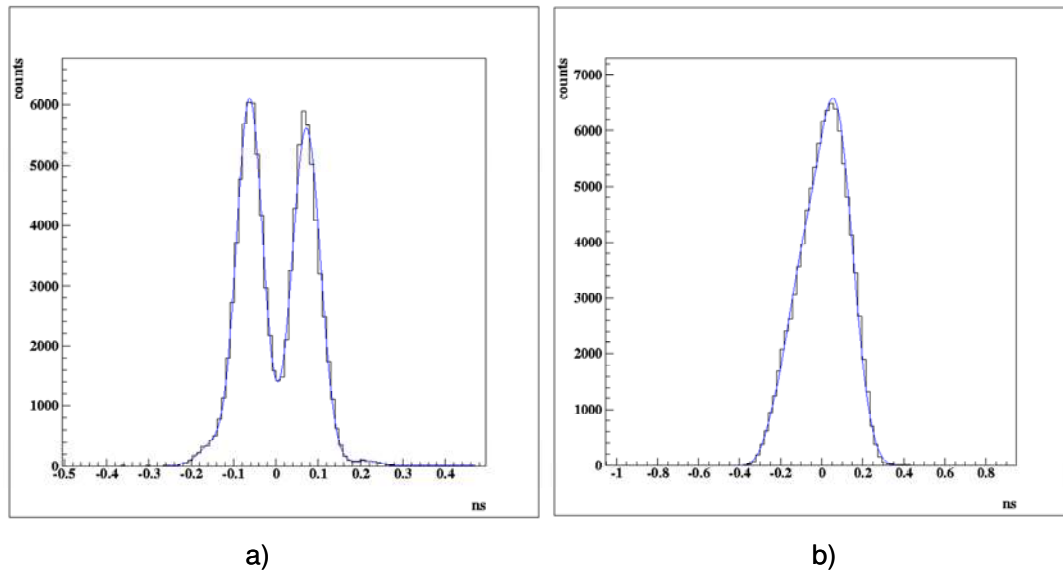
Figure 6-15 a) TTCrx cycle-to-cycle signal output jitter, b) TTCrx long term signal output jitter.

the chip with the second peak being removed, hence halving the jitter value. Without this improvement the TTCrx is only just within the DLL jitter input tolerance of +/-300 ps, which does not give enough margin for use within the experiment. However, no detrimental effects were observed at the x2 output clock from the DLL with the Version-3 TTCrx clock connected

to its input. The cycle-to-cycle and long term jitter measurements taken at the output of the DLL x2 show that there is a maximum 35 ps jitter when using the clock from the TTCrx input.

To measure the effects on the DLL lock signal and x2 clock output in the case of missing input clock pulses, a pattern generator was used to generate a vector of input clock pulses. In this way one or more of the clock pulses could be removed from the vector. Measurements of changes in the lock signal showed that it takes 100 μs to lose the lock signal after the clock pulses are turned off, i.e. no system clock or an irregular 14 out of 16 input clock pulses can be missing before the out of lock signal is given. The x2 output responded to missing input pulses by a doubling in missing output pulses. From this it can be concluded that the DLL will pass on missing clock pulses and that the lock signal does not recognise this condition. Therefore the lock signal can be used for indicating that the DLL has locked-on but not for loss of lock.

To port VHDL algorithms into an FPGA, the Xilinx designer software is used. The software first translates the VHDL into a net list of components and interconnect wires (the net list file can be viewed like a schematic drawing). If requested to do so, the software will then automatically take each component and interconnect, and find a physical placement within the FPGA hardware. However, the DLLs cannot be automatically placed when using more than 2 DLLs within a FPGA chip. There was also a need to maximise the speed performance of the critical logic cells within the multiplexer. Therefore some physical layout within FPGA-1 is
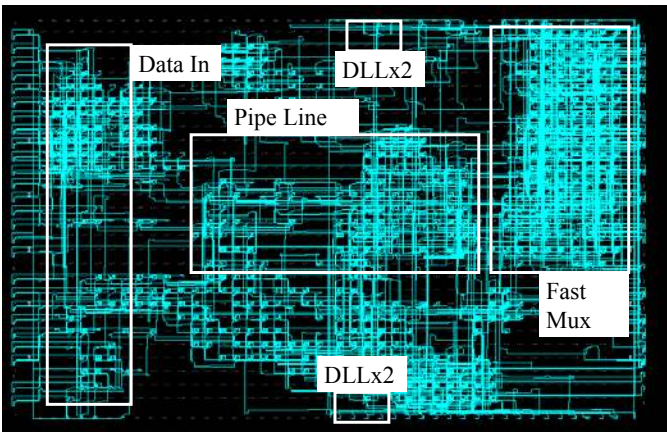

Figure 6-16 Internal Layout of FPGA-1.

