UNIVERSITY OF PADOVA

Department of Electronics and Computer Engineering (DEI)

Ph.D in:
Electronics and Communication Engineering

years 1996-1999

Dissertation

Development of the digital read-out system for the CERN Alice pixel detector

Coordinator: Full Prof. Silvano Pupolin

Supervisor: Full Prof. Enrico Zanoni

Ph.D student: Tullio Grassi

December, the 31st 1999

UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Elettronica e Informatica della Facoltà di Ingegneria

DOTTORATO DI RICERCA IN:
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

XII CICLO

Tesi:

Sviluppo del sistema digitale di lettura del rivelatore a pixel di Alice (CERN)

(titolo originale: Development of the digital read-out system for
the CERN Alice pixel detector)

Coordinatore: Ch.mo Prof. Silvano Pupolin

Supervisore: Ch.mo Prof. Enrico Zanoni

Dottorando: Tullio Grassi

31 dicembre 1999

Acknowledgement

Over the three years in which the work presented in this thesis was done I had the chance to work in three different groups. My original team is the Microelectronic group of the University of Padova. Here I thank very much my Ph.D. supervisor Prof. Enrico Zanoni, but also the professors Sandro Paccagnella and Andrea Neviani for the interest showed in my activity. I do not forget the help and friendship received by Andrea Candelori, Marco Ceschia and the other students of the group.

Thanks also to the group III of the Padova section of INFN, leaded by Prof. Morando, as they first introduced me to the field of the electronics for high energy physics. From this group I mention Nicola Carrer as he was able to explain to an engineer the mysteries of the quark nature! But also Gianfranco Segato for the time spent in the laboratory and Sandro Martini for the time spent outside the lab!

For what I learned in term of circuit and system level design, I owe a special thank to Fabio Formenti and the components of the ED/EP group of CERN: Luciano Musa for the precious advice that he often gave me, Michel Morel and Mike Burns for helping me when I felt lost facing a mountain of cables and instrumentations, Prof. Jose Toledo "Curro" for our useful battles and for teaching me electronics in Spanish, Alex Kluge for the review of my thesis, Ken Wyllie and the members of the CERN/MIC group for our fruitful "interface".

Finally I would like to thank my family for their affectionate and unselfish support

and

my girlfriend Ludmila who could not care less about my work, showing that one can live (very good) also without electronics and physics.

Table of Contents

Abstract

In order to gain new experimental insight at the TeV energy scale, CERN (Geneva) will build the Large Hadron Collider (LHC), a new collider machine operating at a maximum center-of-mass energy of 14 TeV (in the p+/p+ interactions). The accelerator can operate in a heavy ion collision mode achieving a center-of-mass energy of ~5.5 TeV. The experimental environment at LHC is characterized by a high crossing rate of the particle bunches (one every 25 ns for p+/p+) and high levels of radiation. Therefore stringent requirements are imposed on the performance of detectors at LHC. Such a particle physics environment calls for dedicated hardware/software solutions with specific constraints, such as radiation tolerance, limited amount of material and limited power dissipation.

One of the particle physics experiments carried out in LHC is ALICE (A Large Ion Collider Experiment). The ALICE detector will face a very high density of tracks of particles (a multiplicity of 8000 charged particles per unit of rapidity, that implies a maximum density of ~90 tracks/cm^2) and it comprises an enormous number of electronic channels (~ 2x10^7). Most of these channels (~15x10^6) come from the two layers of pixel detector, that produce a data rate of 75 Gbyte/s.

The pixel detector system performs the bi-dimensional high-speed detection of the position of the tracks of ionizing particles, with a spatial resolution of ~ 12 µm.

It comprises the silicon detector cells, the mixed front-end electronics (Pixel chip) and a digital module (Pilot system) located in the detector front-end, that accomplishes high level functions, typical of an external Data Acquisition stage.

My main responsibilities were to contribute to the definition of the interfaces between the sub-systems and the design of the Pilot system.

For such a complicated project like the ALICE detector, the definition of the design specifications and of the interfaces between the sub-systems is an important part of the study, in order to guarantee the feasibility of the project. Hence a closed collaboration between designers is often required, including the involvement of a designer in some design issues of the neighboring systems. This is why the pixel chip is extensively presented in chapter 2 (besides several contributions in the measurements discussed in the same chapter).

The design of the Pilot system (described in chapter 3) challenges the processing of the huge amount of pixel data (in the experiment there will be 15 million pixels, managed by no more than 240 Pilot modules); moreover, the system plays the role of the master in the pixel fast-control protocol.

The key idea to handle the pixel data is based on the low probability of a pixel cell being hit by a particle, due to the high granularity of the detector. A hit is repre-

sented by a logic value one in the pixel matrix received by the Pilot system: an on-line zero-suppression operation allows the full address encoding of every hit. This guarantees the required data compression rate, keeping a simple and reliable hardware implementation of the algorithm. After the encoding, an output stage running a CMI-encoded serial stream on a 40 meters copper cable transmits the data to the following stage (the router, that will be located outside the detector) at a bit rate of 155 MHz, thus minimizing the number of output links. The proposed architecture allows to reduce the clock frequency in the rest of the system. This avoids the risks and side-effects of the high-frequencies in such a harsh environment. This goal is reached also using state-of-the-art technologies, like the recent LVDS standard, for low-voltage differential binary transmission.

As mentioned above, the interfaces between the detector sub-systems were an important part of the investigations, that are still in progress. As a consequence the design specifications of the single sub-systems are subject to several modifications in order to test and optimize the detector protocols. That is why, during the present stage of the preparation, a flexible implementation of the Pilot logic is required. Hence the system is prototyped on a board, based on programmable logic devices. In spite of the technology used, a 310 MHz clock has been successfully routed inside a programmable logic device.

A dedicated set-up (based on LabView) for the testing of the board has been built, including two new custom boards (a receiver card for the test of the 155 MHz serial link, and a second one to interface the Pilot board with a system for logic testing). This set-up (described in chapter 4) allowed to check the correct behavior of the logic, in agreement with the Verilog simulations carried out during the design of the system.

Once the specifications will be fixed, the final version of the Pilot system for the Alice experiment (supposed to begin in 2005) is foreseen to be on a single chip (ASIC). For the design migration to the ASIC technology, the use of automatic translation tools is under investigation.

In addition, the board implementation already satisfies the requirements of other experiments (so far, the NA6i experiment at CERN-SPS).

Sommario

Allo scopo di ottenere nuove conoscenze sperimentali a energie dell'ordine dei TeV, il CERN *(Ginevra) costruirá LHC (Large Hadron Collider, Grande collisore di adroni), una nuova macchina acceleratrice operante a una energia del centro di massa di 14 TeV (per interazioni p^+/p^+). Il collisore potrá funzionare in modalitá ioni pesanti, raggiungendo un'energia del centro di massa di ~5.5 TeV. L'ambiente sperimentale di LHC é caratterizzato da un'alta frequenza di transito dei pacchetti di particelle (uno ogni 25 ns per p^+/p^+) e da alti livelli di radiazione. Quindi sono imposte caratteristiche stringenti sulle prestazioni dei rivelatori di LHC. Un tale ambiente di fisica delle particelle richiede soluzioni hardware/software dedicate e con vincoli specifici, quali tolleranza alle radiazioni, e con limiti sulla quantitá di materia e sulla potenza dissipata.*

Uno degli esperimenti realizzati in LHC é ALICE *(A Large Ion Collider Experiment). Il relativo rivelatore affronterá un'altissima densitá di traiettorie di particelle (una molteplicitá di 8000 particelle cariche per unitá di rapiditá, che implica una densitá massima di ~90 traiettorie/cm^2) e comprende un enorme numero di canali elettronici (~2x10^7). La maggior parte di questi canali (~15x10^6) viene dai due livelli di rivelatore a pixel, che producono un data rate di 75 Gbyte/s.*

Il rivelatore a pixel rivela i punti di transito delle particelle ionizzanti, in maniera bidimensionale, ad alta velocitá e con una precisione di ~12µm. Esso comprende le celle di rivelatore al silicio, l'elettronica di front-end analogico-digitale (Pixel chip) e un modulo digitale (Pilot system) posto nel front-end del rivelatore, che realizza funzioni di alto livello, tipiche di uno stadio esterno di acquisizione dati. Le mie responsabilitá principali sono state un contributo alla definizione delle interfacce tra sottosistemi e il disegno del Pilot system.

Per un progetto complicato come il rivelatore di ALICE*, la definizione delle specifiche progettuali e delle interfacce tra i sottosistemi é un'importante parte dello studio, per garantire la fattibilitá del progetto. Quindi é spesso richiesta una collaborazione stretta tra progettisti, compreso il coinvolgimento di un progettista in alcune problematiche dei sistemi prossimi al suo. Per questa ragione il Pixel chip é presentato ampiamente nel capitolo 2 (oltre a vari contributi alle misure discusse nello stesso capitolo).*

Il disegno del Pilot system (descritto nel capitolo 3) affronta l'elaborazione dell'enorme quantitá di dati dei pixel (nell'esperimento ci saranno 15 milioni di pixel, gestiti da non piú di 240 moduli Pilota); inoltre esso ha un ruolo di "master" nel "fast-control" dei pixel (la parte di controllo funzionante simultaneamente alla lettura dei dati).

L'idea chiave per gestire i dati dei pixel é basata sulla bassa probabilitá che una cella

-

di pixel sia attraversata da una particella (evento chiamato hit), a causa dell'elevata granularitá del rivelatore. Un hit é rappresentato da un bit di valore logico uno nella matrice di pixel ricevuta dal Pilot system: un'operazione di soppressione degli zeri realizzata in linea permette la completa codifica per indirizzo di ogni hit. Ció garantisce il rapporto di compressione richiesto, mantenendo una semplice e affidabile implementazione hardware dell'algoritmo. Dopo la codifica, uno stadio di uscita invia un flusso seriale con codifica CMI, su un cavo di rame di 40 metri e trasmette i dati agli stadi successivi (il router, che sará posto fuori dal rivelatore) a un bit rate di 155 MHz, minimizzando il numero di connessioni di uscita. L'architettura proposta consente di ridurre la frequenza di clock nel resto del sistema. Ció evita i rischi e gli effetti collaterali delle alte frequenze in un ambiente cosí ostile. Questo fine é raggiunto anche utilizzando tecnologie allo stato dell'arte, quale il recente standard LVDS, per la trasmissione binaria differenziale in bassa tensione.

Come menzionato sopra, la definizione delle interfacce tra i sottosistemi del rivelatore é stata una parte importante delle ricerche, che é ancora in corso. Di conseguenza le specifiche progettuali di ogni singolo sottosistema sono soggette a parecchie modifiche, per testare e ottimizzare i protocolli del sistema di rivelazione. Questa é la ragione per cui, nella fase attuale della preparazione, é richiesta un'implementazione flessibile della logica pilota. Quindi il sistema é stato realizzato con un prototipo su una scheda basata su circuiti logici programmabili. Nonostante la tecnologia usata, un clock a 310 MHz é stato implementato con successo in un dispositivo logico programmabile.

Per testare la scheda é stato costruito un set-up dedicato (basato su LabView), e che comprende due nuove schede dedicate (una scheda ricevitore per testare il collegamento seriale a 155 MHz, e una seconda scheda per interfacciare il sistema Pilota con un sistema di testing digitale). Questo set-up (descritto nel capitolo 4) ha permesso di verificare il comportamento corretto della logica, in accordo con le simulazioni Verilog realizzate nella fase progettuale.

Una volta che le specifiche progettuali saranno fissate, é previsto che la versione finale del Pilot system per l'esperimento ALICE (inizio pianificato per il 2005) sia su un singolo chip (ASIC).

Inoltre, l'implementazione su scheda giá soddisfa le richieste di altri esperimenti (per ora, l'esperimento NA6i al CERN-SPS).

CHAPTER 1 Introduction: ALICE at CERN

In this introduction an overview on the experimental high-energy physics environment will be provided. The experiments carried out in CERN, and particularly the ALICE experiment will be seen in detail.

It will be shown how the peculiar requirements of this environment lead to develop very specific systems. These systems make use of advanced technologies of many different fields (including hardware and software) and, when required, contribute to new developments in these fields.

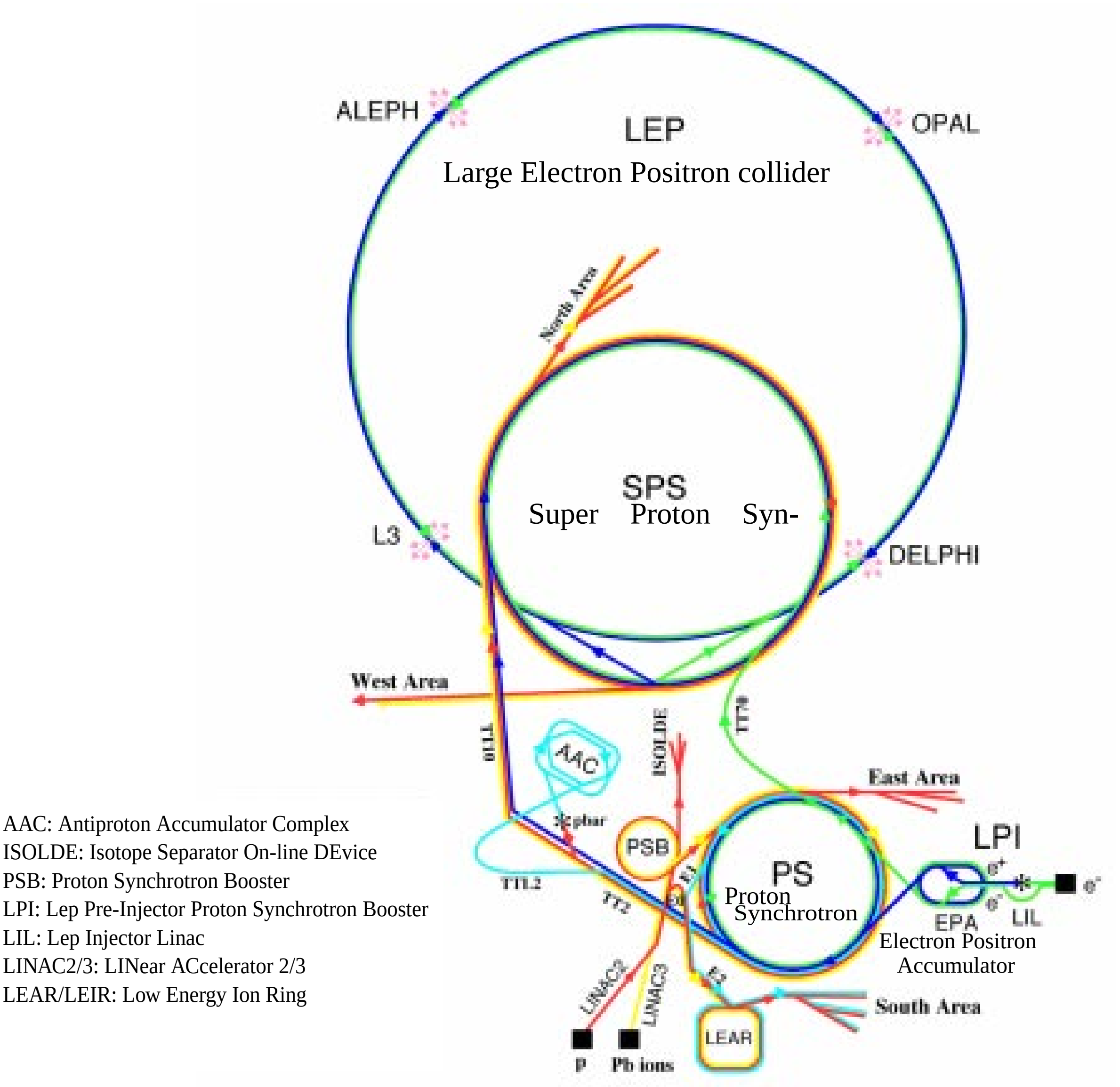
1.1 CERN

CERN is the European Laboratory for Particle Physics, the world's largest particle physics centre. Founded in 1954, the laboratory was one of Europe's first joint ventures, and has become a shining example of international collaboration, with its 20 Member States. CERN explores what matter is made of, and what forces hold it together. On the other hand, this quest for pure knowledge drives new developments in many fields including electronics and other information technologies.

By accelerating particles to very high energies and smashing them into targets or into each other, physicists can unravel the forces acting between them. CERN's accelerators are amongst the world's largest and most complex scientific instruments. Accelerators come in two types, linear and circular. Accelerators use powerful electric fields to push energy into a beam of particles. Magnetic fields are used to keep the beam tightly focused, and in circular machines to steer the particles around the ring. Linear machines push energy into the beam all along the accelerator's length. The longer the machine, the higher the final energy.

CERN's accelerator complex (Figure 1-1) is the most versatile in the world and represents a considerable investment. It includes particle accelerators and colliders, can handle beams of electrons, positrons, protons, anti-protons, and "heavy ions" (the nuclei of atoms). Each type of particle is produced in a different way, but then passes through a similar succession of acceleration stages, moving from one machine to another. The first steps are usually provided by linear accelerators, followed by larger circular machines. CERN has 10 accelerators altogether, the biggest being the Large Electron Positron collider (LEP) and the Super Proton Synchrotron (SPS). CERN's first operating accelerator, the Synchro-Cyclotron, was built in 1954, in parallel with the Proton Synchrotron (PS). The PS is today the backbone of CERN's particle beam factory, feeding other accelerators with different types of particles. The 1970s saw the construction of the SPS, at which Nobel-prize winning work was done in the 1980s. The SPS continues to provide beams for experiments and is also the final link in the chain of accelerators providing beams for the 27 kilometre LEP machine. CERN's next big machine, supposed to start operating in 2005, is the Large Hadron Collider (LHC).

FIGURE 1-1. **CERN Accelerator complex**



AAC: Antiproton Accumulator Complex
ISOLDE: Isotope Separator On-line DEvice
PSB: Proton Synchrotron Booster
LPI: Lep Pre-Injector Proton Synchrotron Booster
LIL: Lep Injector Linac
LINAC2/3: LINear ACcelerator 2/3
LEAR/LEIR: Low Energy Ion Ring

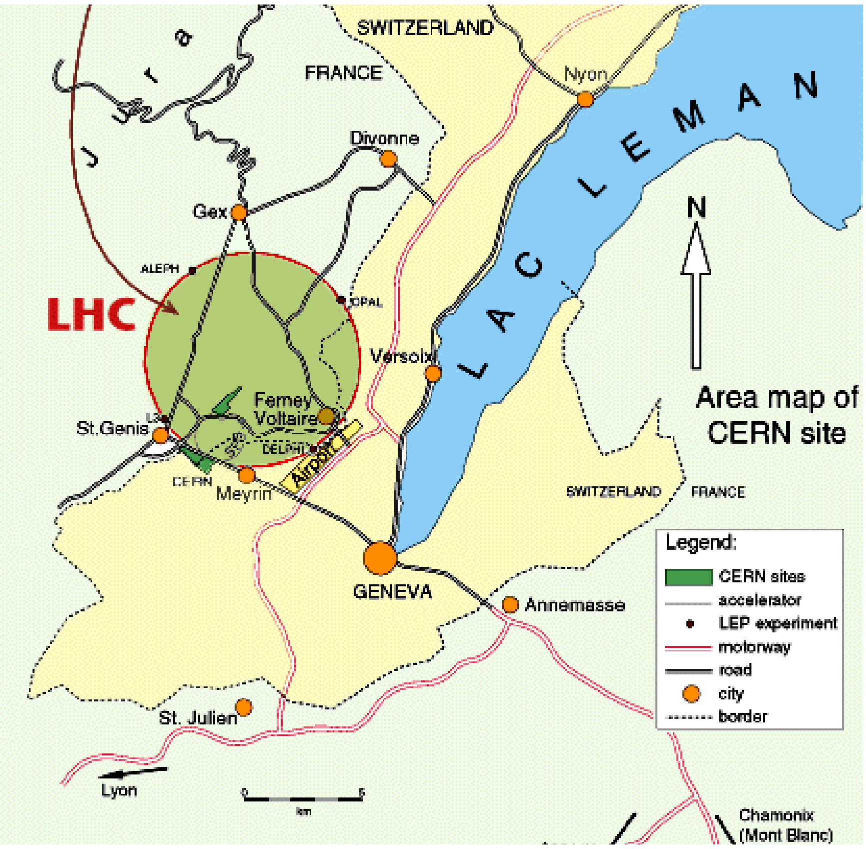
1.2 *THE LARGE HADRON COLLIDER (LHC)*

One of the main events in the field of particle physics at the beginning of the next century will be the construction of LHC. This machine will be installed into the existing Large Electron Positron (LEP) tunnel at the CERN laboratory across the Franco-Swiss border west of Geneva, as shown in Figure 1-2.

The LHC [1] is a two-ring machine with a total length of 27 km which will allow the head-on collision of proton beams, each with an energy of ~7 TeV, and beam collisions of heavy ions, such as lead and calcium, with a total collision energy in centre-of-mass of ~ 5.5 TeV.

The pre-existence of the LEP tunnel and of the CERN accelerator chain for particle injection will make the cost of the machine affordable. The LHC will be the largest hadron Collider in the world and will extend the energy frontier of physics, offering a further insight in the basic mechanisms of nature, as for example the mechanism that gives masses to the fundamental constituents of matter. LHC is a very challenging project because of the engineering constraints and of the physics requirements. As a consequence of the fixed length of the ring, imposed by the choice of using the existing LEP tunnel, the limit to the acceleration of the colliding particles comes from the strength of the maximum bending magnetic field. A new generation of superconducting magnets will provide the magnetic field necessary to achieve the TeV energy range. A large cryostat will keep the magnets at the working temperature of 1.9 K. The statistical significance of the scientific results will be obtained by means of the exceptional luminosity[1] of LHC, which will insure a good number of interesting events per year. On the other hand, it will lead to a very high radiation environment for the particle detectors. Therefore, the survival of the detectors in this harsh radiation environment during 10 years of operation planned for the machine is a challenging issue for the detectors.

FIGURE 1-2. **The LEP/LHC tunnel map.**



[1] Luminosity is the flux of particles per unit of area and per unit of time

In the LHC, the energy available in the collisions between the constituents of the protons (the quarks and gluons) will reach the TeV range, about 10 times that of LEP. The particle cross section (related to the probability of having an interaction) decreases like $1/E^2$, therefore a very high luminosity is required for proton-proton (pp) collision in order to maintain a physics program as effective as in LEP. At LEP the luminosity is around $L = 10^{32}$ cm^{-2} s^{-1}, whereas it will reach $L = 10^{34}$ cm^{-2} s^{-1} at LHC. This value will be obtained by filling each ring with 2835 bunches of 10^{11} particles, injected in the LHC by the Super Proton Synchrotron (SPS) at the energy of 0.45 TeV. The bunch space will be 7.48 m, corresponding to an interval between two successive bunches of 25 ns at the collision energy. An 8.33 Tesla magnetic field provided by 1296 superconducting dipole magnets over a total magnetic length of ~15 km will guide the ions in the LHC orbits. Superconducting quadrupole correctors will make the orbit corrections to the spurious non-linear components of the guiding and focusing magnetic fields of the machine and allow the recovering of the required beam density after interactions. Special orbit correctors will be used, as sextupole, octupole and decapole magnets. The luminosity lifetime will be about 10 hours. The two beam pipes of LHC and the superconducting coils will be hosted in the same superfluid helium cryostat at the temperature of 1.9 K.

1.2.1 Physics at LHC

The LHC machine and the planned experiments have been designed to study the extrapolation of the present knowledge in particle physics to the LHC energy scale and to detect the signatures of a new and possibly unexpected physics above the TeV energy threshold. This will allow scientists to penetrate still further into the structure of matter and recreate the conditions prevailing in the Universe just 10^{-12} seconds after the “Big Bang” when the temperature was 10^{16} degrees. The Standard Model [2] describes very well the results of present experiments. However some fundamental questions remain still unanswered. Four experiments at LHC will study these new physics domains: ALICE (A Large Ion Collider Experiment), ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid) and LHCb (B-physics at LHC).

ALICE is an experiment mainly devoted to the formation of the quark-gluon plasma (QGP) by mean of heavy ion (Pb and Ca) collisions at a centre-of- mass energy of ~5.5 TeV per nucleon pair and a luminosity of about 2 10^{27} cm^{-2} s^{-1}. This conditions will create energy densities of 5-8 GeV fm^{-3}, that are above the QGP threshold [3].

CMS and ATLAS will run pp collision at the maximum centre of mass energy of 14 TeV and up to the maximum luminosity. They will study the origin of the mass of the Z and W vector bosons of the electroweak interactions. The most accredited explanation is the existence of a Higgs field. One of the main goals of LHC is the discovery of the associated gauge boson, the Higgs particle. These two experiments are designed to cover the different physics signatures of the Higgs, consisting in the identification of the predicted decay modes, with statistical significance over all the mass range. This requires very high luminosity because the cross section for heavy particle scales inversely with the square root of their masses.

LHCb is devoted to the study of CP violation in B-meson decays.

1.3 ALICE - A Large Ion Collider Experiment

1.3.1 Purpose of the experiment

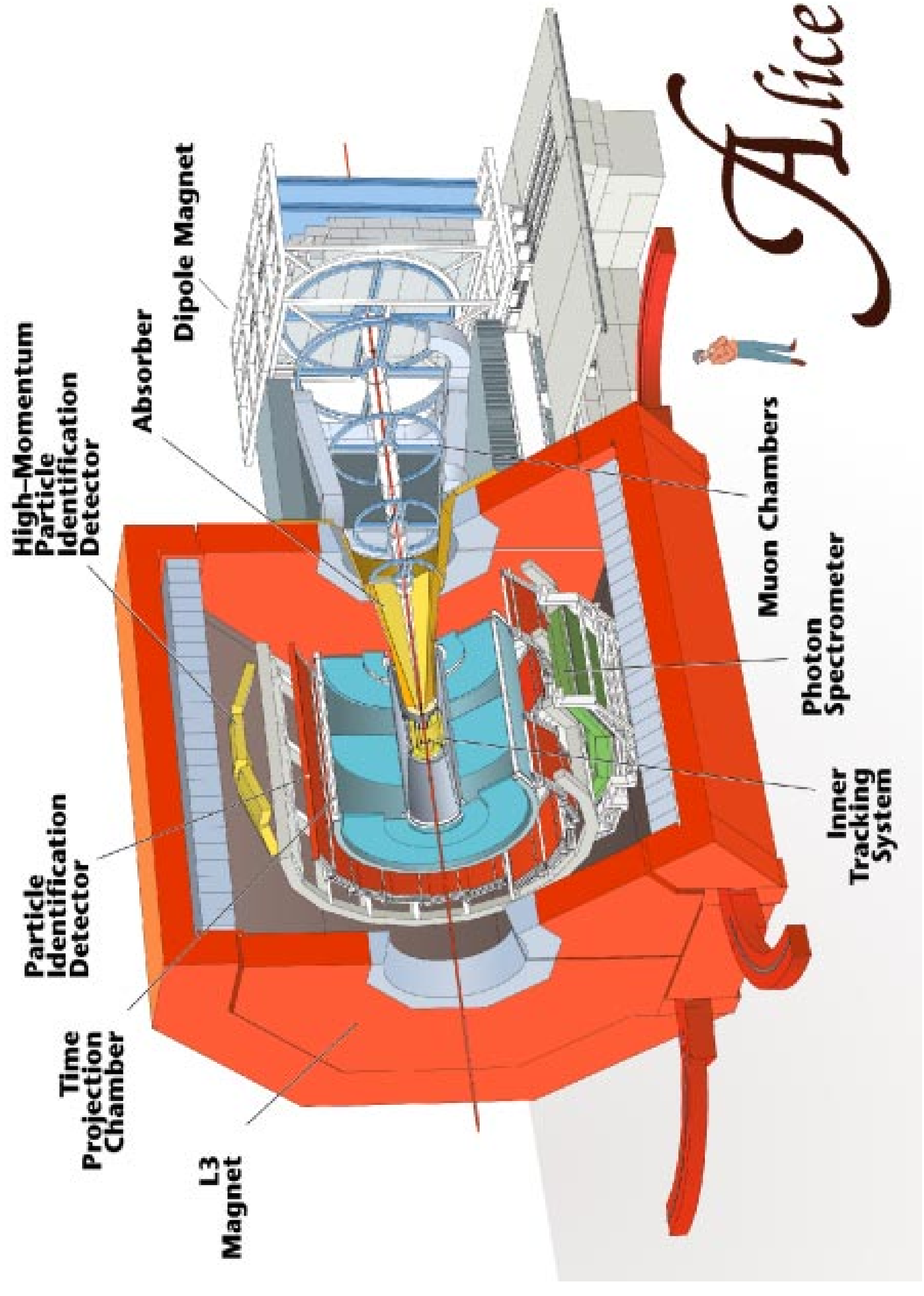
A very simplified idea of this research domain is the will of studying the quarks by breaking a proton (or a neutron). The attempt of having a single, isolated quark in order to study it (like in the past happened for most of the known particles: electrons, protons, neutrons, etc.) would be vain. This is due to the strong interaction that binds the quarks together: this interaction has a minimum at a certain distance between the particles and then increases with the distance. This means that if we begin to supply energy to a system of quarks, they will not go very far from each other. If we increase the energy supplied, when the energy is comparable to the mass of the quarks (being $E=mc^2$), the original system will break, but new quarks will be created from the energy, so that every fragment of the original system will be constituted by more than one quark. The only way to de-confine a quark is in a very dense medium, a plasma, where the quark can “float” together with the gluons, that are the carrier particles of the strong interaction (like the photons are the carrier particles of the electromagnetic interaction). Hence, more complicated approaches and theories are required.

The ALICE Collaboration proposes to build a dedicated heavy-ion detector to exploit the unique physics potential of nucleus-nucleus interactions at LHC energies. The aim is to study the physics of strongly interacting matter at extreme energy densities, where the formation of a new phase of matter, the Quark-Gluon Plasma (QGP), is expected. The existence of such a phase and its properties are a key issue in Quantum-Chromo-Dynamics (theory describing the strong interaction) for the understanding of confinement behaviour. For this purpose, a comprehensive study of the hadrons (like protons and neutrons), electrons, muons and photons produced in the collision of heavy nuclei will be carried out.

1.3.2 The ultra-relativistic heavy-ion collisions

With the advent of ultra-relativistic heavy-ion collisions in the laboratory, at Brookhaven and CERN in 1986, a new interdisciplinary field has emerged from the traditional domains of particle physics and nuclear physics. In combining methods and concepts from both areas, the study of heavy-ion reactions at very high energies denotes a new and original approach in investigating the properties of matter and its interactions. In high-energy physics, the matter involved in interactions consists mostly of single particles (hadrons/quarks). In contrast, at the level of nuclear physics the strong interaction is shielded and can only be described in effective or phenomenological theories, whereas the matter consists of extended systems exhibiting collective features. Combining the elementary-interaction aspect of high-energy physics with the macroscopic-matter aspect of nuclear physics, the subject of heavy-ion collisions is the study of bulk matter consisting of strongly interacting particles (hadrons/quarks).

FIGURE 1-3. **ALICE Detector. The Inner Tracking System contains the Silicon Pixel sub-detector.**



What makes this field particularly interesting is the Quantum-Chromo-Dynamics prediction that at high energy densities matter should undergo a phase transition to an entirely new state, the quark-gluon plasma (QGP). At low energy densities, quarks and gluons are bound (confinement) by the strong force into hadrons (e.g. protons and neutrons). In addition, the quarks acquire a large effective mass via interactions between themselves. When the energy density is increased, either by increasing the temperature ("heating") or the matter density ("compressing"), a phase transition should occur towards the QGP, where quarks and gluons are deconfined. The study of the strongly interacting matter is not only of interest to study and test the Quantum-Chromo-Dynamics, but it might also shed light on such fundamental questions as the nature of confinement itself and on the process of spontaneous symmetry breaking, which is made responsible for the origin of the "effective" quark masses. The early Universe presumably underwent exactly this phase transition 10 microseconds after the Big Bang. Critical phenomena that can occur close to a phase boundary, for example long-range density fluctuations (as in condensing water!), might have a bearing on important aspects of cosmology, such as nucleosynthesis, dark matter, and the large-scale structure of the Universe. In astrophysics, the dynamics of supernova explosions and the stability of neutron stars (10 times more dense than normal nuclear matter) depends on the compressibility (and therefore on the equation of state) of nuclear matter. It is even speculated that the core of neutron stars may consist of cold QGP. The study of extreme states of matter created in high-energy nuclear collisions thus provides us with an opportunity to gain insight into many important aspects in different fields of physics.

1.3.3 The Design of the ALICE detector

ALICE (A Large Ion Collider Experiment, [4]) is an experiment at the Large Hadron Collider (LHC) optimized for the study of heavy-ion collisions, at a centre-of-mass energy ~5.5 TeV per nucleon. The prime aim of the experiment is to study in detail the behaviour of nuclear matter at high densities and temperatures.

A major technical challenge is imposed by the large number of particles created in the collisions of lead ions. There is a considerable spread in the currently available predictions for the multiplicity of charged particles produced in a central Pb-Pb collision. The design of the experiment has been based on the highest value, 8000 charged particles per unit of rapidity[1], at midrapidity (~ center of the system). This multiplicity (related to the track density as a function of the position) dictates the granularity of the detectors and their optimal distance from the colliding beams.

The central part, which covers 45° over the full azimuth, is embedded in a large magnet with a weak solenoidal field. Outside of the Inner Tracking System (ITS), there are a cylindrical TPC (Time Projection Chamber detector) and a large area PID (Particle IDentification) array of time-of-flight (TOF) counters. In addition, there are two small-area single-arm detectors: an electromagnetic calorimeter (Photon Spectrometer, PHOS) and an array of Ring Imaging

[1] The rapidity indicates how the particles coming out from a hadronic collision spread in the direction of the beam (z, that is the axis of the cylindrical Alice detector of Figure 1-3). This unit is introduced because this spread is not homogeneous expressed in angles.

Cherenkov Counters (RICH) optimized for high-momentum inclusive particle identification (HMPID).

1.3.4 Trigger logic in ALICE

In a high energy physics experiment, only a fraction of occurring events provides useful information. The trigger system evaluates the event on-line and provides an accept signal when the event is relevant. Only those event are recorded and are available for later off-line analysis.

The trigger [7] is constituted by a collection of devices, usually a combination of electronics and informatics components, typically associated with some particle detector(s). The event as seen by the trigger must allow one to evaluate conditions that are predicted to be characteristic for interesting events; these conditions are often called the event signature. Conditions may be as simple as identifying a charged track passing through a few scintillation counters within a time gate (typical trigger in a test beam), or as complicated as effective mass criteria between identified leptons that have to be satisfied in high-energy collisions (e.g. the intended triggers at the 40 MHz Large Hadron Collider at CERN).

In many experiments, data taking, through the dead time it causes, is a critical factor limiting statistics and hence physics potential; an efficient trigger system is needed for transmitting data that have a high probability of containing good physics, and rejecting, with respect to the possibilities of the detector, all or most of the background, i.e. trivial physics or non-physics events.

In large experiments, triggers are implemented in multiple levels; typically, a fast and synchronous trigger (“level 1”) identifies candidate events from a subset of events, reducing the rate by some factor. Subsequently, data are digitized, transmitted to more permanent buffers and to the next (asynchronous) trigger, and more complex algorithms based on more complete data reduce the rate again (“level 2”). Eventually, after perhaps a third and fourth level, the entire event is transmitted to permanent storage.

The ALICE trigger system [8] is designed to operate in several different running modes switched by software. It must also allocate trigger types which have widely differing rates, and which do not, in general, read out the same set of detectors. It does so by having a well-specified signal protocol with each detector, treating it independently from all other detectors so as to allow, for example, single detector triggers for test purposes.

The trigger system gives a decision for every bunch crossing. The earliest ($L0_e$) decision is made after 1.2 μs, and is used to strobe the detectors.

Apart from an early interaction trigger signal with very little selective power (which is used in $L0_e$), a trigger signal is not delivered until 5.5 μs after the event has taken place. This poses an important constraint, as it implies that no major rate reduction can occur before this, in order not to lose effective luminosity. However these triggers are subject to a 100 μs past-future protection interval, which means that ultimately ~63% of them will be rejected because of pile-up in Pb-Pb interactions.

At this point, all the currently planned trigger selections will have taken place, and trigger rates can drop to close to their final values. This trigger level is distributed using the RD12 TTC system[6], which is also used to distribute event numbers. A modification has to be made to the standard TTC method, to take into account the fact that a given detector does not contribute to every event. The proposed solution is to label events using the bunch-crossing counter (12 bits) and a 24-bit orbit counter so as to have a period of a few hours during which the combined system of counters does not overflow. The bunch crossing counter is available on the TTCrx chip; the orbit counter can be implemented externally, but a request has been made to include an internal orbit counter on the next version of the TTCrx chip.

The final (level 2) decision is made at the end of the TPC drift time (at 100 µs), so as to ensure that the event is not spoiled by pile-up. An event can be rejected earlier than this, for example if pile-up occurs early in the TPC drift time. Data transfer from the detector electronics to the data acquisition system is initiated by a successful L2 trigger (called L2 Yes or Accept), which means that the data transfer for the event cannot be stopped once it has started.

As mentioned above, the trigger system allows events to be recorded where not all the detectors are read out. The principal physics application of this is to allow more frequent triggering of the dimuon arm. 'Dimuon triggers' need only the pixels and the dimuon arm detectors, and will be taken in parallel with triggers of the whole detector so as to achieve acceptable overall rates for the dimuon arm. The maximum L1 estimated rate is 1.1 KHz, and the maximum L2 estimated rate is 700 Hz.

Table 1-1 shows the expected latency of the trigger levels. Since, in general, the delay to each sub-detector on the return path will be different, the latency has been computed for the case of a signal received at the centre of ALICE, i.e. in the position of the ITS.

TABLE 1-1. Trigger Latencies

Trigger Level	Type	Latency	Action
Level 0	L0e	1.2 µs	Strobe ITS strip
Level 0	L0d	2.7 µs	Strobe ITS drift
Level 1	L1	5.5 µs	Strobe ITS pixel and prepare transfer to DAQ
Level 2	L2Y	100 µs	Transfer data to DAQ via links
Level 2	L2N	≤ 100 µs	Abort or cancel the transfer to DAQ

Each sub-detector communicates a single BUSY signal to the central trigger. The BUSY signal indicates that the sub-detector should not receive further triggers. This is typically because it has no space to register the next event.

1.3.5 Tracking in ALICE

Track finding in heavy-ion collisions at the LHC presents a big challenge, because of the extremely high track density (up to 90 tracks/cm^2). In order to achieve a high granularity and a good two-track separation ALICE uses three-dimensional hit information, wherever feasible, with many points on each track and a weak magnetic field. The need for a large number of points on each track has led to the choice of a Time Projection Chamber as a tracking system (See Figure 1-3), in spite of its drawbacks, concerning speed and data volume. The minimum possible inner radius of the TPC (r_{in} ~ 90 cm) is given by the maximum acceptable hit density.

At smaller radii, and hence larger track densities, tracking is taken over by the Inner Tracking System (ITS).

The ITS consists of six cylindrical layers of silicon detectors. The number and position of the layers are optimized for efficient track finding. In particular, the outer radius is determined by the track matching with the TPC, and the inner one is the minimum compatible with the radius of the beam pipe (3 cm). The silicon detectors feature the high granularity and excellent spatial precision required.

Because of the high particle density, the four innermost layers (r < 24 cm) must be truly two-dimensional devices. For this task silicon pixel and silicon drift detectors were chosen. The outer two layers at r ~ 45 cm, where the track densities are below 1 cm^{-2}, will be equipped with double-sided silicon microstrip detectors.

1.3.6 Design considerations

The basic functions of the ITS are:

- determination of the primary vertex and of the secondary vertices[1] necessary for the reconstruction of particle decays,
- particle identification and tracking of low-momentum particles,
- improvement of the momentum and angle measurements of the TPC.

The following factors were taken into consideration for the design of the ITS:

Material budget: The momentum and track resolution for particles with small transverse momenta are dominated by multiple scattering effects in any existing tracking detector. Therefore the amount of material in the active volume has to be reduced as much as possible. However, the thickness of silicon detectors used to measure ionization densities must be approximately 300 µm to guarantee the required signal-to-noise ratio. In addition the detectors must overlap in order to reach full coverage within the acceptance window. Taking also into account the incidence angles of tracks, the detectors represent a thickness of 0.4% of X_0 (= radiation length[2]). The aim set in the ALICE technical proposal was to reduce the thickness of the additional material in the active volume, i.e. electronics, cabling, support structure and cooling system, to a comparable effective thickness.

Spatial precision and granularity: The granularity of the detectors in the ITS is dictated by the track densities expected. The system is designed for a maximum track density of 8000 tracks per unit of rapidity, the upper limit of the current theoretical predictions. Therefore up to 15000 tracks will have to be detected simultaneously in the ITS (In the case of the innermost pixel detector, this turns out into a density of ~ 90 tracks/cm^2). Keeping the occupancy of the

[1] Primary vertex is the collision point (were the tracks of collision fragments start), secondary vertex is a point were a single particle decays, generating new particles.

[2] The radiation length X_0 is a parameter for the probability of scattering and energy loss of a particle in a medium. These phenomena are proportional to: $e^{-1/(X_0)}$.

system at the level of a few per cent requires several million effective cells in each layer of the ITS. The resolution of the impact parameter measurement is determined by the spatial resolution of the ITS detectors. Therefore the ITS detectors have a spatial resolution of the order of a few tens of µm, with the best precision (12 µm) for the pixel detectors closest to the primary vertex. In addition, for momenta larger than 3 GeV/c, relevant for the detection of some decay products, the spatial precision of the ITS becomes an essential element of the momentum resolution. This requirement is met by all layers of the ITS with a point resolution about one order of magnitude better than that of the TPC, which in turn provides many more points.

Radiation levels: The ionizing radiation dose received by the detector components was calculated using Monte Carlo techniques. The total dose received during the lifetime of the experiment varies from a few krad for the outer parts of the ITS to about 150 krad for the inner parts as shown in Table 1-2. Detailed calculations can be found in Ref. [5].

TABLE 1-2. **Radiation dose and neutron fluence for each of the ITS detector layers, calculated for ten years of operation, including p-p and Ca-Ca runs.**

Layer	Detector	Radius (cm)	Cumulated dose (krad)	Neutron fluence (10^{11}cm^{-2})
1	pixel	4	130	3.2
2	pixel	7	39	3.1
3	drift	15	13	3.5
4	drift	24	5	3.3
5	strip	39	2	3.7
6	strip	44	2	3.3

Each of the sub-detectors is designed to withstand the ionizing radiation doses expected during ten years of operation. The neutron fluence is approximately 3 x 10^{-11} cm^{-2} throughout the ITS, which does not cause significant damage to the detectors or the associated electronics. Where necessary, the components used in the ITS design are tested for their radiation hardness up to the expected doses.

Read-out rate: The ALICE system will be used in two basically different read-out configurations, operated simultaneously with two different triggers. The centrality trigger activates the read-out of the whole of ALICE, in particular all layers of the ITS, while the trigger of the muon arm activates the read-out of a subset of fast read-out detectors, including the two inner layers of the ITS (pixel detector).

1.3.7 Layout of the ITS

The system consists of six cylindrical layers of coordinate-sensitive detectors, covering the central collision region, 10.6 cm along the beam direction (z). The detectors and front-end electronics are held by lightweight carbon-fibre structures. The geometrical dimensions and the technology used in the various layers of the ITS are summarized in Table 1-3.

TABLE 1-3. **Dimensions of the Inner Tracking System (ITS) detectors (active areas).**

Layer	Type	r, cm	± z, cm	Area, m^2	Ladders	Det./ladder	Tot. channels
1	pixel	4	16.5	0.09	80	1	5242880
2	pixel	7	16.5	0.18	160	1	10485760
3	drift	14.9	22.2	0.42	14	6	43008
4	drift	23.8	29.7	0.89	22	8	90112
5	strip	39.1	45.1	2.28	34	23	1201152
6	strip	43.6	50.8	2.88	38	26	1517568

The granularity required for the innermost planes, is achieved with silicon micro-pattern detectors with true two-dimensional read-out: Silicon Pixel Detectors (SPD) and Silicon Drift Detectors (SDD). At larger radii, the requirements in terms of granularity are less stringent, therefore double-sided Silicon Strip Detectors (SSD) with a small stereo angle are used. Double-sided microstrips have been selected rather than single-sided ones because they introduce less material in the active volume. In addition they offer the possibility to correlate the pulse height read out from the two sides, thus helping to resolve ambiguities inherent in the use of detectors with projective read-out. The main parameters for each of the three detector types: spatial precision, two-track resolution, pixel size, number of channels of an individual detector, total number of electronic channels, dissipated power both in the central region and in the end-caps are shown in Table 1-4.

TABLE 1-4. **Parameters of the various detector types. A module represents a single detector chip.**

Parameter	units	Silicon Pixel	Silicon Drift	Silicon Strip
Spatial precision rϕ	μm	12	38	20
Spatial precision z	μm	115	28	830
Two track resolution rϕ	μm	100	200	300
Two track resolution z	μm	800	600	2400
Cell size	μm^2	50 x 400	150 x 300	95 x 40000
Active area per module	mm^2	13.8 x 82	72.5 x 75.3	73 x 40
Read-out channels per module		65536	2 x 256	2 x 768
Total number of modules		240	260	1770
Total number of read-out channels	k	15729	133	2719
Total number of cells	M	15.7	34	2.7
Average occupancy (inner layer)	%	1.5	2.5	4
Average occupancy (outer layer)	%	0.4	1.0	3.3
Power dissipation in barrel	W	1500-2000	510	1100
Power dissipation end-caps	W	--	410	1500

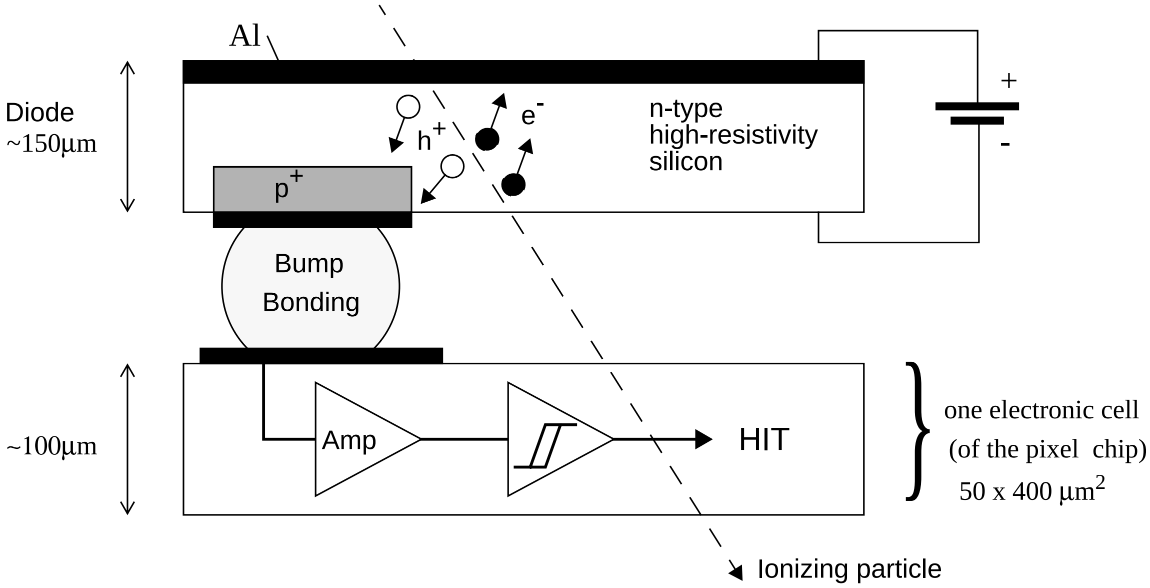
The large number of channels in the layers of the ITS requires a large number of connections from the front-end electronics to the detector and to the read-out. The requirement for a minimum of material does not allow the use of conventional copper cables near the active surfaces of the detection system. Therefore TAB bonded aluminium multilayer microcables are used. The detectors and their front-end electronics produce a large amount of heat which has to be removed while keeping a very high degree of temperature stability. In particular, the SDDs are sensitive to temperature variations in the 0.1°C range. For these reasons, particular care was taken in the design of the cooling system and of the temperature monitoring. A water cooling system at room temperature is the chosen solution for all ITS layers, but the use of other liquid coolants is still being considered. For the temperature monitoring dedicated integrated circuits are mounted on the readout boards and specific calibration devices are integrated in the SDDs. The outer four layers of the ITS detectors are assembled onto a mechanical structure made of two end-cap cones connected by a cylinder placed between the SSD and the SDD layers. Both the cones and the cylinder are made of lightweight sandwiches of carbon-fibre plies and Rohacell^{TM}. The carbon-fibre structure includes also the appropriate mechanical links to the TPC and to the SPD layers. The latter are assembled in two half-cylinder structures, specifically designed for safe installation around the beam pipe. The end-cap cones provide the cabling and cooling connection of the six ITS layers with the outside services.

1.4 Design of the pixel layers

The production of practical silicon pixel devices has been made possible by the continuous progress in the component density achievable in CMOS microelectronics chips and by the development of fine-pitch surface packaging techniques (bump bonding or flip-chip bonding, see Figure 2-34). The pixel technique was developed at CERN in the framework of a dedicated R&D Collaboration (RD19) led by E. Heijne. A two-dimensional matrix (detector ladder) of reverse-biased silicon detector **diodes** (typically rectangles of a few tens of µm by a few hundreds of µm) is flip-chip bonded to several front-end chips: Each cell (see Figure 1-4) on the detector matrix is connected via a solder ball to a cell of the same size on a front-end CMOS chip, which contains the first stage of the front-end electronics. A diode cell reacts to the transit of a ionizing particle generating a weak charge pulse (~4 fC ≅ 20000 e^-), that is transmitted to the electronic cell underneath. Usually, the information provided by the electronic cell is binary: a threshold is applied to the preamplified and shaped signal, and each cell outputs a logical one (called *hit*) if the threshold is exceeded. In certain cases (low threshold or high charge signal) a single particle can fire two cells or more. In this case we speak of double or multiple hits, respectively.

This technique was applied in the development of two generations of Silicon Pixel Detectors (SPDs): Omega2 (with cells of 75 x 500 μm^2) and Omega3 (with cells of 50 x 500 μm^2). Several 5 x 5 cm^2 planes of SPDs were built using Omega2 and Omega3 ladders. At present, 7 planes of Omega2 detectors and 6 planes of Omega3 detectors are in use in the Silicon Pixel Telescope of the experiment NA57 at CERN, for a total of ~ 1.1 x 10^6 channels.

FIGURE 1-4. **Pixel of the Alice experiment**



The pixel layers of the ITS are fundamental in determining the quality of the vertexing capability of ALICE (determination of the position of the primary vertex, measurement of the impact parameter of secondary tracks from the weak decays of particles). They will operate in a region where the track can reach 90 tracks/cm^2. This calls for the use of a detector of high precision and granularity. In addition, the detector must be able to operate in a relatively high-radiation environment (the total dose received in 10 years by the inner layer is estimated to be of the order of 200 krad).

A silicon detector with a two-dimensional segmentation combines the advantages of unambiguous two-dimensional readout with the geometrical precision, double-hit resolution, speed, simplicity of calibration and ease of alignment characteristics of silicon microstrip detectors. In addition, a high segmentation leads naturally to a low individual diode capacitance, resulting in an excellent signal-to-noise ratio at high speed. These are the main motivations that led to the choice of equipping ALICE with a barrel of two layers of SPDs. The price to pay for the use of a silicon detector with very high segmentation is a large increase in the number of connections and electronics channels. The processing of this number of channels is taken in charge by the Pilot chip.