171

done by hand and not left to the automated router. This is achieved by anchoring the group of cells shown in the white boxes of Figure 6-16 and then allowing the automatic router to finish off the rest of the layout. Although the layout of Figure 6-16 looks sparsely populated, with only ~60 % of the chip resources having been utilised, this does have the advantages of flexibility in I/O selection and space for fast routing.

6.3.5 FPGA-2

A Xilinx Spartan II XC2S200 FG456 FPGA is also used for FPGA-2 of the derandomiser area of Figure 6-2. For the present testing of the readout chain, FPGA-2 is only used to interface to the S-Link, and serves the following functions:

- Accepts event-blocks from the QDR memory at a data rate of 40 MHz DDR.

- Strips the 17th bit (parity) and 18th bit (error code) from the event-block and formats them into end words.

- Concatenates the two, now 16 bit-wide event-blocks, into a 32 bit-wide word so as to match the S-Link data-bus width.

- Generates all the necessary control and clock signals for the S-Link.

- Clocks out the 32-bit wide data block to the S-Link at 20 MHz.

6.3.6 System tests

The full demonstrator readout system, described in the previous sections, was tested using a stand-alone LHCBPIX1 chip as an input (an HPD was not available at the time of the test). The LHCBPIX1 chip offers a charge-injection test pulse on each of the readout channels to emulate the charge pulse from an HPD. Due to hardware problems associated with the S-Link and also the need to release the DAQ system for debugging the next iteration of Level_0 board, the demonstrator test at the time was confined to making simple comparisons of the input/output data integrity to verify proof-of-principle. More sophisticated tests involving an

172

HPD and more robust DAQ system were to follow, but with a revised set of hardware elements. These tests will be briefly summarised in the next section.

Figure 6-17 shows an example of a screen dump of a) the input data written into the LHCBPIX1 and b) the data read out through the S-Link/FLIC. A simple bit pattern was input into the LHCBPIX1 chip; the 32x32 pixel array in this case has all the pixels in column 13 selected for test-pulse injection. The event data were then processed through the Level_0 and

a)   b)
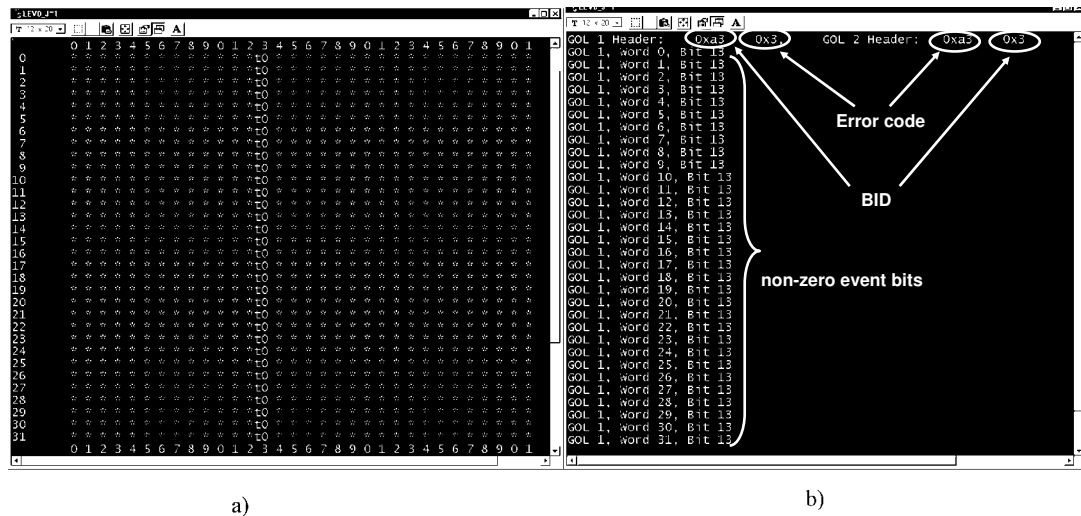
Figure 6-17 a) Screen dump of pulsed LHCBPIX1 test channels, b) the data read out from the Slink.

Level_1 systems, and written to the PC hard disk. The BID number and error code are displayed on-screen, along with any non-zero event bits. Figure 6-17 b) shows the BID, hexadecimal a3 in this case, and error code hexadecimal 3, from two fibres. As expected for correct data transmission, both fibres have the same BID and error code. The error code gives information about the FIFO status in the PINT chip. Also from Figure 6-17 b), it can be seen that the test-bits set in column 13 have been detected in all of the 32 event-data words.

A number of important conclusions are drawn from these tests:

- The PINT is fully operational with the LHCBPIX1, TTCrx, GOLs and VSCELs.

173

- The PINT algorithms are functional. Writing and reading to the LHCBPIX1 chip is proven.