The application of a system of SPDs in ALICE also imposes other tight constraints: The system should be very lightweight and compact and will be inaccessible during operation, thus requiring a powerful and reliable system of remote control. These issues were also addressed. Specific R&D efforts in the area of readout and control have been undertaken, with the development of a fast copper serial data link and of a JTAG accelerator.

1.4.1 Detector overview

The basic building block of the ALICE SPD is the *ladder*, consisting of a pixel detector matrix flip-chip bonded to 8 front-end pixel chips. The detector matrix consists of 256 x 256 cells, each measuring 50 μm in the rϕ direction by 400 μm in the z direction. Each detector ladder measures 13.8 mm (rϕ) x ~82 mm (z). Each front-end chip contains the electronics for the readout of a sub-matrix of 256 (rϕ) x 32 (z) detector cells. The detector is 150 μm thick and the electronics chip 100 μm thick, for a total silicon budget of 250 μm.

Four ladders are aligned in the z direction to form a 33 mm long *stave*. They are mounted (glued and wire-bonded) on a multi-layer thin carrier (stave bus) which contains the bus and power lines. Two pilot chips located at the extremities of the stave bus perform the readout and control functions and transmit the binary data from the pixel cells to a remote router via a serial copper link.

Six staves, two from the inner layer and four from the outer layer, are mounted on a carbon fibre support and cooling sector. Ten such sectors are then mounted together around the beam pipe to close the full barrel. In total, there will be 60 staves, 240 ladders, 1920 chips, 15.7 x 10^6 cells. The staves of the inner (outer) SPD layer are located at an average distance of 4 cm (7 cm) from the beam axis.

The complete system is expected to generate between 1.5 and 2 kW of thermal power. The sectors are equipped with cooling vessels running underneath the staves (one per stave). Cooling collectors are placed at the two extremities of the sectors. Depending on the side, they distribute the cooling fluid to the cooling vessels or collect it from them. In order to avoid radiation of heat towards the SDD layers, which are very sensitive to temperature, an Al-coated carbon-fibre external shield surrounds the SPD barrel.

The average material traversed by a straight track perpendicular to the beam line crossing the SPD barrel corresponds to about 1.7% of X_0 (where X_0 stands for a radiation length ~ 30mm for minimum ionizing particles on silicon).

1.4.2 Front-end electronics

Each front-end chip contains the electronics for the readout of 8192 detector cells. Each cell measures[1] 50 μm x 400 μm. It contains a mixture of analog and digital electronics. A preamplifier-shaper with leakage current compensation is followed by a discriminator with an individual threshold fine tuning. A signal above threshold results in a logical one which is propagated through a delay line during the latency time of the level1 (L1) trigger (5.5 μs).

[1] The original target for the Alice pixel size was 50μm x 300μm [3]. As the requirements of the pixel detector of the LHCb experiment were not very different from the Alice one, it was decided to merge the projects of the two pixel detectors. A trade-off with the new requirements led to a cell size of 50μm x 400μm.

A four-hit deep front-end buffer on each cell allows the event arrival times to be derandomized. When the strobe arrives, the logical level present at the end of the delay line is stored in the first available buffer location.

The periphery contains the JTAG control and biasing circuitry, and the pads for wire-bonding to the stave bus. The front-end chips are being designed with a radiation-tolerant layout technique (enclosed gate geometry) in standard 0.25 µm CMOS. As discussed in Section 2.1, this technique has proven to be tolerant up to at least a few tens of Mrad.

The main requirements for the ALICE SPD front-end chip are listed in Table 1-5.

TABLE 1-5. Main requirements for the ALICE SPD front-end chip

Cell size	50 µm (rϕ) x 300 µm (z)
Number of cells	256 (rϕ) x 32 (z)
Minimum threshold	below 2000 e
Threshold uniformity	200 e
Strobe (L1) latency	up to 10 µs
Strobe duration	200 ns
Clock frequency	10 MHz
Radiation tolerance	500 krad
Individual cell mask	yes
Digital bias adjust (on-chip DACs)	yes
JTAG controls	yes

1.4.3 Readout and control

On the arrival of a Level 2 Accept Trigger signal (latency of 100 µs), the data contained in the front-end buffer locations corresponding to the first (oldest) strobe are loaded onto the output shift registers. Then, for each chip, the data from the 256 rows of cells are shifted out during 256 cycles of a 10 MHz clock. At each cycle, a 32-bit word containing the hit pattern from one chip row is output on the 32-bit stave data bus, where it is read out by the pilot chip mounted at the end of the carrier. A full front-end chip is read out in about 25 µs. The 16 chips of two ladders (one half-stave) are read out sequentially in a total time of about 400 µs. The 120 half-staves are read out in parallel. The dead time introduced by the readout of the SPD is estimated to be below 10% in the worst case, corresponding to ALICE running with Ca-Ca beams at high luminosity, with an L1 rate of 2.5 kHz. On the Pilot chip (described in chapter 3), data are zero-suppressed, reformatted and sent through a 40 m long copper serial link to a router, where a second level of multiplexing is performed before the data are finally shipped to the DAQ on 20 DDL optical links. The test and control system, down to the loading of the parameters on the front-end chip, is implemented using the JTAG protocol. Like the front-end chip, the Pilot chip will be realized in enclosed gate 0.25 µm CMOS.

Design of the pixel layers

CHAPTER 2

The Silicon Pixel Detector

The present chapter contains a description of the design of the ALICE Silicon Pixel Detector (SPD) system, together with a discussion of the main results from the ALICE SPD R&D program.

The technology and architecture for the front-end chip are discussed in Section 2.1. Detector ladders and detector-front-end chip assembly are dealt with in Section 2.2. Section 2.3 contains a description of the bussing, read-out and control systems. Assembly, mechanics and cooling are discussed in section 2.4. Finally, the power distribution scheme is described in the last section.

2.1 Front-end chip

The front-end chip for the ALICE SPD is the direct descendant of the Omega series of pixel front-end chips developed in the CERN Microelectronics Group for the RD19 Collaboration. A sketch of the behaviour of a pixel silicon detector is reported in Fig. 1-4.

The Omega2 and Omega3 chips satisfy many of the requirements for the ALICE SPD, but they are rather sensitive to radiation. Since it is required a process with both high component density and a moderate tolerance to radiation, deep submicron CMOS was explored, with the design of the Alice1Test and Alice2Test chips. The experience gained from these test chips has led to the design of the full ALICE prototype chip called Alice1. In the following, after a brief recall of the main features of the Omega2 and Omega3 chips, we outline the physics motivation for the choice of deep submicron CMOS and explain the design approach. We then describe the Alice1Test and Alice2Test chips and the results obtained with them. Finally, we describe the full prototype Alice1 chip.

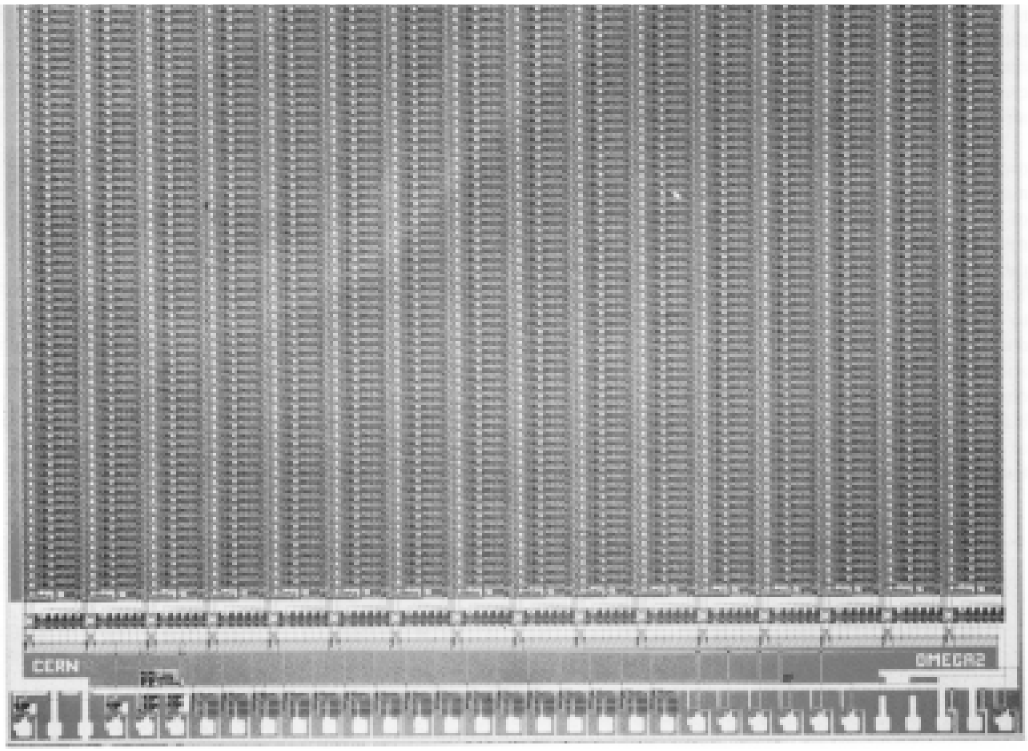
2.1.1 The Omega2 front-end chip

Circuit description

The OmegaD front-end chip [9] was the first member of the Omega series of pixel front-end chips. The pixels are arranged in a matrix of 16 x 63 active pixels. Each pixel cell has a bump-bonding pad and contains a leakage current compensation circuit, a preamplifier, a comparator, a delay line, and coincidence logic. Like the predecessor LAA chip [10], all members of the Omega series of pixel front-end chips are binary, i.e. the information from each pixel cell is a logical one if the signal has exceeded a preset threshold, or otherwise a logical zero. A telescope of three OmegaD chips, each connected to a separate detector chip, was tested in the WA94 experiment in 1993 [11].

The Omega2 chip [12], which is a slightly modified version of the OmegaD, was designed to enable larger area coverage using multi-chip ladders. Like the OmegaD, the Omega2 was produced using the 3 µm Self-Aligned Contact CMOS (SACMOS) process of Faselec, Zurich, which provided a component density equivalent to that of a standard 1.5 µm CMOS process. Several 5 x 5 cm^2 planes were built [13][14][15] using Omega2 chips. Seven of these planes have been used in the WA97 telescope [16] and are currently employed in the NA57 telescope [17].

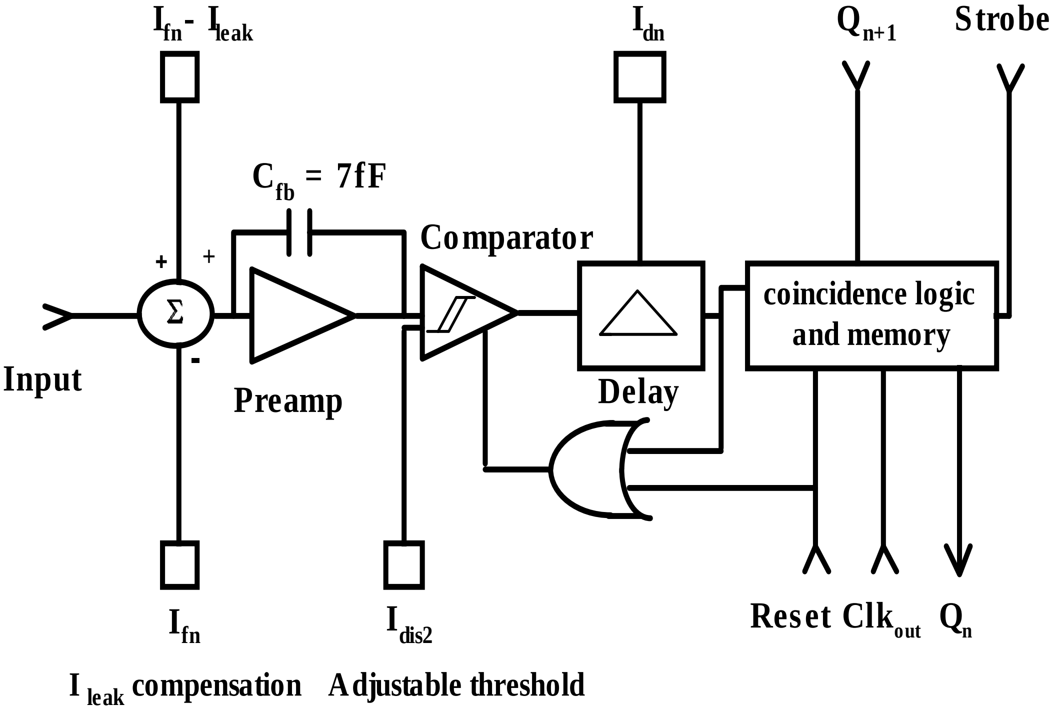
FIGURE 2-1. **A photograph of the Omega2 front-end chip.**



The Omega2 front-end chip is arranged as a matrix of 16 x 63 active pixels. A photograph of the chip is shown in Fig. 2-1. The Omega2 pixel cells measure 75 x 500 μm^2. As opposed to a cell of square shape, a rectangular one provides a precise measurement of the track coordinate in one dimension and at the same time allows a more convenient layout of the electronics components needed inside the cell itself. In the WA97/NA57 telescope, two orthogonal orientations of the planes are used, providing a precise measurement in two coordinates.

The functional blocks contained in each Omega2 electronics cell are shown in Fig. 2-2. The front-end amplifier and the comparator have been designed according to ideas of Vittoz [18] and Krummenacher [19]. The preamplifier is a folded cascode with a small feedback capacitor (7 fF). The gain of 125 mV/fC (500 mV/MIP) is high, in order to diminish the influence of comparator threshold variations. The detector leakage current, I_{leak}, flows via the d.c. connection into the preamplifier circuit. To compensate for an increase of I_{leak} during operation of the detectors, an extra cell at the bottom of each column contains a circuit that senses the leakage current of a dummy detector cell. This current is subtracted from an external reference current, Ifn, which is initially set to 2 nA. The difference is used to cancel the effects of the leakage current in each pixel of the column. If I_{leak} exceeds 2 nA per pixel, the reference current has to be adjusted. The comparator is asynchronous in view of its use in an experiment with an external random trigger. A synchronous design would consume more power and create more problems due to the presence of the clock in each cell (clock feed-through).

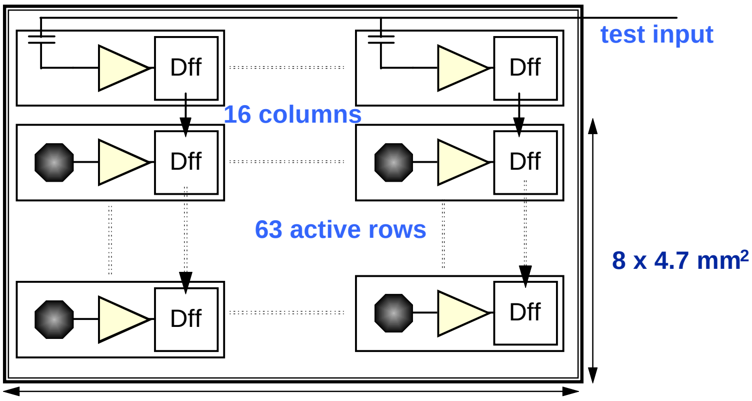
FIGURE 2-2. **Schematic diagram of the pixel cell in the Omega2 chip.**



After a reset, the comparator is in sensitive mode until a sufficiently large signal causes it to switch. The threshold of the comparator is determined by the current I_{dis2}. To enable this electronics to be used with an external, delayed trigger, each cell is equipped with coincidence and readout logic. The delay of the discriminated signal is implemented by three inverters. Their speed is controlled by I_{dn}. In coincidence with the external trigger (strobe), this digital signal is stored in a binary memory inside the cell. The strobed data are read out by using the Clkout signal to shift the contents of the memory to the adjacent pixel in the same column (see Fig. 2-3) or, for the bottom pixels, towards the tri-state output buffer. Data are clocked out only when the chip is enabled by the readout system. In this way, many chips can share the same control signals and data bus [12]. The maximum clock frequency for the Omega2 chip is about 20 MHz. A trigger during readout has to be avoided since entries for the new event would be added to those already being shifted. Therefore, the dead time of this system is the time needed to read out all the pixel cells. The readout and control logic for the Omega2 system are described in Section “Bussing, readout, and control for the Omega2 and Omega3 systems” on page 69.

As shown in Fig. 2-3, the top row of the pixel matrix is used for test purposes. Instead of being connected to the bump-bonding pad, the electronics of these cells is connected to a capacitor which can be externally connected to a step function generator. This test row can be used to measure the functionality and performance of the electronics inside the cells, and the functionality of the shift register of bare dies and assembled detectors.

FIGURE 2-3. **Schematic diagram indicating the readout architecture of the Omega2 chip. The Dff boxes indicate the data flip-flop.**



Performance

Several lessons were learned from the use of this chip in test beams and in the WA97 experiment [13][14][15]. The chip performed to the specifications of the experiment, enabling over 500 million triggered events to be stored on tape thus far (WA97 + NA57). Very high detection efficiency (the efficiency of the working cells is essentially 100%) and the absence of electronic noise allowed the experiment to collect high quality data [20]-[25]. However, there were some aspects of the performance of the full detector which would make it unsuitable for application in the more demanding LHC environment. In particular, differences in the delays from pixel to pixel meant that within one full chip, it was necessary to use a long strobe of about 200 ns for an internal delay of 1 µs. The length of this strobe had to be extended to about 1 µs in the experiment due to systematic differences in the internal delays of the chips as a result of power supply drops on the single layer ceramic support. In addition, only one hit per pixel could be delayed at any given time. Another limitation of the chip was the lack of testability of the individual channels. These issues were addressed in the design of the Omega3 chip.

2.1.2 The Omega3 front-end chip

Circuit description

The Omega3 / LHC1 front-end chip [26] comprises a matrix of 16 x 127 active readout cells of 50 x 500 µm^2, covering a total sensitive area of 8 x 6.35 mm^2. It was manufactured in the 1 µm SACMOS process of Faselec, Zurich, which offered a high component density, equivalent to that of a 0.6 µm standard CMOS process. Unlike the Omega2 chip, in the Omega3 chip every cell can be addressed individually for electrical testing and masking, and cell delays can be individually adjusted with a 3-bit code.

FIGURE 2-4. **Schematic diagram of the Omega3 pixel cell.**

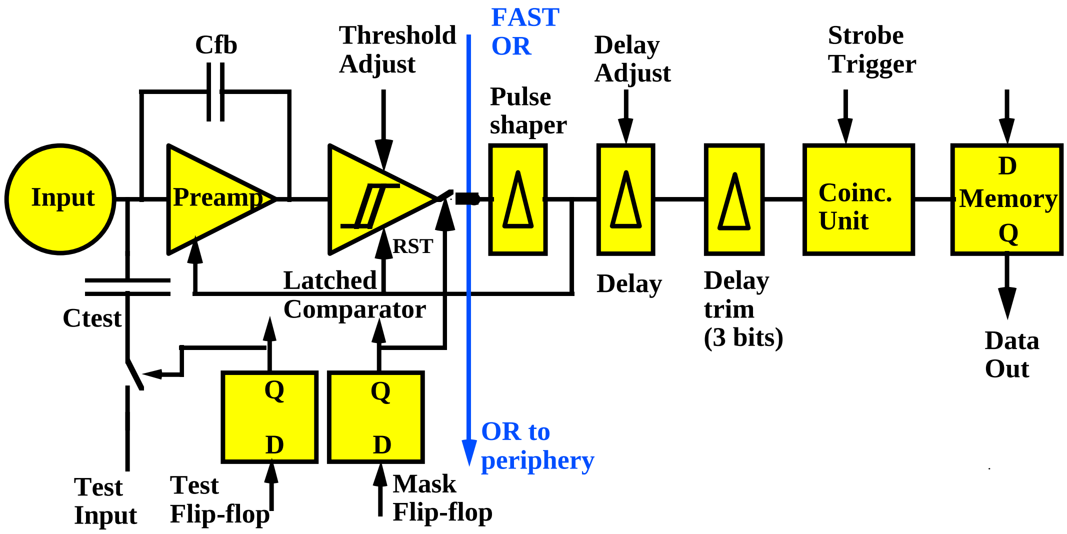


Fig. 2-4 shows the block diagram of one pixel cell. The pixel readout chain contains, from left to right, the successive functions: flip-flop for connection of an analog test signal (length 25 µm), bump pad (octagon in Metal 1 and 2 with 22 µm diameter), preamplifier with leakage current compensation, asynchronous comparator with externally adjustable threshold and fast-OR output (preamplifier and comparator 225 µm), the masking flip-flop (25 µm), a globally adjustable delay (100 µm) with local 3-bit fine-tuning (100 µm), coincidence logic and memory cell (25 µm), and bus lines.

The charge preamplifier is based on a folded cascode circuit similar to that of the Omega2 chip and is designed to consume 19 µW. The feedback capacitance is about 3.5 fF. The feedback resistance is non-linear to limit the swing for high input signals. Following the extraction of parasitics and resimulation, a source follower was added at the output to improve the speed. The signal rise time is designed to be approximately 80 ns. An additional reset has been provided to force a fast return to zero in order to decrease the dead time of the cell after a large signal.

The comparator, shown in Fig. 2-5, is architecturally the same as that of the Omega2 chip and is designed to operate at a power of approximately 15 µW. The current in the bistable non-linear load determines the threshold and can be varied globally for all the cells via an external bias. In the state after reset this current runs in one of the two branches of the non linear load. A signal above threshold will cause the bistable load to change state, switching the current to the other side. This change is detected by a fast sensing circuit and latched. Attention was paid to threshold uniformity by using a symmetric layout of the two branches and the non linear load, and by proper dimensioning of the transistors. Matching data provided by Faselec on the SACMOS 2 µm process were extrapolated to the SACMOS 1 µm process. A trade-off with timewalk had to be made, since a long load transistor improves matching, but slows down the response due to increased capacitance and lower transconductance. The reset of the comparator, together with the preamplifier reset, can be external or from the feedback from the delay chain. The comparator provides a fast-OR and it can be masked, i.e. inhibited to operate, if the preamplifier or the comparator prove to be too noisy or defective. The fast-OR outputs of all cells in a column are connected together in a wired-OR configuration.

FIGURE 2-5. **Schematic diagram of the Omega3 comparator circuit.**

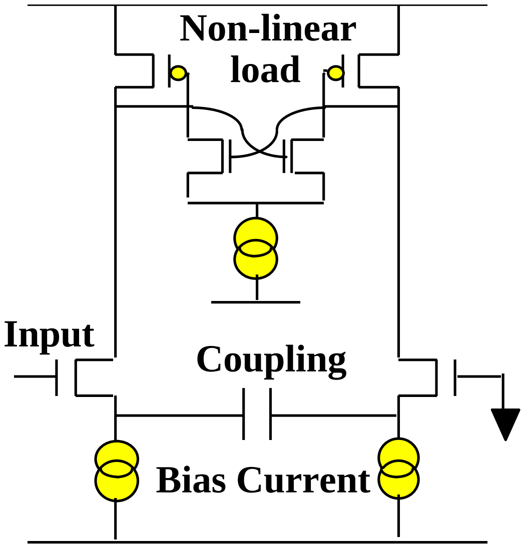


FIGURE 2-6. **Schematic diagram of the delay chain used in the Omega3 pixel cell.**

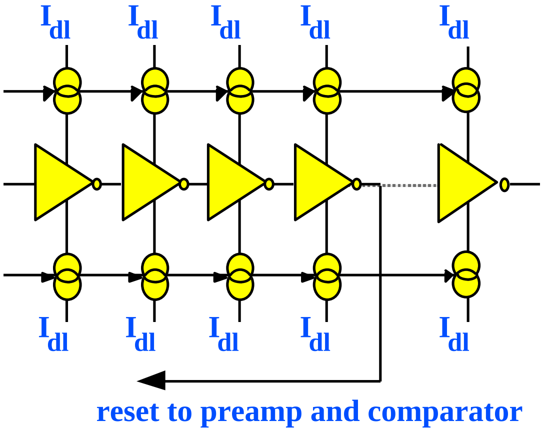
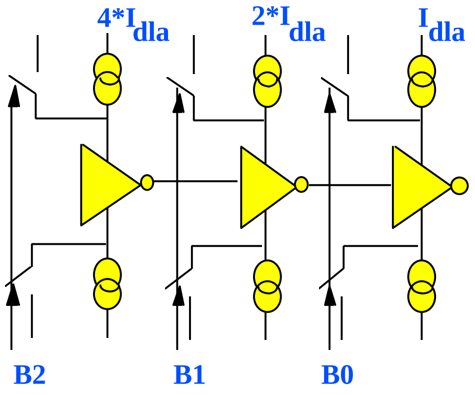


FIGURE 2-7. **Schematic diagram of the delay tuning circuit in the Omega3 pixel cell.**



The delay chain, shown in Fig. 2-6, consists of 36 stages of current-deprived inverters. This type of design, combined with a feedback after four stages to reset the front-end, allows several hit pulses to propagate consecutively in the delay line and decreases the dead-time to below 250 ns. This chain is followed by a 3-bit digitally-controlled delay tuning: it contains three more inverter stages which can be made current-deprived or not by a switch across the current source, as indicated in Fig. 2-7. The absolute value of the range of the fine adjust is determined by the value of I_{dla}. A fourth inverter is used to restore a proper edge of the signal before it is used in the subsequent coincidence logic. The coincidence logic will write a logical one into the data flip-flop if a rising edge at the end of the delay line is detected during the externally provided strobe. As with the Omega2 chip, the data flip-flops of all pixels in one column are configured as a shift-register during data readout.

Like the Omega2 chip, the Omega3 chip contains a one-sided peripheral region. The chip size is 8.72 x 9.12 mm^2, of which 63% is sensitive area. All the 16 cells of the test row are permanently connected to the analog test input, whilst the active pixels are optionally connected to the test input by digital control of the test flip-flop. While this allows, in principle, any test pattern in the matrix to be written, in practice it is undesirable to address many cells at the same time, since this may cause an excessive load on the power supplies, which would alter the characteristics of the individual pixel cells. The peripheral functions are naturally organized according to a column structure. Because of the need to butt together several chips on a ladder, only the bottom side is used for these functions. Here, the biasses for each column are regenerated and the logic signals are buffered. The periphery of the chip also contains bi-directional tristate output buffers to read data and to write and read the contents of the various registers.

Performance

Compared to its predecessor, the Omega3 chip provided improved spatial resolution in the short dimension. Once again, high detection efficiency and the lack of electronic noise provided the experiment with clean data (over 300 million events so far). The internal timing was also improved: during electrical testing with large input signals a strobe width of only 25 ns was enough to achieve full efficiency with an internal delay of 2 µs. In beam tests using single chips, however, a strobe length of 35 ns was used in order to compensate for timewalk in the front-end. At the level of the arrays (See “Bussing, readout and control” on page 69.), a 75 ns strobe was used for electrical testing. The extra length was needed to incorporate random chip-to-chip variations. In the experiment, for practical reasons, a long strobe of about 800 ns was used, although full efficiency could already be obtained with a strobe of 100 ns (in the ALICE SPD, the use of on-chip DACs for bias setting should allow us to compensate for these residual chip-to-chip variations). The addition of flip-flops for pixel testing and masking greatly improved the test coverage. More details of the performance of the Omega3 chip are reported in Ref. [26].

2.1.3 Gate-all-around CMOS design

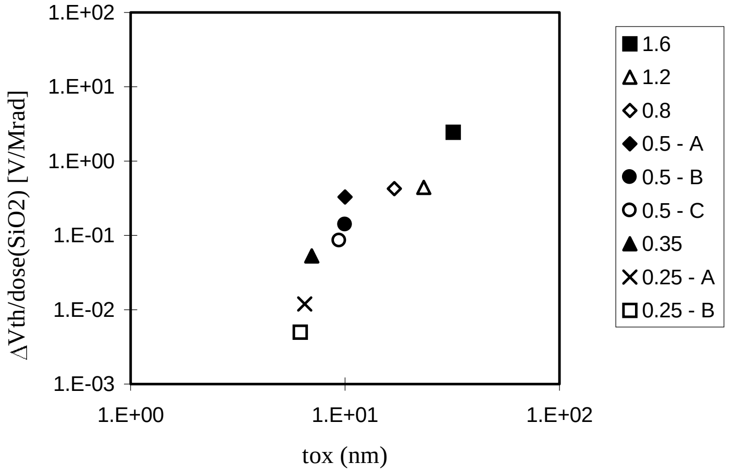
Introduction

A major concern with respect to the application of pixel detectors in ALICE is radiation tolerance. Systematic irradiation studies on front-end chips of the Omega family were first performed at CERN using X-rays, radioactive beta sources and particle beams. The tests were then continued in Rome using a gamma source, in Catania using an electron beam and in Legnaro using a proton beam [27][28]. These studies allowed to establish a common protocol for the irradiation of the submicron technology prototypes. Both the Omega2 and Omega3 chips fail around a dose of 30 krad.

The total dose delivered to the inner layer of the ALICE SPD is estimated, in the standard ALICE 10-year running scenario, to be of the order of 200 krad. A primary concern for the ALICE SPD project is thus to obtain sufficiently radiation-tolerant front-end chips. Indeed, radiation tolerance of the integrated circuits is a primary concern for all the planned LHC experiments. Within the LHC community there is an ongoing effort to implement circuits using dedicated radiation-hard technologies (such as the DMILL or the Honeywell SOI processes). These technologies are expensive, and are, in general, limited with respect to the density of the components. Therefore, we decided to investigate the use of deep submicron CMOS as an alternative.

Irradiation measurements on MOS capacitors performed in the early 1980s [29][30][31] showed a significant decrease of the radiation-induced oxide-trapped charge and interface states for oxides thinner than about 10 nm. Gate oxides in present-day submicron CMOS technologies are in this range. Fig. 2-8 shows the measured threshold shifts per Mrad for a high dose-rate (3 to 4 krad/min.) irradiation on transistors implemented in various standard CMOS technologies. These measurements on transistors confirm the earlier capacitor measurements.

FIGURE 2-8. **Illustration of the sharp reduction of the radiation-induced threshold shift as a function of the gate oxide thickness measured on transistors in commercially available submicron technologies. Four data points were taken from Ref. [31]. The legend gives the minimum gate length for the technologies, in microns.**



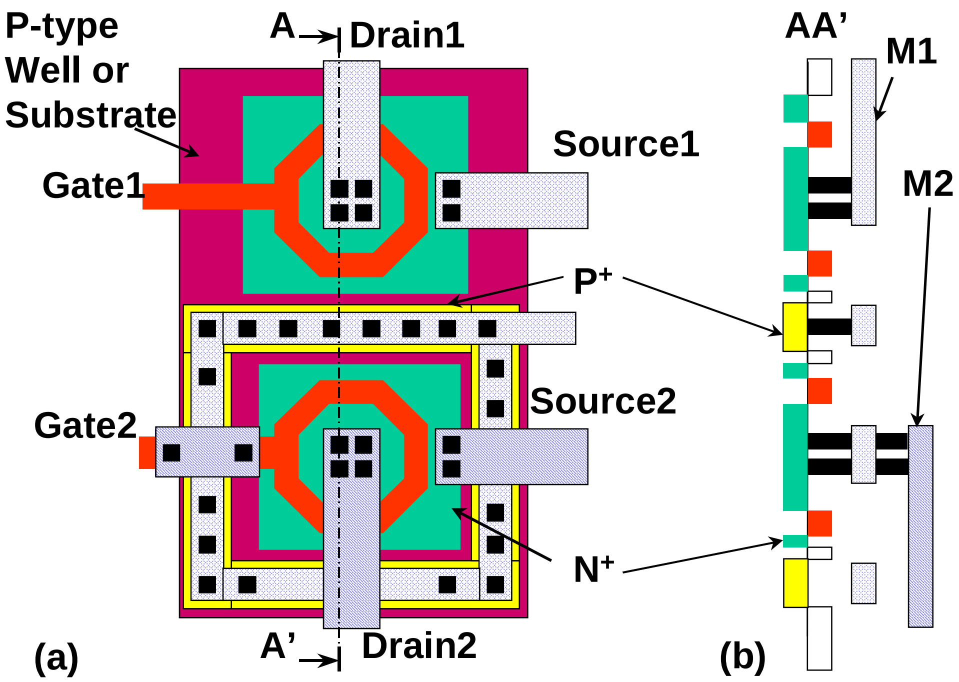
Ionizing radiation can still lead to leakage for the n-channel devices. This can be avoided by designing all NMOS transistors in enclosed (edgeless) geometry and implementing p^+ guardrings wherever necessary (see for instance, Ref. [32]). Recently, due to the advent of deep submicron technologies, this approach has become very interesting: their increased density offsets, at least partially, the increased area requirements of edgeless devices.

Description of the layout approach for individual transistors

Radiation induces NMOS transistor leakage through the formation of an inversion layer in the p-type substrate, or p-well underneath the field oxide, or at the edge of the active area. This inversion layer forms because of the radiation-induced accumulation of positive charge in the silicon dioxide and leads to source-to-drain leakage and inter-transistor leakage between neighbouring n^+ implants. Source-to-drain leakage can be avoided by forcing all source-to-drain current to run underneath the gate oxide, using a closed gate. This is illustrated in Fig. 2-9. The two NMOS transistors shown in the figure are drawn in enclosed geometry, and any current between their sources and drains has to flow underneath the gate. There is therefore no possible current path underneath the field oxide or along the edge of the active area. Often it is not allowed to make a contact with a transistor gate above the active area, and therefore a polysilicon strip is brought out on top of the field oxide for contact.

Inter-transistor leakage (from one n^+ diffusion to the next) is caused by the formation of an inversion layer in the p-type substrate or p-well underneath the field oxide. Increasing the doping level of the p-type substrate or p-well increases the threshold for inversion to a very high level, such that the positive charge generated in the oxide is no longer sufficient to invert the silicon at the silicon-silicon dioxide interface. This can be carried out by implementing a p^+ guardring which is uninterrupted and separates the n^+ implants which one would like to maintain isolated from each other. This is also illustrated in Fig. 2-9.

FIGURE 2-9. **Transistors laid out in enclosed geometry to prevent transistor leakage. The implementation of the p$^+$ guard ring prevents leakage between the two transistors. (a) shows a top view. (b) shows a cross-section along the line AA.**



It should be pointed out that it is not necessary to take the same precautions for PMOS transistors, as with these the positive charge accumulated in the oxide will push the n-substrate or n-well more into accumulation without the danger of the formation of an inversion layer. As can be seen from Fig. 2-9, the p$^+$ guardring is covered with a Metal 1 layer with contacts to the p$^+$ guard. This reduces the sheet resistance of the guard, which, although not necessary for the prevention of leakage, can be useful to prevent radiation-induced latch-up. In the Alice1Test and Alice2Test chips described below, all p$^+$ guards are connected by metal to the substrate potential, but there is not always a continuous piece of Metal 1 covering the guard, so that the use of Metal 1 for connections is not excessively limited.

Experimental proof of the layout approach

The RADTOL Collaboration performed X-ray irradiations on individual transistors designed by the Microelectronics (MIC) group at CERN in order to verify the effectiveness of these layout techniques [33][34]. The irradiations were performed at room temperature using a SEIFERT X-ray generator available at CERN, similar to the more widespread ARACOR [35]. The X-ray energy was 10 keV, the dose rate was 4 krad/min., and the devices were biased in worst-case conditions. The measurements were carried out immediately after the irradiation.

FIGURE 2-10. **Log(Id) versus gate voltage before (thick line) and after 2Mrad irradiation (thin line) for a W/L = 10/0.5 μm traditionally laid out transistor, showing a prohibitive increase of the leakage current. The leakage was measured to be unacceptable already at 40 krad.**

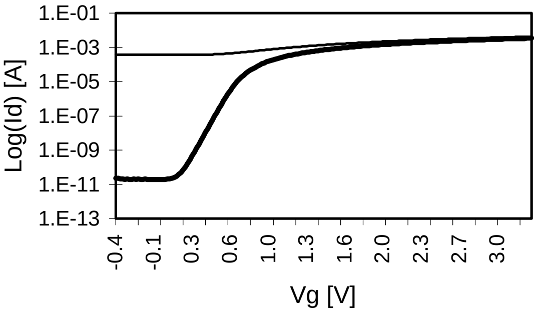


FIGURE 2-11. **Log(Id) versus gate voltage before (thick line) and after 2Mrad irradiation (thin line) for a W/L = 10/0.5 μm enclosed transistor which does not show the leakage problem.**

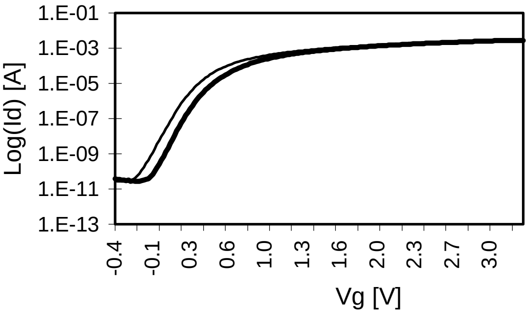


Fig. 2-10 and Fig. 2-11 show the major difference in behaviour between a standard and an enclosed transistor in 0.5 μm technology (gate oxide thickness ~ 10 nm) after a 2 Mrad exposure. The standard transistor shows an unacceptable leakage, which, as was verified in a separate measurement, is already present after 40 krad. The enclosed device remains acceptable up to 2 Mrad and only shows some threshold and sub-threshold slope change.

Similar behaviour was observed with field leakage currents. A p^+ guardring eliminates field leakage. Moreover, previous work [36][37][38] shows that guardrings are very effective against latchup. In tests conducted by the MIC group at CERN, it was possible to induce latchup electrically (by a pulse on V_{dd}) on standard ring oscillators in 0.5 μm technology, but not on ring oscillators with guardrings. No radiation-induced latchup was observed for either type of ring oscillator up to an incident LET of 60 MeV cm^2mg^{-1} (iodine, 240 MeV). Single event upsets (SEU) were not investigated for the 0.5 μm technology.

Layout for radiation tolerance: design implications

Having established that enclosed gate NMOS devices are radiation tolerant, it is important to consider the implications of their implementation in the design of pixel front-end electronics. It is clear from simple geometric considerations that it is impossible to make enclosed transistors with very low width/length (W/L) ratios (precise models for these devices have been developed in Ref. [39]). Therefore, accurate NMOS current mirrors for low current are impossible to design. In addition, our results on pixel threshold dispersion in the Alice1Test chip (see below) indicate that enclosed transistors exhibit higher mismatch than standard devices.

The RADTOL Collaboration studied the noise of PMOS and enclosed NMOS transistors, and its evolution with the received dose, for several 0.5 µm technologies [33]. Only the white noise of the PMOS transistor increases with the dose (by about 10% after 1 Mrad), whilst its 1/f noise remains unchanged. No change with dose is observed for the whole noise spectrum of the NMOS up to a dose of 1 Mrad.

If circuit topologies which rely on accurate NMOS current mirrors are avoided, no significant density penalty is incurred for analog circuitry. For digital circuitry, the use of deep submicron technologies, compared to coarser traditional radiation-tolerant technologies, offsets the area penalty by a finer line-width and the availability of more interconnect layers. For the digital circuitry in the Alice1Test chip, implemented in 0.5 µm with enclosed NMOS devices, the area taken is about the same as in the traditionally laid-out Omega3/LHC1. This is due to the availability of one more metal layer in the 0.5 µm technology.

Other important implications with respect to the use of edgeless transistors are that standard libraries cannot be used, and that extraction and verification routines need to be modified for these devices. These tasks have been undertaken by the Microelectronics Group at CERN for a 0.25 µm technology. Cadence compatible Design Rule Check (DRC), Extraction and Layout Versus Schematic (LVS) rule files have been written and tested. A limited radiation-tolerant library is also available to HEP users of this technology.

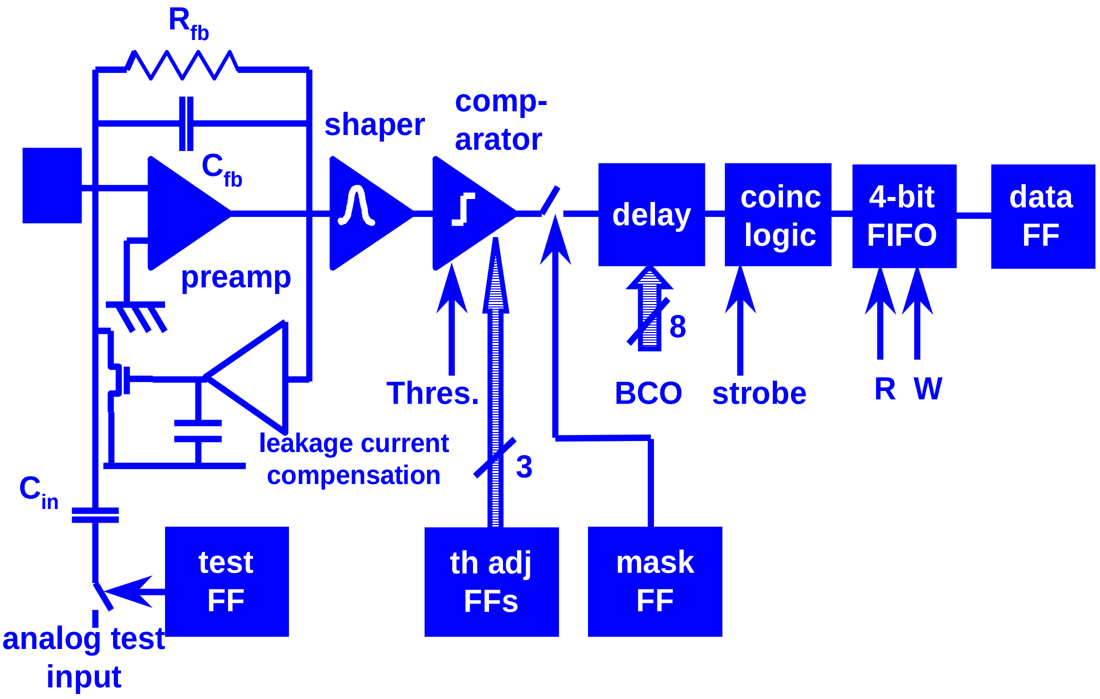
2.1.4 The Alice1Test chip

Circuit description

The Alice1Test is a small test chip designed and produced in order to evaluate the improvements over the Omega2 and Omega3 front-end chips with regard to both front-end performance and radiation tolerance. It is a matrix of 65 x 2 identical pixel cells each measuring 50 x 420 µm^2. A block diagram of the pixel cell is shown in Fig. 2-12.

Each cell comprises an input structure (160 µm long) to simulate a single detector element, a preamplifier, a shaper, a comparator with a variable threshold, and a data flip-flop. The circuit can work both with positive and negative input charges. The preamplifier feedback contains a detector leakage compensation circuit: a low-frequency feedback which adjusts itself to the leakage current coming from the detector. Unlike the Omega2 and Omega3 chips, the leakage current compensation is now carried out on a pixel-by-pixel basis.

FIGURE 2-12. **Block diagram of the Alice1Test pixel cell. The rectangles labelled ‘x’ and ‘y’ in the detector leakage current compensation are voltage level shifters to prevent the NMOS and PMOS transistors which should absorb the leakage from simultaneously conducting current.**



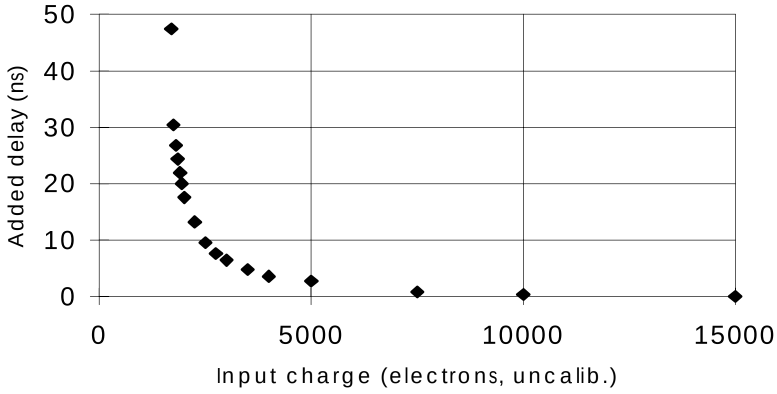
The circuit is a modified version of a circuit proposed in Ref. [19]. It allows the leakage current of both polarities to be compensated. As shown in Fig. 2-12, the output of the preamplifier is amplified and filtered (by a large capacitance on a high impedance node) to make a low-pass filter. This node controls the gate of two transistors, one n-channel and the other p-channel, which have their drain connected to the preamplifier input, and therefore controls the current to the preamplifier input. The circuit works such that if the output of the preamplifier starts to drift due to a change in the detector leakage current, the voltage on the high impedance node will change, and the current in the n- and p-channel devices will change in order to absorb the variation in the leakage current. This scheme has the advantage that not all the detector leakage current has to pass through the feedback resistor. In the design, voltage level shifters (source followers, which are schematically represented in Fig. 2-12 by square blocks, labelled ‘x’ and ‘y’) have been included between the high impedance node and the gates of the n- and p-channel devices. This guarantees that both devices cannot conduct current simultaneously, as this would be detrimental for the parallel noise. To minimize the parallel noise, long and narrow devices should be used to absorb the leakage current. This was done for the PMOS, but was not possible for the NMOS, which had to be designed in enclosed geometry.

To minimize the noise contribution of large detector leakage currents, the shaping time of the first-order semi-gaussian shaper was reduced to 23 ns. The shaper output current is presented to a current comparator. The comparator threshold is dependent on the accuracy of the current sources. The lower current source is implemented by a large NMOS current mirror: it occupies about 35 x 105 μm^2. The contents of the test flip-flop determine whether or not an analog input signal is applied to the preamplifier input across an injection capacitance. Changing the test flip-flop pattern allows one or several pixels to be addressed simultaneously in an arbitrary way during testing. A mask flip-flop allows a pixel to be disabled should it be too noisy or defective. If the comparator output changes polarity when the strobe is high, a logical one is written into the data flip-flop (a flag tells the readout logic which polarity of the comparator output corresponds to a logic one, depending on whether positive or negative input charge is collected). A delay circuit was not implemented in this test chip. During readout, the data flip-flops of pixels in one column are configured as a shift register for sequential readout. Every cell contains about 200 transistors, the total chip contains about 25 000 transistors in a 10 mm^2 area.

Measurements prior to irradiation

The injection capacitance could not be calibrated, but only estimated from data on layer-to-layer capacitances provided by the vendor. All numbers given in absolute electron charge (e$^-$) are based on this estimate. The average threshold charge (over all the 130 cells) can be changed by the threshold-setting voltage from -15 000 e$^-$ to +15 000 e$^-$. The observed threshold spread of 400-500 e$^-$ r.m.s. over the full chip is larger than expected. This is probably due to a poor modelling of the mismatch of the large NMOS current mirror in the comparator. As discussed in the previous section, this is a consequence of the enclosed geometry layout. Solutions which avoid large NMOS current mirrors have been developed for the Alice2Test chip described below. The threshold does not vary by more than 1% for leakage currents ranging from -200 nA to +200 nA. The average noise is about 200 e- r.m.s. at low detector leakage current. The input structure adds about 100 fF to the preamplifier input to simulate the detector capacitance. Care was taken to inject the leakage current into the front-end with the proper noise spectral density (2qI) for a detector. A negative 200 nA detector leakage current is absorbed by a long narrow PMOS transistor, and increases the average noise to 350 e$^-$ r.m.s. A positive detector leakage current is absorbed by a short and wide edgeless NMOS transistor in weak inversion. Therefore, in the positive case, the effect is a bit worse: the noise increases to 400 e- r.m.s. at 200 nA. The timewalk of the circuit is shown in Fig. 2-13: due to the fast shaping, all hits more than a few hundred e- above threshold fall within a 25 ns time window.

FIGURE 2-13. Timewalk performance of the test chip. The threshold is set to about 1650 electrons. Only a few hundred electrons above threshold are required to have an extra reaction time of less than 25 ns compared to a big input signal.



Irradiation measurements

The radiation tolerance of the chip was measured for different irradiation sources.

10 keV X-rays The X-ray irradiations were carried out under the same conditions as for the individual transistors (measurements of Fig. 2-10 and Fig. 2-11). Fig. 2-14 shows the evolution of the average pixel comparator threshold and its r.m.s. dispersion with dose. The chip only starts to degrade significantly after 600 krad. The large oscillatory changes at low dose are due to significant annealing effects immediately after the irradiation, despite efforts to minimize the measurement time. Fig. 2-15 shows that the analog power consumption remains unchanged and that the digital power consumption decreases. The latter can be explained by radiation-induced transistor threshold shifts. This indicates, on a full circuit scale, that enclosed NMOS devices and guardrings prevent radiation-induced leakage.