- The BID is added correctly at Level_0 and is maintained throughout transmission to the PC.

- The error flags, at both the Level_0 and Level_1, are added correctly and give the correct information.

- Data are successfully sent from Level_0 to Level_1 via optical fibre at 800 Mb/s.

- The Level_1 fibre-optic receiver card works correctly.

- Both FPGA-1 and 2 are functional with TTCrx operation.

- FPGA-1 can multiplex and store data into the buffer QDR.

- FPGA-2 can interface to the de-randomiser QDR and interface to the Slink.

6.4 Recent test beam results

Following the successful operation of the first demonstrator system, further iterations of Level_0 and Level_1 boards have been designed to conform to final LHCb requirements. The boards have been refined in terms of physical size, more compact design and doubling the data-transmission speed, but in all other aspects the new system is fully compatible with the original demonstrator system. The author has been fully involved in the development of the new system (with particular attention to the Level_0 region). This work culminated in the successful readout of a single HPD in a test beam in October 2004 [PAT04].

Figure 6-18 shows the preliminary results of Cherenkov rings observed. The plot shows hits from the Pixel HPD, integrated over many events. Cherenkov rings resulting from the passage of 10 GeV/c pions through a ~1m nitrogen radiator at atmospheric pressure are clearly
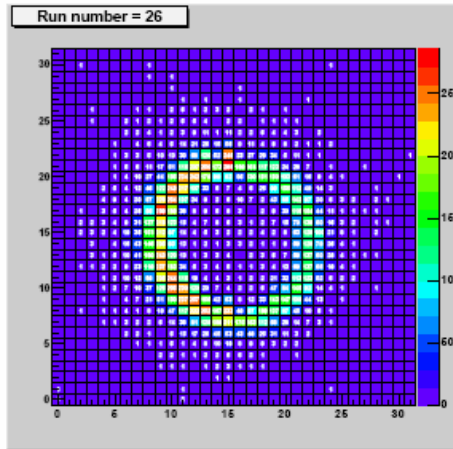
Figure 6-18 Cherenkov rings observed from 10 GeV/c pions in an N_2 gas radiator. The x and y axis represent the column and row number, respectively, of the pixel HPD.

seen. The left and right asymmetry clearly seen in Figure 6-18 is a result of the PINT double sampling the incoming data on some channels due to a timing issue. With the double sampling accounted for a total of 9.1 ± 0.1 photons are observed per event, compared to 7.9 ± 0.8 expected. The results indicate the electronics readout system is operating at full efficiency and with low background.

6.5   Summary

A complete demonstrator readout system, from LHCBPIX1 chip to DAQ, has been successfully developed and operated.   Error detection methods have been incorporated into both the Level_0 and Level_1 regions. The Level_0 board can perform all the necessary data formatting, and operates successfully with the I/O interfaces. Algorithms for the Level_1 FPGAs have been written using VHDL, synthesised and downloaded using JTAG. These algorithms have been tested at a modular and system level. The Level_1 board can store incoming events at 40 MHz DDR, perform the necessary formatting for the S-Link and send data out to a PC on receipt of a Level_1 trigger. The delay lock loops are compatible with TTCrx operation. The optical link from Level_0 to Level_1 has been verified.

In summary, the demonstrator described here has provided a proof-of-principle and test-bed for the final readout system of the LHCb RICH detectors.

175

Summary

In this thesis, the LHCb detector, its global electronic scheme and triggering system have been introduced. A description of the role in which the Ring Imaging CHerenkov (RICH) detectors play within LHCb, and the electronics readout requirements, have been given. Two tiers of dedicated electronics, the Level_0 and Level_1 regions, which filter data into a farm of commodity processors making up the High Level Trigger (HLT), have been described.

The following major contributions have been made by the author to the development of the LHCb experiment:

- An evaluation of the Multi-Anode Photo-Multiplier Tube (MAPMT) as a candidate for the RICH photo-detectors.

- The development and successful testing of a customized readout ASIC, the BeetleMA, which has been optimised for MAPMT readout.

- The evaluation and implementation of the BeetleMA bias generator, the current and voltage DACs, and front-end amplifier.

- The development and operation of a demonstrator of the full RICH readout system, focusing especially on the VHDL algorithms used for the Level_1 electronics.

A vital influence on the performance of the LHCb RICH detectors is the choice of suitable photo-detectors. Both the MAPMT and Hybrid Photon Detector (HPD) have met the RICH requirements of spatial resolution and single-photon sensitivity. Both detectors require specialised radiation-hard ASIC chips that can capture and store the signals, while being fully compatible to the global electronics scheme of LHCb.

The 12-dynode R5900-00-M64 H7546B MAPMT photon detector has been studied by the author. The device has performed to the manufacturer's specifications and is a suitable photon detector for the LHCb RICH detectors. The relative efficiency for single photons converted at the MAPMT photo-cathode has been measured to be $74 \pm 3\%$, reduced from full efficiency by the focussing and gain structure of the device. The author has demonstrated that the readout amplifier of the MAPMT must be able to accommodate a dynamic range of the input signal of 9 and give a SNR ratio of ~40 for a typical single-photon response.

Studies have been made on the most appropriate front-end amplifier for capturing photon signals from an MAPMT. The best approach was found by incorporating a new front-end amplifier design into the architecture of the exiting Beetle ASIC by optimizing its gain performance. The use of a deep-submicron process technology, with thin oxide and enclosed NMOS transistors, makes the BeetleMA tolerant to ionising radiation. CMOS latch-up is suppressed by the use of guard rings. Robustness against single event upset is achieved with triple-redundant logic.

The test setup for evaluating the BeetleMA has been described and the measurement results given. The output has a rise-time of 10 ns and ~30 % of the signal remains after a further 25 ns after the peak, dependent on the value of the feedback resistance. With a test-pulse to represent the MAPMT signal, the output of both the front-end and pipeline readout remained linear up to 10 single photon signals i.e. $\sim 3\text{x}10^6$ electrons at the nominal operating voltage. Simulations were made of the occupancy and crosstalk effects and it was found that the BeetleMA has, in general, less than 1 % crosstalk and can operate up to an occupancy approaching 100%. From the measured and simulated results it was concluded that the BeetleMA, in analogue mode, is a suitable chip for reading out the MAPMT.

A fully integrated bias generator that supplies all of the necessary current and voltage biasing for the BeetleMA chip has been designed, fabricated and tested. The bias generator contains a current source, voltage V-DACs and current I-DACs. The 10-bit I-DACs use a binary-weighted current-source scheme with 1023 PMOS transistors. The necessary voltage references are generated with current mirror scaling circuits. The mean dynamic range of the 36

I-DACs measured is from 0.1 μA to 2.101 mA (with an LSB of 2 μA and a gain error of 2.6%). These values are very dependent on the supply voltage and temperature. The differential and integrated non-linearity has been measured to be ± 0.5 LSB and ± 1 LSB, respectively. This I-DAC is suitable for the Beetle chip requirements.

The 10-bit V-DACs are of the *R-2R* type utilising 3 $k\Omega$ N$^+$ diffusion OP resistors. The mean dynamic range measured for ten V-DACs is from 1.57 mV to 2.498 V with a gain error of 0.5 %. The mean LSB is 2.43 mV and is better than 99% compliant with a load resistance > 70 $k\Omega$. The differential non-linearity is ± 1.4 LSB making the V-DACs non-monotonic, although this is not important for the final Beetle design. The effects of temperature are easily managed. This V-DAC is suitable for Beetle chip use but benefited in performance when the resistors were changed from N$^+$ diffusion to non-silicided polysilicon for the Beetle fabrication submission.

The LHCBPIX1 front-end ASIC is used to read out the HPD photon detector. A complete demonstration readout system has been successfully developed and operated from LHCBPIX1 chip to DAQ. Error detection methods have been incorporated into both the Level_0 and Level_1 regions. The Level_0 board and PINT chip perform all the necessary data formatting, and operate successfully with the I/O interfaces. The data are transmitted to the Level_1 region down fibre-optics using the 800 Mb/s G-Link protocol.

The demonstrator Level_1 region stores data while awaiting the Level_1 trigger accept at an average rate of 40 kHz. The incoming events are stored for a variable latency time into QDR memory chips. QDR control and data formatting is achieved with an FPGA. The data capture rate into the QDR is at a DDR rate of 40 MHz. Data stored in the QDR are sent out to a PC on receipt of a Level_1 trigger using fibre-optics and an S-LINK protocol. The necessary formatting for the S-LINK is achieved using a second FPGA on the Level_1 board. The algorithms for both FPGAs are written using VHDL, synthesised and downloaded using JTAG. These algorithms have been tested at a modular and system level. The timing and control chip, the TTCrx, is compatible with the delay-lock loops used on the Level_1 board.

The results reported here show that the MAPMT and BeetleMA can be successfully used for the LHCb RICH detectors. From the first operation of the Level_0 to Level_1 demonstrator system, it can also be concluded that events can be successfully formatted, stored and transmitted through each tier of the LHCb electronic regions.