FIGURE 2-14. **Evolution of the average comparator threshold and its variation with increasing X-ray dose.**

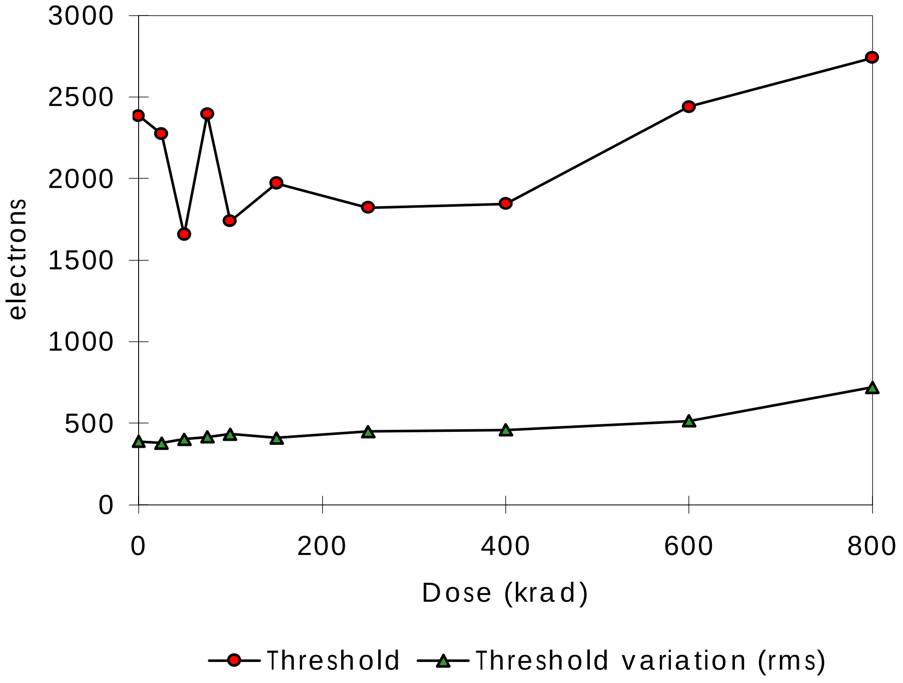
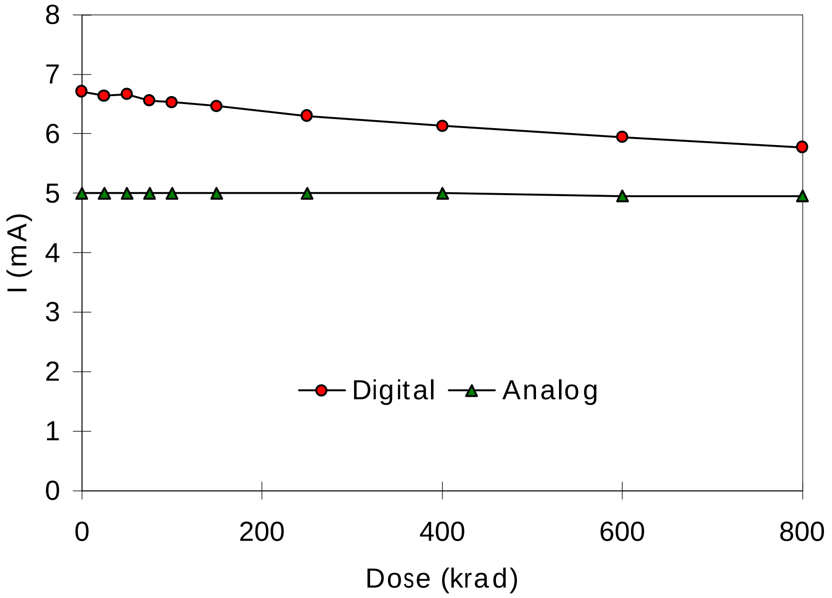
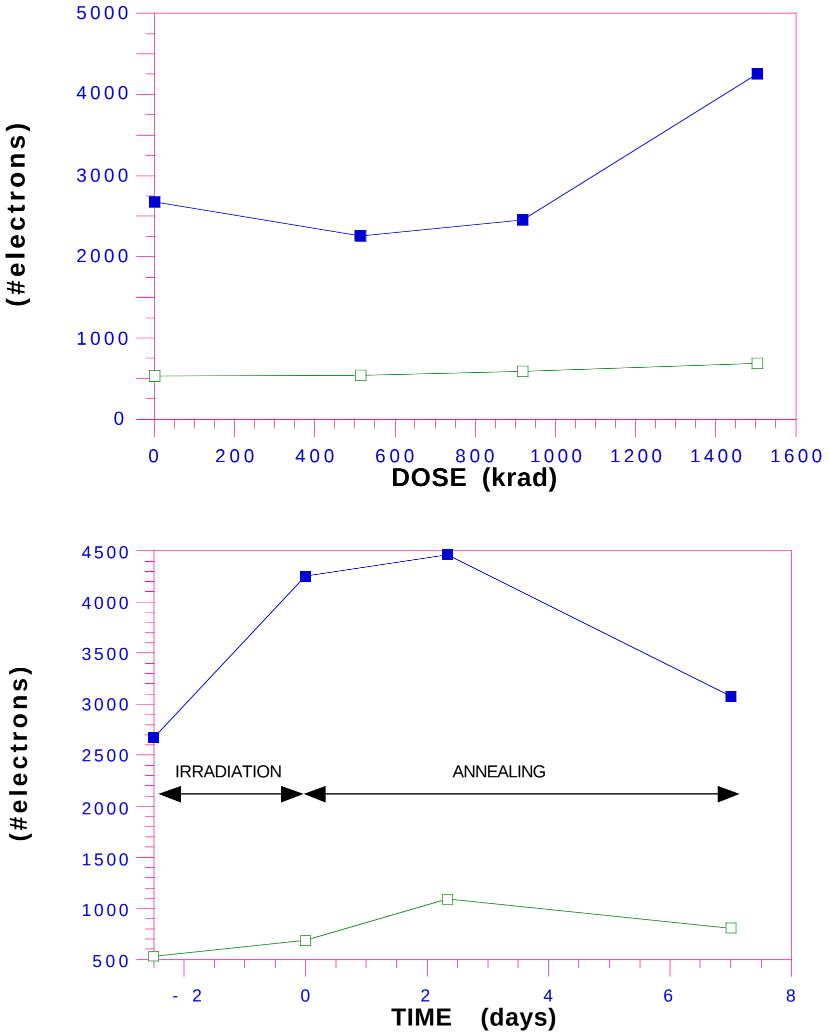


FIGURE 2-15. **Evolution of the supply currents of the chip with increasing X-ray dose.**



^{60}Co rays The gamma irradiation was carried out at the National Health Institute (Istituto Superiore di Sanità, ISS) in Rome, Italy, using a standard source (Gammacell 220) of 1.173 and 1.332 MeV γ rays from ^{60}Co. In this case, the dose rate was 610 rad/min. The dose was calibrated and did not vary by more than 5% over the chip. The evolution of the comparator threshold and its dispersion with the accumulated dose, shown in Fig. 2-16, are similar to those observed for X-rays. The chip is fully functional up to a dose of about 1 Mrad, where degradation sets in. At a total accumulated dose of 1.5 Mrad only 10% of the pixels respond. Partial recovery is evident during annealing (positive time scale in the figure). The supply currents showed a behaviour similar to that observed in the X-ray irradiations (see Fig. 2-15.).

FIGURE 2-16. **Irradiations with ^{60}Co. a) Value (filled squares) and dispersion (open squares) of the discriminator threshold as a function of the accumulated dose. b) Same, as a function of time. t = 0 indicates the end of the irradiation / beginning of the annealing.**



High-energy particle irradiation (electrons) A chip was placed in the NA50 experiment right behind the target, but slightly offset with respect to the particle beam. This resulted in an irradiation of the chip primarily by electrons with an energy of 1 MeV or above. The dose was both calculated by a GEANT simulation [40], and measured using alanine dosimeters placed in the proximity of the circuit. Fig. 2-17 shows the average and r.m.s. spread of the threshold and noise, as well as the number of pixels responding below 20000 e. The drop in the number of responding pixels from 130 to 126, and the rise back to 130, is an artefact caused by a timing problem in the experimental set-up. The drop to zero after 55 h is real. The degradation after 1.7 Mrad (52 h) is evident. There was some recovery around 43 h when the beam was off for a couple of hours. The irradiation was continued up to 2.6 Mrad, well beyond the point where no pixels were responding below 20000 e any more. Fig. 2-18 shows the annealing under bias: after one week at room temperature some pixels start to respond again. After a month at room temperature the average threshold has come down again to about 7500 e, and the average noise to about 500 e. The last week of annealing was carried out at 100C and caused the threshold to drop to about 5000 e.

FIGURE 2-17. **Irradiation in the NA50 experiment (mainly electrons). Evolution of (a) average and spread of pixel threshold, (b) average and spread of pixel noise, and (c) number of pixels responding to an input signal below 20000 electrons as a function of irradiation time and dose. After 52 h, corresponding to about 1.7 Mrad, the chip starts to degrade significantly.**

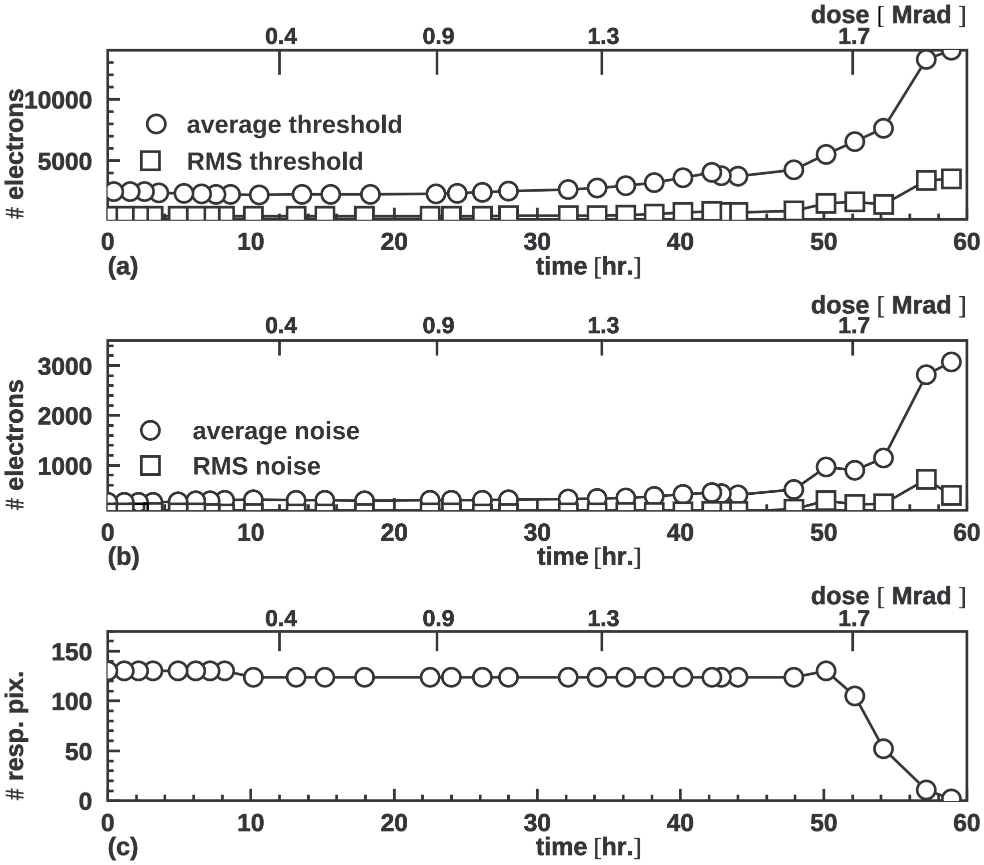
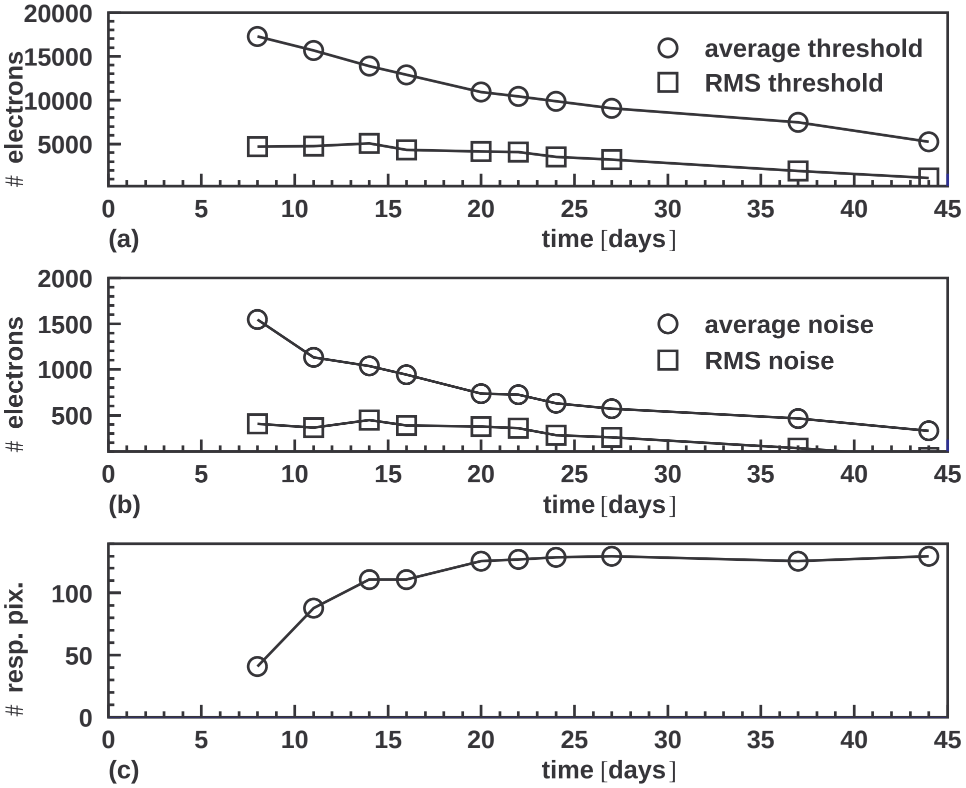
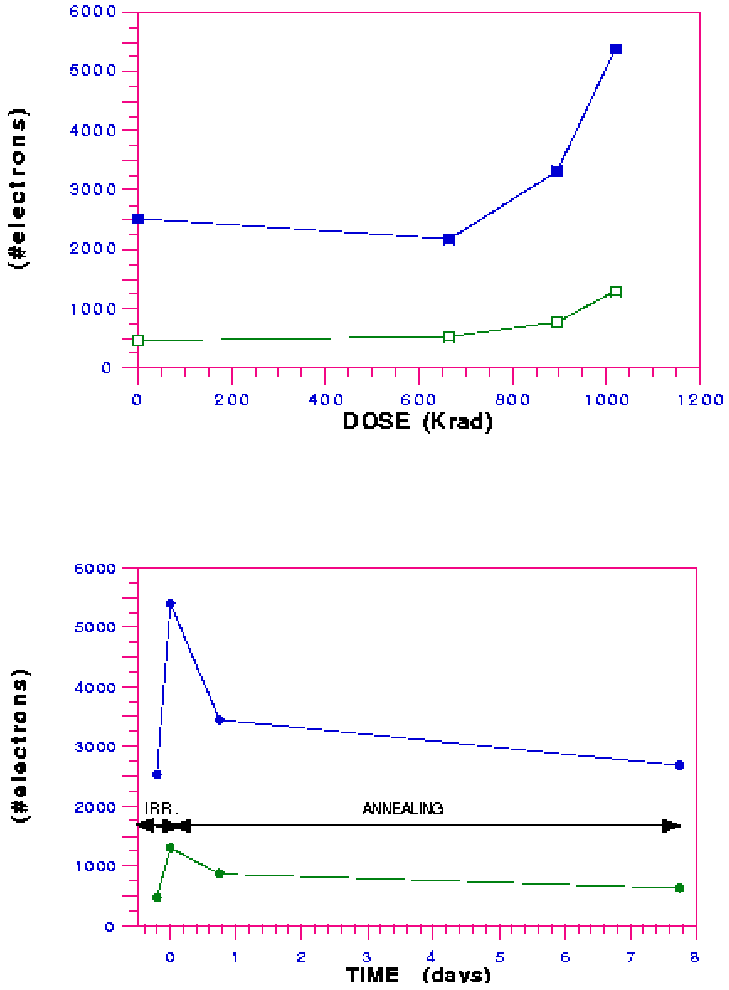


FIGURE 2-18. **Annealing under bias after electron irradiation. Evolution of (a) average and spread of pixel threshold, (b) average and spread of pixel noise, and (c) number of pixels responding to an input signal below 20000 electrons as a function of annealing time. The last week of monitored annealing was carried out at 100C. Partial recovery is evident.**



6.5 MeV protons The tolerance to charged hadronic particles was investigated using 6.5 MeV protons at the Van de Graaf accelerator in the National Laboratory of Legnaro, Italy. The dose was calibrated and did not vary by more than 10% over the chip area. The combination of the proton flux with the energy loss on the chip (for electromagnetic processes only) [41] is used to evaluate the total dose. We only quote the dose corresponding to the Rutherford peak, the background contribution is estimated to be about 20% of the dose quoted [42]. The estimated dose rate is about 12 krad/min. Fig. 2-19 shows the evolution of the average and dispersion of the pixel threshold. Serious degradation occurs above 1 Mrad. Again, partial recovery is evident during the annealing phase. The analog outputs and the power consumption showed a behaviour similar to that observed in the other measurements.

FIGURE 2-19. **6.5 MeV proton irradiation. a) Average and dispersion of the discriminator as a function of the dose. b) Same as a function of time, t = 0 indicating the end of the irradiation and beginning of the annealing.**



Discussion and degradation mechanism

In general, it is difficult to quantitatively compare damage caused by ionizing dose from different radiation sources based only on the amount of absorbed dose. Nevertheless, we observed a qualitatively similar behaviour for all radiation sources. In order to study whether we could reproduce the observed degradation, we introduced the measured transistor threshold shift with increasing X-ray dose (under worst-case conditions) into the circuit simulator. In the simulations, the degradation was similar to the one observed in the measurements, but it sets in earlier, at about 500 krad. This can be explained by the fact that in reality not all the transistors in the circuit are biased in worst-case conditions. In fact, the simulated 500 krad and measured 1 Mrad preamplifier and shaper outputs were in good agreement. The input transistor of the shaper is gradually pushed out of saturation due to the cumulative effect of the radiation-induced threshold voltage shifts of two NMOS devices (V_{tn}), and one PMOS device (V_{tp}). This degrades the shaper gain very quickly, leading to a very high effective pixel threshold and ultimately to a non-functional front-end.

The partial recovery can be explained by the fact that the V_t shift of the NMOS transistors readily anneals out. There is an irreversible effect because the PMOS V_t shift is not cured by annealing. Note that, since the operating margin of this circuit depends on some transistor thresholds, random transistor threshold variations will have a direct influence on the dispersion of the radiation tolerance amongst different chips.

In conclusion, the use of NMOS devices in enclosed geometry and guardrings brought the total ionizing dose tolerance of this prototype, in standard 0.5 μm CMOS, up to 0.6 or 1.7 Mrad, depending on the radiation source. These values are much larger than those observed for the Omega2 and Omega3 front-end chips. The failure mechanism is no longer the leakage, but radiation-induced transistor V_t shifts.

2.1.5 The Alice2Test chip

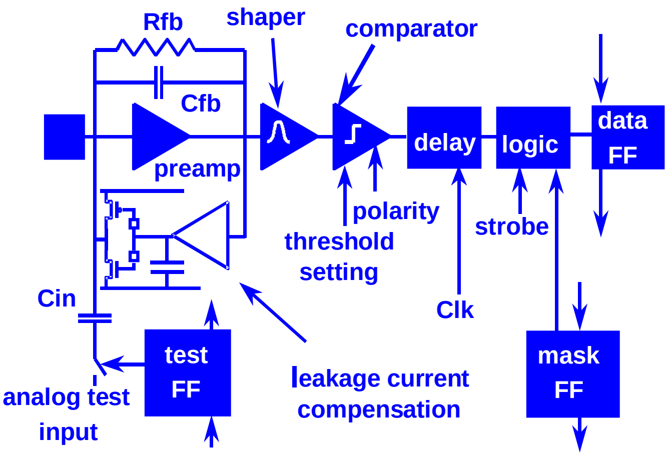
The performance of the analog front-end and the radiation tolerance of the Alice1Test chip would have been adequate for application in the ALICE SPD. However, it became clear that the 0.5 μm technology could not offer the required component density for implementing the full functionality required for the ALICE SPD application within the target cell size of 50 x 300 μm^2. For this reason and because of CERN-wide standardization, we decided to select an even deeper submicron technology (0.25 μm) for the ALICE SPD application. This led to the design of a second pixel test chip, the Alice2Test. This enabled us to characterize the final technology selected for the ALICE front-end chip with a low cost test. In the light of the experience gained from the Alice1Test chip with respect to the matching properties of enclosed NMOS current mirrors, we implemented a new scheme which allowed us to avoid their use. The design of a second test chip also enabled us to test some additional features that we planned to implement in the full ALICE front-end chip, and which had not been tested on the Alice1Test (notably, an individual pixel threshold adjust, and a delay line based on the use of a counter).

Circuit description

The Alice2Test chip was designed in a commercial 0.25 μm process. Like the Alice1Test, it is a matrix of two columns, each containing 65 identical cells, occupying 10 mm^2, but this time containing about 50 000 transistors.

Each front-end cell (see Fig. 2-20) comprises a preamplifier, a shaper filter, a discriminator, a delay line, and readout logic. The circuit can operate with both positive and negative input charges. In each cell, an input structure simulates detector capacitance, capacitive coupling between pixels, and detector leakage current. The preamplifier feedback circuit allows both polarities of leakage current. The feedback capacitor is implemented as a parasitic capacitance between metal 1 and metal 2. The comparator has a 3-bit threshold fine-adjust. It is controlled by a 3-bit bus directly linked to the outside. In the full-scale version of the ALICE SPD front-end chip, the fine adjust will be controlled by flip-flops directly implemented on the pixel cells. The delay element consists of an 8-bit counter and some control logic. A flag tells the delay control logic which polarity of the comparator output corresponds to a logical one, depending on whether positive or negative input charge is collected. If the comparator fires, the counter in the delay is started. The counter can be preset to an arbitrary value, and the carry of the most significant bit is used to generate the end-of-count signal. If the end-of-count signal is in coincidence with the externally applied strobe, a logical one is written into the data flip-flop.

FIGURE 2-20. **Schematics of the Alice2Test pixel cell.**



In order to compare different designs, this counter was implemented in static logic in one column and in dynamic logic in the other. The clock is propagated along the columns using full-swing differential CMOS logic. The content of the test flip-flop determines whether or not an analog input signal is applied to the preamplifier input across an injection capacitance. A mask flip-flop allows a pixel to be disabled should it be too noisy or defective. Each cell consumes ~ 50 µW and the chip operates on an analog supply of 2.2 V and a digital supply of 1.6 V.

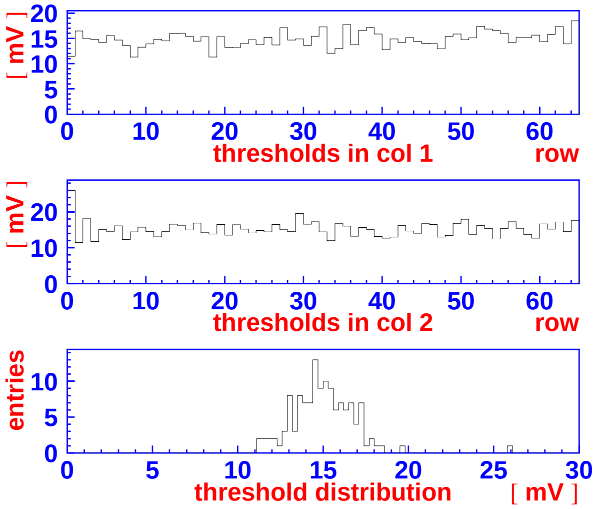
The pitch of the pixel cell in the short dimension is 50 µm, as needed for the final ALICE cell. The cell layout is based on that previously implemented in the Alice1Test chip in the 0.5 µm technology. The 0.25 µm front-end is only 125 µm long. The counter in the delay takes 40 x 60 µm^2 for the static case, and 40 x 35µm^2 for the dynamic case. The delay control logic measures about 20 x 25 µm^2. The rest of the digital part was directly taken from the Alice1Test 0.5 µm chip and was not shrunk to 0.25 µm design rules.

Measurements prior to irradiation

The chip was characterized electrically prior to irradiation using the analog test input and is fully functional. The injection capacitance could not be calibrated, but only estimated from data on layer-to-layer capacitances provided by the vendor: as in the case of the Alice1Test chip, all numbers given in electrons (e$^-$) are based on this estimate.

The threshold can be adjusted linearly using an external bias voltage from -20 000 e$^-$ to +20 000 e$^-$. Fig. 2-21 shows the distribution of the threshold across the chip. There is no systematic dependence of the threshold on the position of the pixel within the chip. The minimum threshold, about 1500 e$^-$, is determined by crosstalk between the analog and digital parts of the circuit. At this threshold setting, the average over all the 130 cells is about 1500 e$^-$ and the spread is typically 160 e$^-$ r.m.s. (although this was as low as 130 e- r.m.s. on some chips). The pixel noise is ~ 220 e$^-$ r.m.s. The behaviour is almost identical for both polarities of input signal. We studied the sensitivity of the circuit to detector leakage current. For leakage currents of ± 100 nA per pixel there was no change in the average threshold; the threshold dispersion was unchanged and the noise degraded by less than 20%.

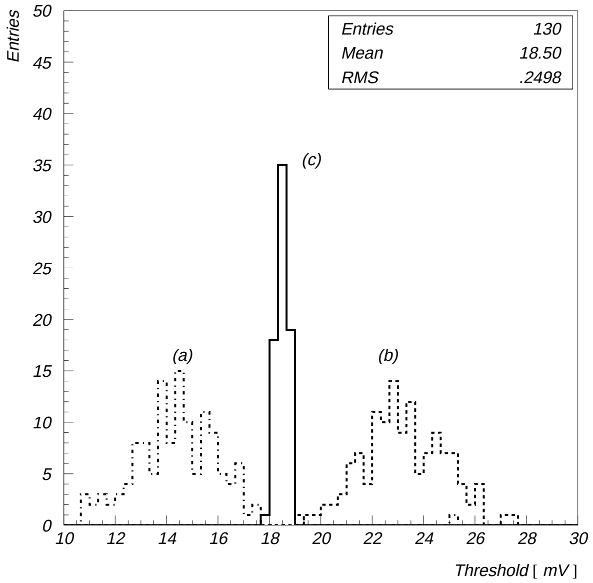
FIGURE 2-21. **Measurement of the pixel thresholds. 1 mV corresponds to 100 e$^-$.**



Pixel detectors normally have a larger capacitance between neighbouring elements than to ground. The input structure has the option of connecting neighbouring pixels together using a 30 fF capacitor or connecting the input node to ground using a 60 fF capacitor. In order to verify the sensitivity of the circuit to capacitive cross-coupling, neighbouring pixels were connected together, the average threshold of the array was set to 1500 e$^-$, and a pixel was stimulated while its neighbour was observed. On average, it was necessary to inject 29 000 e$^-$ into one pixel in order to produce a false hit in the neighbour. The measurement was repeated without the coupling capacitors and this time 30 500 e$^-$ were necessary for the neighbour to react. This indicates that the front-end is rather insensitive to capacitive cross-coupling and that probably much of what was measured was due to parasitic effects in the electrical injection or coupling through the power supplies.

As a verification of the threshold adjust circuit, the thresholds of all pixels were measured for every value of the 3-bit adjust. Fig. 2-22 shows the distribution of the thresholds for minimum added threshold, maximum added threshold, and tuned threshold. The tuned threshold was generated off-line, since the threshold control flip-flops have not yet been integrated on the pixel cells. On this particular chip, the tuning reduced the threshold variation from ~ 160 e$^-$ r.m.s. to ~ 25 e$^-$ r.m.s. Further refinement of the tuning algorithm could lead to even smaller values.

FIGURE 2-22. **Distribution of the thresholds for (a) minimum added threshold, (b) maximum added threshold, and (c) tuned threshold. 1 mV corresponds to 100 e$^-$.**



We observed no difference in analog behaviour between the pixels with the static counter and the pixels with the dynamic counter. In the final chip, however, we will use static logic because of its reduced sensitivity to Single Event Upset [43].

Irradiation measurements

10 keV X-rays As with the Alice1Test, the first measurements of the radiation tolerance of the chip were carried out using an X-ray machine with a peak energy of 10 keV. Fig. 2-23 shows the evolution of the power supply currents with increasing X-ray dose. The absence of any increase in power consumption with dose confirms once more on a full-circuit scale that enclosed NMOS devices and guardrings prevent radiation-induced leakage.

FIGURE 2-23. **The evolution of the analog and digital supply currents with X-ray dose.**

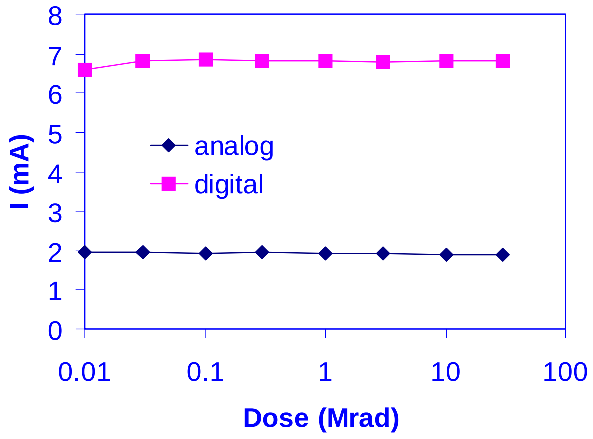


FIGURE 2-24. **Evolution of the average pixel threshold with total dose.**

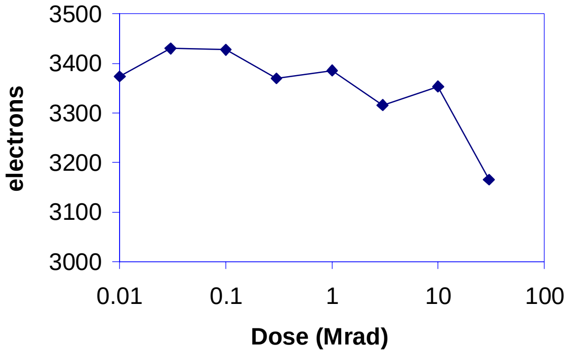


FIGURE 2-25. **Evolution of the threshold variation and noise with total dose.**

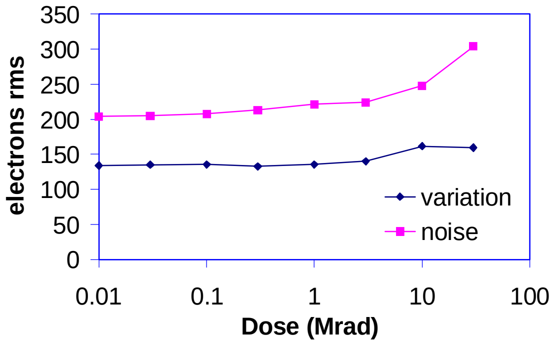
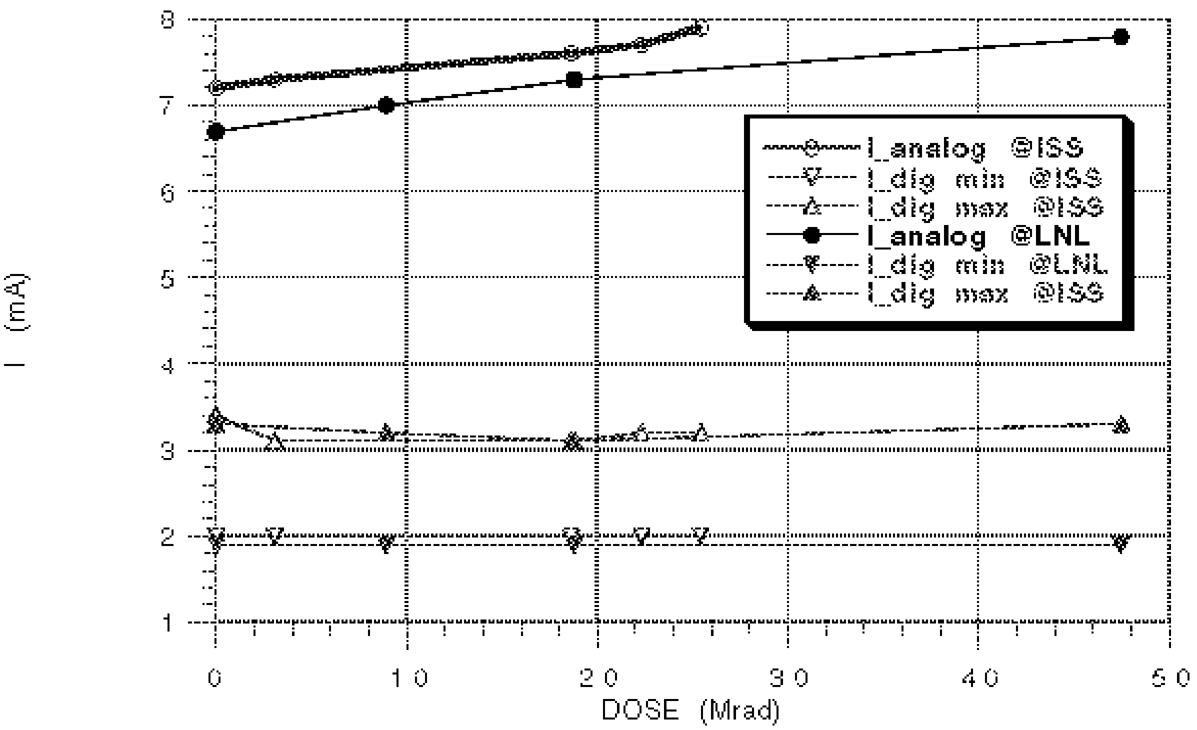


Fig. 2-24 shows the evolution with dose of the average pixel threshold. Fig. 2-25 shows the threshold variation and pixel noise for the same irradiation. For this particular chip, a minor bias adjustment was necessary after 30 Mrad to prevent premature signal clipping in the preamplifier. Apart from this, all other biasses were kept constant. These results illustrate that the chip remains fully functional up to 30 Mrad. After 24 h under bias at room temperature the parameters were unchanged. Following a subsequent annealing for one week at 100°C the average threshold remained the same, the threshold variation degraded slightly to 190 e^- r.m.s. and the pixel noise returned to 230 e^- r.m.s. Further annealing under bias at room temperature did not affect the circuit parameters.

^{60}Co γ rays The gamma irradiation was carried out at the National Health Institute (Istituto Superiore di Sanità, ISS) in Rome, Italy, using the same source used for the irradiation of the Alice1Test chip. The dose rate was 540 rad/min. The chip was irradiated in steps to doses of 3, 19, 23 and 26 Mrad. The results for the power consumption are indicated in Fig. 2-26. In this case, a slight increase in the analog power supply was recorded. For the other parameters, the results of this irradiation closely mirrored those of the X-ray irradiation discussed above.

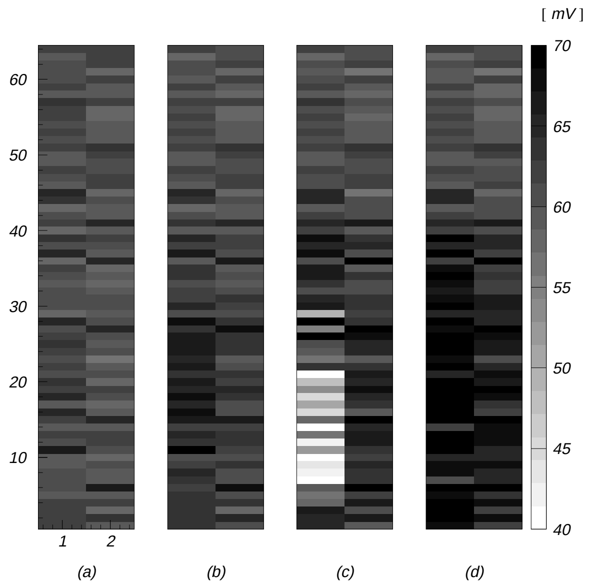
FIGURE 2-26. **Analog and digital supply currents during and after the proton and ^{60}Co irradiation. The results indicated with ISS are those obtained with the ^{60}Co source. Those indicated with LNL were obtained with the 6.5 MeV proton source. The digital power consumption is unchanged with dose, whilst the analog consumption increases by around 10%.**



High-energy particle irradiation (protons) A further test was made with high-energy protons at the NA50 experiment on the CERN SPS machine. The chip was used as a target with the beam focused on a region roughly 2 mm across. In total, the chip received 3.6 x 1013 protons over a 12 h period, corresponding to about 9 x 1014 protons/cm2. The chip was kept under bias all the time and read out between spills of the machine. Fig. 2-27 shows the evolution of the thresholds during irradiation and annealing. During irradiation the threshold of the irradiated pixels was reduced and the noise increased to ~ 1000 e- r.m.s. by the end of the exposure. During annealing at room temperature, the threshold recovered and even increased slightly, whilst the noise returned to its pre-irradiation value. The pixels outside the target region remained unchanged throughout the test. In addition, there was no increase in power consumption.

6.5 MeV protons As for the earlier prototype chip, the tolerance to charged hadronic particles was investigated using 6.5 MeV protons at the Van de Graaf accelerator in the National Laboratory of Legnaro, Italy. Chips were irradiated with doses of up to 48 Mrad. One chip was irradiated in steps to 9, 19 and 48 Mrad. It ceased to function at 48 Mrad. A second chip was irradiated to 37 Mrad. The evolution of the power consumption with the dose is shown in Fig. 2-26. The analog outputs and the power consumption showed a behaviour similar to that observed in the other measurements.

FIGURE 2-27. **Pixel thresholds during and after the proton irradiation: (a) thresholds before irradiation, (b) thresholds after 8 x 1012 protons, (c) thresholds after 6 x 1013 protons and 4-hour annealing, (d) thresholds after 6 x 1013 protons and 20 hour annealing. 1 mV corresponds to 100 e-.**



2.1.6 The Alice1 front-end chip

The measurements described in the previous section indicate that the technology used for the Alice2Test chip fulfils the requirements of the ALICE SPD both in terms of performance and component density. The radiation tolerance of this technology, if gate enclosed NMOS design techniques are used, is well in excess of the ALICE requirements. These results confirmed the choice of this technology for the design of the full ALICE SPD chip.

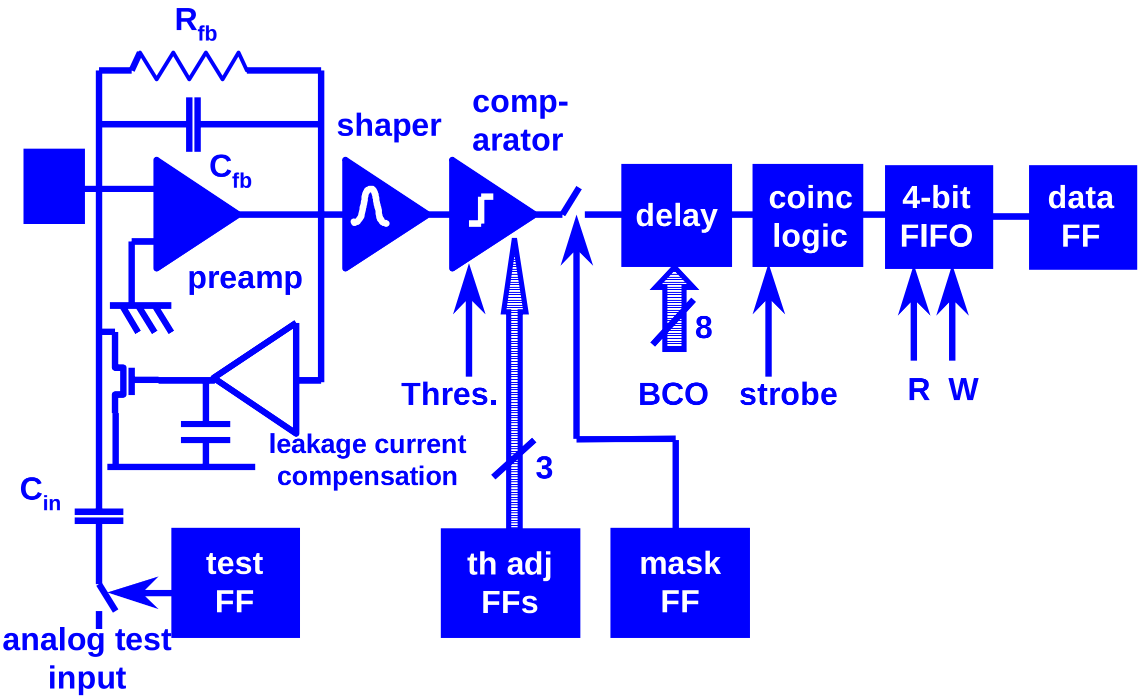
The Alice1 front-end chip, currently being designed, is a matrix of 32 x 256 cells, each one occupying an area of 50 x 400 μm^2 ; all the readout logic and the local control functions are integrated on one side of the matrix.

Since the requirements for the LHCb-RICH application are compatible with those of the ALICE SPD, the chip is designed to allow two modes of operation [44], selected by a hard-wired input: ALICE mode and LHCb mode. In the following, only the ALICE mode of operation is described.

Cell electronics

A schematic diagram of the Alice1 cell is shown in Fig. 2-28. Many of the building blocks used in the design are similar to those used in the prototype chips, but a number of changes and additional features have been introduced.

FIGURE 2-28. Schematic diagram of the Alice1 cell



Signal amplifier One major difference between this design and the previous pixel chips is that we have chosen to implement a differential front-end amplifier. This change is the result of detailed simulations of the expected crosstalk through the substrate and the power supplies as a result of switching in the digital part of the circuit during data acquisition (the detection and readout phases are not separated in time as they were in our previous front-end chips). These simulations indicate that a single-ended front-end would be extremely difficult to operate at the threshold and noise levels required by ALICE, given the significant simultaneous digital activity on the chip. Owing to the strong improvement in the power supply rejection ratio, a differential front-end is much more suited to this application and allows operation at much lower signal thresholds, enabling the use of thinner silicon detectors without compromising the detection efficiency. The penalty incurred is an increase in the analog power consumption: the preamplifier consumes about 30 μW, more than half of the total analog power consumption.

Another modification compared to the Alice1Test and Alice2Test chips is that the front-end returns rapidly to its quiescent state after a hit. The peaking time is 25 ns, and the pulse returns to zero in less than 200 ns. This minimizes pile-up at the level of the front-end, and eliminates the need for extra signal clipping.

As before, the front-end compensates detector leakage currents of up to ~ 100 nA. Unlike that of the prototypes, the front-end is no longer capable of accepting both positive and negative input signals, and has been optimized for positive input charges alone.

The input to the preamplifier is connected to an injection capacitor, which can be used for calibrating and testing every cell, without requiring a connection to a detector.

Discriminator The discriminator is architecturally similar to that used in the Alice2Test chip. The output of the shaper is connected to one of the terminals of a differential pair of transistors, whilst the other terminal is connected to the threshold-setting bias voltage. The current which flows in the differential pair is connected to a current comparator with a rise time of ~ 4 ns. All the hits should fall within a 20 ns window provided they deposit 160 e^- above threshold. Three bits are stored locally in each pixel cell to allow a fine threshold adjustment. These bits control an offset current of the comparator which changes the threshold of the individual pixels. The output of the discriminator is connected to a synchronization circuit which detects a negative going edge and produces a pulse with a duration of half a clock cycle. This pulse can be masked if the cell is noisy or defective. This short pulse is connected to two circuits, each of which sends a pulse of 2 µA to the bottom of the column. Here, the pulses are sensed and used to generate the Fast-OR and Fast-multiplicity signals (see below).

Delay circuit The delay circuitry consists of two registers made of eight latches. Each register is connected to a digital comparator. In this way, up to two hits can be delayed during the LVL1 trigger latency, while minimizing power consumption and the silicon area. One counter at the periphery of the chip counts continuously up to n and then down to zero, where 2n corresponds to the number of clock periods required by the LVL1 trigger latency. This counter is 8-bit long providing up to 512 clock cycles of delay. The counter contents are transmitted to the columns in the form of a Gray code which limits the number of digital transitions to one per clock cycle. This minimizes power consumption and the risk of noise injection into the analog circuitry. A simple state machine is included which determines which delay register is free and rejects a third hit during the latency interval. When a pixel is hit, the short pulse at the output of the synchronization circuit latches the 8-bit Gray code value, which is stored in one of the registers on the cell. This number is then compared with the counter value at each clock cycle. Each register has its own logic which identifies the second true comparison and produces a short coincidence pulse. If this pulse coincides with the arrival of the strobe signal (corresponding to the LVL1 trigger), a logical one is saved in the first available FIFO latch (see below). The falling edge of the pulse from the comparison is used to reset the delay register, so that it is ready to record the next hit.

If the required delay does not correspond to an even number of clock cycles, the strobe must be delayed by one extra clock cycle before being applied to the chip. This delay is provided by the pilot electronics.

Readout circuit The strobed data are stored in a 4-bit deep FIFO buffer. The read and write pointers of the FIFO are controlled by the peripheral logic and ultimately by the pilot chip. The purpose of the FIFO is to enable a derandomization of the data to be transferred out of the chip (see next section). When a Next Event Read signal is received, the contents of the FIFO location enabled by the read pointer are transferred to the output data flip-flop. The read pointer is incremented by one. The data flip-flops, one per cell, are of the fully static master-slave type, and are connected together as a shift register. If a pixel has a hit, the flip-flop will latch a logic one. Otherwise, it will remain in its previous state. This is due to space constraints. This feature requires that the flip-flops be reset to zero prior to loading. A total of 256 clock pulses are needed to read out one chip.

The readout sequence can be aborted while the data is still in the front-end chip, either in the FIFO or on the data output shift register. If the data is still on the FIFO, a Next Event Read pulse has to be given in order to transmit the data from the FIFO to the data shift register. Once the data is on the shift register, this can be reset by an Abort signal.

Configuration flip-flops Each cell contains five unupsettable flip-flops, the Matrix set-up Registers (MRs), controlling the test input, the mask, and the three-bit threshold adjust.

Peripheral control electronics

There are two 4-bit counters which, in the ALICE mode, are configured as modulo-4 counters. They are used to generate the values of the write and read pointers of the FIFO. The strobe signal increments the write pointer whereas the Next Event Read signal increments the read pointer. The coincidence of these two pointers is prohibited by the pilot chip.

The periphery also contains the modulo-n up-down counter for the delay. The contents of the counter are Gray-scale encoded and latched by the clock before being buffered and sent along the columns.

Fast-OR and Fast-multiplicity There are two circuits at the bottom of the column which sense the accumulated pulses from one column. The input to both circuits is a buffer which conveys the current from a high capacitance node to a low capacitance node. With the Fast-OR sensing circuit, the next stage is optimized to detect a small signal with high precision. The output of this circuit is a logic pulse which is Or-ed with the outputs of the other columns and is transmitted off chip.

With the Fast-Multiplicity sensing circuit, there is an amplifier which is optimized for the dynamic range. The output of this circuit is a current pulse which is added to the pulses from other columns. The sum is transmitted off chip.

Biassing circuitry Experience with the Omega2 and Omega3 telescopes taught us that most of the non-uniformity in the behaviour of large systems comes from systematic or random chip-to-chip variations in biassing. Although systematic variations can be eliminated by careful design of the power supplies, random chip-to-chip variations can only be reduced by individual adjustment of the chip bias voltages. In the Alice1 chip the analog biassing is implemented using DACs integrated on the periphery of the front-end chip. We have designed an 8-bit DAC which generates current locally. The current is then mirrored to the pixels. The range of currents is determined by the aspect ratio of the output PMOS pair.

Chip Configuration Joint Test Action Group (JTAG) circuitry For the configuration of the chip we employ the JTAG protocol, which is described in the next section. A standard state machine is used to address either the Instruction Register (IR) or one of a number of Data Registers (DR) on the front-end chip. The contents of the IR determine whether the chip is to be by-passed or which of the DRs is to be addressed. There are separate JTAG registers for the global variables (DAC values, reference voltages, digital control bits) and for the configuration bits inside the pixel cells. We refer to them respectively as Global Registers (GRs), and Matrix Registers (MRs). The columns will be configured one at a time. A 5-bit enable register (ENBL) selects which column is to be configured, or which global variable is to be configured (see Section).

I/O Pads and logic Most of the pads at the periphery of the chip are for digital I/O. There are five pads dedicated to JTAG and a 32-bit Output bus. Gunning Transceiver Logic [45] was chosen as the standard for the digital I/O of the chip. This logic standard is relatively low-swing (0.4 V to 1.2 V) and, as there is a general reference voltage which determines the mid-point, these levels can be adapted in case of need. GTL is single ended and open drain. As a result of this and of the sparse nature of the data, we have chosen to employ negative logic, to minimize power consumption. It was impossible to use differential logic off-chip because of a limit in the number of lines on the support bus. However, a great effort has been made to control power supply bounce by providing variable slew-rate control on the output buffers.

2.2 Detector modules

The detectors to be used in the ALICE SPD will be very similar to those already used for the Omega2 and Omega3 planes in the WA97 / NA57 telescope. The first part of this section is dedicated to the description of the silicon detector substrates: first we recall the properties of the Omega2 and Omega3 detectors and then we describe the detectors we intend to use for ALICE. We then move on to discuss flip-chip assembly of front-end chips and detector ladders. We conclude with an outline of the component qualification steps for the module production.

2.2.1 Omega2 and Omega3 detector ladders

The basic unit of the Omega2 / Omega3 planes is called a ‘ladder’. It is a hybrid assembly of six CMOS front-end chips on a high-resistivity silicon detector substrate. Omega2 (Omega3) detectors measure 54 x 5.8 mm^2 (55 x 8.3 mm^2). The Omega2 (Omega3) detector substrate is a 300 μm thick matrix of 96 columns x 63 rows (96 columns x 127 rows) of p$^+$/n, ion-implanted rectifying diodes surrounded by a guardring. An additional row of dummy cells at the bottom, near the I/O pads of the front-end chip, is connected to a leakage current-sensing circuit for each column. The guardring is connected to ground through the front-end chip.

The diodes have a pitch of 75 μm for the Omega2 (50 μm for the Omega3) in the short dimension and 500 μm in the long dimension, matching the dimensions of the front-end cells. At the boundary between two front-end chips the front-end cells are connected to 1 mm long detector cells. This allows for a gap of several hundred μm between the front-end chips without loss of detector coverage. The choice of a different overall size for the detector and the electronics dies is due to yield considerations. The dimensions of the front-end chip are limited by the maximum reticle size allowed by the foundry and by the expected yield for a given chip size. The Si detector processing is much simpler and in this case the whole detector is made using one set of large masks.

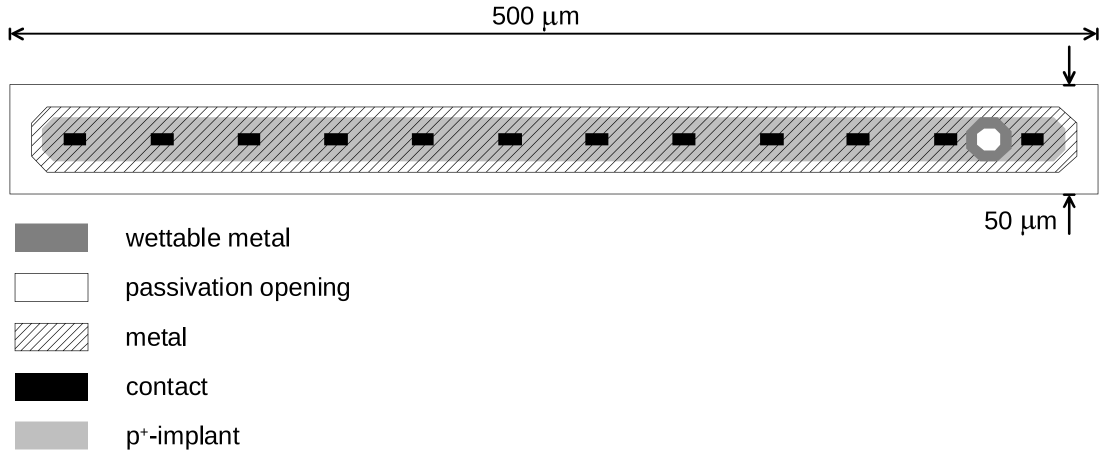
The detectors are built on a high-resistivity silicon substrate. The diodes are p^+-implants created by ion implantation. This allows good control over the junction depth. The implants are entirely covered by a metal layer with an overlap of several microns. A passivation layer covers the whole detector except where a connection to the front-end chip is foreseen. The guardring is a large diode surrounding the matrix of detector diodes. It protects the active matrix from the influence of defects (cracks) introduced when cutting the wafer into chips which would otherwise induce surface leakage. Some technical details of the processing carried out by Canberra [46] are shown in Table 2-1 on page 62.

TABLE 2-1. Properties of Canberra processing for Omega2, Omega3 detectors.

Wafer resistivity	17500 Ω cm
Wafer N_D - N_A	2.5×10^{11} cm^{-3}
Wafer thickness	300 µm
Wafer diameter	4” / 5”
p^+-implant thickness	500 nm
Al layer thickness	800 nm
Passivation thickness	1-2 µm
Backside n+-implant thickness	500 nm

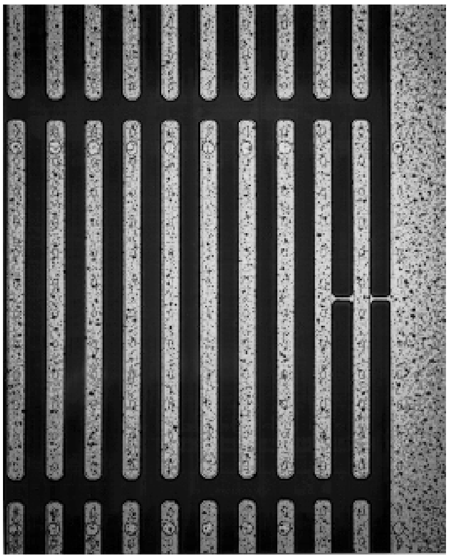
As the pixel capacitance has an influence on the noise behaviour, we attempted to minimize it. For the Omega3 detectors, two different designs were produced: one called the ‘conventional’ (c-type) and one called the ‘advanced’ (a-type), as described below.

FIGURE 2-29. Layout of a conventional (c-type) detector cell.



In the conventional layout, shown in Fig. 2-29, the implant covers a large fraction of the cell area and stops 15 µm short of the cell boundary. A metal layer covers and overlaps each implant by 5 µm, leaving a 10 µm-wide area around uncovered. The edges of the implant and metal layers are rounded in order to avoid too high electric fields at the edges. The opening in the otherwise continuous passivation layer has a diameter of 10 µm and the octagonal wettable metal pad on top of it measures 20 µm across. A photograph of some cells at the bottom of one column is shown in Fig. 2-30. The two lowermost and the two uppermost cells in a column are connected to the guardring to create a homogeneous electric field for the entire sensitive area. For the same reason, the guardring on the left and on the right also mimic the shape of the basic cell.

FIGURE 2-30. **Photograph of the bottom region of one column of a c-type Omega3 detector.**



The advanced layout is based on an attempt to reduce the inter-pixel capacitance as much as possible by increasing the distance between the implants and metal layers of adjacent cells. A single cell consists of five minimum-sized implants, measuring 25 x 20 µm^2, connected with a metal strip. The metal strip has a minimal width except where it widens up to cover the implants. A photograph of some cells at the bottom of one column is shown in Fig. 2-31. Again, the two lowermost and the two uppermost cells constitute a part of the guardring which, on the left and right sides, resembles the cell structure. Owing to the irregular implant structure in this detector, the electric field near the surface will be distorted. This can result in collection time variations for charge deposited at different positions if the detector bias is too low [47]. As we measured no appreciable difference in the noise of the advanced detector assemblies compared with the conventional assemblies, we decided to discontinue studying the advanced type.

FIGURE 2-31. **Photograph of the bottom region of one column of an a-type Omega3 detector.**

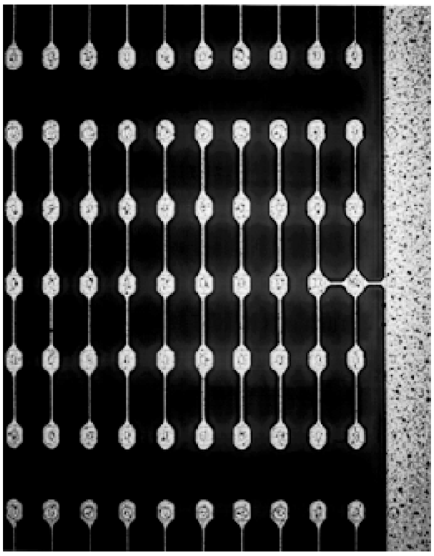
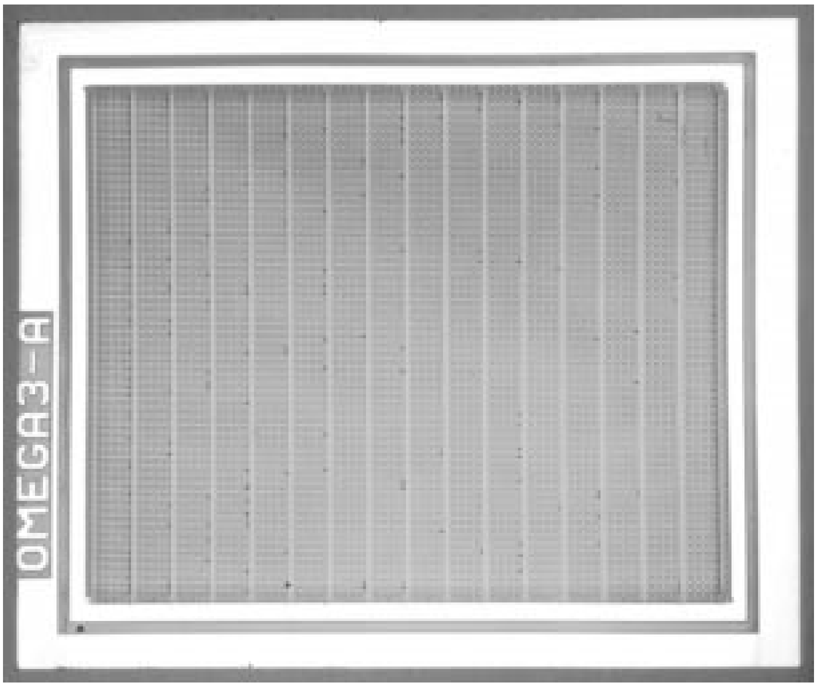
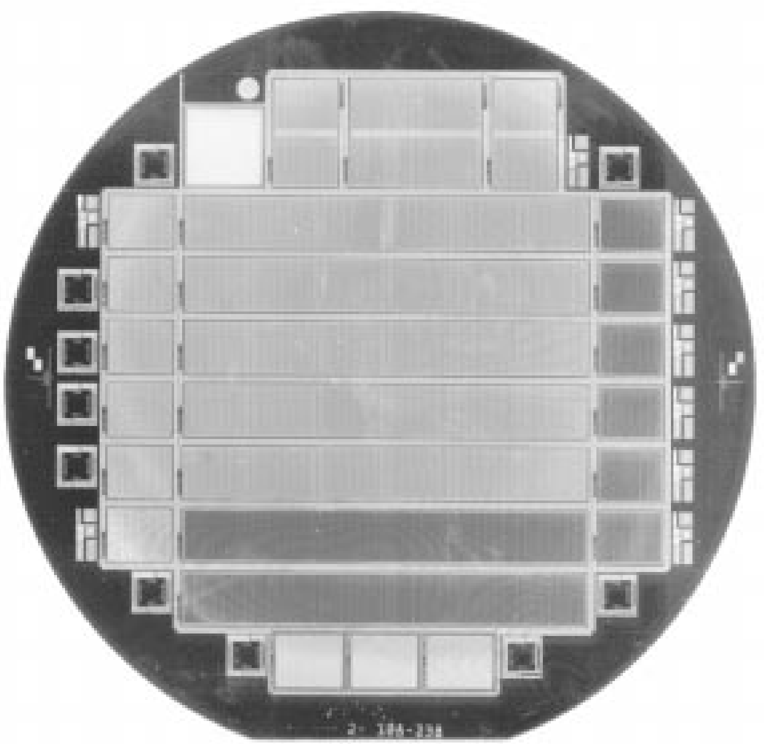


FIGURE 2-32. **Photograph of a single Omega3 c-type detector chip.**



For use with a single front-end chip the cells are arranged in a matrix of 128 rows and 16 columns, surrounded by an approximately 500 μm-wide guardring comprising the aforementioned dummy cells. A photograph of a single chip is shown in Fig. 2-32. The device measures 8.3 x 10 mm^2. Two Omega3 wafer designs were submitted to Canberra: a first design for a 4” wafer in spring 1995, and a slightly revised design for a 5” wafer in May 1996. The change to the larger wafer diameter allowed a higher number of detectors to be produced at only slightly higher costs. The 4” wafer contained seven ladders (two of type a, five of type c). The 5” wafers contained 12 ladders along with a certain number of singles and test structures. A photograph of the 4” wafer is shown in Fig. 2-33. The company’s change to 5” wafers also allowed some necessary changes in alignment marks for the flip chip bonding process to be implemented. These alignment marks in wettable metal were not easily distinguishable from the underlying metal layer in the first design.

FIGURE 2-33. **Photograph of a 4” Omega3 detector wafer.**



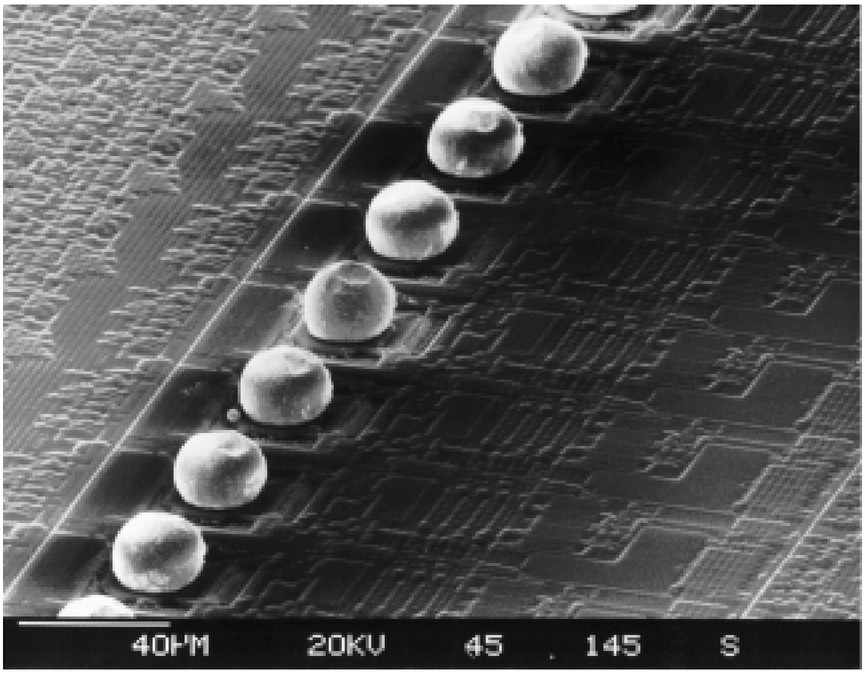
2.2.2 ALICE detector ladders

The ALICE detector ladders will look similar to those used for the Omega3 telescope. Since the anticipated total dose for the pixel layers is about 200 krad, we will continue to use the simplest and therefore cheapest material which is p^+ on n. Owing to the new leakage current subtraction scheme implemented in the ALICE front-end chip, a special row of cells for leakage current sensing will not be necessary. A schematic drawing of a detector is shown in Fig. ch2fig:a_b. The ladder will comprise a matrix of 256 rows and 256 columns of 50 x 300 μm^2 pixels, and will be flip-chip bonded to eight front-end chips, each one serving a sub-matrix of 256 rows x 32 columns. At the boundary between two front-end chips, there will be two columns of pixels with a size of 50 µm x 600 µm. This reduction in the size of the boundary columns with respect to that of the Omega2 and Omega3 chips is the result of using a 5-layer CMOS process, which eliminates the requirement for power supply routing around the sensitive electronics. The active area will be surrounded by a fiducial area of a width of 500 µm all around, comprising the guardring. The total size of the detectors will be 82 x 13.8 mm^2, for a thickness of 150 µm.

2.2.3 Flip-chip assembly

The electrical and mechanical connection between the detector and front-end chip is established by solder balls in a flip-chip solder-bonding process. The original solder flip-chip process was developed by IBM around 1970 [48] as a more reliable alternative to wire-bonding. It allows a much higher density of connections per unit area and offers the advantage of reduced capacitance and inductance. The average bond size in this process was of the order of 100 µm with a pitch of about 250 µm. Other manufacturers developed this technology further for finer pitches [49]. For both the Omega2 and the Omega3/LHC1 assemblies a fine-pitch solder-bump technology developed by GEC-Marconi Ltd. was chosen [50]. A scanning electron microscope photograph of a solder-bumped Omega3 chip is shown in Fig. 2-34.

FIGURE 2-34. **SEM photograph of solder bumps on an Omega3 chip.**



Yield considerations

An analysis of the yield figures for the Omega2 assemblies can be found in Ref.[51]. For the production of Omega2 ladders, a full functional test of the chips before assembly was not performed. As a result, a large fraction of the ladder yield loss was due to bad electronics. Nevertheless, an indirect calculation of the intrinsic bump-bonding yield, corrected for losses due to faulty chips, was performed, resulting in a ladder yield figure of around 80%. At present, about 0.5 M channels of bump-bonded Omega2 pixels are employed in the NA57 pixel telescope.

The assembly of Omega3-type ladders turned out to be more problematic. Part of the problems were traced down to the poor reflectivity of the Omega3 wafers, which caused problems at the level of the optical alignment of the detector-chip assemblies for the bonding operation. These were overcome by modifications to the illumination system. Another important issue has been bump oxidation: a different bump shape was adopted for the Omega3 bumps, which resulted in a different behaviour of the oxide formed at the bump surface after reflow. This in turn resulted in poor electrical contact on a sizeable number of bumps, and therefore in yield problems. While these problems at the bonding level now seem to have been overcome, there are still problems of fluctuating yield which are probably due to a poor control of the wettable metal deposition. GEC-Marconi is commissioning a new wafer processing line where the control of the metal deposition should be much improved. In the process of understanding these problems, we experienced large fluctuations from batch to batch in the Omega3 ladder yield, which, even for the best batches, never exceeded 70%. At present, a total of about 0.6 M channels of bump-bonded Omega3 pixels are employed in NA57, bringing the total number of channels in the pixel telescope to about 1.1 M.

An important lesson was learned from the Omega3 campaign. Owing to the complexity of the pixel ladder systems, the task of qualifying the produced assemblies could not be carried out directly by the bump-bonding vendor: relatively large batches of ladders were assembled and shipped for testing to CERN or other collaborating institutes, where the ladders were tested under probe stations. As a result of this, most of the assembly problems were detected only after a full batch had been delivered, resulting in large component losses. For the ALICE SPD, we will implement a bump-bonding test pattern around the sensitive area of the front-end chip and the detector ladder, with a pitch corresponding to the small dimension of the matrix. A simple continuity test on this pattern should enable the bump-bonding supplier to detect a large fraction of assembly problems and have immediate quality control feedback for optimizing the process. Only ladders which pass the continuity test will then be shipped for the full probe station test.

Component thinning

The total thickness of the ALICE detector and the electronics assembly has to be limited due to material budget considerations. We are aiming for a total material budget of 250 µm of Si and are planning to assemble electronics chips thinned down to a thickness of 100 µm on a 150 µm-thick detector.

Omega2 and Omega3 wafers were already thinned down, prior to shipment, from the typical wafer thickness of ~600 µm to a thickness of ~300 µm. In the framework of the RD19 Project [52], attempts were made to go below the 300 µm thickness. Although a large number of dies were lost in the process, a few specimens of thin detector + electronics (150 µm + 80 µm) Omega2 assemblies were produced and tested with good results, demonstrating the possibility of obtaining thin working prototypes.

Subsequently, an Omega3 wafer was thinned down from 300 µm to ~120 µm. The thinning operation was performed successfully, but problems were encountered at the wafer-processing stage during the wettable metal deposition, resulting in a large loss of dies.

GEC-Marconi is confident that these problems will be overcome by processing the wafers prior to thinning them down, and proposes to include these steps in the process. This company has experience with this type of operation from other products. Although this has not been tried on the Omega3 at wafer level yet, GEC has successfully thinned down diced processed Omega3 chips to well below our target thickness of 100 µm. Within the next few months, we should receive some prototype Omega3 assemblies made with thin electronics chips for testing.

Suppliers for the assembly of the ALICE SPD ladders

A non-negligible restriction in the selection of vendors is that the vendor must be able to process 8” front-end wafers from our 0.25 µm CMOS supplier. GEC-Marconi has so far delivered to us over a million working fine-pitch bumps and has experience in component thinning. Besides GEC-Marconi, we are considering two other potential suppliers of fine-pitch flip-chip bonding.

One possible alternative is IZM, Germany, who is already working with the ATLAS experiment and has 8” capabilities. Alenia, Italy, also working with ATLAS, is another candidate, although so far this company only has 6” processing facilities. A few other potential vendors are coming into the market, and we are of course watching closely what other pixel detector projects are doing in this field.

Component qualification

One key element in the production of such a detector is the testing and qualification of the components. All the groups involved in the ALICE SPD project are equipped with semiautomatic or automatic probe stations. In this section we shall review the various levels of testing which will be required to build the full system. The procedure closely matches the one developed and employed for the production of the Omega3 detectors, with the important addition of the continuity test at the bump-bonding supplier site. The detector wafers will be tested by the manufacturer.

From experience, we know that a sizeable fraction of the front-end chips on a wafer is defective. It would therefore be too costly to assemble front-end chips without testing them beforehand. The first step will consist of probing the 8” front-end chip wafers. The test features implemented in the ALICE SPD chips will allow full functional tests to be performed at this stage, connecting the front-end chip pads to a VME readout or to a general-purpose tester card via the probe needles. The wafer tests will permit Known Good Die (KGD) maps to be established at the wafer level.

Only KGD will be selected for flip-chip assembly to the detector wafers. As discussed above, a bump-bonding test pattern should allow the supplier to reject most of the defective assemblies via a simple continuity test. Besides improving the optimization procedure, this will increase the reliability of the ladders delivered. This first stage of ladder tests, however, will not eliminate the need to probe-test the ladders before using them in the experiment.

A second stage of ladder tests will be performed on the shipped assemblies. This time, full functional tests will be performed in two steps: first the test protocol used for the wafer-level tests will be repeated, to check that the chips are still behaving as they did before wafer processing, wafer dicing, and bonding. Then, a final test will be performed exposing the ladder to a source of beta particles, in order to verify the detection functionality, and all the bump-bonding connections. Only the assemblies which successfully pass these final tests will be selected to be assembled in staves (See “Assembly, mechanics and cooling” on page 90.).

A final full test will be performed once the ladders are assembled on the stave carriers, before the staves are assembled on the SPD support sectors.

2.3 Bussing, readout and control

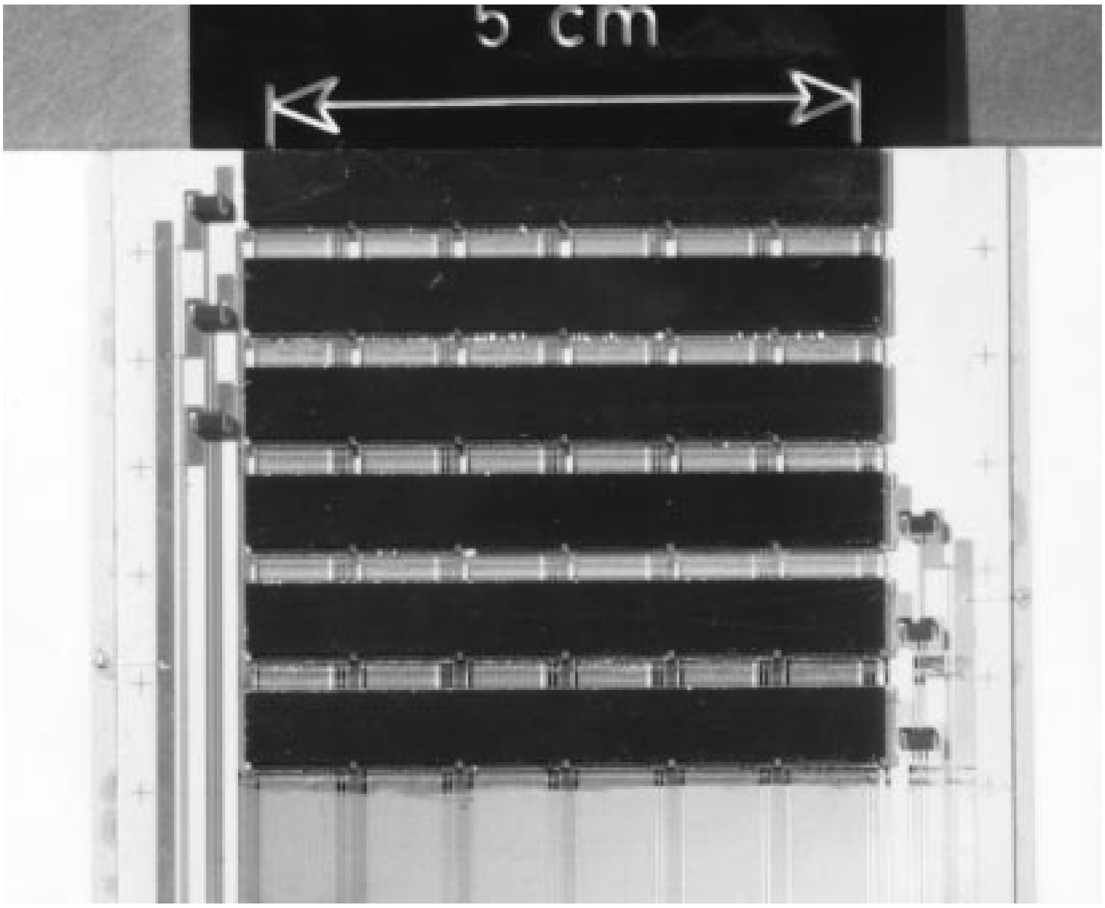
The present section deals with the areas of readout, control, and bussing for the ALICE SPD. Here, too, we shall start by describing the main features of the systems successfully employed with the Omega2 and Omega3 pixel detector planes and the experience gained from them. The complexity of the ALICE SPD and consequently the number of detector parameters are larger than those found in the previous systems. As a result, a notable difference between the system described here and the previous systems is the introduction of a powerful architecture for tests and controls based on the JTAG protocol. This is described in Section 2.3.2 together with the specific R&D activity undertaken in this field with the development of a JTAG accelerator to allow testing and control over a long distance. We shall then move on to describe, in some detail, the SPD readout logic. We shall conclude with a description of the data paths: the stave bus connecting the front-end chips with the pilot logic located at the end of the stave, and the short serial data link developed for the transmission of the data off the SPD barrel.

2.3.1 Bussing, readout, and control for the Omega2 and Omega3 systems

Omega2 and Omega3 array carriers

Six of the Omega2 pixel detector ladders, each bump-bonded to six front-end chips were mounted together on a 300 μm-thick ceramic carrier as shown in Fig. 2-35 resulting in what we call an ‘array’ [14]. One layer of gold lines for data and bias connections was deposited on the carrier surface. The pitch of the lines was 250 μm, matching that of the wirebonding pads on the front-end chips. The 36 front-end chips were then connected to this bus by means of ultrasonic wirebonding. A thin ceramic spacer was glued to the top of the bus under each ladder. The front-end chips were glued with conductive epoxy to the top surface of the spacer, which was metallized. This provided a very low inductance connection to V_{dd} for the chip substrate, ensuring low noise operation of the array. The back of the carrier was also coated with gold, in order to provide a ground plane for the assembly. Two of these ‘arrays’, suitably staggered, were used to build up each of the ‘logical planes’ of the Omega2 telescope [14], as shown in Fig. 2-36. Each ‘plane’ hermetically covers a 5 x 5 cm^2 area with pixel detectors and contains 72576 active pixel cells.

FIGURE 2-35. **A photograph of an Omega2 array.**

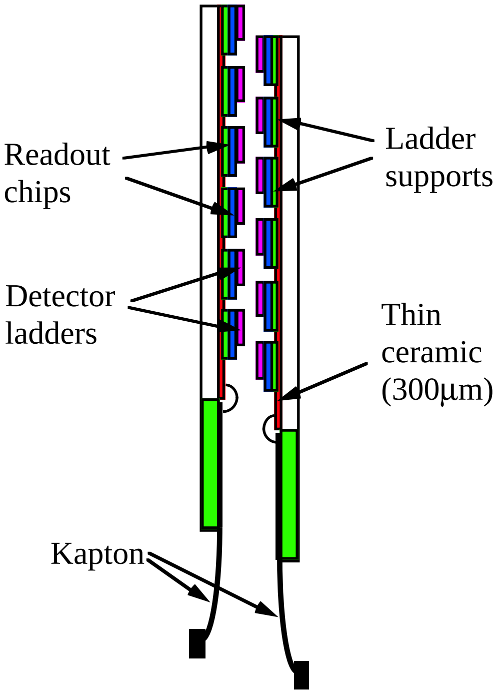


Two problems became evident during the operation of the first prototypes of the Omega2 arrays in test beams organized by the CERN RD19 collaboration and in the WA97 experiment:

1. the amount of local decoupling was not sufficient to maintain the power supply constant when resetting the digital part of the front-end chips,

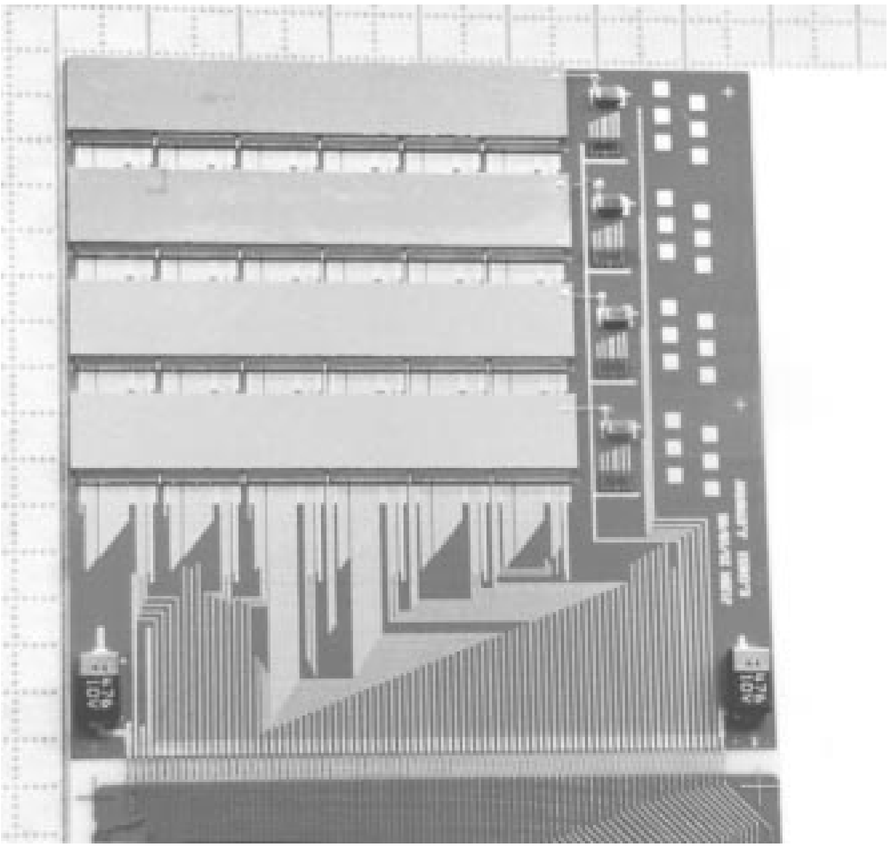
2. there was a systematic delay non-uniformity caused by voltage drops on the power supply lines (six front-end chips in a column were connected in parallel to the same line carrying the delay control current [53]).

FIGURE 2-36. **A schematic diagram of the assembly of two Omega2 arrays forming one logical plane.**



Both of these problems were addressed when building the carrier for the Omega3 pixel front-end chips. This time, a more advanced technology, allowing 200 µm line pitch and multiple conductive layers was chosen [47][54]. The Omega3 array (Fig. 2-37) consists of four detector ladders, each bump-bonded to six front-end chips. Five Al layers interleaved with polyimide dielectric are stacked on an alumina substrate, for a total thickness of ~ 400 µm. The Al layers provide x and y connectivity, V_{ss} and V_{dd} planes, and a V_{ss} plane on top of which the back of the front-end chips is glued with conductive epoxy. Compared to the Omega2 array, this construction results in lower mass, improved local decoupling, and the possibility to draw a single delay-control line per chip. The delay-control current is generated on the array itself from a common adjustable voltage using high-stability thin film integrated resistors of 0.5% accuracy, directly glued onto the carrier. SMD electrolytic and ceramic capacitors provide, respectively, more supply decoupling at low frequencies, and filtering for the detector bias lines. Assembly accuracy and speed were improved by using the automated ultrasonic bonding facility of the ECP/MIC group at CERN. As for the Omega2 system, two Omega3 arrays were staggered to cover a 5 x 5 cm^2 surface, forming a logical plane.

FIGURE 2-37. **A photograph of an Omega3 array. As well as the ladders, the carrier holds two large decoupling capacitors, and four integrated resistor chips (see text).**



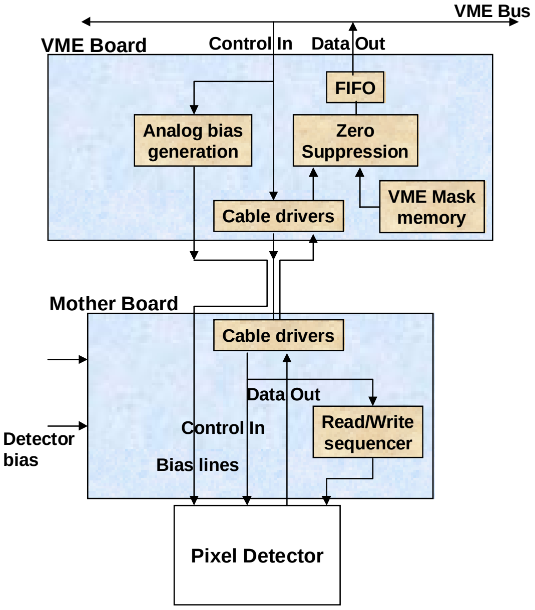
Omega2 and Omega3 read-out and control logic

Omega2 and Omega3 arrays are read out by a VME system, consisting of two PCBs equipped with programmable logic and standard CMOS and TTL ICs. One card, the ‘motherboard’, is directly connected to the array via a flex cable. This motherboard is then connected to a VME card, located in the control room, via a 30 m long cable consisting of 50 twisted pairs.

The Omega2 readout system is rather simple. The arrays are read out in parallel. The 36 chips of each array are read out sequentially in pairs. At the beginning of the readout, the first pair of chips is enabled. Then, 64 clock cycles are applied, during which the data from the 64 rows of the chips are shifted out and combined in 64 32-bit data words. The procedure is then repeated on the next pair of chips, and so on, until all the chips of the array have been read out. The entire readout requires 1152 clock cycles. The shifted words are transmitted to the VME board where a zero suppression logic skips those which contain only zeroes and produces a progressive address for the useful data. Words and addresses are then stored in a FIFO which is accessible from the VME bus, to be transferred to the VME processor which controls the system. The readout speed is 2 MHz, mainly limited by the fact that the readout control sequence is generated on the VME card, and has to be transmitted through the long cable to the motherboard. This results in a total readout time of 576 μs.

The mask memory present on the VME board allows each single pixel on the array to be masked and to exclude it from the readout should it be too noisy or defective. The VME card also contains a set of five 8-bit DACs, controlled via the VME bus. They provide the adjustable bias voltages needed to operate the Omega2 chips, to control the pixel threshold, and to set the internal delay of the pixel cells.

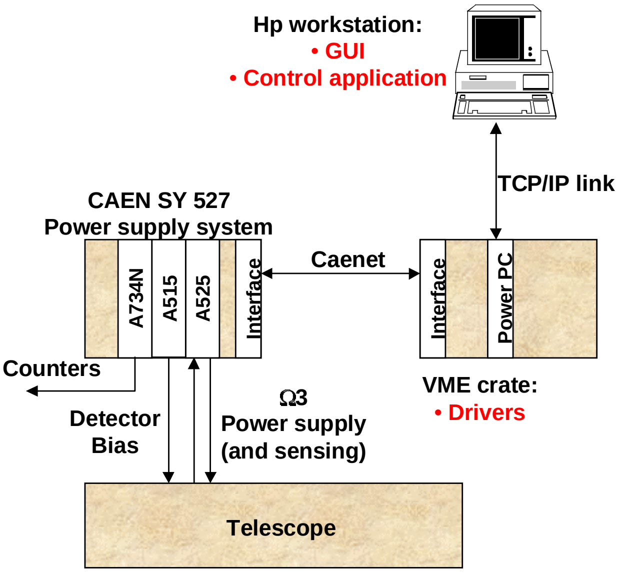
FIGURE 2-38. **Schematic diagram of the VME system used to read out the Omega3 arrays.**



The Omega3 VME system (Fig. 2-38) operates along the same lines. In addition, it features a bi-directional link between the VME bus and the pixel array. It is possible to download the values of the different control registers into the Omega3 pixel chips and to read them back. This allows single pixel cells in the Omega3 chip to be connected separately to a test input. This results in an improvement of the quality of the electrical tests. In the Omega3 system the readout sequence is controlled locally on the motherboard, allowing a significant improvement in the readout speed. The readout clock is also generated on the motherboard, and is activated only during readout, in order to minimize coherent noise coupling to the sensitive pixel front-end. The readout frequency is limited to 4 MHz by the speed of the programmable logic. Each card reads an array (49152 pixels) in about 400 μs. Zero suppression and FIFO buffering are very similar to those of the Omega2 system, as well as the DACs which provide analog biasses. Although it is possible to mask single pixels at chip level, a second level of masking is available in the VME.

In order to ensure reliable operation of the Omega3 detectors during data-taking, a cooling system and a slow control system were also designed and implemented in the NA57 experiment. We observed that local heating of the detector by the integrated readout electronics can cause a significant rise in the leakage current. This increases the electronics noise, so that more channels become permanently active and the temperature tends to rise further. This dangerous thermal runaway can be effectively circumvented by cooling the detectors. A simple scheme, in which each Omega3 plane is fluxed with ~ 100 l/h of N_2 at a temperature of 5-10°C proved to be effective.

FIGURE 2-39. **Scheme of the slow control system for the NA57 pixel telescope.**



The slow control system (Fig. 2-39) provides the detectors bias for all the Omega2 and Omega3 planes and the power supply for the Omega3 integrated electronics. All leakage currents and the power consumption of the Omega3 arrays are constantly monitored. If any of the currents exceeds their preset limit, the array affected by the fault is switched off and an alarm is sent to the control room, where a recovery procedure can be launched. A graphics interface developed under Labview provides an easy way to check all the parameters, keep a logbook of their history, and change the defaults.

2.3.2 JTAG control

Introduction

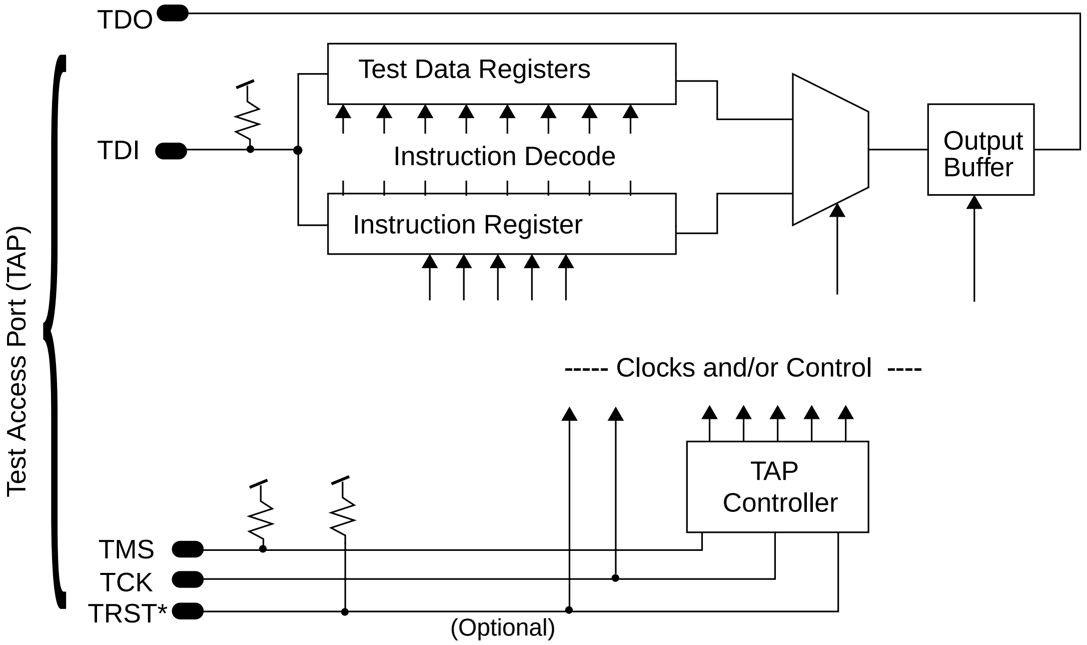
With the increasing use of electronic devices of high integration and densities (surface mounted devices, VLSIs, hybrids, dense packages, etc.), the Joint Test Action Group (JTAG) was set up in 1985 to define a standard methodology for testing their electronic systems. These devices cannot be easily tested with traditional probing techniques.

The JTAG was initially promoted by European electronics companies. Later, US companies also joined, and the step was made towards the organization of an international committee. By the summer of 1988, the JTAG Committee had produced a final version of an IEEE standard (IEEE 1149.1 [55]), which was approved late 1989. Today, this standard is well established in the electronics field and several components equipped with JTAG test features exist on the market.

The aim of IEEE 1149.1 is to define the test architecture to be embedded into a digital circuit in order to provide a standardized approach for testing electronics devices at different levels of hierarchy (an integrated circuit, a component mounted onto its board, the interconnections between functional blocks that are part of the system).

The standard test logic, shown in Fig. 2-40, consists of a Test Access Port (TAP), made of a TAP controller (a state machine of 16 states), a unique Instruction Register (IR), and some test Data Registers (DRs). Four mandatory signal lines entirely define the protocol: Test ClocK input (TCK), Test Mode Select input (TMS), Test Data Input (TDI) and Test Data Output (TDO). Optionally, a fifth signal line can be included: Test ReSeT input (TRST*).

FIGURE 2-40. IEEE1149.1 standard test logic.



The very high level of design complexity and the dense component integration of the pixel detector system, where several ICs are physically mounted together to build a standalone device (the ladder), require a careful design of the test features. The test protocol should allow simplification of component verification at each assembly step.

In addition, the test protocol is an attractive solution for performing detector control tasks, essentially without adding extra resources: in practice, part of the test Data Registers (DRs) are effectively detector set-up registers.

Such a solution offers three fundamental benefits:

1. detector control is embedded into the same minimal-size architecture used for testing,

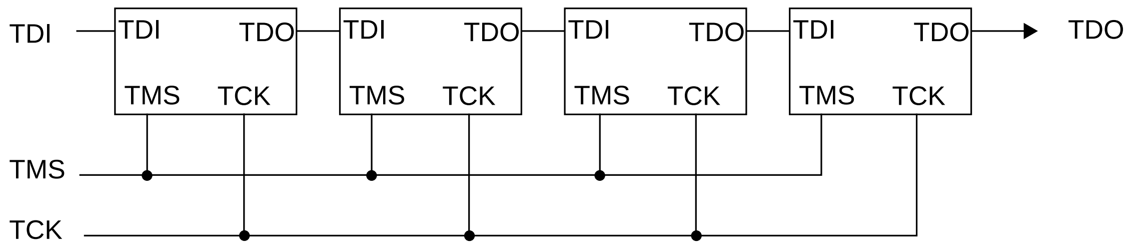
2. the control interface is completely standard (i.e. commercially-produced JTAG controllers and basic JTAG software can be purchased),

3. the controls are managed via a dedicated bus with a minimal number of lines.

Fundamentals of the JTAG specification

The JTAG architecture is based on a serial chain data path where several physical components are connected by daisy chaining Test Data Outputs (TDOs) onto Test Data Inputs (TDIs). A bus formed by TCK, TMS and TRST* provides the distribution of control signals for the TAP controller, as shown in Fig. 2-41.

FIGURE 2-41. **Basic daisy chain of JTAG slave devices.**



IRs and **DR**s are shift register-based circuits, which can be equipped with a latch to save the shifted data. The IR register is loaded with the instruction that selects the DR to be addressed, corresponding to the test to be performed or the setup to be downloaded. The DRs are a bank of registers each related to a specific test action or a setup configuration.

TCK is a clock signal dedicated to the synchronization of the JTAG operations. Its rising edge is used to sample values of the input signals (TMS and TDI) into the TAP controller. Its falling edge is used to drive the values of the output signal (TDO) of the TAP controller. The TCK can be run in a burst mode, and its duty cycle can be different from 50%.

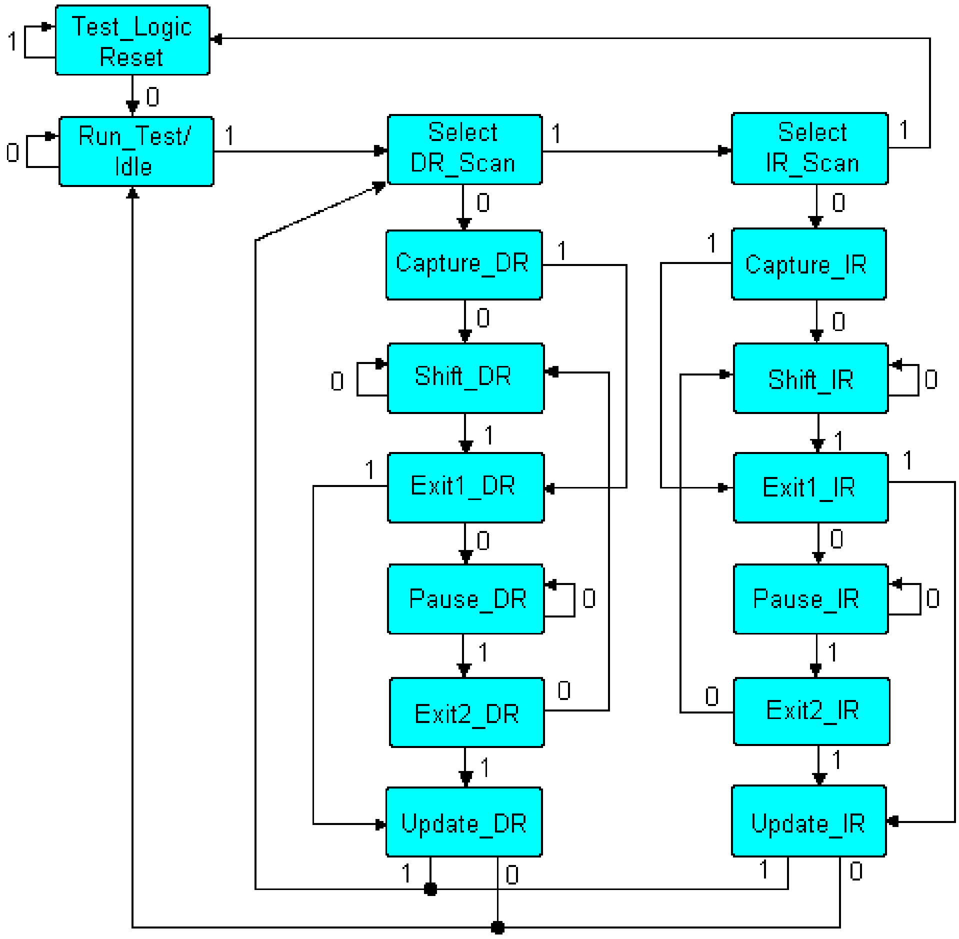
TMS is an input signal which changes state at the falling edge of the TCK. It consists of a sequence of logical 0s and 1s. The sequence is used by the TAP controller state machine to step through its own states. A pull-up resistor, or an equivalent circuit, ensures a logical high state whenever the TMS is not driven properly. In this way, the state machine will always end up in a reset state (inactive) after a maximum of five clock periods.

TDI is an input data signal which changes state at the TCK falling edge. It is used to feed data into the IR or the DRs of the TAP controller. It is connected to the preceding TDO signal line. As with the TMS, a pull-up resistor, or an equivalent circuit, holds a high logic state if there is a problem from the TDO connection. Such a faulty situation forces the TAP controller to always address the by-pass register, i.e. a dummy data register which does not corrupt the full daisy chain data.

TDO is an output data signal which changes state at the TCK falling edge. It shifts the data out of the selected register (IR or a selected DR). The TDO data must not be inverted with respect to the TDI data. When data are not being shifted out, the TDO is set to an inactive drive state, i.e. high impedance.

TRST* is an optional negative true input signal. It asynchronously resets the TAP internal state to a predetermined dummy configuration and brings the TAP controller state machine in the TEST-LOGIC-RESET state (see below). It is equipped with a pull-up, or an equivalent circuit, to keep it inactive if there are problems. When the TRST* is active, the whole TAP controller must not interfere with the normal operation of the device: TRST* can also be seen as an ‘Enable Test Mode’ signal. The TRST* rising edge, i.e. the end of the asynchronous reset state, must take place when the TMS is high, in order to ensure a race-free operation of the TAP controller state machine.

FIGURE 2-42. **State diagram of TAP controller** (input: TMS on rising edge of TCK)**.**



The TAP controller is a finite 16-state machine which operates synchronously with the TCK rising edge and according to the TMS logic values (see Fig. 2-42).

The states are defined as follows:

TEST-LOGIC-RESET: This is the reset state. The TAP controller is disabled. This state must be forced at power-on, or any time the TRST* is activated. The IR must be initialized either to the value of the optional ID code, or to the mandatory Select By-Pass register instruction.

RUN-TEST/IDLE: This state allows to run an IR-defined Built In Self Test (BIST). If no internal self tests are selected, this is just an idle state.

CAPTURE-DR (or **CAPTURE-IR**): This state preloads into the selected DR (or into the IR) the value of the corresponding parallel latch register. If no parallel latch register is implemented, the DR may retain its current value. (The IR preloads a fixed and optionally design-defined value.)

SHIFT-DR (or **SHIFT-IR**): In this state the shift between TDI and TDO takes place synchronously with the TCK rising edge. Each instruction must select only a single DR (or the IR). During the shift operation the TDO output must be enabled, whereas it must be disabled for all other states.

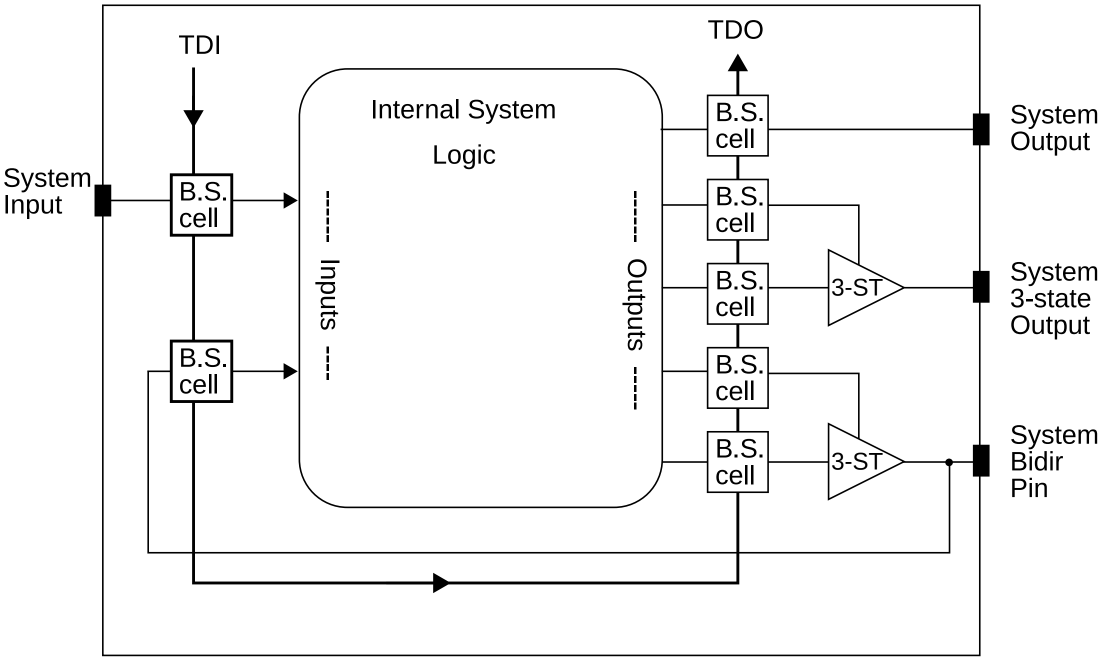
UPDATE-DR (or **UPDATE-IR**): This state completes the shifting process by validating, at the TCK falling edge, the shifted DR data (or the IR data) onto the parallel latches. If no updating is required, this state has no effect. (The IR takes the new validated instruction which becomes the current instruction.)

PAUSE-DR (or **PAUSE-IR**): This state allows the shift process to be temporarily halted, for instance to allow the JTAG controller to fetch data from the disk unit.

SELECT-DR (or **SELECT-IR**), **EXIT1-DR** (or **EXIT1-IR**), **EXIT2-DR** (or **EXIT2-IR**): These states are decision points which allow the TAP state machine to step through the state diagram.

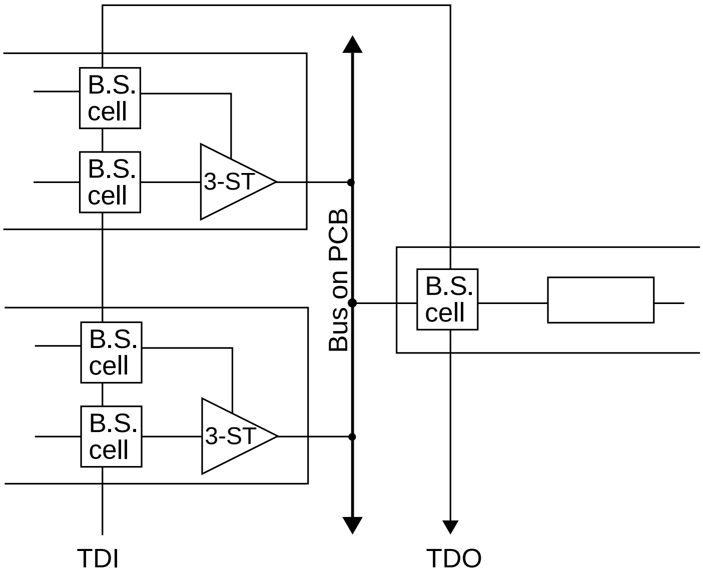
The mandatory TAP controller registers are: only one IR and two DRs, i.e. a By-Pass (BP) register and a Boundary Scan (BS) register. The BP register is a single bit shift register that gives a minimum data path length between the TDI and the TDO. It loads a logical zero at CAPTURE-DR, if selected by the BYPASS instruction. It has no parallel latches, therefore the UPDATE-DR has no effect on this register. The BS register consists of a number of Boundary Scan cells connected to the I/O pins and to their tristate enables if the output has the tristate feature, and/or to the direction controls if the pin is bidirectional. A typical configuration is shown in Fig. 2-43.

FIGURE 2-43. **Boundary scan cells.**



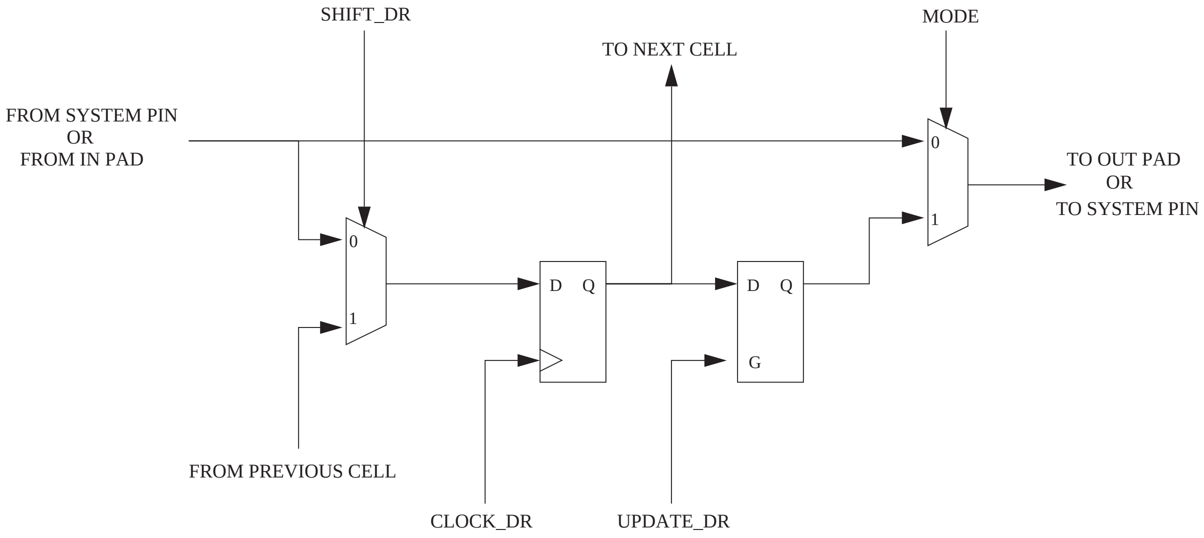
Through the BS test the full board-level electrical connection between any pair of input/output pins can be checked, validating the data path (see Fig. 45). Note that this includes the whole connection, i.e. from the inside of the driving device to the inside of the receiving device.