Glossary of terms

AMS	**A**ustria **M**ikro **S**ystem International
ASIC	**A**pplication **S**pecific **I**ntegrated **C**ircuit
Beetle	ASIC detector readout chip
BeetleMA	Modified Beetle MAPMT readout chip
BID	**B**unch **ID**
BILBO	**B**uilt-**I**n-**L**ogic-**Ob**server
Bulk	Also body. The ASIC substrate
CLB	**C**ontrol **L**ogic **B**locks
C_{ox}	Capacitance of FET gate oxide
CSA	**C**harge-**S**ensitive **A**mplifier
CTRW	**C**ontinuous-**T**ime **R**andom **W**alk
Derandomiser	Temporary data storage area
DDR	**D**ouble **D**ata **R**ate
DLL	**D**elay **L**ock **L**oops
DNL	**D**ifferential **N**on-Linearity
DMILL	**D**urci-**M**ixte sur **I**solant **L**ogico-Lineaire
ECAL	**E**lectromagnetic **CAL**orimeter
ECS	**E**xperiment **C**ontrol **S**ystem
ELT	**E**nclosed **L**ayout **T**ransistor

ENC	**E**quivalent **N**oise **C**harge
ESD	**E**lectro**S**tatic **D**ischarge protection
FIR	**F**inite **I**mpulse **R**esponse filter
FSM	**F**inite **S**tate **M**achine
FWHM	**F**ull-**W**idth-**H**alf-**M**aximum
g_m	Transconductance
GOL	**G**igabit **O**ptical **L**ink
HCAL	**H**adronic **CAL**orimeter
HDMP-1034	Fibre-optic receivers
HL	**H**igh **L**evel
HLT	**H**igh **L**evel **T**rigger
HPD	**H**ybrid **P**hoton **D**etector
IT	**I**nner **T**racker
JTAG	**J**oint **T**est **A**dvisory **G**roup
L	Length of FET gate region
Latency	Data storage time
LEP	**L**arge **E**lectron **P**ositron collider
LET	**L**inear **E**nergy **T**ransfer
Level_0	The first level of LHCb trigger and storage
Level_1	The second level of LHCb trigger and storage
LHC	**L**arge **H**adron **C**ollider
INL	**I**ntegral **N**on-**L**inearity

LSB	**L**east **S**ignificant **B**it
LVDS	**L**ow **V**oltage **D**ifferential **S**ignal
(M1-M5)	**M**uon system
MAPMT	**M**ulti-**A**node **P**hoto-**M**ultiplier **T**ube
MIM	**M**etal **I**nsulator **M**etal capacitors
MPW	**M**ulti-**P**roject **W**afer
MSB	**M**ost **S**ignificant **B**it
MWPC	**M**ulti **W**ire **P**roportional **C**hambers
NA	**N**umerical **A**perture
OT	**O**uter **T**racker
PINT	**P**ixel **INT**erface chip
PM3705	JTAG interface unit
Poly	Polysilicon semi-conductor material
pp	**P**roton-**P**roton
PS	**P**re-**S**hower
QDR	**Q**uad **D**ata **R**ate
RAM	**R**andom **A**ccess **M**emory
RICH	**R**ing-**I**maging **C**herenkov
r_o	The incremental output resistance of a FET
$\overline{rps}$	QDR SRAM **R**ead **P**ointer **S**ignal

182

RS	**R**eadout **S**upervisor
SA	QDR **S**RAM **A**ddress bus
SEE	**S**ingle **E**vent **E**ffects
SEL	**S**ingle **E**vent **L**atch-up
SEU	**S**ingle **E**vent **U**pset
SNR	**S**ignal-to-**N**oise **R**atio
SOI	**S**ilicon-**O**n-**I**nsulator
SPD	**S**cintillating **P**ad **D**etector
SRAM	**S**tatic **R**andom **A**ccess **M**emory
SRBF	**S**ilicon **R**esin **B**onded **F**ibre
TAP	JTAG **T**est **A**ccess **P**ort
TFC	**T**iming **F**ast **C**ontrol
t_{ox}	Thickness of FET gate oxide
TT	**T**rigger **T**racker
TTC	**T**iming **T**rigger **C**ontrol
TTCrx	**T**iming **T**rigger **C**ontrol receiver
VCSEL	**V**ertical **C**avity **S**urface **E**mitting **L**asers
VELO	Silicon **VE**rtex **LO**cator
VETO	Pile-up **VETO** detector

V_{FB}	FET gate **F**lat**B**and voltage
V_{GS}	Voltage between FET gate and source terminals
V_t	FET V_{GS} threshold voltage
W	Width of FET gate region
$\overline{wps}$	QDR SRAM **W**rite **P**ointer **S**ignal

Bibliography

[ALB98]
E.Albrecht et.al. "First observation of Cherenkov ring images using hybrid photon detectors". NIMA 411 (1998).

[ALB01]
E.Albrecht et.al. "A prototype RICH detector using multi-anode photo-multiplier tubes and hybrid photo-diodes". NIMA 456 (2001).

[ALB02]
E.Albrecht et.al. "Performance of a cluster of multi-anode photomultipliers equipped with lenses for use in a prototype RICH detector". NIMA 488 (2002).