FIGURE 2-44. **Boundary scan cells connected at bus level.**



A general-purpose BS cell that could be used for both IN and OUT pins is shown in Fig. 2-45. Simpler versions with limited features can also be designed.

FIGURE 2-45. **General purpose boundary scan cell diagram.**



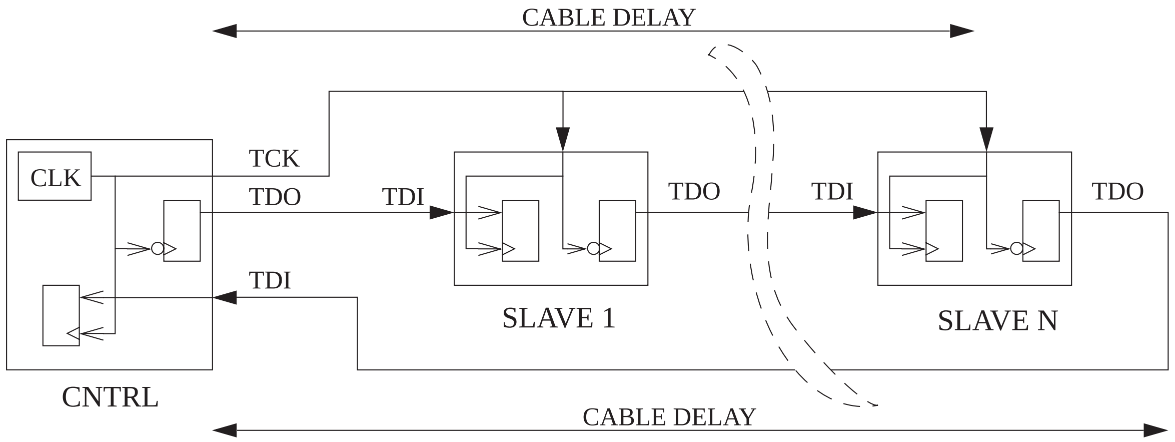
R&D issues

The minimum cable length between the SPD detector and the external control electronics (~ 40 m) would put a serious limitation on the data rate of a standard JTAG system. This is mainly due to the fact that JTAG has not been designed to transmit data over long distances, but rather to provide a reliable way to test functional modules. We have developed a solution to overcome this limitation.

The IEEE 1149.1 specification does not define the physical level. This can be a conductor cable (either single ended or differential) as well as an optical fibre. Very often the manufacturers of JTAG controllers do not take into consideration the distance issue. Today, it is impossible to find commercial modules which function correctly in a high-speed environment.

Another important problem concerns the JTAG specification. The standard requires that data being shifted out of a device make a transition on the negative-going edge of TCK, while data captured into a device are sampled on the positive-going edge of TCK. In addition, the TCK is not circulated back following the full daisy-chain connection: the same signal line on which the TCK is generated by the JTAG controller is used both to clock out the TDO from the JTAG controller and to clock in the TDI data coming back from the last device in the chain (see Fig. 2-46).

FIGURE 2-46. **Effects of long cable delay on JTAG.**



In our configuration, the problem would occur at the JTAG controller side, since the delay to get the TDI signal back to the controller approaches 1/2 of the 10 MHz TCK clock period (here, the TCK and TDI phases are not compensated). Any deformation of the TCK signal would worsen the problem.

In general, this problem would be solved by reducing the TCK frequency. This, however, is not our preferred solution, as we wish to use JTAG also for a fast downloading of the front-end chip parameters.

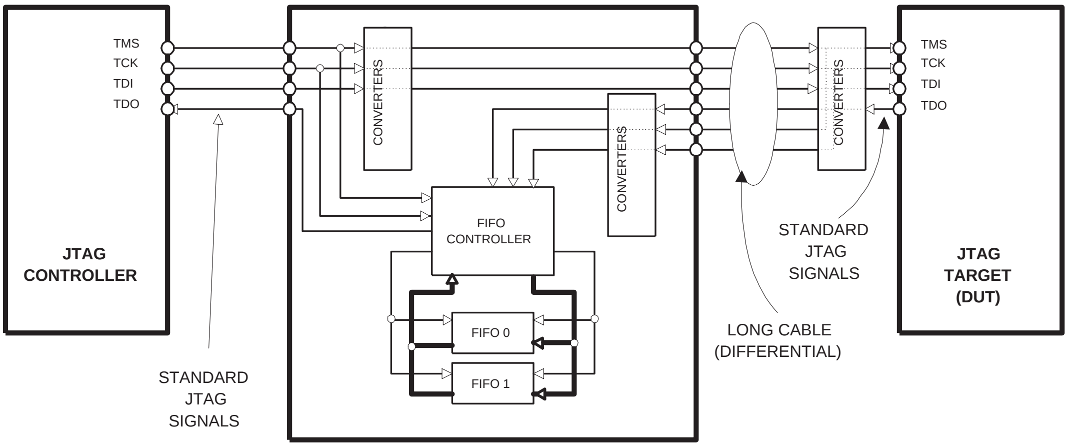
The solution we have developed consists in the introduction of a JTAG Accelerator Unit as close as possible to the JTAG controller. It is used to transmit the signals differentially over a long distance. At the target side, the signals are reconverted into single ended ones. Three additional lines are added: a TCK-ret (TCK returned), a TDO-ret (TDO returned) and a TMS-ret (TMS returned). The JTAG connector has been modified in a backward compatible way.

The Accelerator Unit, schematically shown in Fig. 2-47 has two FIFOs working in swapped mode: one (the W-FIFO) receives the serial data from the farthest device on the JTAG chain, the other one (the R-FIFO) serially transfers the previous data to the nearby JTAG controller. At Capture action, the content of the W-FIFO is moved into the R-FIFO (i.e. the two are swapped). At the same time the W-FIFO is zeroed.

The Accelerator Unit also has two TAP ports: one driven by the signals from the nearby JTAG controller, the other driven by the signals returned from the far end.

In this way, all the long-distance signals (TDI, TMS, TCK and TDO-ret, TMS-ret, TCK-ret) travel in the same direction, preventing any skewing among them. The penalty introduced by the use of two FIFOs is the need of an extra shift cycle when data have to be read back. If the data have only to be written and what is read back is discarded, there is of course no penalty. This solution requires a minor departure from the IEEE 1149.1 specification, which can be accounted for in the software.

FIGURE 2-47. **JTAG accelerator card block diagram.**



The additional time required by the extra shift cycle is amply compensated for by the increased clock frequency (we have achieved a factor of 20). We have been able to transmit data at a rate of 25 Mbit/s (the maximum possible with our JTAG controller) over a distance of 50 m.

JTAG architecture for the ALICE SPD

Besides the mandatory DRs, in the ALICE front-end chip we are implementing a number of additional DRs to store the set-up information. The whole architecture is shown in Fig. 2-48.

There are Matrix (MR) and Global (GR) set-up Registers. The GRs are located in the periphery of the chip and control global bias voltages/currents and global thresholds. The MRs are located inside the cells and control individual cell threshold, test enable, and cell mask (See “Peripheral control electronics” on page 60.).

The MRs and the GRs are selected by a common five-bit Enable Register (ENBL), in combination with the IR decoding. For the MRs, only one column of 256 cells can be enabled at any time. For the GRs, one register only is addressed at any time.

Besides the standard TAP state machine (JTAG-SM), there is a second slave state machine: (JTAG-CHK). The JTAG-CHK has two TDI inputs: TDI0 and TDI1. For chip number i (see Fig. 2-48), TDI0 is connected to the TDO of chip number i-1, while TDI1 is connected to the TDO of chip number i-2, as shown in Fig. 2-49. If chip i-1 is faulty (i.e. the JTAG daisy chain is interrupted), this condition is detected by the JTAG state machine, which skips it and switches to chip i-2.

This non-standard hardware extension of the JTAG specification is transparent and the software is standard.

FIGURE 2-48. JTAG architecture within the Pixel chip.

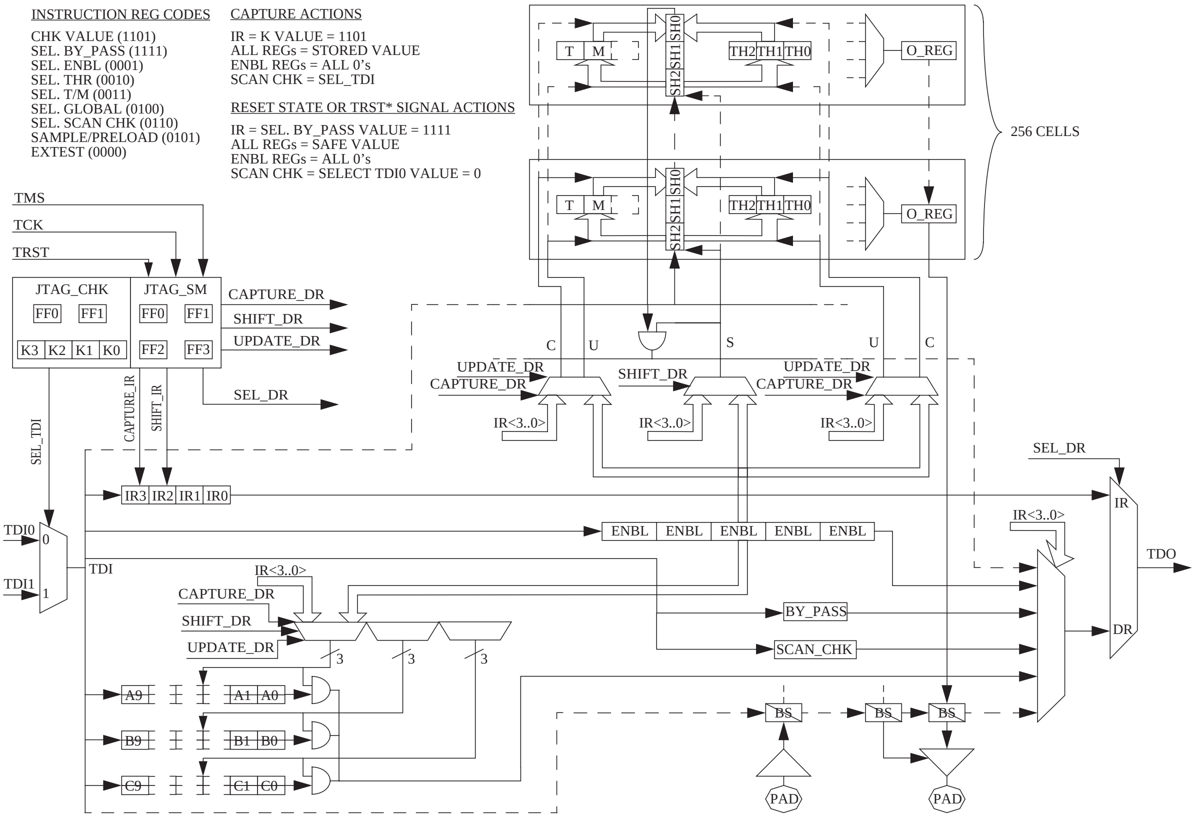
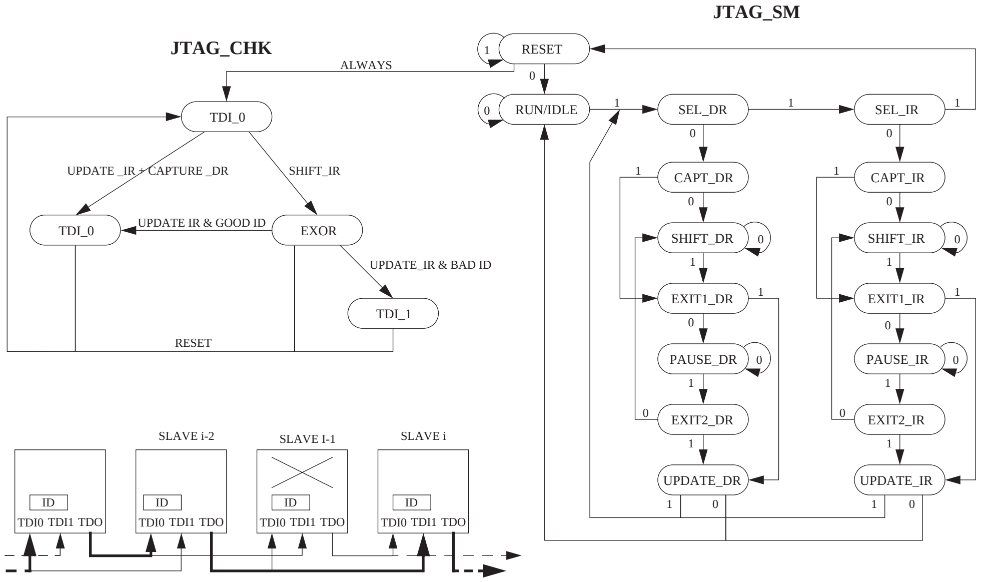


FIGURE 2-49. Daisy chaining pixel chips and JTAG state machines.



RULE: WOULD EXECUTE TDI / TDO CHECK ONLY ONCE AFTER EXITING JTAG RESET STAUTS.

The basic principle is the following: at TRST* active or in the TEST-LOGIC-RESET state, the TDI0 input is selected. Every time the JTAG-SM enters the SHIFT-IR state coming from the TEST-LOGIC-RESET state, a cycle of IR shift is executed. At CAPTURE-IR, each device loads the shifted pattern (CHK value) into its IR. At UPDATE-IR, each device checks the received CHK value against the one it contains internally. If the check is successful, TDI0 is selected; if the check fails, TDI1 is selected. During the subsequent JTAG cycles, no further checks are executed until the JTAG-SM comes back to the initial TEST-LOGIC-RESET state. Fig. 2-50 gives the times required to download some typical set-up registers.

FIGURE 2-50. **Typical shift delays.**

IR (16 chips) = 3 x 16 = 48 CKs	=> @ 10MHz => 4.8µs
ENBL for T/M&THR&GBL on 1 chip ON and 15 OFF = 5 + 15 = 20 CKs	=> @ 10MHz => 2.0µs
1 THR column on 1 chip ON and 15 OFF = 3 x 256 + 15 = 783 CKs	=> @ 10MHz => 78.3µs
1 T/M column on 1 chip ON and 15 OFF = 2 x 256 + 15 = 527 CKs	=> @ 10MHz => 52.7µs
1 GBL reg of 10b on 1 chip ON and 15 OFF = 10 + 15 = 25 CKs	=> @ 10MHz => 2.5µs
Write 1 column of THR = 48 + 20 + 48 + 783 = 899 CKs	=> @ 10MHz => 89.9µs
Write 32 columns of THR = 899 x 32 = 28768 CKs	=> @ 10MHz => 2.88ms
Write all 32 THR columns of 16 chips = [(48 + 5 x 16) + (48 + 3 x 256 x 16)] x 32 = 398848 CKs	=> @ 10MHz => 39.9ms
Write 1 column of T/M = 48+ 20 +48 + 527 = 643 CKs	=> @ 10MHz => 64.3µs
Write 32 columns of T/M = 643 x 32 = 20576 CKs	=> @ 10MHz => 20.6µs
Write all 32 T/M columns of 16 chips = [(48 + 5 x 16) + (48 + 2 x 256 x 16)] x 32 = 267776 CKs	=> @ 10MHz => 26.8ms

2.3.3 Readout logic

Multi-event buffering

The time available for the readout of the Silicon Pixel Detector (SPD) is limited by the requirement that the dead time introduced must stay below 10% in the worst case of Ca-Ca running at high luminosity (LVL1 rate of 2.5 kHz).

In the absence of an event buffer on the front-end, and in order to comply with the dead time specifications of the SPD, each parallel readout channel would have to complete the readout within about 200 µs. At a clock frequency of 10 MHz, this would limit the maximum number of front-end chips which can be multiplexed on a single readout channel to eight. In such a scenario, the readout channels would not be used optimally, as they would have to move data rapidly for short periods of time. At the expected LVL1 trigger rate (about 1 kHz for Pb-Pb collisions), the readout channels would be idle most of the time.

With the implementation in the front-end chips of a multi-event buffer acting as a data derandomizer, the read-out system can be dimensioned for the average rate, rather than the peak rate. In order to avoid that the front-end buffers are filled faster than they can be emptied the readout cycle should never exceed the average LVL1 period.

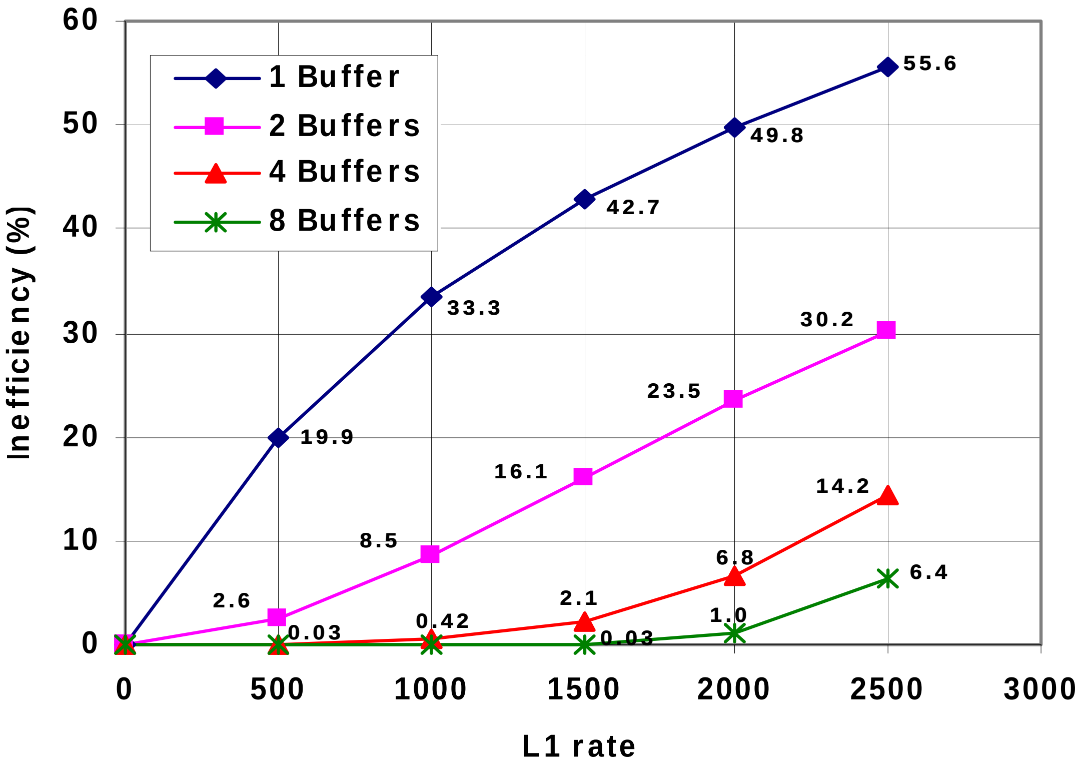
To stay below a dead time of 10%, the readout now has to be completed in 400 µs. At a clock frequency of 10 MHz, this allows to multiplex data from 16 front-end chips instead of eight. The number of buffer cells depends on the statistics of the process and on the maximum acceptable buffer inefficiency (i.e. the percentage of events lost due to the buffer being full). In order to fix the buffer depth for the ALICE front-end cells, a simulation was performed. The results are displayed in Fig. 2-51, which shows the inefficiency of the multi-event buffer versus the strobe rate, for several buffer depths (1, 2, 4 and 8 events). The strobes were generated using Poissonian statistics.

The calculation is made under the following worst-case assumptions:

- LVL2Y (r/o) latency = 100 µs
- LVL2N (abort) = None (all strobed events are good)
- r/o time = 400 µs

As can be seen from the figure, for Ca-Ca at high luminosity (LVL1 rate of 2.5 kHz), a buffer depth of four results in an inefficiency of 14% in the extreme case that all LVL1 become LVL2-yes. So we decided for this solution.

FIGURE 2-51. **Inefficiency (=dead time) of the multi-event buffer versus the strobe rate.**

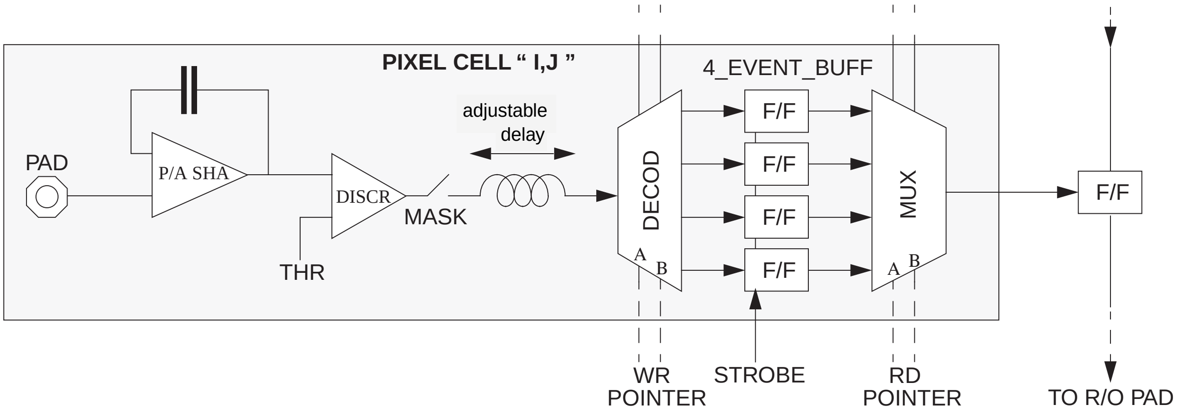


The front-end chip has 8192 cells organized in 256 rows of 32 cells. The architecture of each cell (see Fig. 2-52) comprises an analog part (preamplifier/shaper, discriminator) and a digital part: a LVL1 trigger latency delay line, a 4-hit deep multi-event buffer FIFO organized as a circular memory, and an output shift register (256 bit long, composed by the right-most FF of every cell in a column).

The analog part of the cell is always active. When the pixel chips receive (in parallel inside a half-stave) a strobe signal, a discriminated binary signal (*hit*, produced by the discriminator), suitably delayed to compensate for the LVL1 trigger latency (up to 6μs) is written into the first available location of the event buffer memory. At the falling edge of each strobe, the write pointer is updated to point to the next location. This operation is called *acquisition*.

A *read-out operation* is responsible for sending out the oldest data of the event-buffer memory, to the outside of the chip, through the output shift register. It happens when the pixel chips receive an event-readout signal from the Pilot chip (see following section).

FIGURE 2-52. **Multi-event buffering in the Pixel cells.**



Pilot chip

The 120 half-staves of the SPD barrel are read out in parallel. Each half-stave (two ladders, 16 front-end chips) is served by a pilot chip, that is responsible for the read-out and control, according to the trigger signals.

The pilot system reads parallel data (32-bit wide), encode them in a few 25-bit words, which are serialized and sent through the short data link (see below) to the router unit, located in a VME crate about 40 m away from the barrel. The full Silicon Pixel Detector will require 120 short links, 60 per side, each one receiving multiplexed data from one half-stave (two ladders).

The readout electronics up to the short serial link has been prototyped on a 6U VME card, using FPGAs and standard commercial ICs. This will allow enough flexibility to adapt the readout logic for testing the full Alice1 front-end chip currently being designed. A detailed description of the Pilot prototype is given in Chapter 3: “The Pilot System” .

In the final version, this Pilot electronics will be integrated in an ASIC using the same radiation-tolerant technology (enclosed-gate 0.25 μm CMOS) used for the front-end chip, with a full custom design. The pilot chip is expected to measure about 20 mm^2, and to contain of the order of 100 I/O pads. Most of the area is occupied by the radiation tolerant 2k x 47 bit FIFO (static RAM). The device will be packaged for tests and wired-bonded for use on the barrel.

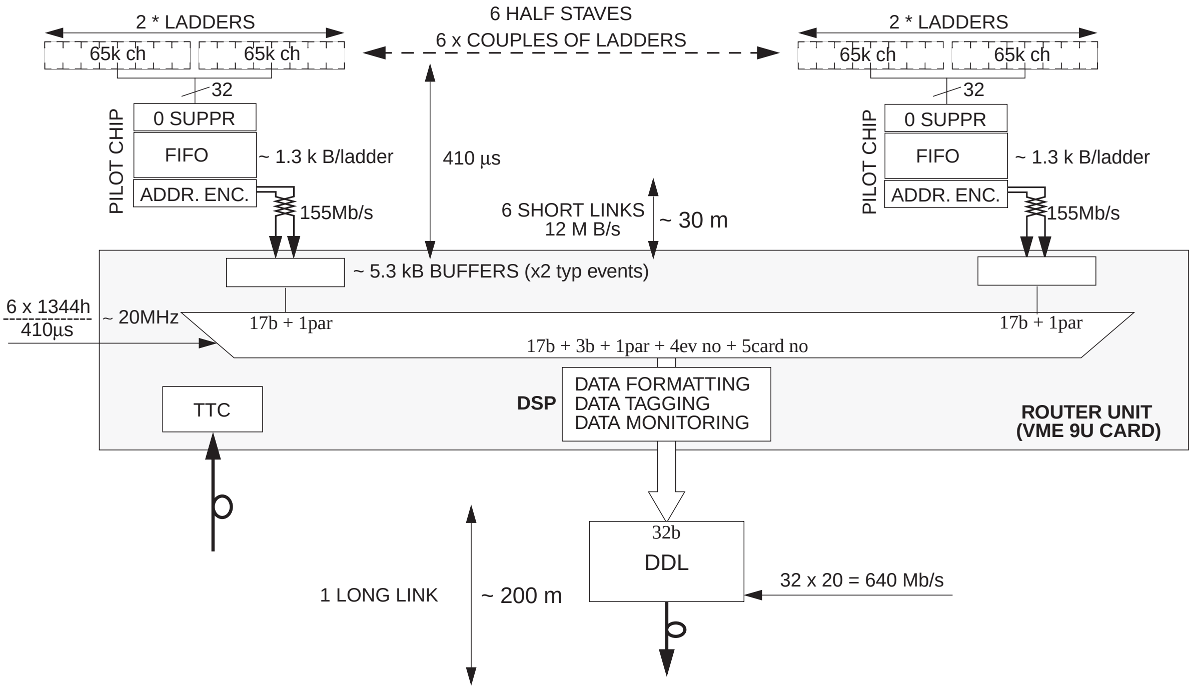
Router

The data from six short links are received on a router unit (a 9U VME card), where they are merged, formatted, tagged and sent to the DAQ via the ALICE standard DDL optical link. The router also interfaces the detector electronics to the trigger system (through the TTC) and provides local monitoring.

A schematic drawing of the architecture of the router is shown in Fig. 2-53.

The data received by the router are written into six input buffers large enough (10 kbyte) to ensure that at least two typical-size events can be stored.

FIGURE 2-53. **Local Concentrator block diagram.**



As discussed above, the data coming from the pilot chips are enclosed between a header and a trailer word. Once all the trailer words of the event have arrived, the router starts reading out the buffers, multiplexing their data and formatting them for transmission to the DDL. To sustain the 400 µs readout cycle at the maximum expected occupancy, the multiplexer must run at 20 MHz. To avoid data congestion, we are planning to use a safety factor of two, employing a 40 MHz clock.

The router also provides a monitoring facility consisting of a VME-accessible local memory which stores a copy of the current event.

The core of the router is a fast DSP, complemented by external hardware, tailored to maximize the efficiency of the algorithm. The use of a DSP will ensure the flexibility to run different algorithms.

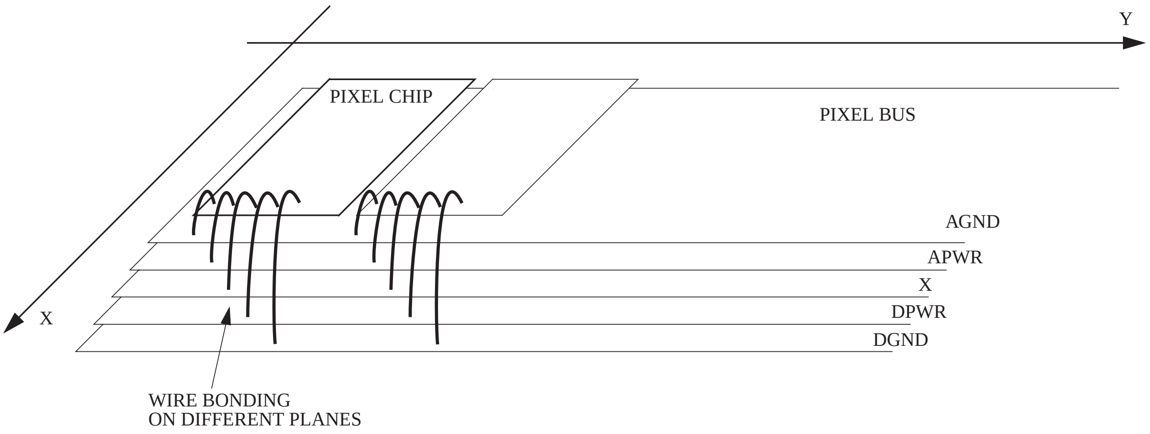
A solution based on the Texas Instrument TMS320C6202 DSP is being investigated. This device has a 32-bit data bus which reads the data from the input buffers, enough internal memory to store the reformatted data, and an internal DMA controller to pass the data on to the DDL via an independent 32-bit expansion bus.

A total of 20 routers will be needed to equip the full SPD barrel.

Stave bus

The four ladders from each stave will be glued and wirebonded onto two thin multilayer carrier busies (stave busies, see Fig. 2-54). Each stave bus connects the electrical signal and power lines of two ladders (16 front-end chips) to those of the Pilot chip mounted at the end of the stave bus itself. The stave bus terminates in a flexible pigtail for external connection to the long cables going out of the detector. For the full SPD, there are 120 pigtails in total, i.e. two per stave (one for the right side and one for the left side).

FIGURE 2-54. **Schematic representation of the stave bus.**



We have chosen to employ a polyimide circuit in order to minimize the material in the active detector area, and to allow for the possibility of directly prolonging the carrier into a flexible pigtail. Aluminium will be used as the conductor material.

The technology allows us to have a pitch of 100 m, therefore on one layer there can be enough room for up to 100-120 signals. Two layers (x, y) are required to provide the connection from the front-end chip bonding pads, all located on one long side of the stave bus, to the bonding pads of the Pilot chip, located at one end of the stave bus. In addition, four full plane layers are needed to provide the analog and digital power and ground connections, giving a total of six layers.

The assembly of the six layers is carried out respecting a few simple rules:

1. signal lines configured as a microstrip structure (because of controlled impedance and shielding);

2. power and respective ground planes very close to one another to provide maximum capacitive decoupling (by-pass capacitors would not fit on the stave bus);

3. only ground planes on the external sides of the sandwich.

We foresee the following stacking: ground analog plane, power analog plane, x signal layer, y signal layer, power digital plane, ground digital plane. The two internal layers, for x and y signal lines, are made in traditional polyimide technology with electrical vias. The four layers for powers and grounds are made starting with a polyimide foil initially covered with thick aluminium which is then thinned down to a thickness of 25 µm. This is the minimum thickness which provides an acceptable electrical resistance (about 15 mΩ). These four planes are glued onto the two x and y layers, with no vias in between.

A schematic drawing of the stave bus is shown in Fig. 2-54. The width of the layers will decrease from the bottom to the top of the circuit, so that all the power and ground layers will be accessible for wire bonding.

A total of 70 signal lines are printed on the x and y layers: data(31...0), ce(15...0), clk(1,0), fast-or, fast-mult, load-shreg, strobe, abort, data-rst, cntrl-rst, test-in, test-out-a, test-out-d, tdi0, tdi1, tdo, tms, tclk, trst, dac-ref-in, dac-ref-out, gtl-ref-in, gtl-ref-out.

Clk(1,0) is the 10 MHz clock by the Pilot chip sent to readout the pixel chips, transmitted on a differential line. This relatively slow clock allows to have low rise and fall times, thus reducing the spectrum of frequencies. In this way no termination is required on the bus, and this is very important in order to reduce the power dissipation.

Short data link

The lack of space between the beam pipe and the SDD layers and the need to minimize the total material put severe constraints on the overall mechanical structure of the SPD, which needs to be as compact as possible. Within such constraints, there is no possibility for cross-connections between the different stave busies at the barrel level. Therefore, independent power, data and control connections are needed for each half-stave. In particular, the multiplexed data from each half-stave have to be sent out to the DAQ without further multiplexing on the barrel.

Owing to real estate, radiation environment and budget constraints, it is not possible to do this by mounting a standard ALICE Detector Data Link (DDL) per half-stave at the barrel level.

As mentioned above, the multiplexed data from each half-stave will be serialized on the pilot chip and transmitted on a serial simplex copper link (Short Data Link, SDL).

The Short Data Link physical layer is a shielded twisted-pair type AWG 24, with a length of about 40 m with a data rate of 155.52 Mbit/s. A full description of the link is in Chapter 3: “The Pilot System” .

Upon reception of the readout command issued by the DAQ, all the Short Data Links start sending data. In case no detector data are present, just the header and trailer are sent. In this way faulty channels can be detected.

2.4 Assembly, mechanics and cooling

2.4.1 Introduction

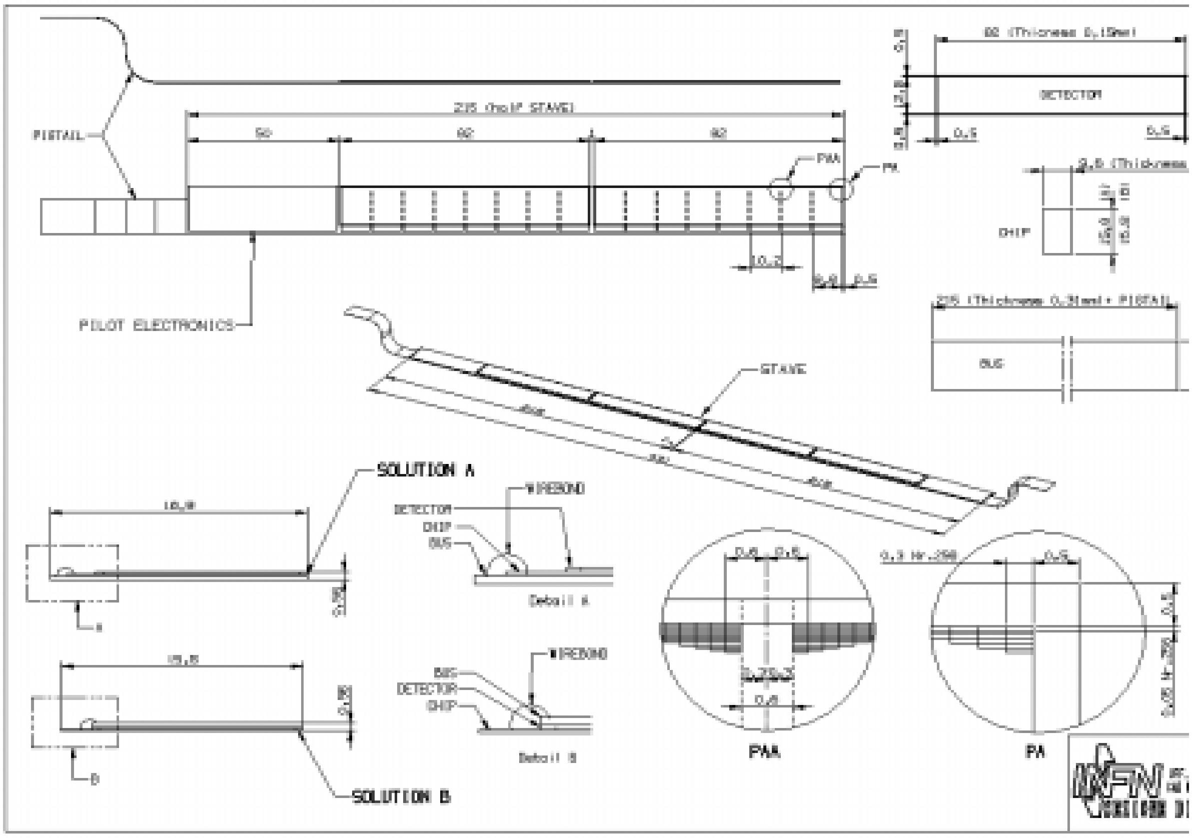
Four silicon pixel detector ladders are glued and electrically connected to a carrier bus to form a stave (Fig. 2-55).

Two layers of pixel staves are arranged in space, at a radial distance from the beam axis of about 40 mm for the first, and 70 mm for the second. We assume that a positioning precision of around 100 µm should be adequate in order to allow the final alignment with tracks.

The staves will be glued onto 10 carbon-fibre support sectors (CFSS). Each sector supports two staves from the first layer and four staves from the second layer. The choice of closed-structure support sectors is dictated by the need to maximize the stiffness within a limited material budget. The modularity is the result of a compromise between the conflicting requirements of stiffness (which is larger for larger closed geometry sector sizes) and maintenance/reworkability (which of course would call for a low number of ladders mounted on each independent support structure). In the solution adopted, we have six staves (24 ladders) mounted on each support. This number is still rather large, and particular care will be required in the assembly procedure to ensure the possibility of removing and replacing each half-stave (two ladders) should the need for maintenance arise. The choice of the turbo layout for the outer layer is dictated by the need to maximize the support base for the very delicate ladder-stave assemblies. The superposition of the staves is such that no particle can go undetected through the openings above a momentum cutoff of 27 MeV/c.

Additional constraints are imposed by the very delicate procedure of installation of the Silicon Pixel Detector (SPD) barrel inside the ALICE detector, where we will have a very limited space available for manipulations in close proximity to the thin Be beam pipe and to the outer detectors. The presence of the muon arm on one side also complicates the access and service to the SPD from that side.

FIGURE 2-55. **ALICE SPD stave. Two stacking scenarios are being considered (See "Stave assembly" on page 91.).**



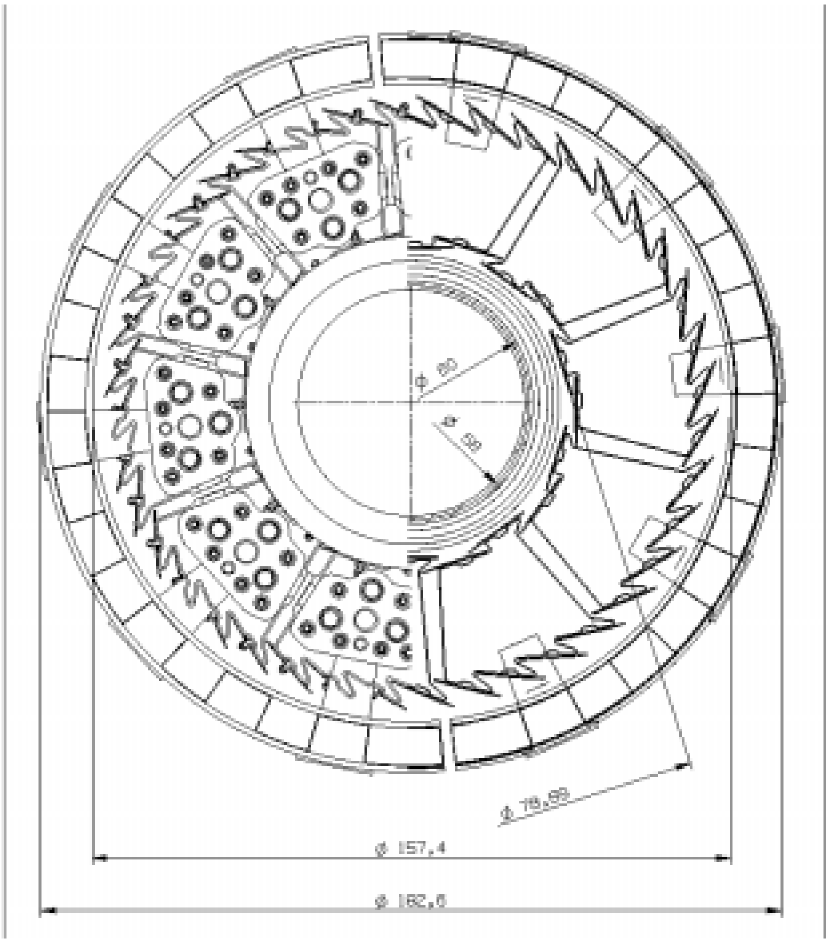
An External Shield (ES) made of Al-coated carbon fibre acts as a thermal screen towards the temperature-sensitive SDD planes and provides support and protection for the SPD barrel during the delicate installation operation. A general view of the barrel is displayed in Fig. 2-56.

A total of between 1.5 kW and 2 kW of power will have to be removed from the two-layer pixel barrel. The plans are to do this with cooling vessels embedded in the carbon-fibre structure. We are considering deionized water as the cooling medium. C_6F_{14} is being investigated as an alternative.

2.4.2 Stave assembly

Two options are evaluated for the geometrical stacking of the ladder modules and the stave bus. This is shown in Fig. 2-55. Solution A is our baseline option: the backplanes of the chips are glued on the stave bus, which has to protrude out of the chip by 1 mm to allow the space for wirebonding connections. The total width of the stave will therefore be 16.8 mm. In this scenario, the thermal connection between the front-end chips and the cooling channels which run on the surface of the CFSS goes through the stave bus, which is in thermal contact with the cooling channel through a layer of thermal grease.

FIGURE 2-56. **General view of the ALICE SPD barrel (distances in mm).**



An alternative solution (B) where the stave bus is glued on top of the detector ladder is under investigation. This would allow us to place the backs of the chips in close thermal contact with the cooling channel through a layer of thermal grease, and to reduce the total stave width to that of the front-end chip (15.8 mm), thus saving 1 mm in the total footprint. This would be a bonus given the tight space constraints for the stave geometry. This solution imposes a more stringent constraint on the maximum allowed width of the bus, which is now determined by the width of the ladder (12.8 mm), and implies the need to wirebond to a stave bus which is mechanically supported by the back of the ladder module. Since the backs of the chips have to be connected to ground, this solution also requires a good grounding of the CFSS and cooling channel system, and good electrical properties from the thermal grease.

We have investigated the effects of wire bonding to the back of the detector ladders, as would be required by solution B, on a few Omega3 ladder modules, in order to test whether this action could be damaging to the bump-bonding connections between the detector and the front-end chips. We tried this on high quality assemblies, with essentially 100% working contacts, and on low quality assemblies, where some detector areas were not responding properly, indicating problems in the bump-bonding connections. While the wirebonding caused no visible damage on the high quality assemblies, on the low quality assemblies we observed an enlargement of the inefficient areas as a result of the wirebonding tests. These tests were performed on rather thick assemblies (300 μm detector + 300 μm chip), and will have to be repeated on thinner assemblies.

The staves will be assembled by gluing the chip and the detector ladders on the carrier, using an optical positioning system. The chips will then be wirebonded to the carrier.

2.4.3 Carbon-Fibre Support Sector

As recalled above, the choice of employing a limited number of independent supports with closed geometry was dictated by the strict material budget requirements imposed by the ALICE physics programme.

The choice of the materials for the support sector was determined by the following requirements:

- high local and global stiffness while minimizing the material in the sensitive area: high modulus, low Z material;
- good workability of the support material, to allow the arrangement in a complex shape;
- short- and long-term stability in terms of absolute deformations and induced distortions;
- low coefficient of thermal expansion (CTE);
- limited CTE mismatch of the adopted materials, and, where not possible, introduction of soft materials at the coupling surfaces for reduced coupling constraints;
- good thermal conductivity;
- good radiation tolerance;
- reduced creep effects.

For the support sector material, beryllium, carbon-carbon and high-modulus carbon fibre (HMCF with epoxy or ester-cyanate polymers) were considered. We decided to use HMCF. The ALICE requirements are not so demanding in terms of the radiation tolerance and service temperature as to require the use of carbon-carbon (like, for example, in ATLAS). However, for the ALICE SPD, the chosen material has to have good workability properties, on account of the complex support shape to be realized within a strict material budget.

The Carbon Fibre Support Sectors (CFSS) are made by winding two layers of unidirectional, high-modulus, 100 μm thick carbon-fibre tapes, with fibres respectively parallel and perpendicular to the beam axis, around a metallic mandrel. A mechanical drawing of the CFSS is shown in Fig. 2-57.

A pictorial view of the CFSS is shown in Fig. 2-58. Each CFSS provides two support planes for inner layer staves and four support planes for outer layer staves.

Each support plane has a groove for lodging a cooling duct. The cooling channels are stainless steel tubes with an external diameter of 2 mm and a wall thickness of 35 μm. They are squeezed to a thickness of 0.6 mm in the stave region, in order to maximize the heat exchange surface. Structural analysis and prototyping of the CFSS are reported in [4, pages 71-75]

FIGURE 2-57. **Carbon-fibre support sector (CFSS): mechanical drawing and exploded view.**

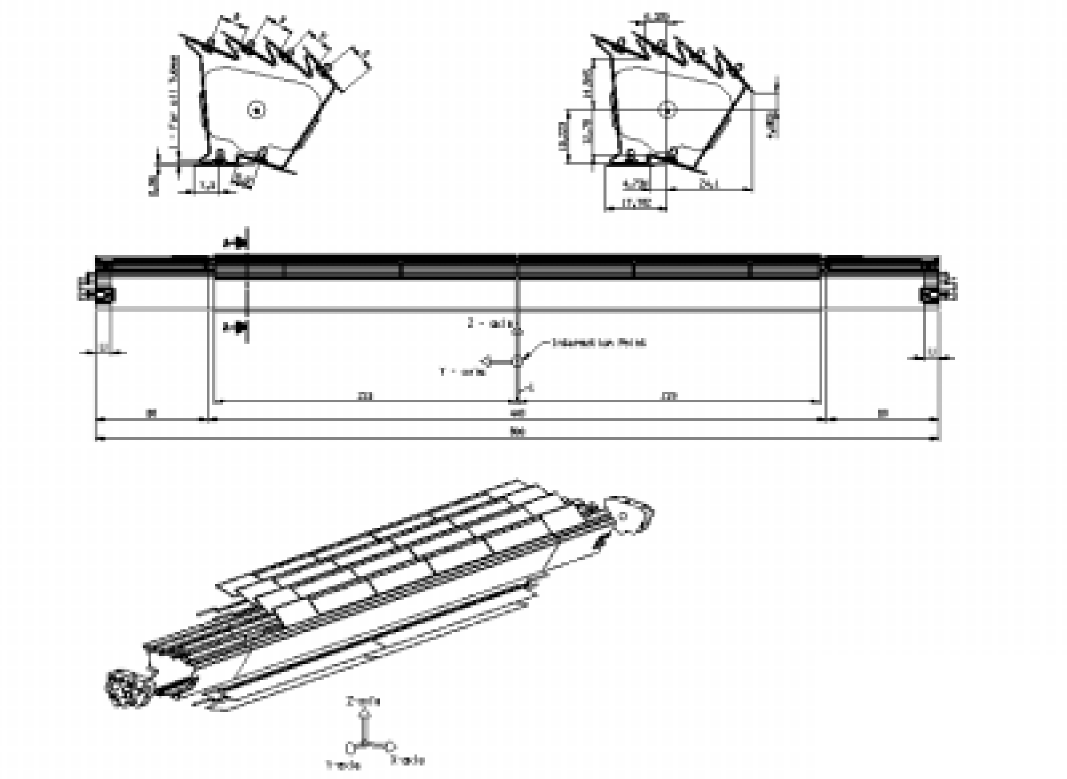
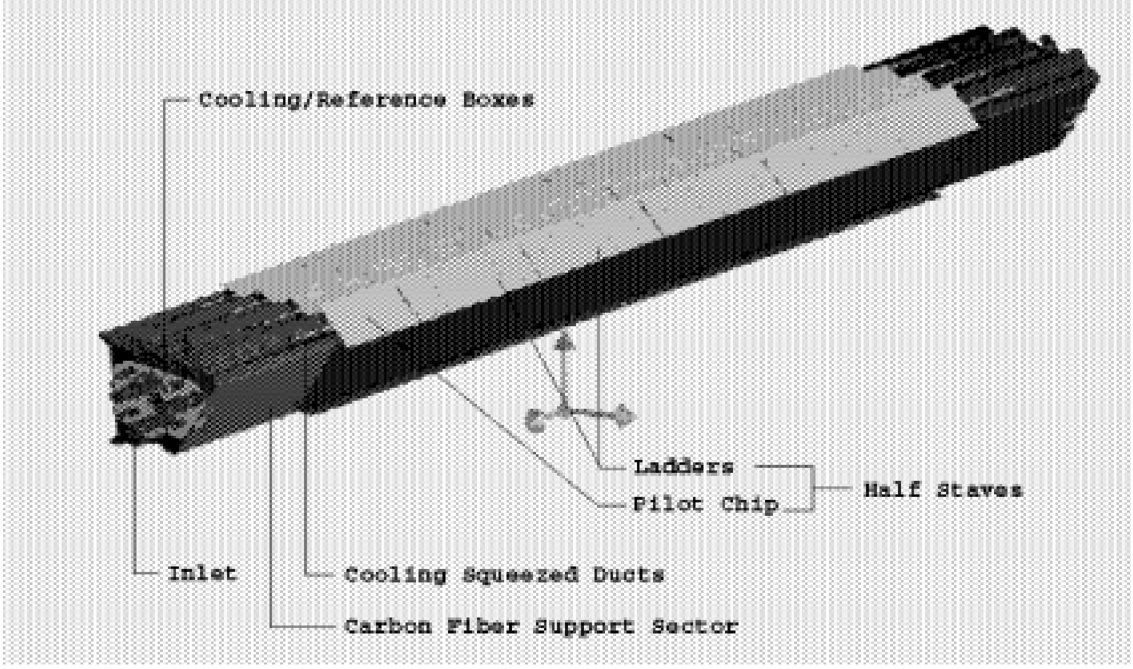


FIGURE 2-58. **Pictorial view of a Carbon Fibre Support Sector.**



2.4.4 Sector assembly

The dimensions of the CFSS will be checked to verify that the design tolerances are respected. We plan to select and equip with detectors a minimum of twelve CFSS (ten to be mounted on the experiment plus two spares). Twelve assembled and wire-bonded half-staves will be mounted on each CFSS using a three-dimensional measuring machine equipped with an assembly station which is still in the design phase. The half-staves will be held in position by a suction system, while the CFSS will be displaced to the different half-stave assembly positions. A thin film of thermal grease will be placed on the CFSS surface, in order to guarantee a full thermal contact between the half-staves and the carbon-fibre cooling duct assembly. Several points of UV glue will be placed on the CFSS in small islands left clean of thermal grease for this purpose. While we think that such a scheme should enable us to perform a careful replacement of one half-stave in the laboratory, should the need for maintenance arise, we have not yet had practical experience on this point.

2.4.5 Cooling system

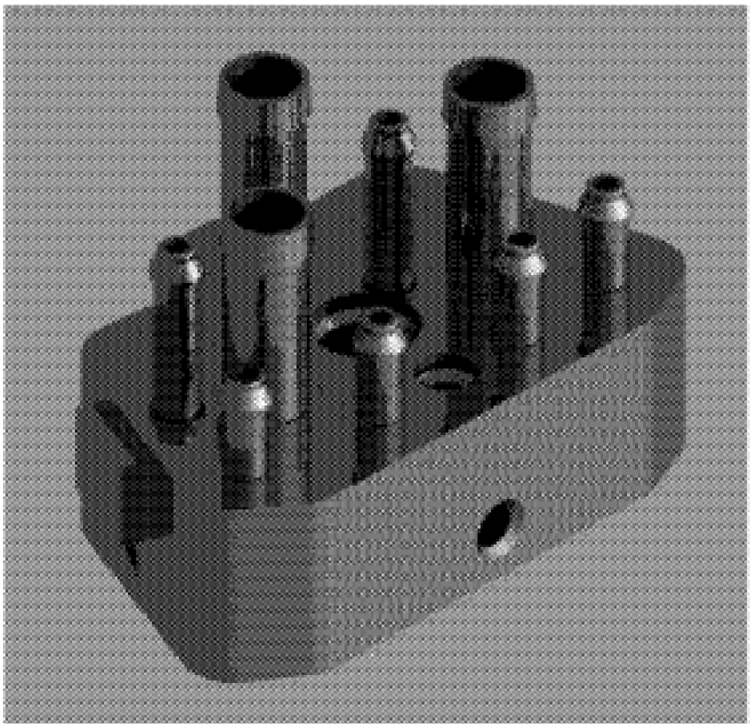
The front-end chips are expected to generate a heat load of 25-30 W per stave. The additional heat load coming from the pilot chips is expected to be negligible. The cooling system has to remove this heat from the SPD barrel. Silicon pixel detectors are not very sensitive to temperature. We will operate the SPD at around room temperature. We aim at keeping the temperature spread on the barrel within about 10°C.

Each stave will be put in thermal contact with a cooling duct mounted in a groove on the CFSS. So far we have concentrated our R&D on a ‘leakless’ [56] deionized water cooling system (below atmospheric pressure). As an alternative, we are considering C_6F_{14}.

Unavoidably, some fraction of the power will tend to propagate away from the cooling ducts. An external shield will ensure that no heat is irradiated towards the external ITS layers. A moderate flow of dry air through the detector volume and through ducts in the external shield should be effective in removing the residual power.

The cooling ducts will be obtained from stainless steel tubes with a wall thickness of 35 μm and an initial external diameter of 2 mm, squeezed down to an external thickness of 600 μm. The choice of steel as opposed to aluminium should prevent dilution problems and guarantee an adequate strength of the vessel in the shaping phase and during operation below atmospheric pressure. The duct is bent near the two ends to pass through two holes in the CFSS, and connected via silicon tubes to the cooling collectors (see below). The duct is fixed at one end and free to elongate inside the groove at the other end. Thermal strains between the carbon-fibre cooling duct assembly and the detector modules are avoided by the interposition of a soft thermal grease.

FIGURE 2-59. **Pictorial view of the cooling collector.**



Each sector is equipped with two cooling collectors at the extremities, one functioning as an inlet and the other as an outlet. The inlet and outlet collectors are identical. We are considering both steel and aluminium as material for the cooling collectors.

A pictorial drawing of the collector is shown in Fig. 2-59. Each collector has three independent inlet (outlet) circuits, each servicing two staves in parallel. In total, the SPD barrel is serviced by 30 inlets on one side and 30 outlets on the other. The external connections are made using flexible tubes with an inner diameter of 10 mm, reduced to 3 mm at the inlet (outlet), at the connection with the collector nipples.

The temperature is monitored on the SPD barrel at both extremities of each of the 60 cooling ducts. The pressure is monitored for each of the 30 inlet and 30 outlet lines. Each line is serviced by electrovalves which can be used to stop the flow in case of problems.

2.5 Power distribution

The SPD needs three supplies: the analog power (1.6-2.0 V), the digital power (1.6-2.0 V) and the detector bias (~ 50 V). While the latter does not require large currents (reversed biased diode current), the analog and digital currents are estimated to be 5 A each for two ladders together (one half-stave, 16 front-end chips).

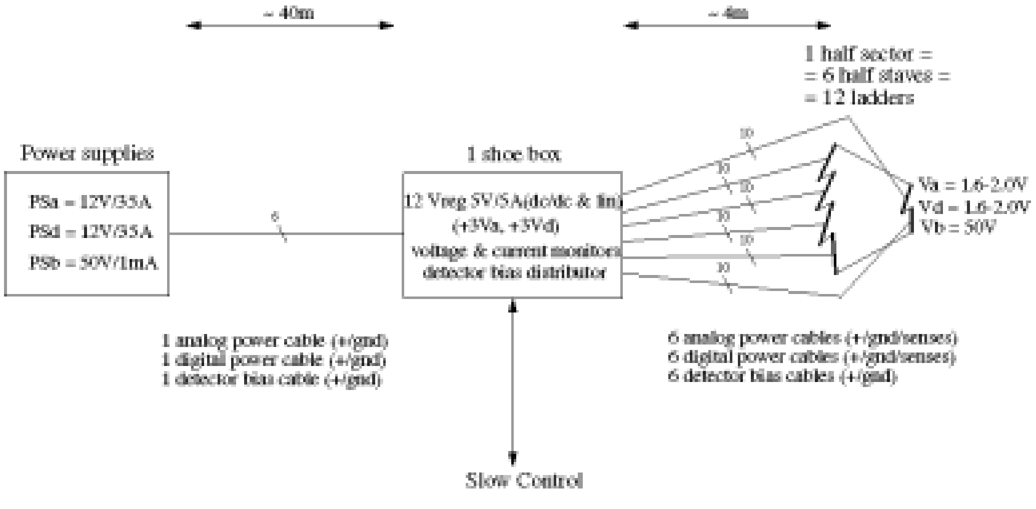
The voltage regulation of the analog and digital supplies is very important. In order to reduce the power dissipation, small operating voltages will be applied to the front-end chips. This implies that the voltage regulation should preferably be done as close as possible to the detector. Given the impossibility to do this on the barrel itself, the nearest position is about 4 m away, on the endcaps of the TPC detector, where we will install the “shoeboxes” that will receive the main power cables and locally distribute all the necessary supplies, individually to each half-stave. A scheme of the SPD power distribution is shown in Fig. 2-60.

On the detector, each half-stave receives the following power lines: analog power, analog ground, two sense wires for the analog supply, digital power, digital ground, two sense wires for the digital supply; detector bias and detector ground. In total 10 lines per half-stave. Sense wires are necessary to guarantee the correct voltage setting on the barrel, about 4 m away from the distribution.

Each shoebox supplies six half-staves. The supplies of the same type within a shoebox have a common main power line coming from the respective power supplies. These are located about 40 m away, outside the L3 magnet. In total, six lines are coming from the power supplies: two power/ground pairs with low voltages and high currents (12 V and about 35 A) and a power/ground pair with high voltage and low current (50 V and about 1 mA).

Since the regulation is done on the shoeboxes, the voltage and current monitoring are also performed at the level of the shoeboxes. Voltage setting and V/I monitoring are taken care of automatically by the slow control system.

FIGURE 2-60. **Block diagram of the SPD power distribution. In each side we will have 10 half sectors, 10 shoe boxes and 10 power supply groups.**


~ 40m
~ 4m
1 half sector =
= 6 half staves =
= 12 ladders
Power supplies
PSa = 12V/35A
PSd = 12V/35A
PSb = 50V/1mA
6
1 shoe box
12 Vreg 5V/5A(dc/dc & lin)
(+3Va, +3Vd)
voltage & current monitors
detector bias distributor
10
10
10
10
10
10
Va = 1.6-2.0V
Vd = 1.6-2.0V
Vb = 50V
1 analog power cable (+/gnd)
1 digital power cable (+/gnd)
1 detector bias cable (+/gnd)
6 analog power cables (+/gnd/senses)
6 digital power cables (+/gnd/senses)
6 detector bias cables (+/gnd)
Slow Control

CHAPTER 3

The Pilot System

This chapter describes the development of the prototype of the Pilot logic. The framework in which the Pilot system operates has been described in the section "Readout logic" on page 84. Data Acquisition and read-out are independent activities that are executed in parallel. They are both controlled by the pilot logic, while the front-end chips run as slave devices.

This pilot electronics up to the short serial link has been prototyped on a 6U VME card, using programmable devices and standard commercial ICs. This allows enough flexibility to adapt the logic to the Alice1 front-end chip currently being designed (and not yet fully defined with its interface) and to the other neighbouring sub-systems. As the final version of the system will be integrated on a chip, we will use sometime the expression "Pilot chip", as a general reference to the system under development.

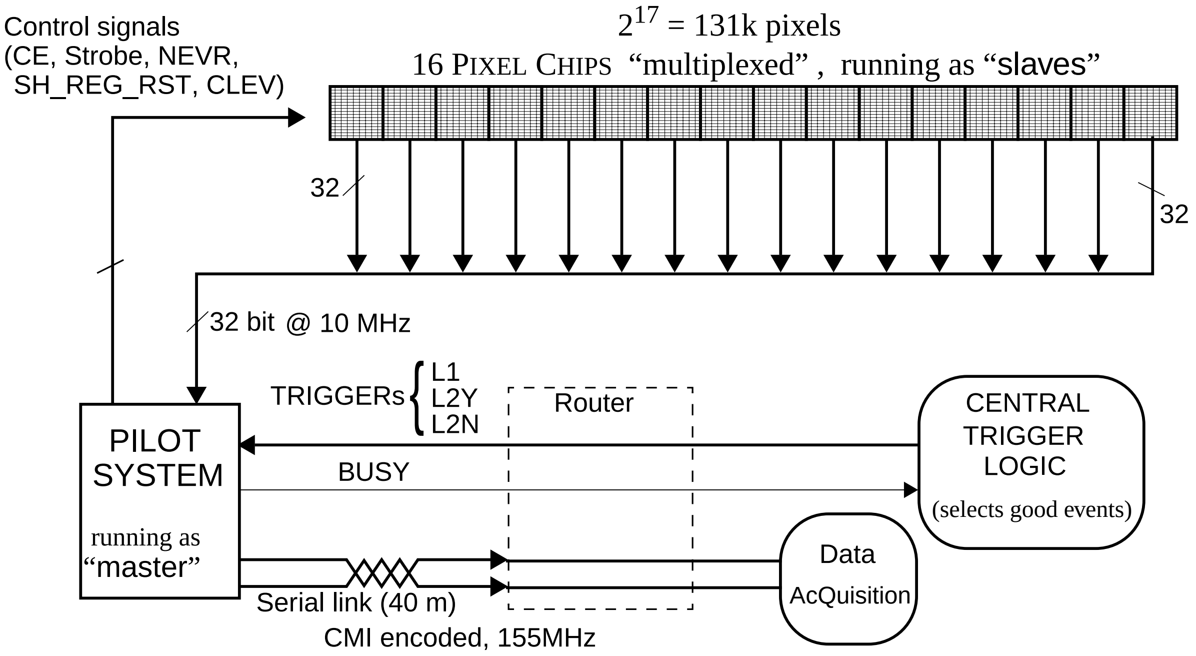
A single Pilot chip is able to manage the data coming from 2^{17}=131072 pixel cells; in such a way that only 240 Pilot chips will manage the 15.73 million pixels of the Alice experiment. As most of the pixels (~99%) will send out a "0" binary signal, the spare hits (i.e. bits set to "1") are fully encoded in the Pilot chip. This operation is carried-out on-line by the Pilot hardware, without any software intervention.

3.1 Problem description

The Pilot system is responsible for the read-out of 16 pixel chips (= 2 ladders = half-stave), for encoding and for serializing the pixel data to be sent out of the detector. The serialization is imposed by the limited material budget inside the detector (including cables), thus the encoding is due to the huge amount of pixel data. Moreover, the run-time control[1] of the half-stave resides in the pilot system, and not in the pixel chips. This is to avoid duplications and because the pixel chip is a mixed-mode circuit, where every digital activity should be minimized.

While the general performances of the detector are imposed by the application, the detailed specifications and above all the sharing of the duties between the different sub-systems (Fig. 3-1) were an important part of the study.

FIGURE 3-1. Overview of the sub-systems linked to the Pilot chip



When the pilot logic receives a LEVEL1 trigger signal (See "Trigger logic in ALICE" on page 20.), it sends a strobe signal in parallel to all the 16 front-end chips of the half-stave, if the system is not full (i.e. busy).

After the release of a LEVEL2 ACCEPT signal (L2Y), the Pilot chip begins a read-out cycle addressing sequentially the 16 front-end chips. For each chip, for all the pixel cells, the data from the oldest unprocessed strobe present in the multi-event-buffer-FIFO are loaded into the output shift registers and the read pointers are updated to point to the next location. Subsequently, a sequence of 256 clock cycles at 10 MHz outputs the pixel data onto the read-out bus,

[1] Run-time control includes all the control functionalities running simultaneously with the data acquisition/readout. It is also called fast-control, in contrast with the slow control, active when the acquisition is stopped (see section 2.3.2).

at the rate of one 32-bit word per clock cycle. Each word corresponds to one row of the pixel matrix (= 256 rows x 32 columns), the hits being represented by a logical 1 (a hit is a bit set to "1" in the output data of the pixel chips. It is strictly related to the transit of a particle in the detector, see Fig. 1-4). In this way the read-out time of a pixel chip is 256/10MHz ≅ 25 µs. The Pilot system sequentially reads 16 pixel chips in a total read-out time of ~16 x 25µs = 400 µs. The number of 16 is a trade-off between the total read-out time (to be kept smallest than possible and proportional to the number of Pixel chips "multiplexed" by a Pilot chip) and the physical and mechanical constraints limiting the material (including the number of Pilot chips and their serial outputs) in the detector. In case of a LEVEL2 NO occurrence, the event is simply discharged.

This way of organizing the read-out resulted from the evaluation of other approaches. In a first design, the multi-event-buffer (see Figure 2-52 on page 86) in the pixel cell was not conceived, and the Pilot chip should have started the read-out at the LEVEL1 occurrence. In case of a subsequent LEVEL2 NO an abort sequence would have interrupted the current read-out, with some logic complications and additional constraints. The introduction of the multi-event-buffer resulted from the investigations aimed to reduce the dead-time of the detector and to improve the interface between the Pilot and the Pixel chips.

The maximum expected occupancy of the pixel matrix is of the order of a few per cent (for the first Silicon Pixel Detector layer, for central Pb-Pb collisions, at the maximum expected multiplicity, [4]). This means that the output data of the pixel chips are zeros at ~ 99%. Therefore, we decided to zero-suppress the information and fully encode the hit address on the Pilot chip. In principle this encoding function could have been implemented in the Pixel chip itself; but this approach would make the detector less flexible in case of unforeseen changes in the operation conditions.

Let us examine the full hit encoding. The row size is $32=2^5$ bits, so 5 bits are needed to indicate the position of a hit in a row. Similarly 8 and 4 bits encode the row and the pixel chip, for a total of 17 bit. The encoding process is split into two hardware steps (Fig. 3-2). In the first one the zero-words are eliminated from the data flow, and the chip/row address is attached to every non zero word. At this point the rate of words is non constant, but on average it is lower than 10 MHz. In order to work at this lower data rate a FIFO is used as a derandomizer, storing in this FIFO only words containing at least one hit. The second step (i.e. the complete encoding of every single hit) is done after the FIFO. The FIFO makes also easier the operation of hit encoding: in case of several hits in the same row the hit encoder requires several clock cycles to process a data word. The FIFO depth is a trade-off between the occupied area and the time the system is stopped because the FIFO is full. In addition, eight service bits are added resulting in a 25-bit word per hit.

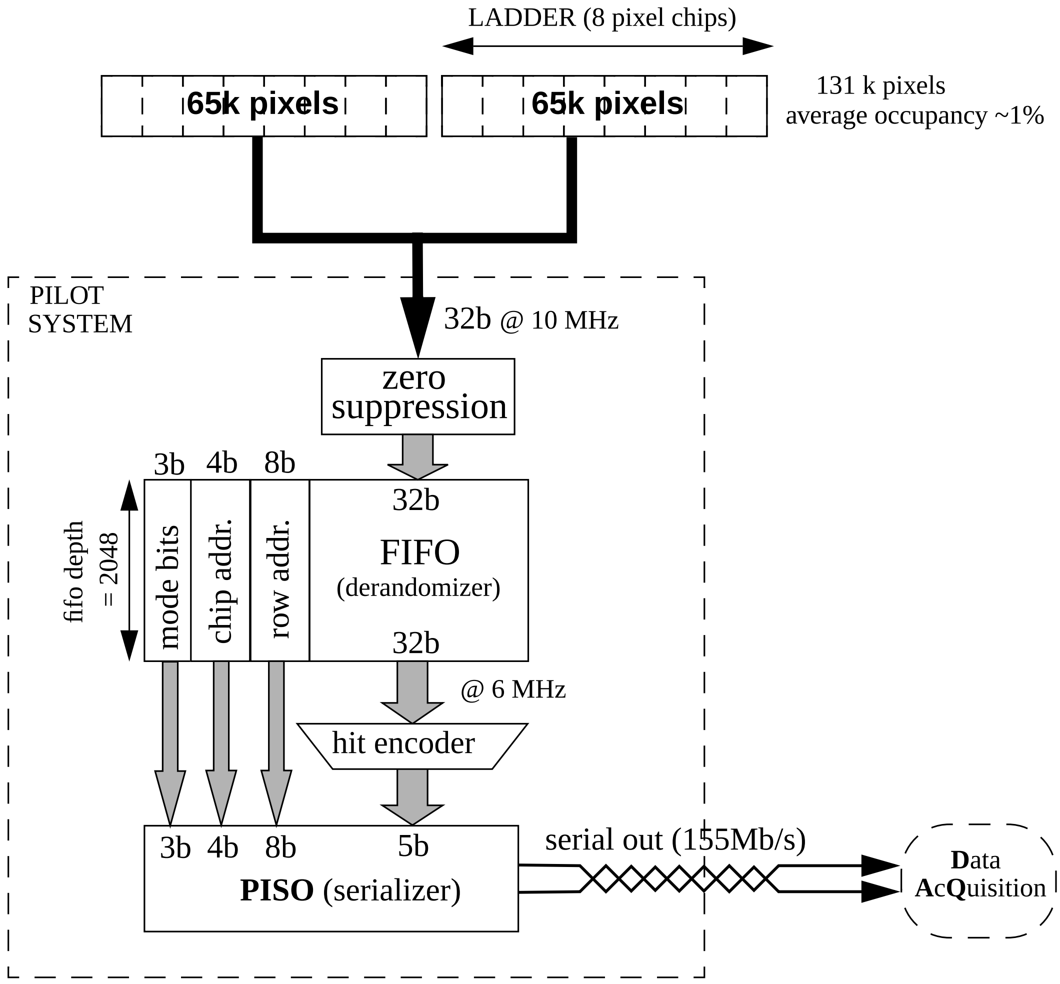
This 25-bit word is serialized and sent through the short data link (running at 155.52 Mbit/s, on differential pair, CMI encoded, see Fig. 3-20) to the router unit, located in a VME crate about 40 meters away from the barrel and then to the DAQ[1]. The choice of the output bit rate comes from the requirement of reading-out a half-stave in 400µs, in the expected case of 1% hit occupancy. This implies that: 16 (chips) x 8192 (pixel/chip) x 100^{-1} (hits/pixel) = 1310 hits are transmitted in 400µs.

[1] DAQ stands for Data AcQuisition system. It is the system located outside the Alice detector, and receives data from it.

As a hit is encoded in a 25-bit word, a word-rate of 1310/400µs = 3.28Mword/s is required. This leads to a bit-rate of: 3.28 Mword/s x 25 bit/word = 81.9 Mbit/s. A safety factor of 2 is taken into account. Last, the frequency of 155.52 Mbit/s is an industry standard, and this improves the availability of commercial components.

In terms of architecture, the design is synchronous: the Pilot system provides the Pixel chips with the same 10 MHz clock used internally, and the readout protocol is synchronous with such a clock.

FIGURE 3-2. **Block diagram of the read-out functions of the Pilot System**



3.1.1 Detailed Specifications

Apart from the main task related to the processing of the pixel data, the Pilot system has other duties, mainly related to the control of the detector, and the interface between the Trigger Logic (common to all Alice sub-detectors), the Pixel chips and the data acquisition stage.

The list of the input/output signals of the Pilot system (as in the PCB prototype) is in table 3-1.

The pilot system can enable or disable individually the pixel chips, through 16 Chip Enable (CE) lines. The CE signal affects only the synchronous readout functionality of the pixel chips, not the acquisition (analog part + delay + writing in the event buffer). The CE signal validates the NEVR, SH_REG_RST and CLEV control signals. This is to keep the system insensitive to spikes on these control signals. Moreover, the CE starts the shifting-out of the pixel data.

TABLE 3-1. Inputs and Outputs of the PCB Pilot System

Signal or bus name	Direction (Pilot point of view)	Connection	Notes
PIXBUS<31..0>	input	pixel bus	Data lines
CE<15..0>	output	pixel bus	Chip Enable lines
STROBE	output	pixel bus	Store a hit in the Multi-Event-Buffer
NEVR	output	pixel bus	New Event Readout
CLEV	output	pixel bus	CLear EVent (= Abort signal)
RST	output	pixel bus	Reset of everything but JTAG circuitry
SH_REG_RST	output	pixel bus	Reset of the shift-registers in the pixel chip
CK10	output	pixel bus	10 MHz clock
TESTIN	output	pixel bus	analog pulse, not yet implemented
TESTOUTV	input	pixel bus	analog equivalent of JTAG (not yet used)
TESTOUTI	input	pixel bus	analog equivalent of JTAG (not yet used)
FASTOR	input	pixel bus	Asynchronous digital signal (not yet used) Asserted if at least 1 hit exists.
FASTMULT	input	pixel bus	analog signal ∝ hit count (not yet used)
SPAREIN<4..1>	input	pixel bus	future use
SPAREOUT	output	pixel bus	future use
LVL1	input	trigger logic	Level 1 Trigger
LVL2Y	input	trigger logic	Level 2 Trigger Yes
LVL2N	input	trigger logic	Level 2 Trigger No
REMOTE_RST	input	trigger logic	Global reset
BUSY	output	trigger logic	Pixel Multi-Event-Buffer full
CMIOUT	output	router unit	155 MHz serial link data output

The Pilot system receives three signals from the trigger logic:

- LEVEL1 Trigger (LVL1)
- LEVEL2 Trigger YES (LVL2Y)
- LEVEL2 Trigger NO (LVL2N)

The trigger signals (plus a reset signal) are pulses longer than 1µs. As in the final version of the system they will arrive from a cable ~40 m long passing through the full ALICE detector (See Figure 1-3 on page 18), they can collect noise from cross-talk with other sub-detectors. Hence, it is important to filter out the spikes, as explained in section 3.2.1.

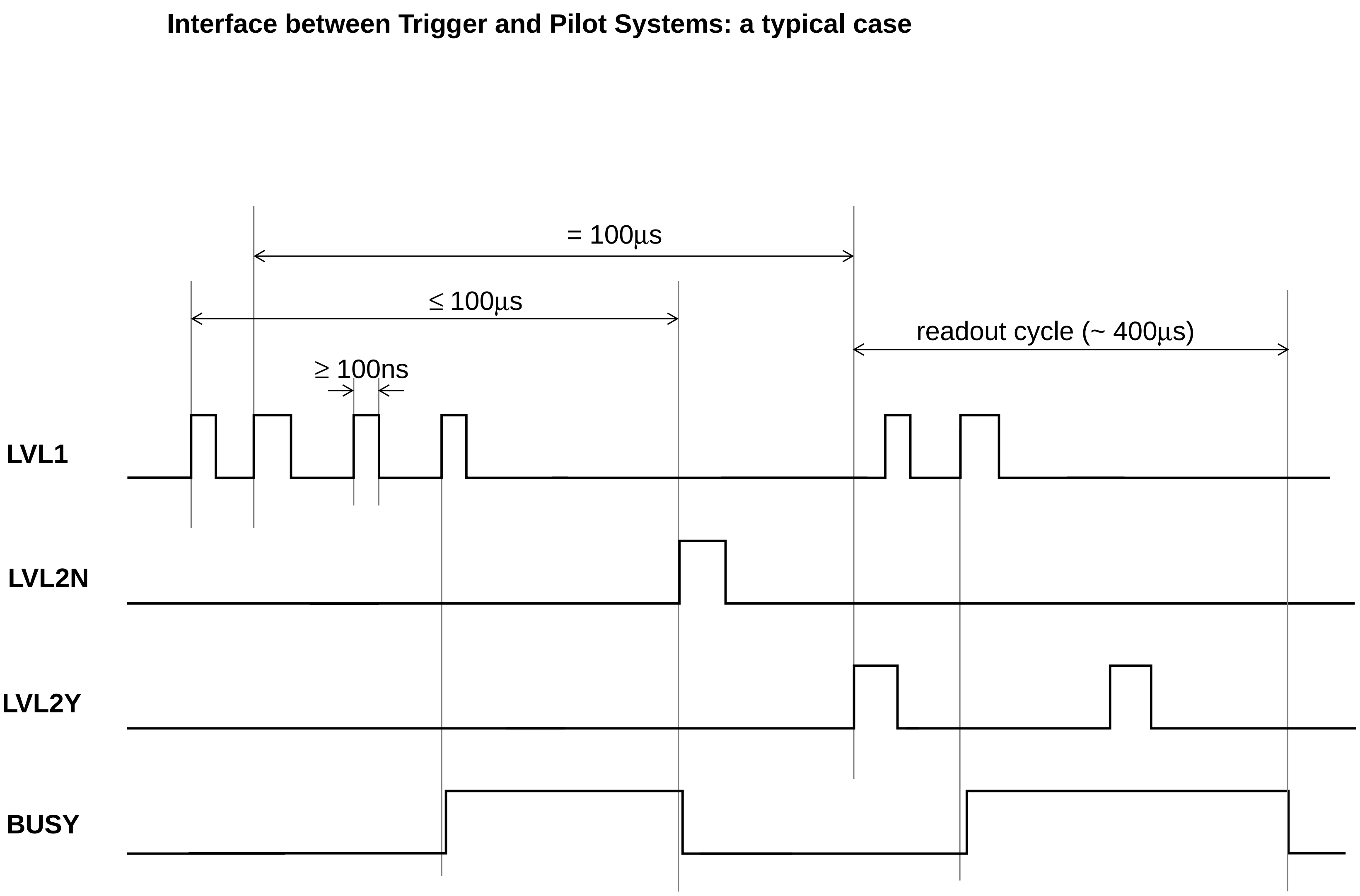


FIGURE 3-3. Interface between Pilot and Trigger Systems: a typical case.

The LEVEL1 Trigger indicates that an event is to be stored in the multi-event-buffer (see section 2.3.3) of the pixels. Hence the pilot system generates a strobe signal and sends it to the pixel chips. The STROBE has to be a pulse two-clock-period long, but it has no timing correlation constraint with the other pixel control signals (CE, NEVR, SH_REG_RST, CLEV). This means that the Pilot can generate a STROBE in any moment (provided the system is not busy).

The arrive of a LEVEL2 Trigger signal (YES and NO) indicates if the oldest event present in the multi-event-buffer is to be read-out or discharged. In the first case the Pilot logic starts a New Event Readout cycle (with the NEVR signal), in the second one it generates a CLear EVent (CLEV) signal (with the protocol explained later in this paragraph).

The Pilot issues a BUSY signal when the number of LEVEL 1 triggers exceeds by four the number of served LEVEL 2 triggers (see Fig. 3-3), that is when the pixel multi-event-buffer is full. In this case the Trigger Logic must not send any other LEVEL 1 trigger, until the BUSY signal is disactivated. If the Pilot chip receives a LVL1 when it is busy, it fires an error bit in the data frame (See BV in header in Fig. 3-7). This error flag is removed only after a global reset. In fact, such an error can cause a loss a event synchronization between LVL1 and LVL2 for an unlimited time, and this error cannot be fixed inside the Pilot system.

At the beginning of a new readout cycle the SH_REG_RST and NEVR signals (see timing diagram on Fig. 3-4) are sent sequentially in parallel to the 16 pixel chips with CE<15..0>=11...11, when the previous readout is finished. After that the pixel chips will be enabled one-by-one and each one of them will react shifting out the 256 data words, in a time of ~ 25 µs. The SH_REG_RST is not strictly necessary, but is sent for safety reasons: it resets to zero the shift-registers of the pixel chips before loading the data with NEVR.

Also the CLEV signal is sent once in parallel to the 16 pixel chips (see Fig. 3-5). In principle it could be sent in any moment, independently from the readout process as this signal affects the multi-event-buffer in the pixel cell, and not the shift-register. But in order to keep a simple and safe interface the Pilot chip send a CLEV only when no reading out is in process.

The situation of coincident NEVR and CLEV should be avoided by the Pilot system, as it has no logical meaning. In this case the pixel chips would react like in the CLEV case, due to the internal architecture.

When the internal FIFO of the Pilot system gets almost full, due to the high hit occupancy, the Pilot system interrupts temporarily the readout (freezing situation), in order not to loose data. It is possible to interrupt the readout in any moment (even during the read-out of a single pixel chip), but for reasons of safety we preferred to do it only when switching from a pixel chip to the next one (Fig. 3-6).

FIGURE 3-4. Interface between pilot and pixel chip: basic readout case

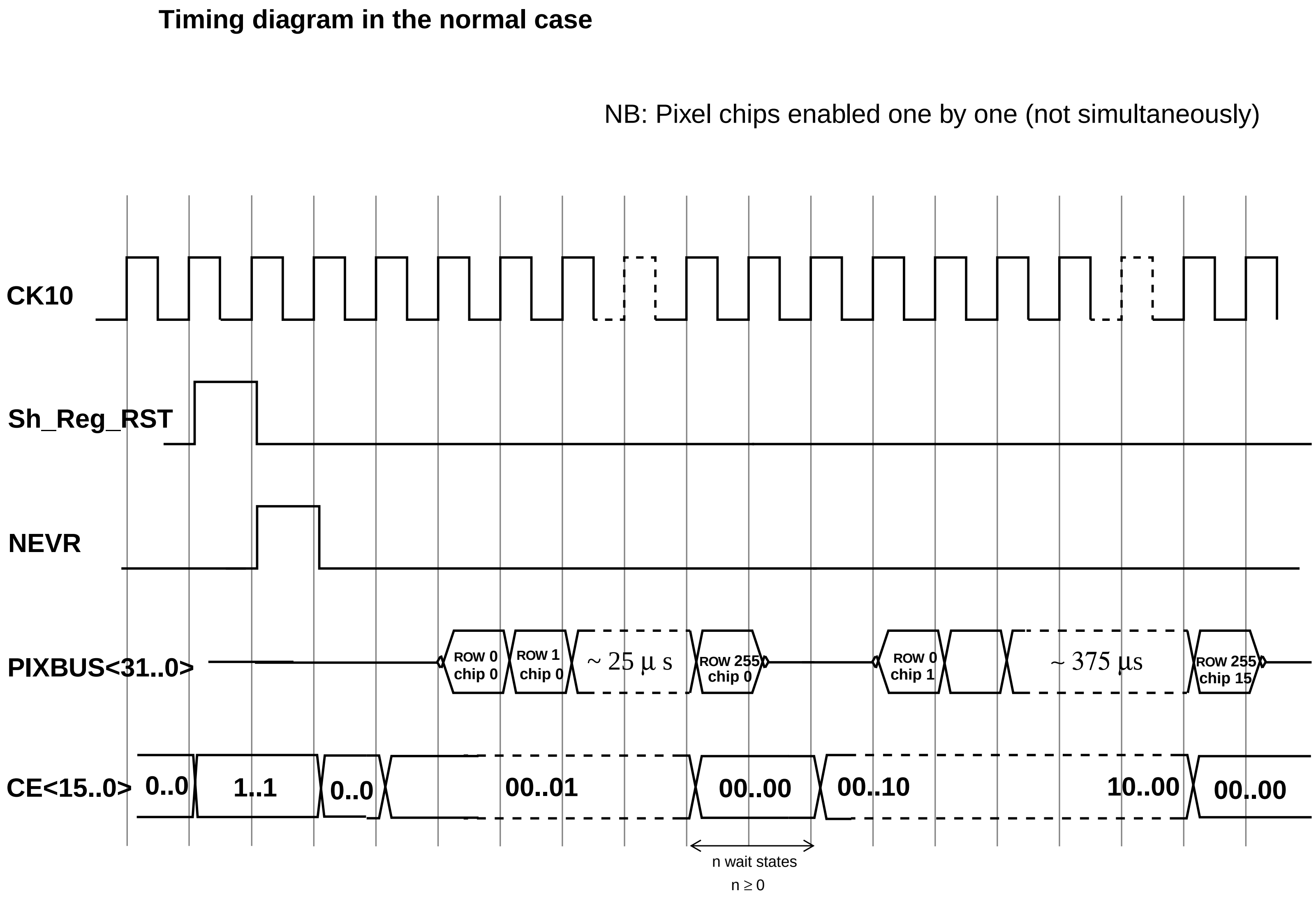
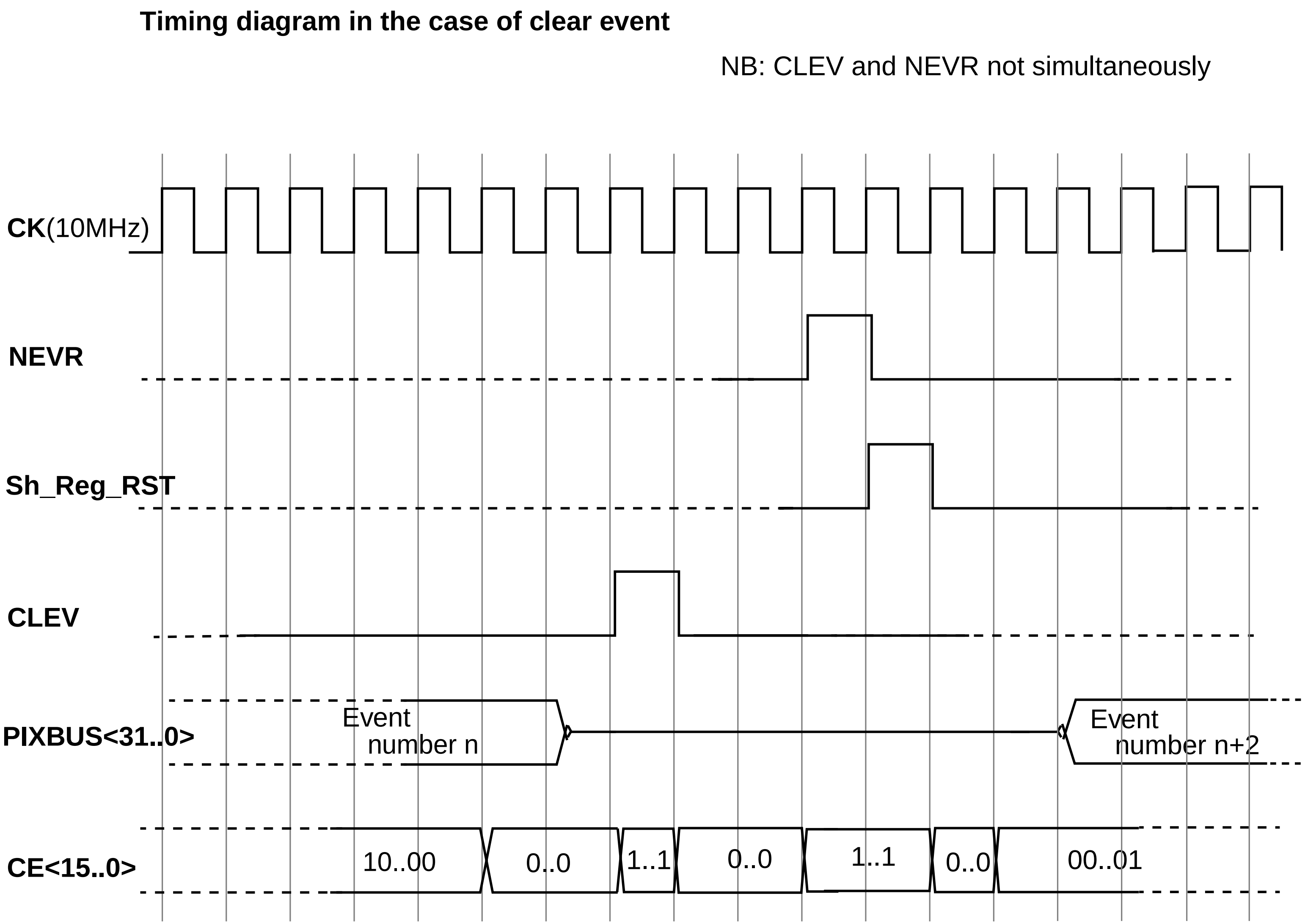


FIGURE 3-5. Interface between pilot and pixel chips: case of clear event



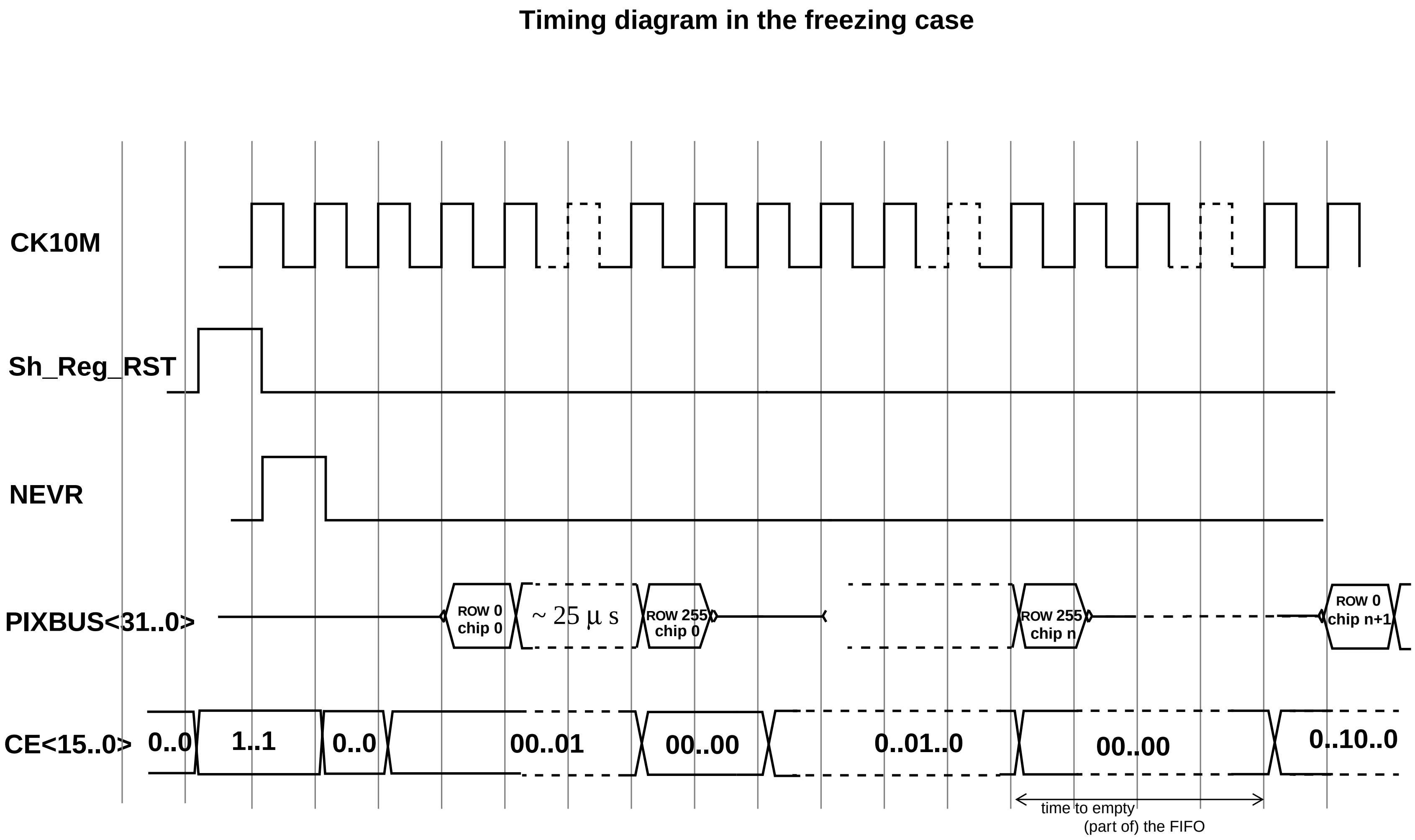


FIGURE 3-6. Interface between pilot and pixel chip: freezing case

The Pilot system has to create some additional information:

- event count (5-bit count increased when a LEVEL2 trigger is served);
- hit count (17-bit count, performed by a system of adders).

The event count information is used by the Data Acquisition system (DAQ) to align properly the data coming from different pilot chips. The hit count is the total amount of hits contained in a single event, as received by a pilot chip. This information will be used by the DAQ for a cross-check with the number of received words (as every hit is transformed into a word).

Moreover, the pilot system compares the current hit count with an externally adjustable threshold in order to skip the read-out of a pixel chip if this is too full. This function can be useful if for any reason the input pixel data contain a very high number of hits. In this case the data have no physical meaning, but they would saturate the Pilot logic and the serial link.

The same happens on the row level: the number of hits in a row is compared with an other adjustable threshold allowing to eliminate the row from the data processing. The default values (on power on) of the thresholds is the maximum, in such a way that this feature is not active unless is configured.

The output format contains, apart from the full encoding of the hits (data word), two more words. One of them is sent before the data word and is called header. It includes the event count information. The other one is the trailer, added at the end of the data words and containing the hit count. In case of LVL2NO (i.e. when the current event has to be eliminated), we decided to send anyway a header followed by an abort trailer. This is useful for the data acquisition system that receives the data. In this way a cross-check is provided by the following equality:

$$\text{Number of LVL1} = \text{Number of LVL2Y} + \text{Number of LVL2N}$$

The three most significant bits of every word (*mode bits*) distinguish the different kind of words. The link protocol (shown in Fig. 3-7) adds a minimum overhead, and is defined “ad hoc” for the pixel system. Each word includes also a start bit, always at logical 0, and four odd parity bits (one every fifth bit), for a total of 25 bits. As mentioned before, the 25-bit words have to be sent out serially, at a bit rate of 155.52 MHz (this means a word rate of ~6 MHz). The odd parity bits assure the presence of “1” also in an zero word.

FIGURE 3-7. Link format

Data frame related to a LEVEL2_TRIGGER_YES

Header
Data
...
Data
Trailer

$0 \leq$ N. of Data words $\leq 2^{17}$

Data frame related to a LEVEL2_TRIGGER_NO

Header
Abort trailer

	b19	b18	b17	b16	b15	b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	b0
Header	0	1	x	EC	EC	EC	EC	EC	BV	x	x	x	x	x	x	x	x	x	x	x
Data	1	x	x	CA	CA	CA	CA	RA	RA	RA	RA	RA	RA	RA	RA	HA	HA	HA	HA	HA
Trailer	0	0	0	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC
Abort	0	0	1	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC

NB: Parity bits are not represented in the figure.

b19, b18, b17 = Mode Bits.

TC	Transmitter word Counter (count up from 2 +)
EC	Event synchronization Counter
CA	Pixel Chip Address
RA	Pixel Row Address
HA	Hit Address
BV	BUSY-Violation error flag
x	Don’t care

We decided to add some more features for **testability** purposes.

A Pixel Emulator: a FIFO device on the same board can transmit to the data inputs of the Pilot system a bit flow emulating the pixels; it is called TXFIFO, it has a size of 32 x 8192 bits. Data are written from the VME bus and read from the Pilot system (see Table 3-2, “Addressing of the different functions of the vme Pilot card,” on page 127).

Another FIFO (20 x 16384 bits) receives the processed data right before the serialization. We call this device RXFIFO. It is very useful during debugging to decouple problems in the serializer from problems in the previous part of the system. In order to use the same board for a different application (pixel readout of the LHCb experiment, see section 1.2.1), the width of RXFIFO has been expanded to 32 bits (instead of 20). RXFIFO is read from VME.

It is worth to note that the kind of testing supported by the TXFIFO and RXFIFO is an issue addressed with the JTAG boundary scan (See “JTAG control” on page 74.) in most industrial applications. In our case the bit rate was to fast for the JTAG capabilities, thus we developed a custom approach.

A control bit (CNTRLSEND) allows to emulate the condition of FIFO almost full, to check the behaviour of the system under a massive amount of data. This control bit is ANDed with the fifo_empty flag and sent to the hit encoder (see pseudoFIFO_EMPTY signal on Fig. 3-16).

3.2 Design of the Pilot System

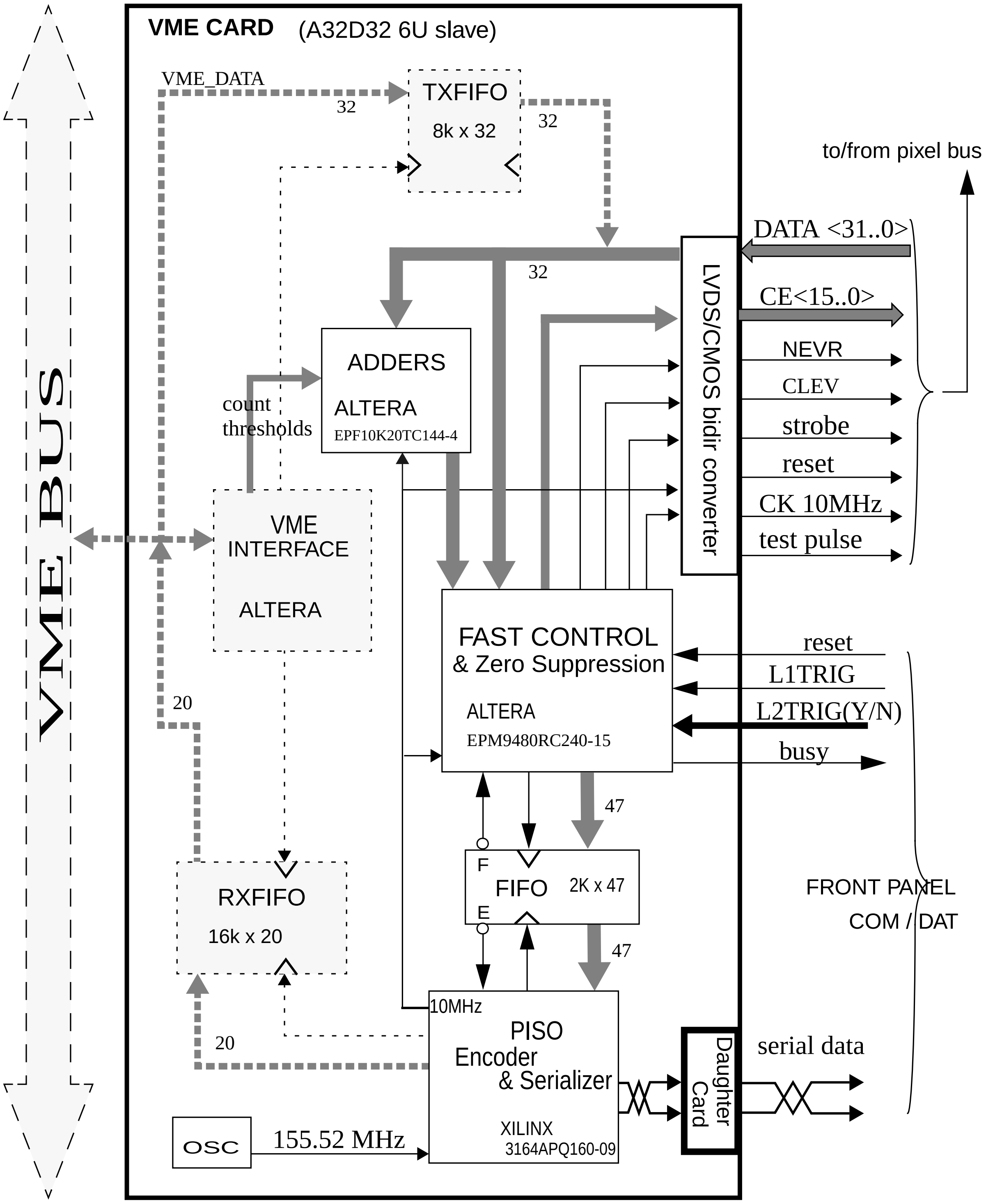
The first prototype of the Pilot system (Fig. 3-8) includes 4 programmable circuits (described in the following sections):

• FAST CONTROL	Fast-control, zero-row suppression, address generation.
• ADDERS	System of adders and comparators for the calculation of the hit-count.
• PISO	Hit encoding + Parallel-In-Serial-Out module.
• VME INTERFACE	The interface with the VME bus.

for a total number of logic gates of about 12000. These programmable circuits contain the logic functions mentioned in Fig. 3-2.

The system is on a VME [60] board. It uses the VME connectors J1 and J2, but not the J3 (i.e. it is a 6U module), as these kind of boards are more portable between different crates. According to the VME terminology, this board is a A24D32 slave, that is, it receives 24 address lines, use 32 data lines and it occupies the VME bus only when required by other modules (master boards). In principle, the final application and the specifications do not require to design the prototype as a VME board; but this is useful for testing purposes. In fact the main data stream of the Pilot system can receive and send data (See dotted bus in Fig. 3-8) to the VME bus, allowing to test independently different parts of the system. The dotted components and lines in Fig. 3-8 improve the testability, but will not be in the final version of the system.

FIGURE 3-8. **Diagram of the PCB prototype of the Pilot System**



The main output of the system is the serial data link (bottom-right in Fig. 3-8), running at 155.52 MHz. As a clock recovery is performed by the receiver of this link, it is useful to encode it with a code that guarantees some transitions also in case of long series of zeros. We decided to use the CMI standard (See Fig. 3-20), as used on the STS-3 electrical interface for the Synchronous Optical NETwork (SONET) standard. The drawback of this encoding is that it requires a clock of twice the frequency (~ 310 MHz). Hence this clock is generated inside the PISO module from the 155.52 MHz clock. A daughter card hosts the circuitry for the driver of the serial link. This is to improve the flexibility and allow future modifications, while keeping the VME mother-board as it is. The Short Data Link (= serial link) physical layer is a shielded twisted-pair type AWG 24, with length of about 40 meters.

The reset logic allows to reset the system from the VME reset signal, with a control register and with a push-button. Power-on reset circuitry exists as well.

3.2.1 Description of the module of Fast Control

This circuit is synchronous with the 10 MHz clock generated by the PISO device (See Fig. 3-8). The Fast Control module (See Fig. 3-9) has two main functions: zero row suppression and fast control (see footnote 1, on page 99). The first one eliminates from the data flow the rows containing no hits. This is carried-out with an architecture based on a 32 bit input logic OR gate. Only when the OR output is at level “1”, the generation of the pulse to write the 32-bit data into the FIFO is enabled, synchronously with the 10 MHz clock (some additional circuitry handles synchronization problems).

When a write operation is performed, the data to be loaded (FIFO_DAT_IN in Fig. 3-9, word format in Fig. 3-10) in the FIFO are provided by a system of multiplexers. When the multiplexer is in position D (Data) it provides the 32-bit row from the pixels and some additional information: 8 bit for the row address (indicating one row out of 256), 4 bit for the chip address (one chip out of 16) and 3 mode bits (indicating the kind of data sent out). The address information comes from two counters driven by the control logic. Each one of them generates as well a flag (not in figure) to mark the end of the read-out of a single pixel chip and of the set of 16 pixel chips. The mode bits (as well generated from the control logic) distinguish a data word from the special words: a Header (position H of the mux) and one or more Trailers (position T of the mux). The header precedes the data words; it is useful to mark the beginning of an event and it contains a cyclic count of the event (easily processed in the counter named EVENT_CNT of Fig. 3-9). The trailer follows the data words; it marks the end of an event and contains the total count of hits present in the full read-out cycle. This information (hit count in Fig. 3-9) requires complex calculations that have been implemented in a separate chip (called “Adders”). The hit count is on a 17-bit bus. Twelve of these bits are placed on the lines normally occupied by the addresses. The remaining 5 bits are placed in 5 lines normally occupied by the 5 most significant bits of the pixel data.