[ALItp]
ALICE. "Technical proposal for a large ion collider experiment at the CERN LHC". CERN/LHCC/95-71 (December 1995).

[AMS]
Austria-micro-systems. European headquarters, AG A-8141 Schloss Premstatten, Austria. http://www.ams.co.at/.

[ANE00]
G.Anelli. "Conception et caracterisation de circuits integres resistants aus radiations pour les detecteurs de particules du LHC en technologies CMOS submicroniques profondes". Doctoral thesis, CERN (2000).

[ATL94]
W.Blum, H.Kroha, P.Widmann. "Development of an optical alignment monitoring system for the ATLAS Muon Spectrometer". ICHEP conference Glasgow (July 1994).

[ATLtp]
"ATLAS Technical Proposal". CERN/LHCC (December 1994).

[BAK98]
R.Baker et al. "CMOS circuit design, layout and simulation". IEEE Press ISBN 0-7803-3416-7 (1998).

[BAU03]
D.Baumeister. "Development and characterisation of a radiation hard readout chip for the LHCb experiment". Doctoral thesis, Universitat Heidelberg, Germany (2003).

[BEE04]
"The Beetle reference manual". Chip version 1.3, 1.4 and 1.5, document version 1.51 (23.06.04). This manual can be found at http://wwwasic.kip.uni-heidelberg.de/lhcb/Documentation.html.

[BRA00]
E.Brandin. "Development of a prototype read-out link for the Atlas experiment". Master thesis, CERN (June 2000).

[CHA91]
Z.Chang et al. "Low noise wide band amplifiers in bipolar and CMOS technologies". Kluwer academic publishers (1991). ISBN 0-7923-9096-2.

[CHA03]
M.Charles, "The performance of SCT128A ASICs when reading out silicon sensors and a study of $B_s^0 \rightarrow D_s^\pm \pi^\mp$ at LHCb". Doctoral thesis, University of Oxford (2003).

[CHR01_L0]
J.Christiansen. "Requirements to the L0 front-end electronics". LHCb Technical Note second version, revision 1.0. LHCb 2001-014, created July1999, last modified, (July 3 2001).

[CHR01_L1]
J.Christiansen. "Requirements to the L1 front-end electronics". LHCb technical note revision 2.0. LHCb 2003-078, created June 2001, last modified, (August 15 2003).

[CHR01_L1a]
J.Christiansen. "Requirements to the L1 front-end electronics". LHCb technical note revision 1.0. LHCb 2003-078, created June 2001, last modified, (August 15 2003).

[CHR_E]

J.Christiansen. "Electronics in LHCb". http://lhcb-elec.web.cern.ch/lhcb-elec/. Also see [CHR01_L0], [CHR01_L1] and [CHR01_L1a].

[CHR_s]

J.Christiansen, I.Alfonso. "Simulation of the LHCb front-end". LHCb technical note revision 0.1 LHCb FE 99-047, created December 1st 1999, last modified, (December 1st 1999).

[CMStp]

"CMS technical proposal". CERN/LHCC (December 1994).

[DAM02]

G.Damerell. Private communication. Particle and Nuclear Physics dept, Oxford University.

[ECS]

"LHCb online system data acquisition and experiment control" TDR CERN/LHCC 2002-40 LHCB TDR 7 (19th December 2001). For a general overview of the ECS see http://lhcb-elec.web.cern.ch/lhcb-elec/html/ecs_interface.htm.

[EIS03]

S.Eisenhardt and F.Muheim. "Performance of multianode photo multiplier tubes at low gain". LHCb note CERN/LHCC/2003-043 (2003).

[EURO_p]

Europractice IC service contact. Kapeldreef 75, B-3001, Leuven, Belgium. Also see the Europractice home page http://www.europractice.imec.be/europractice/on-line-docs/design/homepage_design/homepage_design.html

[FAC98]

F.Faccio et al. "Total dose and single event upsets in a 0.25 μm CMOS technology". LHCb note CERN/LHCC/98-36 (1998) 105.

[FAL98]

W.Fallot-Burghardt. "A CMOS mixed-signal readout chip for the micro-strip detectors of HERA-B". DPhil Inaugural dissertation, Ruprecht-Karls-Universitat, Heidelberg (1998).

[GEI90]

R.Geiger et al. "VLSI design techniques for analogue and digital circuits". McGraw-Hill ISBN 0-07-100728-8 (1990).

[GIR98]

A. Giraldo. "Evaluation of deep submicron technologies with radiation tolerant layout for electronics in LHC experiments". Doctoral thesis, University of Padua (1998).