FIGURE 3-9. **Fast Control Circuitry implemented on an Altera CPLD**

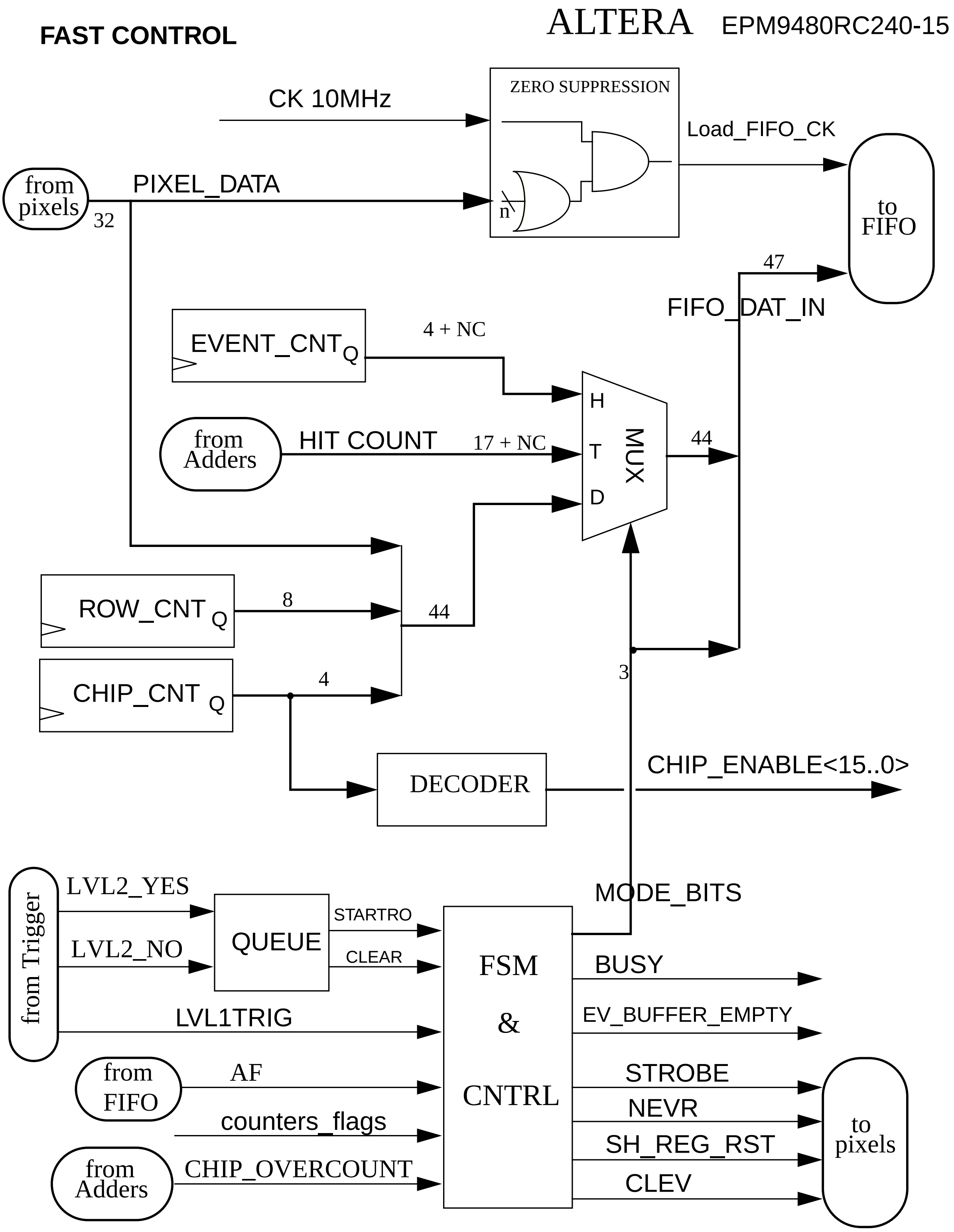


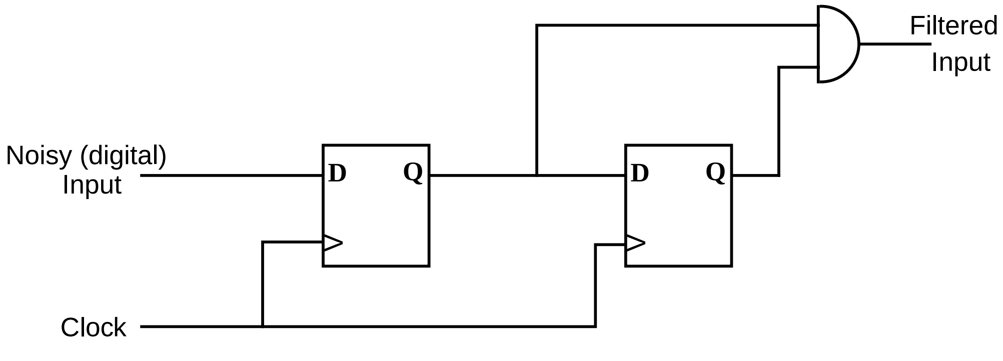
FIGURE 3-10. **Data format on the output of the “Fast Control” module.**

Word type	46 45 44	43	39	32 31	27 0
Header	0 1 0	Ev. Count	BV		
Data	1 x x	Chip Addr.	Row Address	Pixel	row
Trailer	0 0 0	Hit Count			
Abort trailer	0 0 1	Hit Count			

The information of the chip address is also used to generate (through a decoder) the 16 chip enable signals to be sent to the 16 pixels chip. In this way the 16 pixel chips are enabled sequentially to shift-out their data. One setting bit (provided by a jumper on the board, and configured with JTAG in the future chip) allows to configure the system to read-out a single pixel chip (instead of 16); this can be useful for the test of the pixel chips.

The lower part of Fig. 3-9 concerns the control logic (pixel fast-control and internal control). Some of the input signals of this module need to be synchronized and/or filtered. One of them is the AF input; it is the almost-full FIFO flag and goes to a high level when the FIFO contains 2048-256=1792 or more words (2048 is the FIFO depth). This signal is asynchronous with respect to the 10 MHz clock, as the FIFO read clock runs at 6 MHz (see section 3.2.3). Hence it is synchronized with a series of two flip-flops, in order to avoid metastability problems (See Fig. A.12). A similar circuit (Fig. 3-11), is useful to filter noisy signals, like triggers and reset (see section 3.1.1). This circuit is insensitive to pulses shorter than 1 clock cycle (~100ns), like spikes and glitches. The advantage of this approach is the possibility to integrate the filter in a fully digital system (a FPGA in this prototype); moreover it is also useful to synchronize the inputs. After the digital filter the signals are formatted in order to have a fixed duration, independent of the original signals and suitable for the internal use.

FIGURE 3-11. **Simple digital filter**



The LEVEL1 Trigger is simply re-formatted to generate immediately the STROBE signal for the pixels. If a LVL1 is received when BUSY=1, the STROBE is not generated and an error flag is included in all the following headers (see BV in Fig. 3-10), until a global reset is received.

A queue handles the LEVEL2 triggers. They are pushed in the queue as they arrive from the trigger logic and have to be served sequentially, according to the arrival order. A LVL2N is served with a pulse on the CLEV (CLear EVent) output, and it requires a few clock cycles. A LVL2Y is served with a pulse on the SH_REG_RST and NEVR (New EVent Read-out) outputs, that start a full read-out cycle (of ~ 400µs, see Fig. 3-4). The control logic commands the pop operation. An up-down counter with 5 states (0,1...4) is included in the control logic and keeps tracks of the occupancy of the Multi-Event-Buffer in the pixel chips. It can provide the information of BUSY (i.e. Multi-Event-Buffer full, to be sent to the Trigger logic, when counter=4) and empty (when counter=0). A setting bit (provided on the board by a jumper called MEB and by JTAG in the future chip) allows to switch to the configuration with no use of the Multi-Event-Buffer in the pixel cell. This configuration was the first architecture of the pixel system and could be useful in certain situations (e.g. for debugging).

The core of the control logic is a finite state machine (FSM showed in Fig. 3-12 with its inputs) controlling the state of the full pixel system. It generates the reset, the enable and the other control signals for the rest of the device (latches, counters, decoder, multiplexers, etc.) and for the rest of the system. Some attention was required for a proper time alignment of the several processes managed by the control logic.

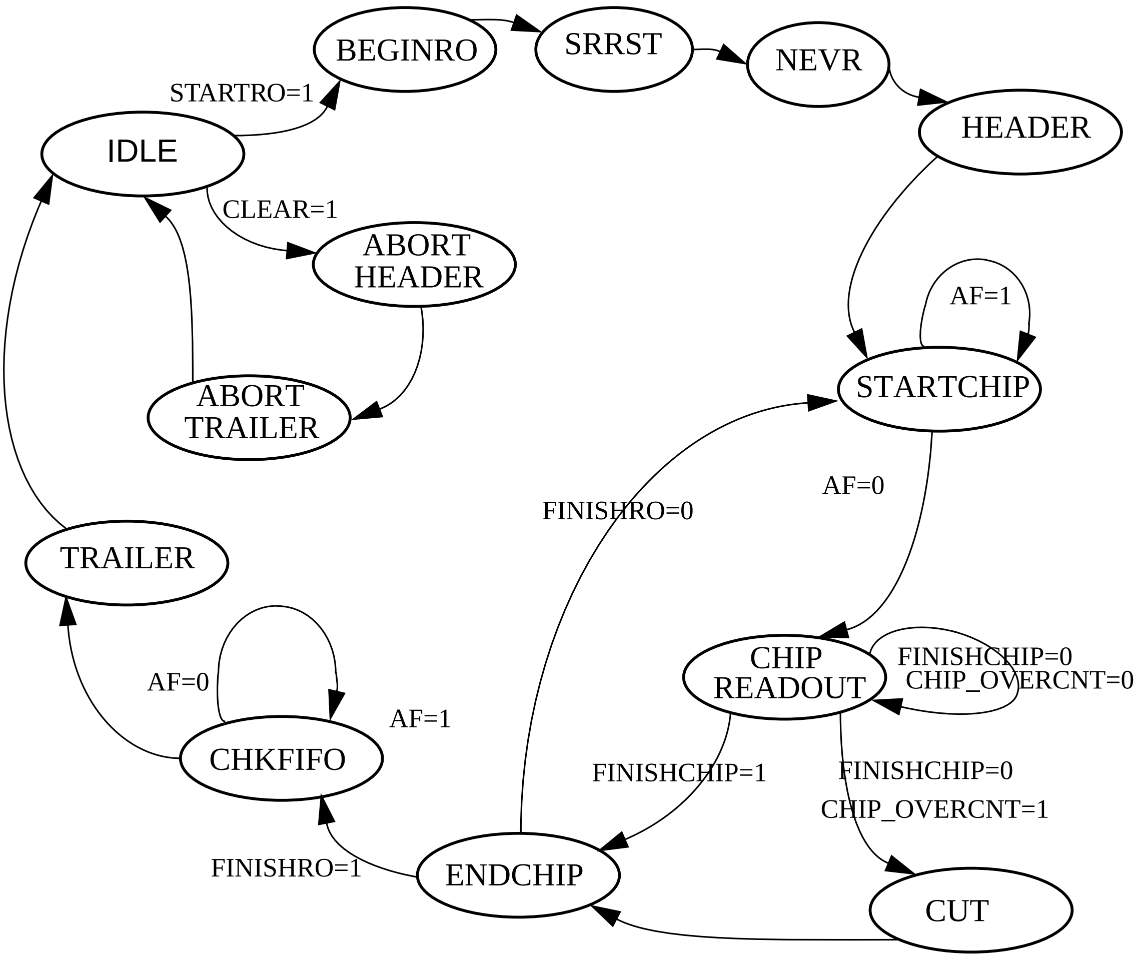
The FSM is initialized in IDLE and waits for CLEAR or STARTRO (outputs of the LVL2 trigger queue, STARTRO = Start ReadOut). In the first case a simple abort cycle is issued. In the second case a read-out cycle starts. A sequence of three states sends the reset for the pixel shift-register (SH_REG_RST) and the New EVent Readout signal (NEVR). Then a header is always produced. The readout of every single chip happens during the CHIP_READOUT state; it stands for 256 clock cycles, as long as the signal FINISHCHIP stays low (FINISHCHIP comes from the 8-bit row-address counter of Fig. 3-9). This state is delimited by STARTCHIP and ENDCHIP states. STARTCHIP evaluates if in the FIFO there is enough space for the next chip data rows: the AF flag has a threshold at (fifo-depth) - 256 words, and 256 is exactly the number of rows coming from a single pixel chip. ENDCHIP evaluates if the full readout cycle of 16 chips is completed (this information is provided by the two row and chip counters, with the FINISHCHIP signal). CHKFIFO evaluates again the status of the FIFO, before the issue of the trailer. The alternative path including the CUT state handles the possibility of skipping the readout of a single pixel chip, if it contains to many hits (the comparison with an adjustable threshold is done in the “Adders” circuit, see CHIP_OVERCOUNT signal of Fig. 3-13).

This module has been implemented on a Altera device, the EPM9480RC240-15. It takes about 50% of the internal logic (= 480 Logic Cells ≅ 10000 gates) and 149 pins on a total of 171 user I/O pins.

FIGURE 3-12. The Finite state Machine, core of the run-time-control.

FSM synchronous with the 10 MHz clock (not on the figure). On the transition arcs, are indicated only the relevant signals:

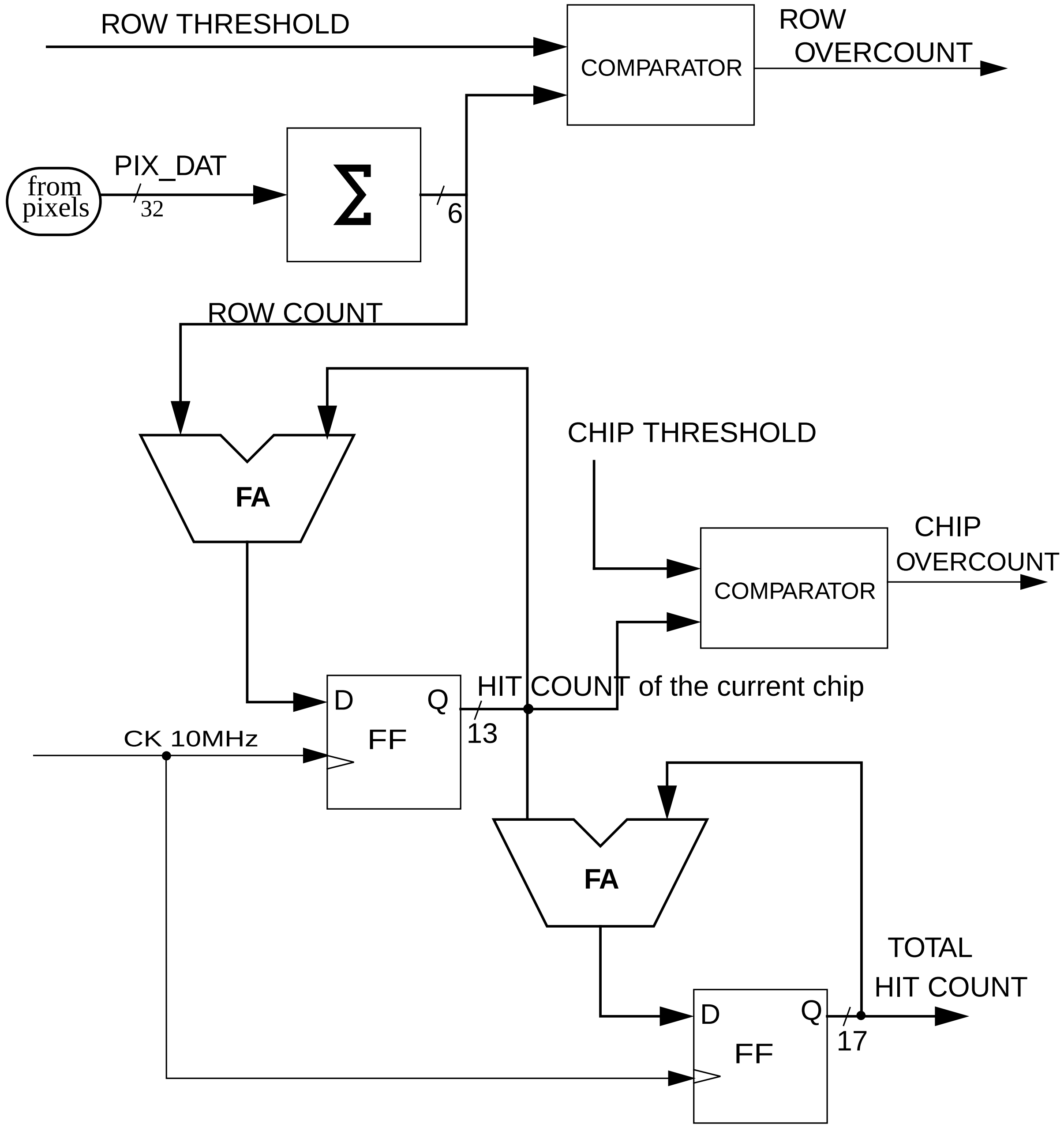
STARTRO= START Read-Out, from LVL2Y (after the LVL2 queue)
CLEAR = Internal clear event signal, from LVL2N (after the LVL2 queue)
AF = fifo Almost Full (fifo flag asserted at fifo_depth-256 words)
FINISHCHIP = Finish of the readout of a pixel chip; from the row (8-bit) counter
CHIP_OVERCNT = Overcount of the hits of a single pixel chip, from "Adders"
FINISHRO = Finish of the Read-Out of a full event (16 pixel chips, normally)



3.2.2 Description of the module Adders

The fast control module requires the calculation of the number of hits in the current read-out cycle. As a row of a pixel chip contains 32=2^5 pixels, the total count of hits in a row can assume 2^5+1 values (0,1,... 32); hence 6 bits are needed to express the count. In a similar way one can see that 14 bits are needed to express the count of a chip (256 x 32 = 2^{13} pixels) and 18 bits to express the count of 16 pixel chips.

FIGURE 3-13. **Adders circuit**

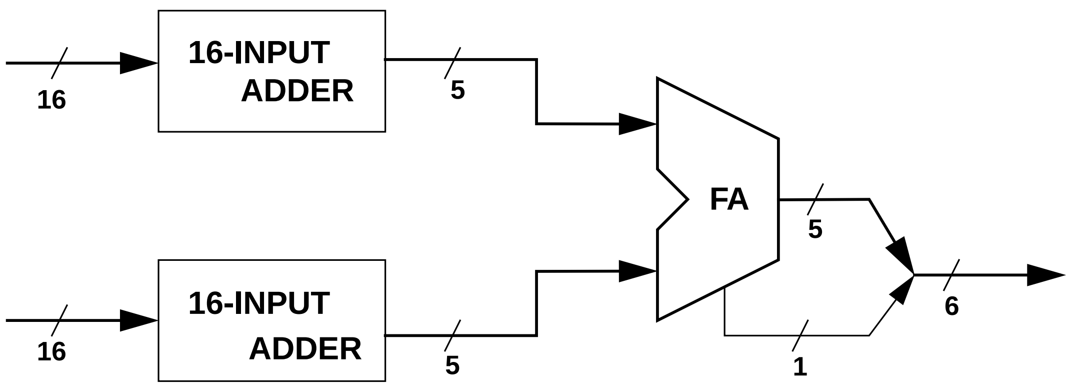


As the case of 16 chips fully occupied is rare and not significant from a physics view point (as the expected occupancy is ~1%), we decided to save 1 bit in the count of the 16 chips, thus using a bus of 17 bits. In this way the situations of 2^{17} and 2^{17}-1 hits are represented by the same pattern (11...11).

Let's now analyse the hardware implementation. To calculate the number of hits in a row a full adder with 32 inputs (each one 1-bit wide) and 6 output is required (See block Σ in Fig. 3-13). In the literature [63] are reported several studies on 2-input adders, but we have not found studies on multiple-input adders. The very first approach one can think for the design of the circuit is to add sequentially the 32 bits, one by one using 31 classical full-adders in series. Using the carry-in and carry-out signals, one can achieve the same results with 15 full-adders. They are to many for a combinatorial circuit as the calculation time would be 15 x T_{FA}, where T_{FA} is the propagation time inside a full-adder. Moreover, in a FPGA based design, such an approach is not recommended, as a combinatorial path would run through several logic cells, increasing the delays. This architecture would strongly limit the clock frequency. To avoid these problems one could create a pipeline, i.e. insert a register after every combinatorial block (the full-adder in this case). This would decrease the computation time to 1 x T_{FA}, but would introduce a latency of 15 ticks, and this is not convenient, keeping into account the rest of the system (in fact to synchronize the system outputs, all the other data paths should be delayed of 15 ticks).

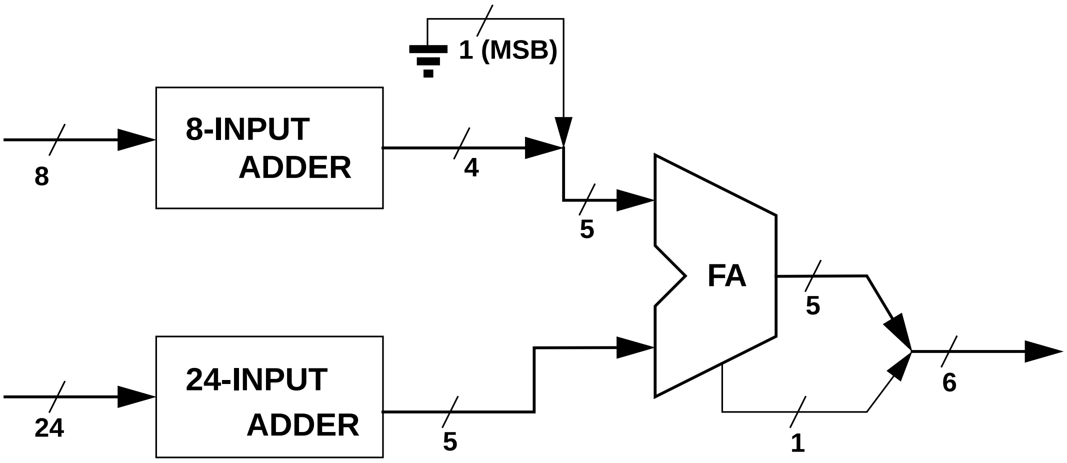
These considerations lead to a more parallel architecture, where all the bits are processed in the same time. Let us approach the problem with the divide-and-conquer strategy, that is splitting the circuit into several simpler circuits, and than combining the results. So one 32-input adder can be built as two 16-input adders whose output are added together by a 2-input adder (See Fig. 3-14), and splitting again the design into adders of 8, 4 and finally 2-inputs. A 2-input adder is the classical full-adder, whose implementation is well known [63].

FIGURE 3-14. **First approach for a 32-input adder using the divide-and-conquer strategy**



With a further analysis one realizes that with this architecture every 2^n-input adder require n+1 output bits, where 1 bit is needed to handle just the limit case (all inputs set to "1"). In fact the balanced architecture is the worst possible. Better results are obtained with an unbalanced partition (See Fig. 3-15), based on an 8-input and a 24-input adders. These two adders are partitioned iteratively, until small adders are used. In order to fit within a logic cell it would be enough to split the design down to 4-input adders, but as this logic is a prototype for an ASIC, the design is split down to 2 and 3-input adders (built with the basic AND / OR logic gates). The architecture of Fig. 3-15 saves about 20% of logic resources compared to that one of Fig. 3-14, keeping the same performances.

FIGURE 3-15. Final architecture for the 32-input adder using the divide-and-conquer strategy



Let us come back to the overall architecture of the “Adders” device. Once the count of one row is ready on a 6-bit bus, it is latched (Flip-flops not shown in Fig. 3-13) in order to avoid a critical path of combinatorial logic. The sequence of latched data is accumulated to generate the count of the current pixel chip (this count is stored in a bench of 13 flip-flops). In a similar way, when the chip count is ready (that is after 256 ticks), it is accumulated to generate the count related to the 16 pixel chips (total hit count in Fig. 3-13). This information is sent to the fast control module.

The counts related to the current row and current chip are used for a second goal. Each one of them is compared to a threshold (defined from outside) in order to provide two flags: chip_overcount and row_overcount. These flags are used by the fast control circuit in order to skip from the read-out one row or one chip when too many hits are present.

This module has been implemented on a Altera device, the EPF10K20TC144-4. It takes about 3000 gates (that is 15% of the internal logic) and 72 pins.

3.2.3 Description of the PISO device

The PISO (Parallel-In-Serial-Out) device (Fig. 3-16) receives words of 47 bits. Each word includes a non-zero pixel row (32 bits), the corresponding chip-row address (4+8 bits), plus 3 mode bits. The main tasks of the circuit are hit encoding, serialization and CMI-encoding.

The first one is the encoding of every hit of the row on a 5-bit information. In this way every 32-bit non-zero input word is split in a number of words, equal to the number of hits contained. These words are now constituted by 20 bits (5, 8 and 4 for hit, row and chip address respectively, and 3 mode bits). This operation is carried-out in the following way. First, the 32-bit input word is stored in an array of FFs, individually resettable. Then a 32-bit priority encoder (a pure combinatorial circuit) encodes the most significant bit of the word in a 5-bit information (coded hit in Fig. 3-16). The coded hit goes in two modules. In the first one it is joined to its chip/row address and sent to the serializer (this is the main data flow). The same coded hit feeds a 5-bit decoder (an other fully combinatorial circuit) that produces a 32-bit word, with a bit set to “1” in the position of the last coded hit. This 32-bit word is used to reset the coded bit from the array of FFs. A new encoding iteration starts: the priority encoder will encode the second most significant bit (as the first one was deleted from the array of FFs), and so on, until the least significant hit of the input word have been encoded. At this point the control logic detects the equivalence signal (as the decoded word matches the content of the FFs) and enable the loading of a new word from the FIFO, generating a unload_fifo pulse signal.

Clearly, the hit encoding must be applied only on the data words, not on the special words (header, trailer, abort trailer). The most significant bit of every word (b46 in Fig. 3-10, that is transformed in b19 in Fig. 3-7) is used to discriminate these two situations. When this bit indicate a special word, the hit encoder is by-passed and the 5 bits placed in the most significant lines of the row bus pass through the first multiplexer and join the remaining 15 bits (addresses and mode bits). Thus the 20-bit word is generated in every situation.

From an implementation point of view, the **32-bit priority** encoder requires a lot of the FPGA resources. The cell structure of the device in use (a Xilinx 3164APQ160-09) is called Configurable Logic Block [66] and includes 5 inputs and 2 outputs (Fig. 3-17 a). If only 4 inputs are used, the 2 outputs can implement any combinatorial function of the 4 inputs, otherwise if 5 inputs are needed, only 1 output can be used. If we give a high-level behavioural description of the priority encoder, the synthesizer fits the circuit in the device, but the results are not optimized. We found a way to describe the 32-bit priority encoder in a way that perfectly matches the cell structure.

We found a way a mapping of the priority encoder tailored to the XC3100A architecture, using the basic cell as in Fig. 3-17 b) and c). With these two cells we can initially build the 16-bit priority encoder of Fig. 3-18.

Then we use two of these 16-bit priority encoders to build the final 32-bit priority encoder, as shown in Fig. 3-19. Again, the additional logic required to merge the two 16-bit encoders fits perfectly in the Xilinx cell structure.

FIGURE 3-16. PISO circuit (Parallel In Serial Out)

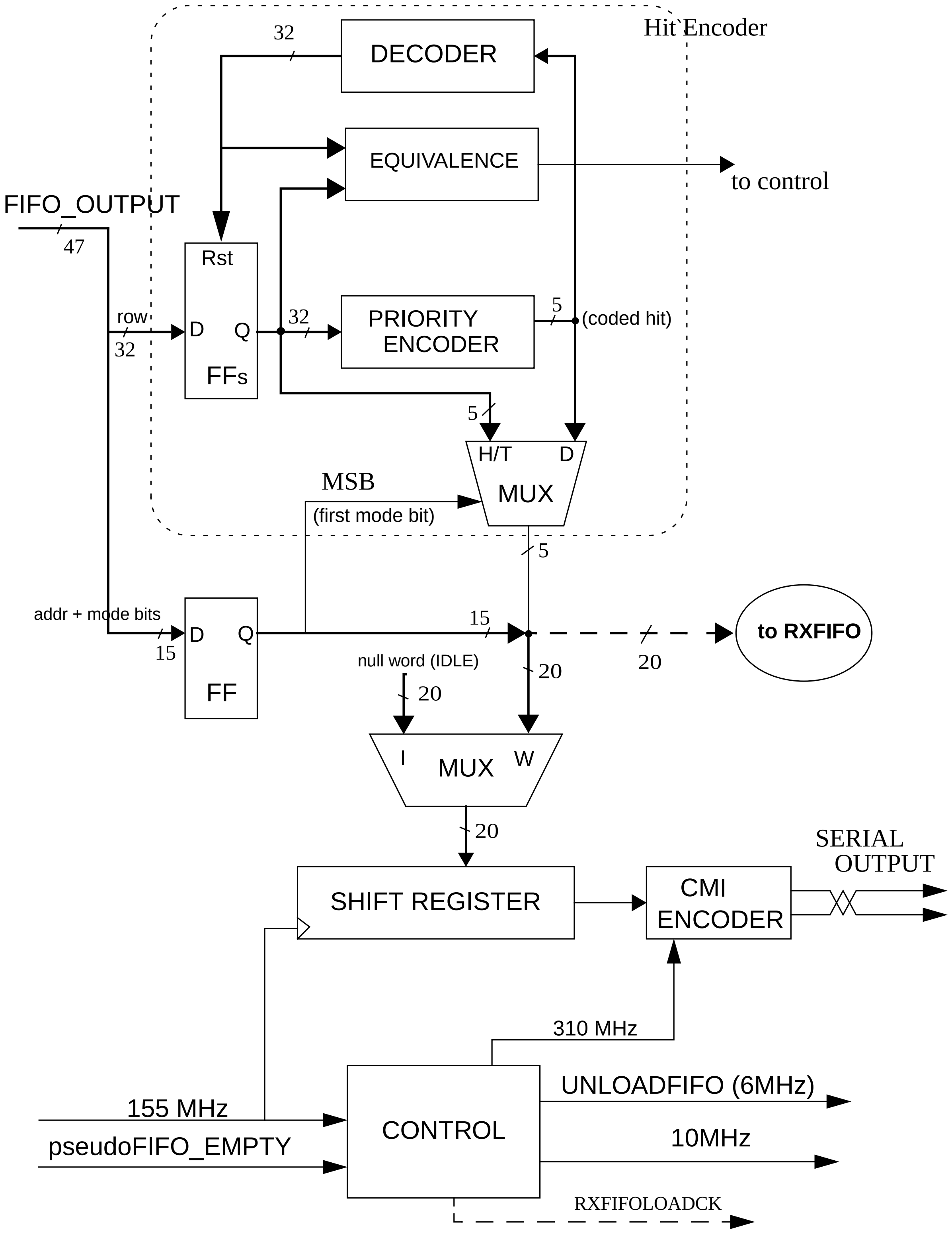


FIGURE 3-17. Use of the Basic XC3100A cell

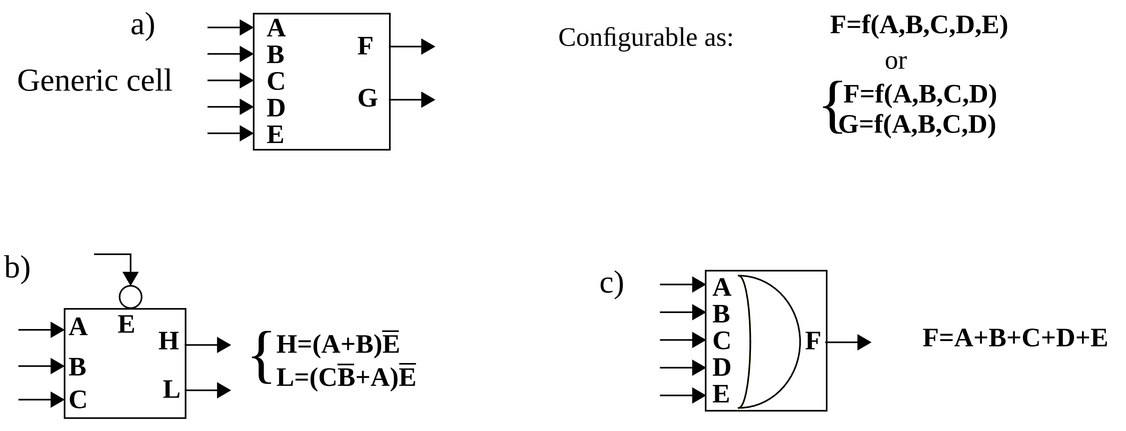


FIGURE 3-18. Design of a 16-bit priority encoder, tailored to the used technology (X31xxA).

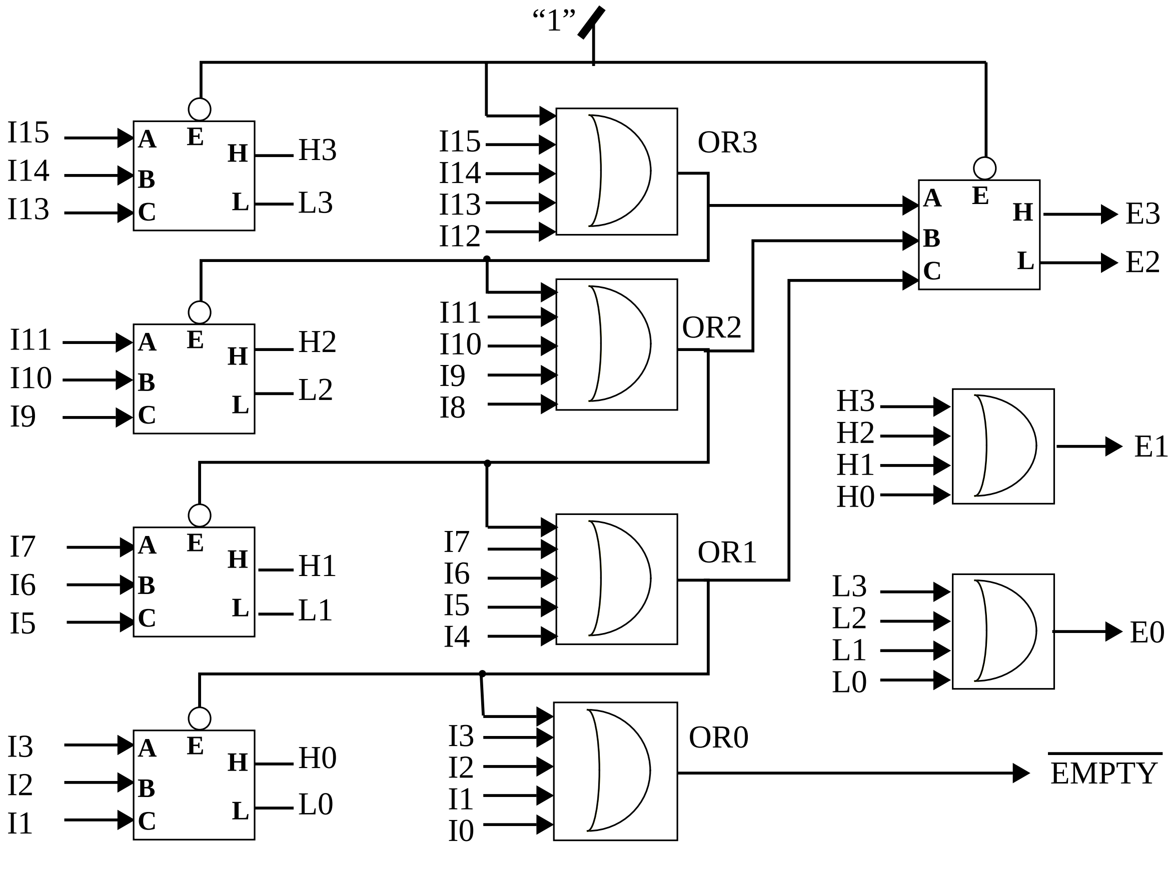
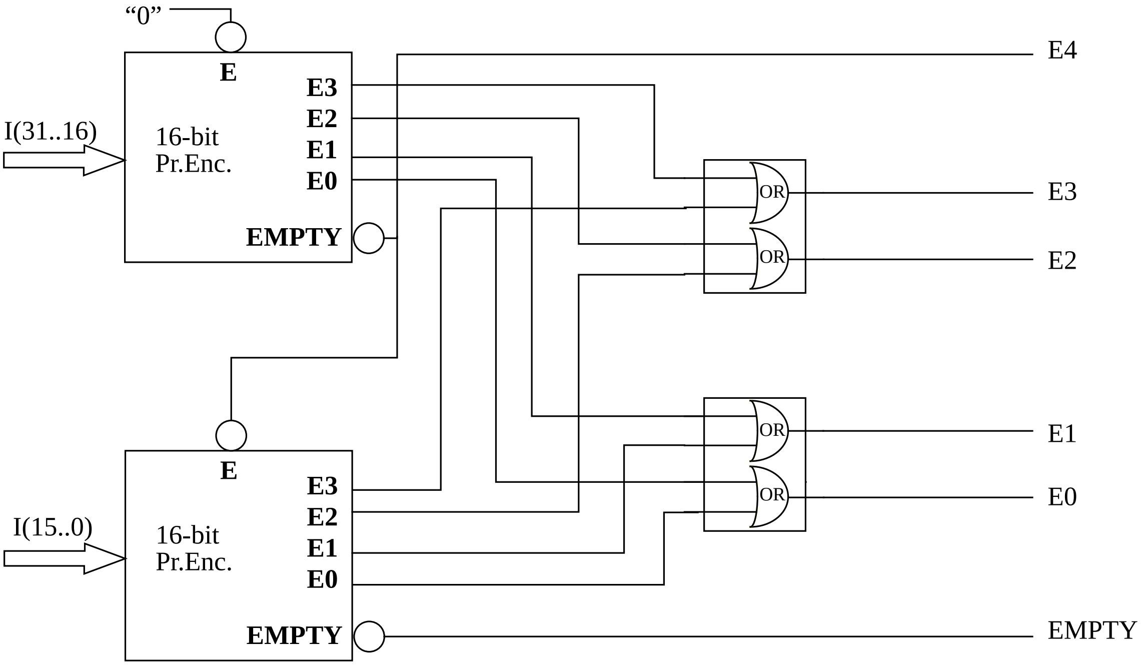


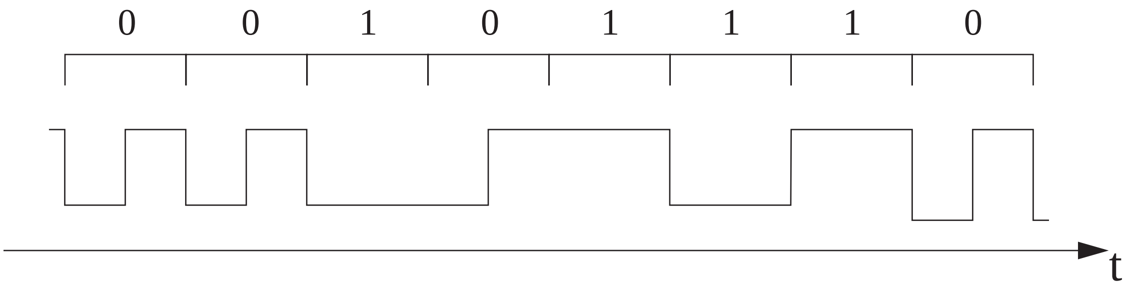
FIGURE 3-19. **Design of a 32-bit priority encoder, tailored to the used technology.**



The second main task of the PISO circuit is the serialization of the 20-bit words. This is done with a parallel-in-serial-out shift-register, clocked at 155 MHz. Due to the high speed clock, the placement of the logic block has been done by hand, to optimize the speed performances. The shift-register also introduces 4 odd parity bits, one every 5th bit. Moreover, when no information is present in its inputs, the shift-register generates an idle word, that is a word with all the information bits set to zero, and as a consequence the odd parity bits sets to one. The idle word sequence is: 000001 000001 000001 000001.

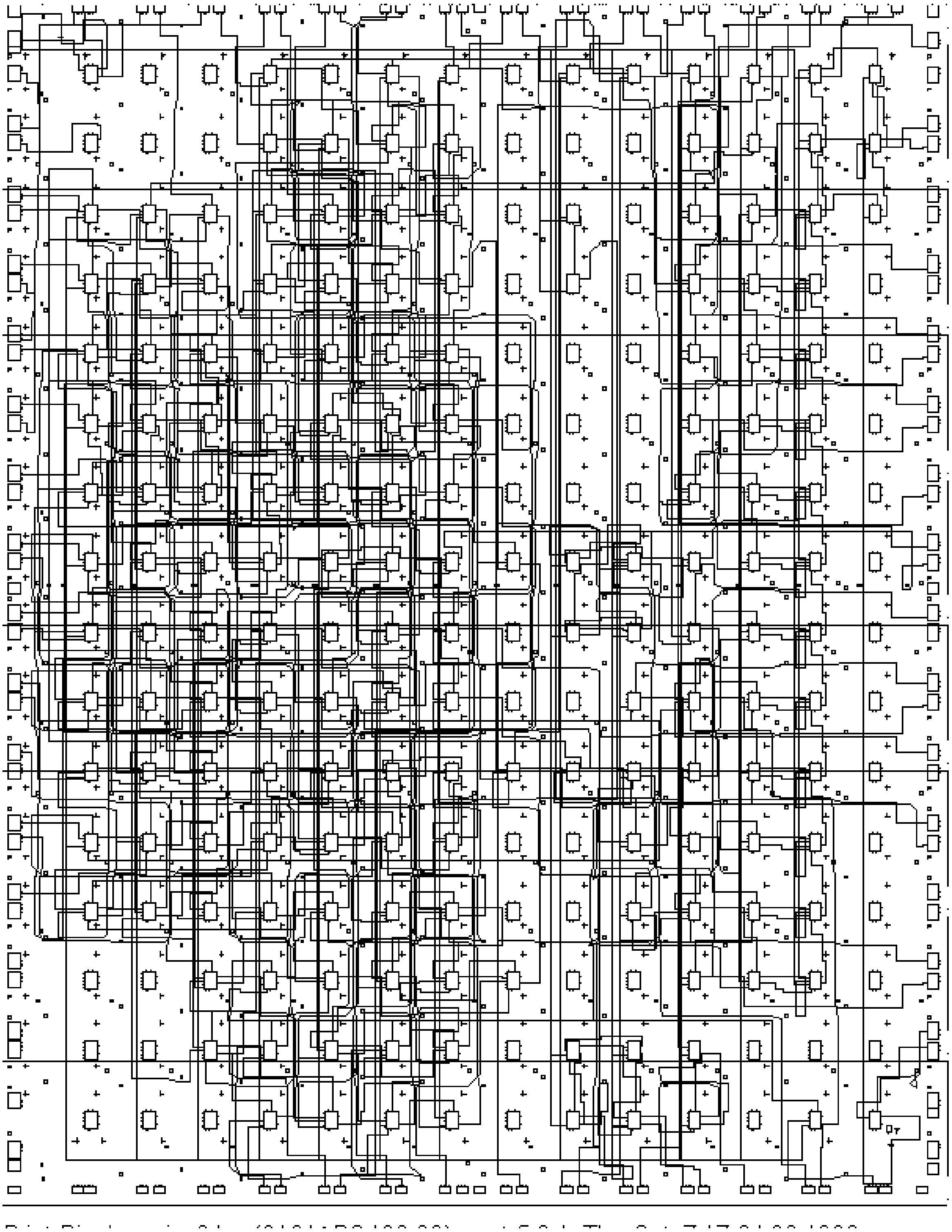
In any case a 24-bit serial stream is sent to the next block, that is the CMI-encoder. This block runs at 310 MHz, so special care was needed on the manual place&route, using a few cells very close to each other, duplicating the signals with a fan-out greater than one (in this context fan-out indicate the number of cells receiving a signals) and matching their timing. CMI encoded signals ensure at least one data transition per 1.5-bit period, thus providing enough edge transitions to ensure an efficient clock recovery. As shown in Fig. 3-20, logical 0s are represented by a low state for half a bit period, followed by a high state for the rest of the period (is exactly this transition that requires a double-frequency clock). Logical 1s are represented by a steady low or high state for the full bit period. The states of the logical 1 periods alternate at each occurrence of a logical 1. The CMI-encoder block also adds a 0 in front of every sequence, for frame synchronization reasons, creating a 25-bit frame that fully represents one original hit.

FIGURE 3-20. **CMI encoded data.**



The PISO circuit (See Fig. 3-21) has been implemented on a Xilinx 3164APQ160-09 device. It takes about 3300 gates (that is 167 Complex Logic Blocks on a total of 224) and 90 pins. Most of the cell placement has been forced by constraints at the schematic level, in order to have the best performances. Moreover, a manual modification directly on the cell array was required to have the correct behaviour of the output circuitry, running with the internally generated clock at 310 MHz. The external clock of 155 MHz reaches the Xilinx device in three different pins, in order to optimize the skew and the routing of the device.

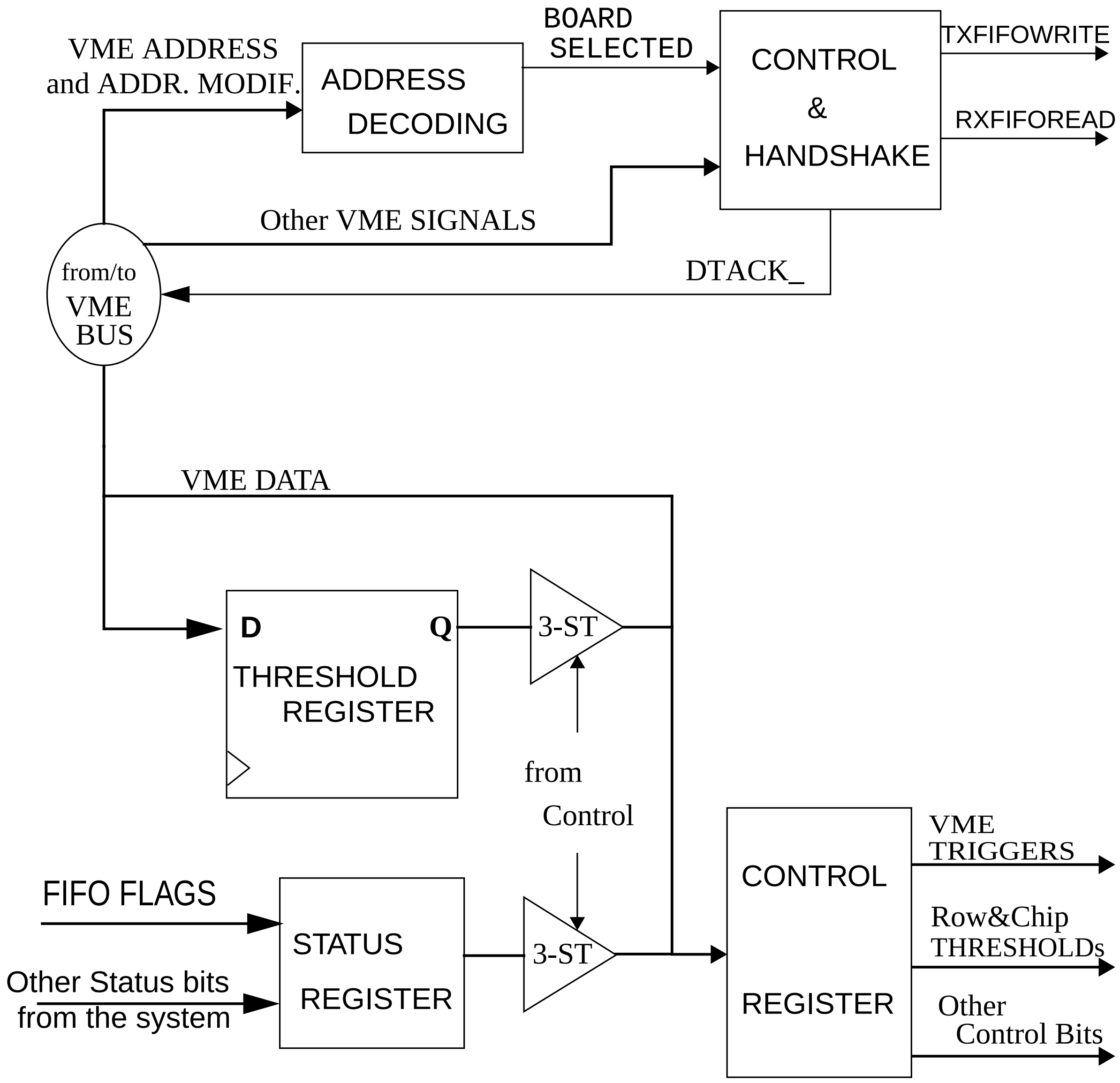
FIGURE 3-21. **Cell placement and routing of the PISO circuit on a 3164APQ160-09 Xilinx device. Some parts on the left side of the array required a manual routing, for the correct behaviour of the output circuitry running at 310 MHz. The circuit uses 167 complex logic block on a total of 224 (usage of 75%).**



3.2.4 The VME Interface Circuit

This module (Fig. 3-22 and box EPM7192 in Fig. 3-23) is responsible for the hand-shaking with the VME bus. Strictly speaking it is not part of the Pilot logic, but it is required to interface it with the VMEbus.

FIGURE 3-22. The VME interface circuit



This circuit monitors the address lines A1-A32 and the address modifier lines AM0-AM5. Once detected its own address range and addressing modes, the interface is ready to execute the VME commands and activate the full board. Once executed the VME command, the interface generates the acknowledgement signal, with the proper timing. The delays required for the correct behaviour are generated with RC delay lines on the board, as the Interface is asynchronous and does not receive any clock. This approach has some advantages as the device will work (allowing VME access) also in case of problems in the clock circuitry.

As a range of address is able to activate the board, every different address in the range can be used to select different part or operations inside the board. From the VME view point, the interface works only with the basic commands of quad-byte (in the VME terminology it means 32-bit data bus) write cycle and quad-byte read cycle. The resulting options are summarized in table 3-2.

TABLE 3-2. **Addressing of the different functions of the VME Pilot card**

$\overline{\mathrm{WR}}$	**A15**	**A14**	**Operation**	**Description**
0	0	X	WRITE IN TXFIFO	The interface generate a single TXFIFO-load-clock, and enables to write the VME data lines in TXFIFO.
1	0	X	READ FROM RXFIFO	The interface generate a RXFIFO-unload-clock and enables the output data lines of RXFIFO to go to the VME data lines.
0	1	1	WRITE IN CNTRL REGISTER	
1	1	1	READ STATUS REGISTER	
0	1	0	WRITE THRESHOLD REGISTER	
1	1	0	READ THRESHOLD REGISTER	

Again, some delay lines on the board are used to generate the trigger pulses and a reset pulse, sent to the rest of the Pilot system.

The control register includes 3 JK-FFs and 4 D-FFs. The VME bus can only access them all together, in the same cycle. The signals related to the JK-FF are:

- CNTRLSEND: If “0” inhibits the PISO circuit from reading out the value of the FIFO. This is to artificially fill up the FIFO and test the behaviour of the system. Note that the piso circuit behaves in the same way when the fifo is empty (in fact it receives the result of the logic operation CNTRLSEND AND $\overline{\text{FIFOEMPTY}}$).
- RUN_TST: If “0” (= RUN) the system is ready to read the data from the pixels; if “1” (= TEST) the system uses the data coming from the pixel emulator (i.e. TXFIFO).
- ENTXRT: If “1”, enables the TXFIFO to retransmit the data already readout. This option can be useful to run the system for a long time with the same input data, in order to get some statistics on the errors.

The structure of the JK-FF avoids the possibility of a mistake from the VME software. In fact in the need of writing a “1” to a certain FF while keeping unchanged the remaining FFs, a word with an asserted bit (in the correct position), and the remaining bits set to “0” has the desired behaviour. Not so if a D-FF is set to “1”, as it would be reset after this write operation.