[GOL01]

"GOL reference manual". Preliminary version March 2001 CERN-EP/MIC, Geneva Switzerland

[HAM_www]

Hamamatsu headquarters, 325-6, Sunayama-cho, Hamamatsu City, Shizuoka Pref, 430-8587, Japan. Also see the Hamamatsu home site, http://www.hpk.co.jp/Eng/main.htm

[HAW89]

J.Hawkes. "Optoelectronics, an introduction". Prentice Hall second edition (1989). ISBN 0-13-638461-7.

[HD_B]

Contact Ulrich Trunk, Physikalisches Institut der Universitaet Heidelberg, Asic Labor, Heidelberg, Im Neuenheimer Feld 227, D-69120, Heidelberg, Germany, http://wwwasic.kip.uni-heidelberg.de.

[HD_U]

Beetle home page, wwwasic.kip.uni-heidelberg.de/lhcb/. Also see [HD_B], [BEE04] and [BAU03].

[HER95]

"HERA-B technical design report". HERA-B collaboration, DESY-PRC 95/01 (1995).

[HER]

HERA-B home page, http://www-hermes.desy.de. Also see [HER95].

[HOR80]

Horowitz and Hill. "The art of electronics". Cambridge University Press 1980. ISBN 0 521 23151 5.

[HP]

"The use of gain-switched vertical cavity surface-emitting laser for electro-optic sampling". Hewlett Packard instrument and photonics laboratory note, HPL-93-76 (1993).

[HU96]

Y.Hu et el. "Design of a low noise, low power consumption CMOS preamplifier shaper for readout of micro-strip detectors in the DMILL radiation hardened process and irradiation measurements". NIM A 378 (1996) 589-593.

[IBM_CMOS]

European headquarters, IBM technology group, IBM Switzerland 48, Avenue Giuseppe-Motta P.O. Box CH-1211 GENEVA 2, Switzerland. Also see IBM-CMOS home page http://www-3.ibm.com/chips/techlib/techlib.nsf/products/CMOS_6SF.

[JOH97]

D.Johns et al. "Analogue integrated circuit design". John Wiley and sons Inc ISBN 0-471-14448-7.

[JON98]

L.Jones et al. "A 128 channel analogue pipeline chip for MSGC read-out at LHC". Proceedings of the fourth workshop on electronics for LHC experiments, Rome, (September 1998) CERN/LHCC/98-0036.

[JOS99]

B.Jost. "Timing and fast control". LHCb technical note, issue 1, revision 1.0, LHCb DAQ 99-001, created 1st of June 1998, last modified (26th January 1999).

[JTAG]

"IEEE Standard test access port and boundary-scan architecture". Institute of electrical and electronic engineers, Inc. IEEE standard test access port and boundary-scan architecture (IEEE Std 1149.1-1990). New York: Institute of electrical and electronic engineers, Inc. (1990).

[KAP98]

J.Kaplon et al. "DMILL implementation of the analogue readout architecture for positioning sensitive detectors a LHC experiments". Proceedings of the fourth workshop on electronics for LHC experiments, Rome, CERN/LHCC/99-27 (1999).

[KAP98n]

J.Kaplon. "SCT128 and SCT128HC". CERN internal note (February 1998).

[LEB99]

"Proceedings of the fifth workshop on electronics for LHC experiments". Snowmass, Colorado, USA. LEB 1999 CERN 99-09 CERN/LHCC/99-33 (29th October 1999).

[LHCbtp]

"LHCb technical proposal". CERN/LHCC 98-4, LHCC/P4 (20th February 1998).

[LOC03]

S.Lochner. Doctoral thesis in preparation. Universitat Heidelberg, Germany (2004).

[LUT99]

G.Lutz. "Semiconductor radiation detectors device physics". Published by Springer (1999). ISBN 3-540-64859-3.

[MAP97]

"Hamamatsu R5900-00-M64" data sheet (September 1997). Also see [HAM_www]

[MAP98]

"Photonmultiplier tube principle to application, photon is our business". Produced and published by Hamamatsu photonics K.K, Japan (1994). Also see [HAM_www].

[McL89]

F.McLean et al. "Electron hole generation, transport and trapping in SiO_2, ionizing radiation effects in MOS devices and circuits". Wiley (1989).

[MOO65]

G.Moore. Cramming more components into integrated circuits. Electronics, volume 38, number 8 (April 19th 1965).

[MOR03]

Nuclear science symposium, medical imaging conference, 13th international workshop on room temperature semiconductor x and gamma-ray detectors, symposium on nuclear power systems, (October 2003), Portland, Oregan, USA. Transparencies.

[MUH00]

F. Muheim et el. "Proposal for multi-anode photo-multiplier tubes as photo-detectors for the LHCb RICH". LHCb 2000-065 RICH (September 4 2000).

[MUH02]

Private communication with Franz Muheim, School of Physics, University of Edinburgh, James Clerk, Maxwell Building, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom.

[NIK_B]

Contact Ruud Kluit, NIKHEF, The National Institute for Nuclear Physics and High Energy Physics, Amsterdam, The Netherlands. Also see

http://www.nikhef.nl/pub/departments/et/vlsi/BEETLE/beetle.html.

[OX_B]

Contact Nigel Smale, Department of Physics, Denys Wilkinson Building, Keble Road, Oxford, OX13RH, UK. Also see http://www2.physics.ox.ac.uk/smale.

[O'CO98]

P.O'Connor et al. "Ultra low noise CMOS preamplifier-shaper for X-ray spectroscopy". Nuclear Instruments and Methods in Physics Research, section A. A 409 (1998) 315-321.

[PAT04]

M.Patel. "System test of a prototype LHCb RICH detector". 2004 IEEE Rome Oct 16-22 2004.

[PHI95]

"The I²C bus and how to use it". Phillips Semiconductors (1995).

[RAD02]

J.Radenmacker. "Evaluation of the LHCb RICH detectors and a measurement of the CKM angle γ ". Doctoral theses, CCLRC, RAL-TH-2002-002.

[RD49www]

CERN RD49 Project. Contact A. Marchioro CERN. Also see http://rd49.web.cern.ch/RD49/.

[RD49_stat]

P.Jarron, G.Anelli et al. "Study of the Radiation Tolerance of IC's for LHC". 3rd RD49 Status Report, CERN/LHCC/2000-003, LEB Status report/RD49 (2000).

[SAK84]

N.Saks et al. "Radiation effects in MOS capacitors with very think oxides at 80 K". IEEE Trans, Nicl. Sci. Vol. 33, No 6 (1986) 1249.

[SED98]

A.Sedra and K.Smith. "Micro-eletronics circuits" fourth edition. Oxford University Press 1998, ISBN 0-19-511690-9.

[SOM03]

L.Somerville. "DPhil status report". Oxford University, Particle physics.

[SEX01]

E.Sexauer. "Development of radiation hard readout electronics for LHCb". Doctoral thesis, Universitat Heidelberg, Germany (2001)..

[SCH02]

M.Schmelling. "The LHCb experiment" presentation. Four seas conference, Thessaloniki, Greece (April 2002).

[TAY98]

B.Taylor, "TTC distribution for LHC detectors". IEEEE Trans. Nuclear Science, vol.45, No3, (June 1998).

[TAY02]

B.Taylor. "Timing distribution at the LHC". 8th workshop on electronics for LHC experiments, Colmar (9-13 September 2002).

[TDRO]

"LHCb Outer Tracker technical design report". CERN/LHCC 2001-024 LHCb TDR 6 (14th September 2001).

[TDRp]

"LHCb TDR" www page. http://lhcb.web.cern.ch/lhcb/TDR/TDR.htm. Also see [TDR00].

[TDR00]

"LHCb RICH technical design report". CERN/LHCC/2000-0037, LHCb TDR 3 (7 September 2000).

[TDR03opt]

"LHCb technical design report, re-optimised detector design and performance". CERN/LHCC/2003-03, LHCb TDR 9 (9 September 2003).

[TDRtrig]

"LHCb trigger system". CERN LHCC 2003-031, LHCb TDR 10 (July 30 2003) Draft 4.0.

[TP98]

"LHC-b technical proposal". CERN/LHCC 98-4 LHCC/P4 20 (February 98).

[TRU00]

U.Trunk. "Development and characterisation of the radiation tolerant HELIX128-2 readout chip for the HERA-B micro-strip detectors". Doctoral thesis, Universitat Heidelberg, Germany (2000).

[TTCweb]

TTC Website, http://ttc.web.cern.ch/TTC/intro.html. Also see [TTCstat].

[TTCstat]

M.Ashton et al. "Timing, trigger and control systems for LHC detectors". Status report on the RD-12 Project, CERN/LHCC 2000-002, LEB Status Report/RD12 (3rd of January 2000).

[VCSEL]

General VCSEL information, www site. http://www.lasermate.com/transceivers.htm. Also see [HP].

[VTDR02]

"LHCb VELO TDR". CERN/LHCC 2001-011 LHCb TDR 5 (31st of May 2001).

[WEL00]

H.Weller et al. "Low-noise charge-sensitive readout for pyroelectric sensor arrays using PVDF thin film". Sensors and Actuators 85(2000) 267-274 (9th February 2000).

[YPS99]

T.Ypsilantis and J.Seguinot. "Evolution of the RICH technique". Nuclear instrumentation and methods in physics research A., accelerators, spectrometers, detectors and associated equipment, volume 433, issues 1-2, 21st August 1999, pages 1-16.

[ZEU]

"The ZEUS detector technical proposal". The ZEUS collaboration (March 1986). Also see the ZEUS home page http://www-zeus.desy.de.