The remaining 4 D-FF are used to generate a pulse of a certain duration, thanks to a self-reset mechanism. They are:

- CNTRLRST: To reset the rest of the system with a regular VME write cycle
- VMEL1: VME emulator of the Level 1 Trigger
- VMEL2Y: VME emulator of the Level 2 Yes Trigger
- VMEL2N: VME emulator of the Level 2 No Trigger

As these signals return to the low level after a few ticks, they do not need the protection mechanism of the JK-FF. The self-reset mechanism is again based on delays from some RC-lines on the board. In this way they can be adjusted simply changing the values of these discrete components.

The operation of readout of the status transfers to the VME bus the following signals:

- CNTRLSEND
- RUN_TST
- ENTXRT
- TXFIFOEMPTY
- TXFIFOFULL
- FIFOHALFFULL
- FIFOEMPTY
- RXFIFOEMPTY
- RXFIFOFULL
- READOUT_FINISHED = $\overline{\text{FIFOEMPTY}}$ AND IDLE

In a similar way the VME bus can read and write the thresholds register. It includes 8 bit for the Chip Threshold (CTH<13..6>) and 5 bit for the Row Threshold (ROWTH<5..1>). The role of these thresholds has been explained in “Description of the module Adders” on page 117). On the global reset, all the bits of the register are set to logic “1”. With this default value the features of row and chip skipping become transparent to an unaware user.

According to the values of the $\overline{\text{VME\_WR}}$ signal and the RUN_TST register, the interface controls the status of the buffers connected to the VME signals (See “Design of the Pilot System” on page 110.).

One more line, $\overline{\text{RXRT}}$ goes to the RXFIFO to control the retransmit option. So far it is not used, but the connection in the PCB allows future modifications simply reprogramming the FPGA, and with no need of changes in the PCB.

3.2.5 Board Design

The layout of the board is shown in Fig. 3-23 and Fig. 3-24.

The design of the board and of the FPGAs has been done taking into account the possibility of future logic changes. This is possible as the FPGA can be easily reprogrammed and as some spare lines are routed (see Table 3-1), while manual soldering and wiring on the PCB is not recommended, neither is convenient doing a new PCB every time some modifications are needed.

The main power supply of the board is +5V, as most of the circuitry works at the standard CMOS logic levels.

Special care has been put on the routing of the critical lines (clocks and high speed signals), minimizing the length (~1cm) of the 155 MHz TTL clock signal, properly routing the differential pairs, introducing Thevenin resistance terminations[1] to match the impedance of the important lines. The standard techniques to improve signal integrity have been applied, like decoupling capacitors between ground and the power supplies (to avoid ground bounces), use of ground and power supplies layers for shielding the signal layers. The board has been placed and routed on 7 layers (top layer, bottom layer, 2 internal layers for the signals and 3 internal layers for ground, power supplies and shielding). One more layer has been added for manufacturing reasons.

In the routing we tried to keep the important lines on the external layers (top and bottom), to improve the testability.

On the right side of Fig. 3-23 and Fig. 3-24 there are the VME connectors (J1 and J2); close and connected to them there is an array of bidirectional buffers. They are required for reasons of signal integrity, thus they are placed as close as possible to the VME connectors.

The front-panel is on the left of Fig. 3-23. It includes a large 128-pin connector (to interface the pixel chips), a small 7-pin connector (to interface the trigger logic), some LEDs. In the bottom-left part there is a daughter card, used for the drivers of the serial output. The daughter card will allows future modifications in the choice of the drivers. Under the daughter-card some lines provide 5V, 10V and ground outputs, for the pixel chips. Again some fuses are used to decouple the internal power supplies from the outputs.

All the trigger interface signals and the pixel interface signals are on low voltage differential signal (LVDS) levels, to improve noise immunity and transmission capability. LVDS is a new data interface standard that is defined in the TIA/EIA-644 and the IEEE 1596.3 standards. It is essentially a signalling method used for high-speed transmission of binary data over copper. It uses a very low voltage swing (350 mV). Since the receivers respond only to differential voltages, they are relatively immune to noise such as common-mode signal reflections. In addition, LVDS emits less electromagnetic interference (EMI) than other data transmission standards.

The LVDS drivers and receivers are closed to the front-panel connectors; they need a 3.3-Volt power supply, that has been generated from the 5-Volt supply. The portion of the power plane under these components has been reserved for this 3.3-Volt supply. Some fuses have been placed in the power lines coming from the VME backplane connector, to protect the components in case of short circuits.In the internal part of the board one can distinguish three benches of fifos (TXFIFO: 4 components on the top-right, internal FIFO: 6 components in the middle, RXFIFO: 4 components in the bottom-right) and the four programmable devices (squared). The pins on the Altera devices have been assigned to minimize the bus length (central busies of Fig. 3-24), while respecting the internal structure of the devices[2].

[1] The characteristic impedance is different for single and differential lines. So different values of the resistors are required.

The Xilinx and the FLEX "Adders" devices require an external eeprom for the power-on configuration. We had some care to put in high impedance all the Xilinx inputs during the configuration, in order to avoid destructive level collisions in case of a bad configuration procedure.

In the bottom-right part of the board, there are place and connections for a second daughter card. This will be designed and plugged to deal with the analog signals from and to the pixel bus (these lines were not fully specified at the time of the board design).

Some spare lines and test-points on the unused pins have been added to allow future modifications.

[2] The following recommendations help to get the best timing performance on FLEX and MAX 9000 devices [61]:
- Assign bidirectional I/O pins and wide buses to ROW pins
- Assign speed critical and low fan-out inputs to ROW pins
- Assign high fan-out pins to COLUMN pins

FIGURE 3-23. **PCB layout of the Pilot card (top layer)**

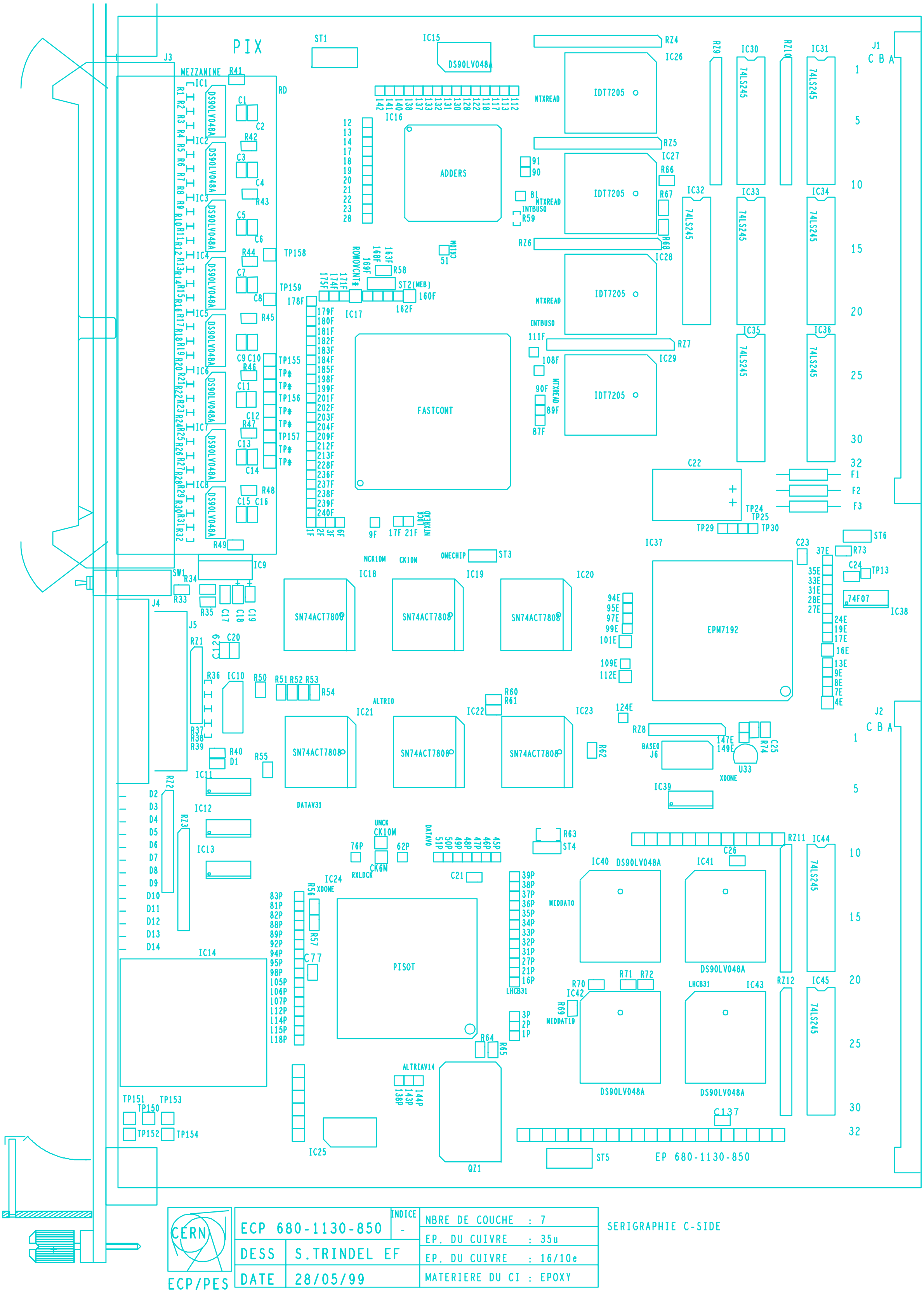
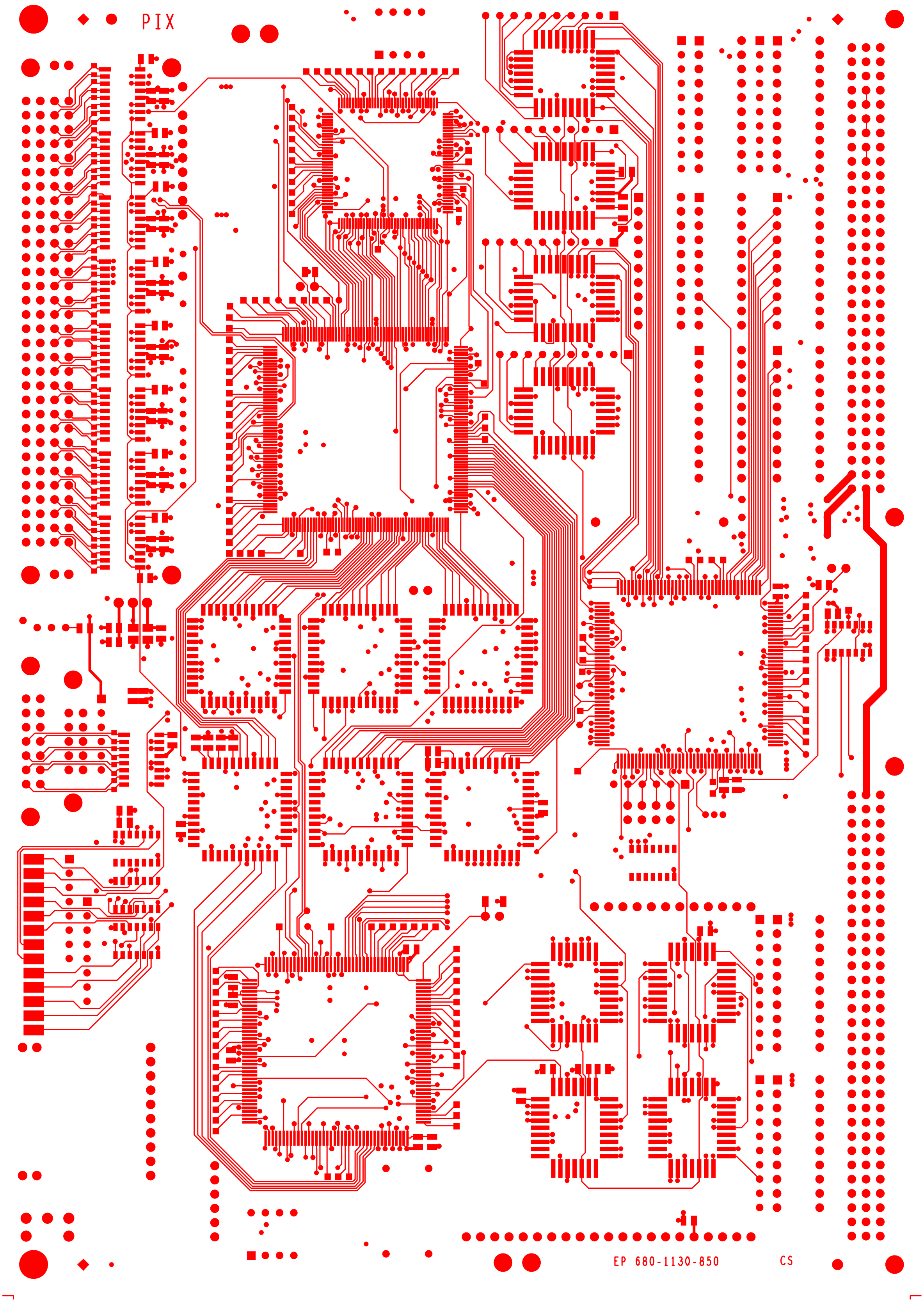


FIGURE 3-24. routing of the PCB of the Pilot card (top layer).



3.3 Design evolution

Once the correct behaviour of the Pilot logic is verified, together with the other modules connected to it (pixel chips, trigger logic, router and data acquisition), an integration of silicon will be faced.

In principle, software tools for automatic synthesis are available. In our case the chip has to be radiation tolerant, and this requirement cannot be solved by the standard design tools and techniques. This requirement can be faced in two ways: at the lay-out level and at the fabrication level. At the level of the lay-out, one can use the techniques described in section “Gate-all-around CMOS design” on page 38; these techniques allows to use a conventional technology in the fabrication of the chip.

Otherwise the problem can be faced with special processes in the fabrication of the chip, available for whatever circuit based on the standard lay-out techniques. These special process are called radiation hardened technologies. They have an advantage: they allows to use the automatic translation and synthesis tools to migrate the present design of the Pilot logic (based on programmable circuits) to an ASIC. Generally, radiation hardened technologies are much more expensive then the conventional commercial technologies. But in the case of the Pilot chip, this will not be so relevant as the experiment will need a very little quantity of about 240 chips.

The use of such a rad-hard technology will ease the migration of the design to an ASIC, as most of the development activity carried out for the Pilot board will be directly transferred into the ASIC.

Another approach under study is the implementation of the Pilot system in a Multi-Chip-Module (MCM). This approach can be convenient if some of the chips can be re-used or shared with other applications or even found in the market.

CHAPTER 4

Testing of the Pilot System

In this chapter we describe the testing of the Pilot card. In order to carry out these tests, we developed dedicated hardware and software facilities. The most important of them is another card conceived for receiving the 155 MHz CMI signal (produced by the Pilot card, see previous chapter). This receiver card (called VME Serlink test card) contains complex circuitries that needed to be tested as a stand-alone module.

Moreover, we developed a testing system, using Labview as software environment. The software is designed to allow low level access to VME registers of tested cards or perform complete test with manual or automatic error detection. The main purpose of this software is to test the integrity of VME cards and of developed algorithms. The software was written in Labview 5.1, using VISA driver to access VME. The software relies on PCI-VME bridge.

Last we produced an interface card to connect the Pilot card to a system of Pattern Generator and Logic Analyzer, in order to emulate as close as possible the real utilization of the Pilot card.

The prototype developed shows the correct functionalities and this proves the validity of the architecture in order to implement the Pilot logic on an Application Specific Integrated Circuit (ASIC).

4.1 Introduction to testing

The complexity of modern digital circuits has dictated that the problems of testing a device must be addressed as an integral part of the design process, and not as an afterthought, as testing costs are to be maintained within acceptable limits. Furthermore, the designer must be aware not only of the techniques of enhancing the testability of the circuit regarding testing immediately after fabrication, but also of the methods to enhance its testability at subsequent stages of system assembly (the full pixel detector and Alice detector). Some of these techniques determined the design of the Pilot system described in the previous chapter and they are review in this introduction to give them a structured presentation.

System-level testability of hardware/software systems

For modern hardware/software systems, system-level testability has become a severe problem. A system is regarded as a collection of co-operating hardware and software modules. System-level testing denotes the testing of the overall system behaviour. This system behaviour is composed of the behaviour of all the individual hardware and software modules. Due to the large numbers of interacting hardware and software modules, exhaustive[1] testing is unfeasible. Moreover, embedded hardware/software modules are difficult to control and observe during testing. Testing systems with real-time behaviour (like the pixel detector system) requires testing both functional and timing behaviour.

The current approaches towards system-level testing are *ad hoc* approaches. They focus on a single testing activity or on a single stage in the system life cycle. Structured approaches towards system-level testing of a generic hardware/software system which cover the entire system life cycle, are still missing.

Experience shows that design errors in hardware and particularly in software are an important failure cause during field operation. These field errors all escaped during system-level testing.

Design for testability includes some guidelines:

- A global reset signal is important, this brings all of the internal memory elements to a known state.
- Long counter chains should be broken. A 10 bit counter needs 1024 cycles to test it fully, if divided into 25 bit counters, only 32 cycles are required. (Plus a few cycles for testing the decode logic.).
- Bring difficult to test internal nodes, out to device pins. This may also be difficult, as pads are usually limited resource.
- Derive all frequencies from a single master clock.
- On board clock generators should be replacable by an external test clock signal, this will allow external control of the clock during test.
- Use synchronous design techniques.

[1] In this context, exhaustive testing means to verify the correct behaviour of the system for every combination of inputs, in any timing relation. The complexity of the problem grows exponentially with the number of inputs.

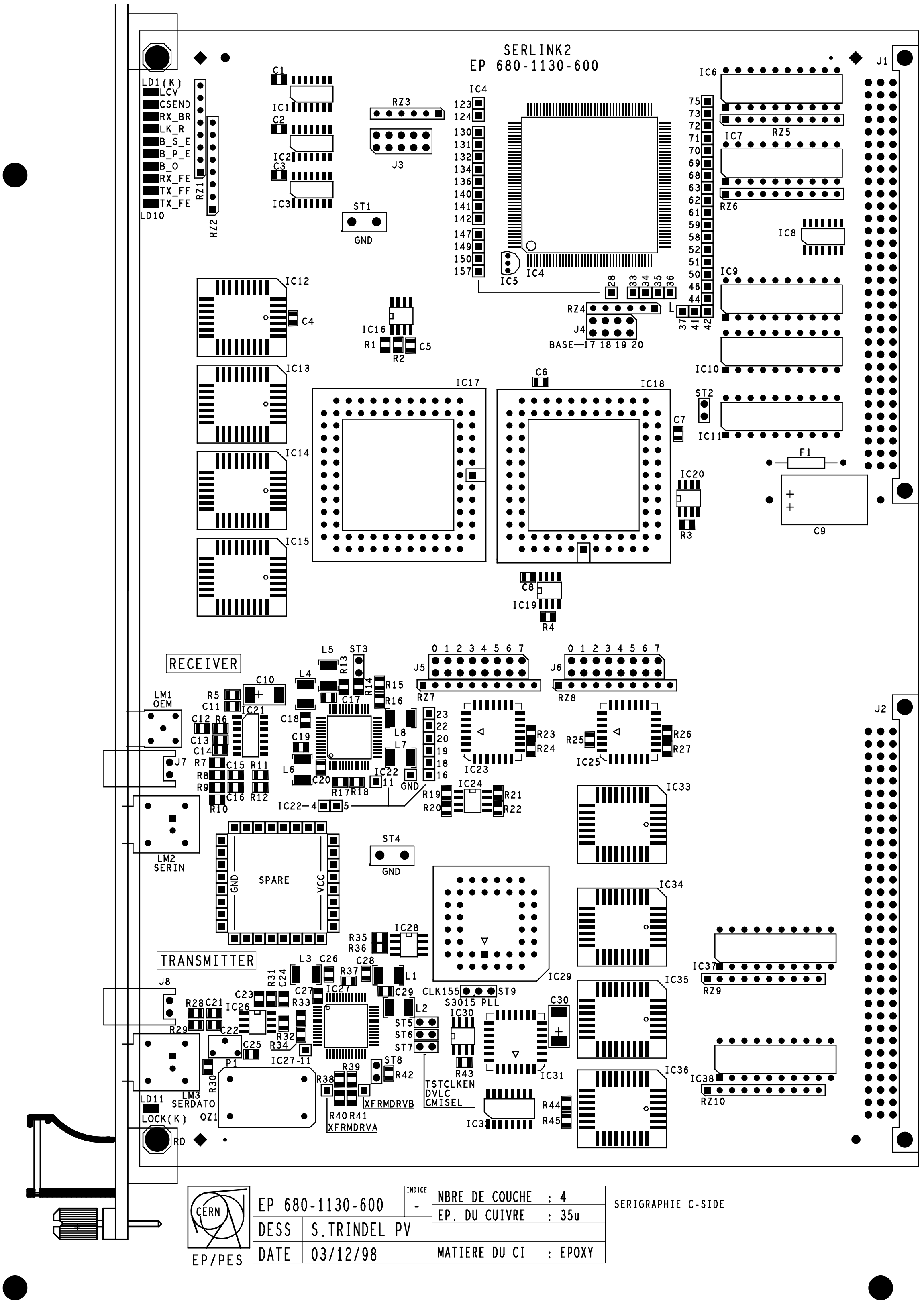
4.2 Short Link Test Card

In order to test the performance of the Short Data Link, we have developed the VME test card shown in Figure 4-1.

The purpose of this card is the testing of the serial output of the Pilot Card, after a cable of about 50 m. This means to use the Test Card as a receiver. To be sure not to introduce any error during the operations of the receiver or along the cable, we decided to fully test this card before connecting it to the Pilot card. For this reason the same card is also able to transmit serial data, in this case with a commercial transmitter, and read back the data at the end of the cable. Thus the card contains both transmitting and receiving circuitry, transmitter and receiver FIFOs, and a A24/D32 VME interface built in a 7192S-10 Altera CPLD. This VME interface is similar to the Pilot card’s interface (see section 3.2.4), as it include a control register, a status register, and handles the writing to a TXFIFO and the reading from a RXFIFO. Three 7064S-5 Altera CPLDs are used for the parallel-to-serial and serial-to-parallel conversions, for the framing and for the link control. For the CMI serial encoding/decoding we have employed two commercially available AMCC devices (S3015 and S3016). A 19.44 MHz quartz device and an integrated PLL (MC88915-16) produce the 155.52 MHz clock.

This Short Link Test Card was tested during several days over 56 m of cable (twisted-pair type AWG 24). No transmission error was detected during the transmission of 7.04 x 10^{10} packets, which corresponds to 1.76 x 10^{12} bits transmitted.

FIGURE 4-1. **PCB layout of the VME short link test card.**



4.3 Testing with the VMEbus

The Pilot board is a VME board. Typically the VME bus allows to access the boards through a VME controller that is placed in the left-most slot of the crate. The controller contains a processor that addresses the other boards as an extension of the its memory. The controller has a standard operating system, and is normally programmed in some high-level language.

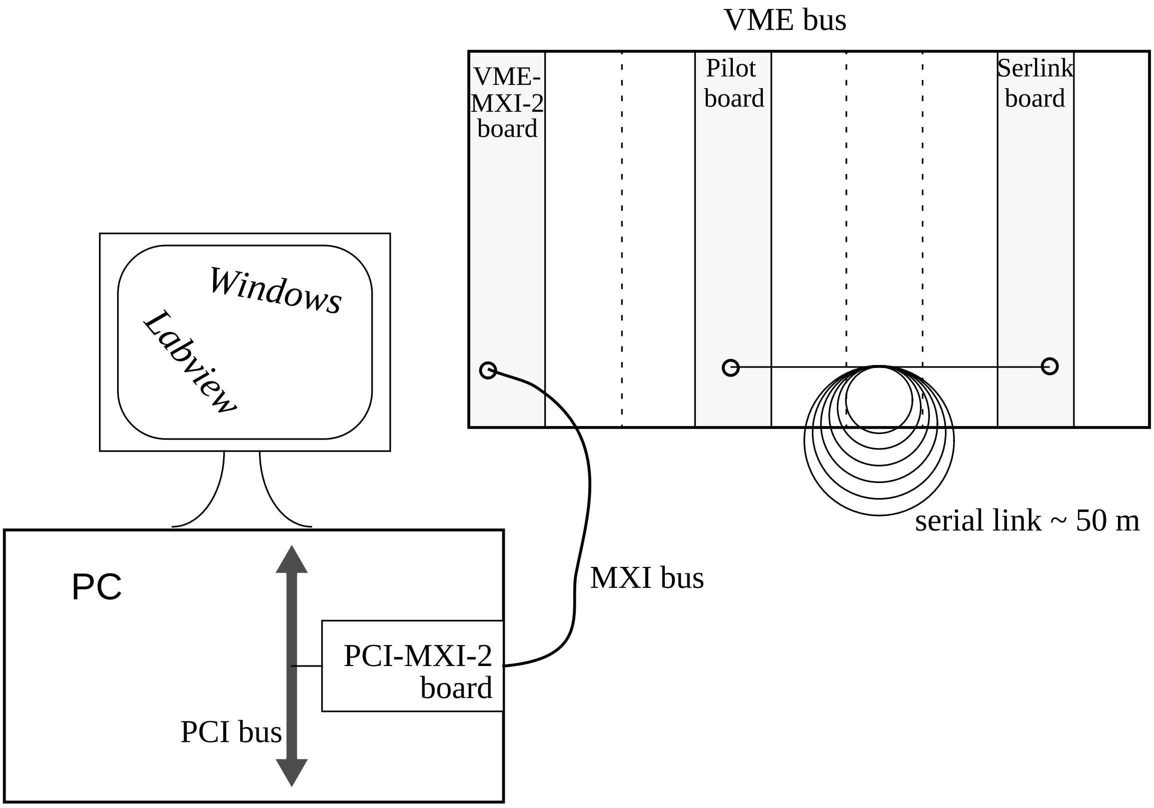
In our case the configuration was slightly different. We used the Labview environment. Labview is an graphical programming development environment. Its source code uses a block diagram approach that works much like schematics and flow charts to solve problems. In addition, Labview is platform-independent, so the programs created on one platform can easily be ported to other platforms. We installed Labview on a PC; then a bridge between PCI and VME, and an interface standard (VISA) allow the PC to control the VMEbus

4.3.1 PCI-MXI2-VME bridge.

The VME-MXI-2 interface board is a 6U, single-slot VMEbus extender based on MXI-2 technology (i.e. MXIbus). It is possible to install the VME-MXI-2 board in any slot of a VME chassis to be the VMEbus system controller. The VME-MXI-2 extends the VMEbus architecture outside of a VME mainframe via the high-performance MXI-2 cable link. The MXIbus was derived from the VMEbus, and is essentially VME on a cable.

The VME-MXI-2 is a solution for VME systems that need high-performance control of VME using an external computer. With the VME-MXI-2, external computers can control the VME backplane directly. This approach delivers the benefits of an embedded computer, such as high-performance data transfers, shared memory communication, and direct control of the VMEbus, while still maintaining the advantages of an external computer, such as flexibility, a wide selection of performance, and efficient use of only one VMEbus slot. In our set-up (Figure 4-2), we linked the MXIbus with the PCIbus of a standard PC, through a PCI-MXI-2 board that plugs into one of the expansion slots in a PCI-based computer.

FIGURE 4-2. **Experimental set-up**

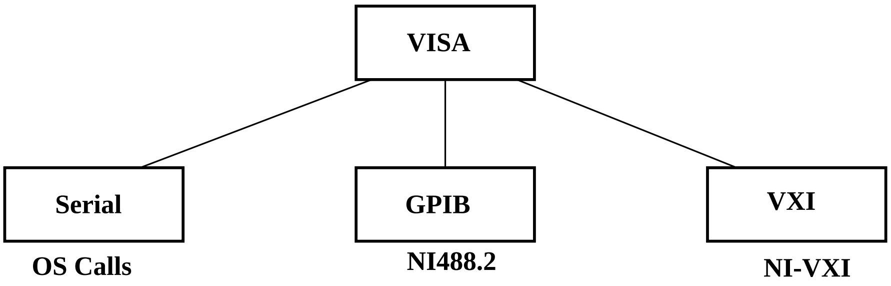


4.3.2 VISA.

The Virtual Instrument Software Architecture (VISA) is an I/O language for instrumentation programming. It is the industry standard for developing instrument drivers. It is a comprehensive package for configuring, programming and troubleshooting instrumentation systems comprised of VXI, VME, PXI, GPIB, and serial interfaces. VISA provides the interface between programming environments such as LabWindows/CVI, and languages such as Labview, C, C++, and Visual Basic. NI-VISA is the National Instruments implementation of the VISA I/O standard, that we use in our testing set-up.

VISA by itself does not provide instrumentation programming capability. VISA is a high-level Application Programming Interface (API) that calls into lower level drivers. The hierarchy of NI-VISA is shown in Figure 4-3.

FIGURE 4-3. **Hierarchy of NI-VISA**



One of VISA’s advantages is that it uses many of the same operations to communicate with instruments regardless of the interface type. For example, the VISA command to write an ASCII string to a message-based instrument is the same whether the instrument is Serial, GPIB, or VXI. Thus, VISA provides interface independence. This can make it easy to switch interfaces and also gives users who must program instruments for different interfaces a single language they can learn. VISA is also designed so that programs written using VISA function calls are easily portable from one platform to another. To ensure this, VISA strictly defines its own data types such that issues like the size, of an integer variable from one platform to another should not affect a VISA program. The VISA function calls and their associated parameters are uniform across all platforms; software can be ported to other platforms and then recompiled. In other words, a C program using VISA can be ported to other platforms supporting C. A Labview program can be ported to other platforms supporting Labview. Another advantage of VISA is that it is an object-oriented language which easily adapts to new instrumentation interfaces as they are developed.

4.3.3 Software description.

The Labview software controlling the VMEbus resides in the PC and consists of several standalone programs. They can be invoked from the main program (called Control Panel). In principle, all of this programs can run in parallel, since each of them uses its own VISA session identifier. However, this is a potential risk, since the cards could be confused, by setting the Control registers in improper way. To access VME registers, the software uses two terms:

- prepared values
- current values

Since the VME master card resides in the PC, the software keeps register and FIFO values in its memory. The software can load data from disk and store them in this local memory (called “prepared” data, in the graphic interface). After writing data to VME, the values will become “current”. So in general, current values show the actual content of VME registers, as achieved during the last write, while prepared values are still in the PC. To change current values one has to first modify the prepared ones and then send them to VME. For read-only registers (RXFIFO or Status register) current values reflect the status after last READ command. So to see the real values one has each time to execute the corresponding READ command first.

Each card has its own Control Panel, which is a program allowing the low level access (i.e. register access). The software has a graphic representation of the FIFOs content, useful to compare data. A more complicate sub-routine (called Test utility) has been developed to carry out complex tests. It allows to generate random input data that emulate the statistics of the data expected in the real Alice experiment (basically adjusting the average occupancy, the number of double and multiple hits, See “Design of the pixel layers” on page 25.).

4.3.4 Execution of the testing with VME access

The testing set-up is sketched in Figure 4-2. The serial output of the Pilot card is connected to the input of the Serlink card. The LabView software generates random data, with a desired statistic distribution, in order to emulate the particle crossing the detector. Three parameters define the distribution: the total number of hits, the percentage of double hits and the percentage of multiple hits. To emulate an event, it is necessary the same amount of data produced by 16 pixel chips, that is 8192 words of 32 bits. Thus the software generates these data an load them in the Pilot TXFIFO. Then it sends the desired sequence of trigger signals (through the VME trigger emulation signals of Figure 3-22), that starts the Pilot activity and the transmission over the 155 MHz link to the Serlink card. Once the Serlink detects the trailer word (see Figure 3-7), it sets a flag and the software reads the Serlink RXFIFO through the VMEbus. The software performs the same encoding algorithm of the Pilot hit encoder, so it can check if the data received by the Serlink card are correct and issue an error message if there are mismatches. Monitoring also the Pilot RXFIFO and the flags of the two cards, the software can identify which section of the system originated the error.

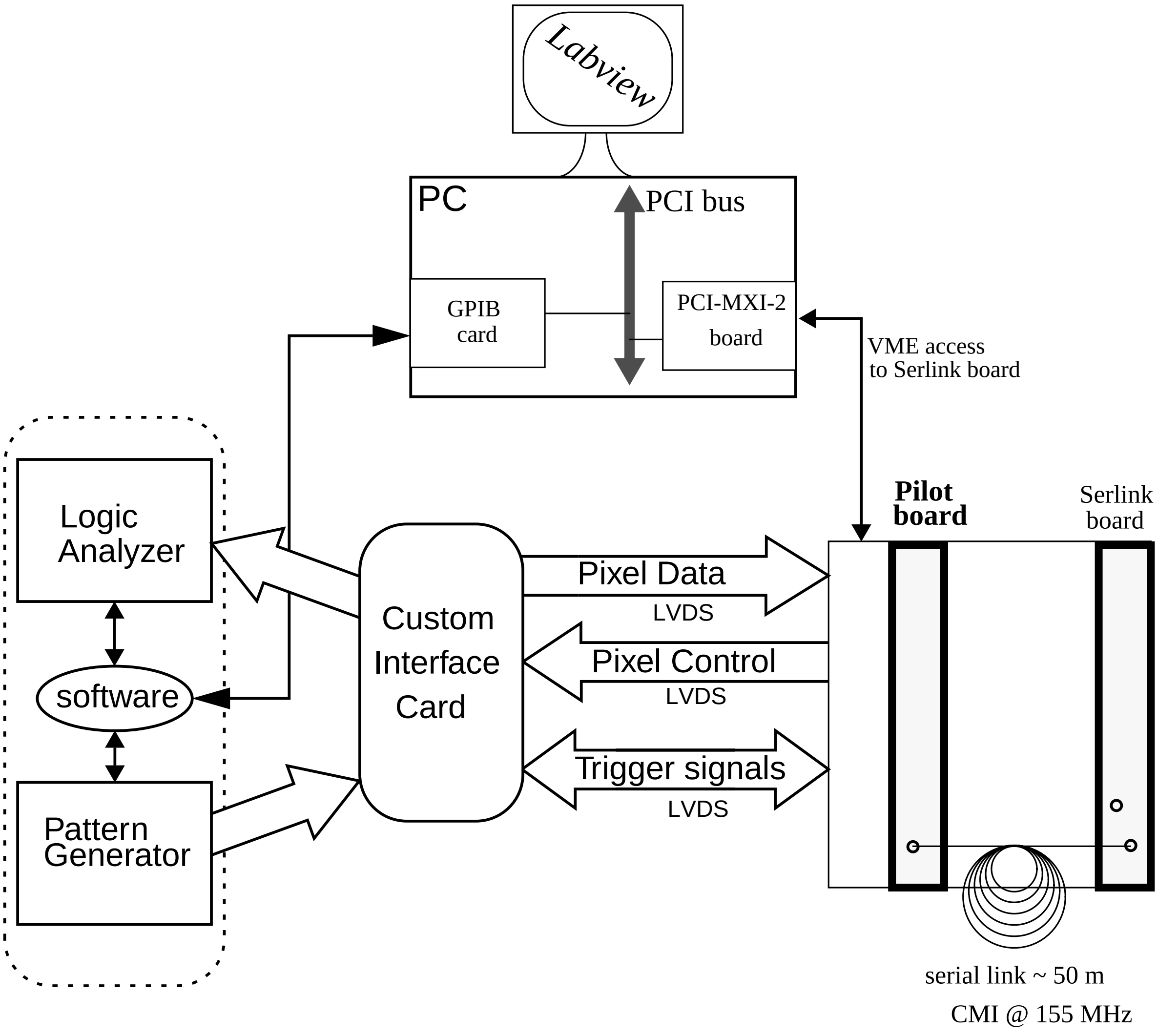
After a careful debugging of the system and an hardware/software co-verification, this testing performed correctly after the emulation of 10^6 events, during several days.

The hardware/software set-up of Figure 4-2 has been used for a preliminary test of the Pilot and Serlink cards. Moreover, with a few adjustments it will be used to test and run the system including the Pixel chips and the silicon detector, once they are produced.

4.4 Testing with Pattern Generator and Logic Analyzer

A testing procedure closer to the real use of the Pilot card requires to exchange data from the front panel and not from the VME bus. In fact the pixel chips will be connected to the front-panel I/Os. The requirement was to build a dedicated set-up for an automatic test of the Pilot card, in view of a small production of these cards. The solution that we adopted foresees to send signals to the front-panel with a pattern generator, and read the Pilot outputs with a logical analyzer. We use the modular logic system HP16500C mainframe including the 16550 logic analyzer, the 16552 pattern generator and a 16533 digitizing oscilloscope. We designed a custom interface card to connect the front-panel connectors with the instrumentation. This card is necessary not only because of the different format and grouping of the signals, but also because the logic system in use cannot manage directly the elaboration of all the interface signals; so some simple pre-elaboration is done by a programmable logic circuit on the interface card. Figure 4-4 shows the situation. A GPIB interface card allows to control the HP16500C mainframe with the LabView software resident on a PC. In this way the same software can control the data and control signals sent to the Pilot card (through the pattern generator), the control signals received by the logic analyzer and the data received by the Serlink card (through the VMEbus).

FIGURE 4-4. **Set-up with the Logic Analyzer/Pattern Generator developed for an automatic testing of the Pilot card production.**



Apart from accessing the Pilot card through the front panel connections, the advantage of this testing procedure is the possibility to check the fast-control output signals (CE, STROBE, NEVR, CLEV, SHIFT_REG_RST, see Pixel Control in Figure 4-4) with the logic analyzer. This check was not possible with the VME testing described in section 4.3.4.

4.5 Conclusions of the testing

The testing activity concluded successfully. The two described set-up allowed to check the correct behaviour of the Pilot card, in agreement with the Verilog simulations carried out during the design of the single programmable devices and of the board. This activity provided some useful feedback on the design of the system.

When the interfaces of the Pilot system with the trigger logic and with the Pixel chip will be confirmed, it will be possible to carry-out a design migration from the FPGA technologies to a CMOS radiation tolerant design of the Pilot logic.

APPENDIX A

Design with Programmable Logic Devices

This appendix surveys commercially available, high-capacity field-programmable devices. It contains a description of the three main categories of PLDs: simple and complex programmable logic devices, and field-programmable gate arrays. Then it gives architectural details and example applications of each type of device.

Programmable logic is loosely defined as a device with configurable logic and flip-flops linked together with programmable resources that control the interconnections. User-programmable memory cells control and define the function that the logic performs and how the various logic functions are interconnected. Thus a wide range of sequential circuits can be implemented on a low-cost PLD. Though various devices use different architectures, all are based on this fundamental idea.

Recently, the development of new types of sophisticated field-programmable devices has dramatically changed the process of designing digital hardware. Unlike previous generations of hardware technology in which board level designs included large numbers of SSI (small-scale integration) chips containing basic gates, virtually every digital design produced today consists mostly of high-density devices. This is true not only for custom devices such as processors and memory but also of logic circuits such as state machine controllers, counters, registers, and decoders. When such circuits are destined for high-volume systems, designers integrate them into high-density gate arrays. However, the high non recurring engineering costs and long manufacturing time of gate arrays make them unsuitable for prototyping or other low-volume scenarios. Therefore, most prototypes and many production designs now use field-programmable devices. The most compelling advantages of field-programmable devices are low start-up cost, low financial risk, and, because the end user programs the device, quick manufacturing turnaround and easy design changes. The field-programmable device market has grown over the past decade to the point where there is now a wide assortment of devices to choose from. To choose a product, designers face the daunting task of researching the best uses of the various chips and learning the intricacies of vendor-specific software. Adding to the difficulty is the complexity of the more sophisticated devices. To help sort out the confusion, we provide an overview of the various programmable device architectures and discuss the most important commercial products, emphasizing devices with relatively high logic capacity.

There are a few major programmable logic families available on the market. Each architecture typically has vendor-specific sub-variants within each type. We will describe later the major types including:

- Simple Programmable Logic Devices (SPLDs)
- Complex Programmable Logic Devices (CPLDs)
- Field Programmable Gate Arrays (FPGAs)

The FPGA term often refers to every kind of programmable devices, as the FPGA architecture was the first one to gain credibility and success (introduced by Xilinx in 1984).

A.1 User-programmable switch technologies

User-programmable switches are the key to user customization of programmable devices. The first user-programmable switch developed was the fuse used in PLAs. Although some smaller devices still use fuses, we will not discuss them here because newer technology is quickly replacing them. For higher density devices, CMOS dominates the IC industry, and different approaches to implementing programmable switches are necessary. For CPLDs, the main switch technologies (in commercial products) are floating gate transistors like those used in EPROM (erasable programmable read-only memory) and EEPROM (electrically erasable

PROM). For FPGAs, they are SRAM (static RAM) and antifuse. Table A.1 lists the most important characteristics of these programming technologies.

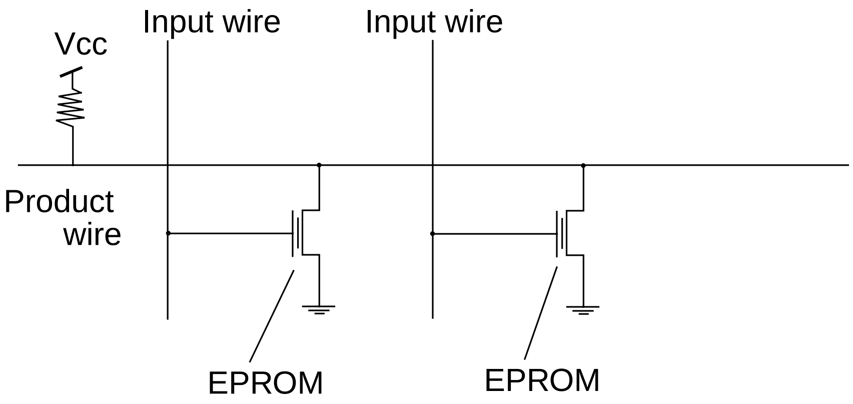
TABLE A.1. **Summary of PLD programming technologies.**

Switch type	Reprogrammable	Volatile	Technology
Fuse	No	No	Bipolar
EPROM	Yes (out of circuit)	No	UVCMOS
EEPROM	Yes (in circuit)	No	EECMOS
SRAM	Yes (in circuit)	Yes	CMOS
Antifuse	No	No	CMOS+
Flash	Yes	No	

A.1.1 EPROM and EEPROM

Electrically Programmable ROM commonly used in SPLD and CPLD are similar to the technology used in standard EPROM memory devices. EPROM cells are electrically programmed in a device programmer. Some EPROM-based devices are erasable using ultra-violet (UV) light if they are in a windowed package. However, most EPROM-based SPLD/CPLDs are in low-cost plastic packaging that cannot be UV erased. The cell of an Electrically-Erasable Programmable ROM is physically larger than an EPROM cell but offers the advantage of being erased electrically.

FIGURE A.1. **EPROM programmable switches**

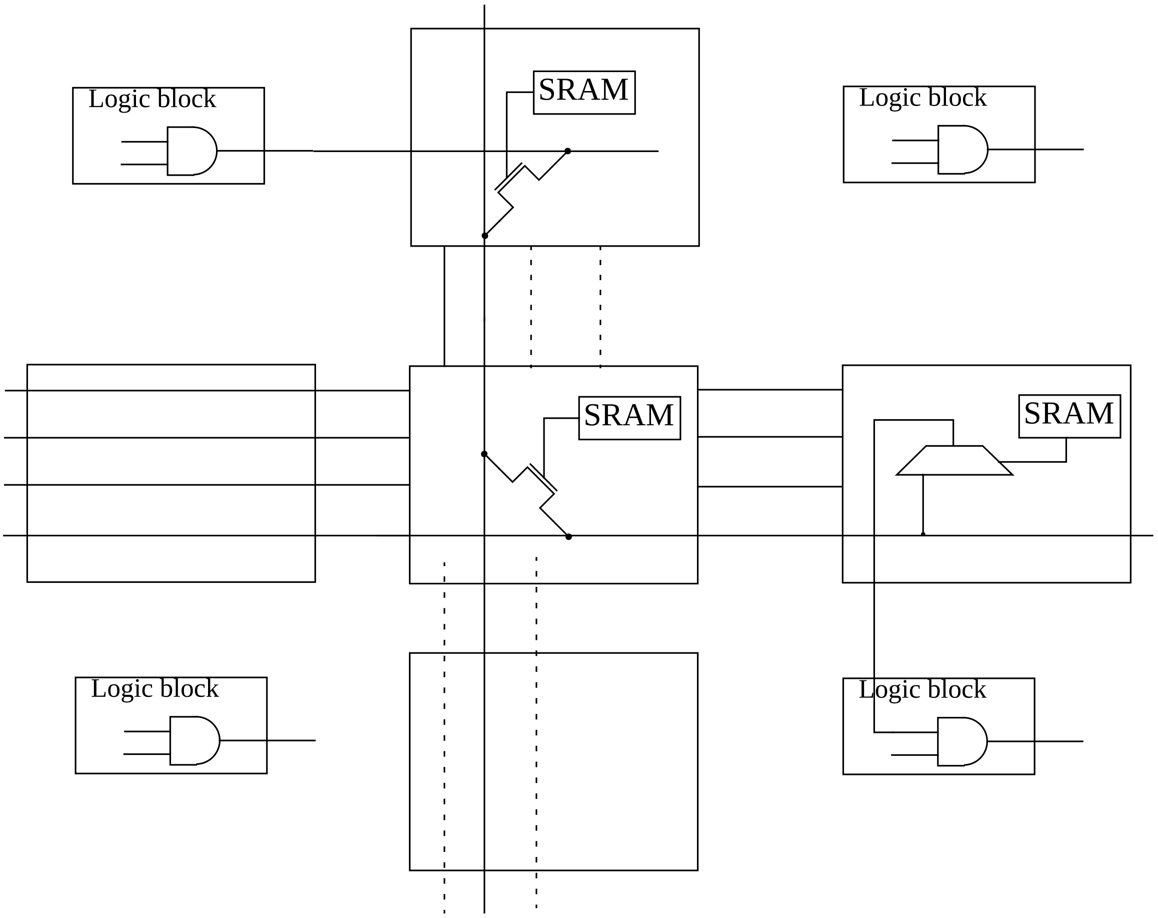


To use an EPROM or EEPROM transistor as a programmable switch for CPLDs (and many SPLDs), the manufacturer places the transistor between two wires to facilitate implementation of wired-AND functions. Figure A.1 shows EPROM transistors connected in a CPLD's AND plane. An input to the AND plane can drive a product wire to logic level 0 through an EPROM transistor, if that input is part of the corresponding product term. For inputs not involved in a product term, the appropriate EPROM transistors are programmed as permanently turned off. The diagram of an EEPROM-based device would look similar to the one in Figure A.1.

A.1.2 Static Memory Technology (SRAM)

SRAM in the form of Look-Up Tables (LUTs) are widely used in FPGAs to implement the programmable combinatorial logic of the device. It is similar to the technology used in static RAM devices but with a few modifications. The RAM cells in a memory device are optimized for fastest possible read/write performance. The RAM cells in a programmable device are usually designed for stability instead of read/write performance. Consequently, RAM cells in a programmable device have a low-impedance connect to VCC and ground to provide maximum stability over voltage fluctuations.

FIGURE A.2. **SRAM-controlled programmable switches**



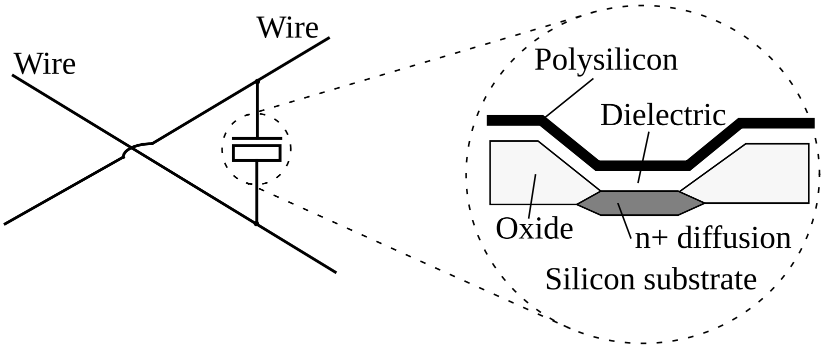
The example of SRAM-controlled switches in Figure A.2 illustrates two applications, one to control the gate nodes of pass-transistor switches and the other, the select lines of multiplexers that drive logic block inputs. The figure shows the connection of one logic block (represented by the AND gate in the upper left corner) to another through two pass-transistor switches and then a multiplexer, all controlled by SRAM cells. Whether an FPGA uses pass transistors, multiplexers, or both depends on the particular product.

Because static memory is volatile (i.e.-the contents disappear when the power is turned off), SRAM-based devices are “booted” after power-on. This makes them in-system programmable and re-programmable, even in real-time. As a result, SRAM-based FPGAs are common in reconfigurable computing applications where the device’s function is dynamically changed. The configuration process typically requires only a few hundred milliseconds at most. Most SRAM-based devices can boot themselves automatically at power-on much like a microprocessor. Furthermore, most SRAM-based devices are designed to work with either standard byte-wide PROMs or with sequential-access serial PROMs. SRAM cells are also used in many non-volatile CPLDs to hold some configuration bits to reduce internal capacitive loading.

A.1.3 Anti-fuse

Antifuses are originally open circuits that take on low resistance only when programmed. They are manufactured using modified CMOS technology. As an example, Figure A.3 depicts Actel’s PLICE (programmable-logic interconnect circuit element), antifuse structure [68]. The antifuse, positioned between two interconnect wires, consists of three sandwiched layers: conductors at top and bottom and an insulator in the middle. Unprogrammed, the insulator isolates the top and bottom layers; programmed, the insulator becomes a low-resistance link. PLICE uses polysilicon and n+ diffusion as conductors and a custom-developed compound, ONO (oxide-nitride-oxide [68]), as an insulator. Other antifuses rely on metal for conductors, with amorphous silicon as the middle layer [69].

FIGURE A.3. **Actel’s PLICE antifuse structure**



They are usually physically quite small and have low on-resistance and capacitance. Consequently, anti-fuse technology has benefits for creating fast, low power programmable interconnect. However, they require large programming transistors on the device. With some modifications in the process, anti-fuse technology can be made radiation tolerant (Actel).

A.1.4 FLASH technology

FLASH-erased (or bulk erased) electrically-erasable programmable read-only memory. FLASH has the electrically-erasable benefits of EEPROM but the small, economical cell size of EPROM technology.

A.2 Families of PLDs

Before describing the major PLD families we define two concepts coming from the technology used in the devices, and useful to further classify the different PLDs.

One-Time and In-System Programmability

A *one-time programmable device* can only be programmed once. Once programmed, it cannot be re-programmed. All fuse and anti-fuse-based devices are one-time programmable. EPROM-based devices in plastic packages are one-time programmable.

Though definitions vary slightly, *in-system programmability* (ISP, also called in-circuit programmability) means that the device can be programmed and reprogrammed while it is mounted on the circuit board with the other components. This has the big advantage of reducing the time-to-market. Moreover, it allows to debug and upgrade the system “in the field”. Most ISP devices are built using SRAM, EEPROM, or FLASH technologies. All SRAM-based devices are inherently in-system programmable because they must be configured on power-up. EEPROM and FLASH processes are erasable technologies. However, not all EEPROM- and FLASH-based are programmable while soldered on the circuit board. In-system programmability (ISP) requires special on-chip programming logic. Not all CPLDs have this logic, even if they are built with EEPROM and FLASH. Those lacking the on-chip programming circuitry are programmed and erased in a device programmer. Those with the programming circuitry either use a vendor proprietary interface or a JTAG (IEEE 1149.1) standard interface to re-program the device.

A.2.1 SPLD - Simple Programmable Logic Device

They include several kind of architectures:

- PLA (Programmable Logic Array)
- PAL (Programmable Array Logic, Vantis)
- GAL (Generic Array Logic, Lattice)

and they are sometime simply referred simply as PLD.

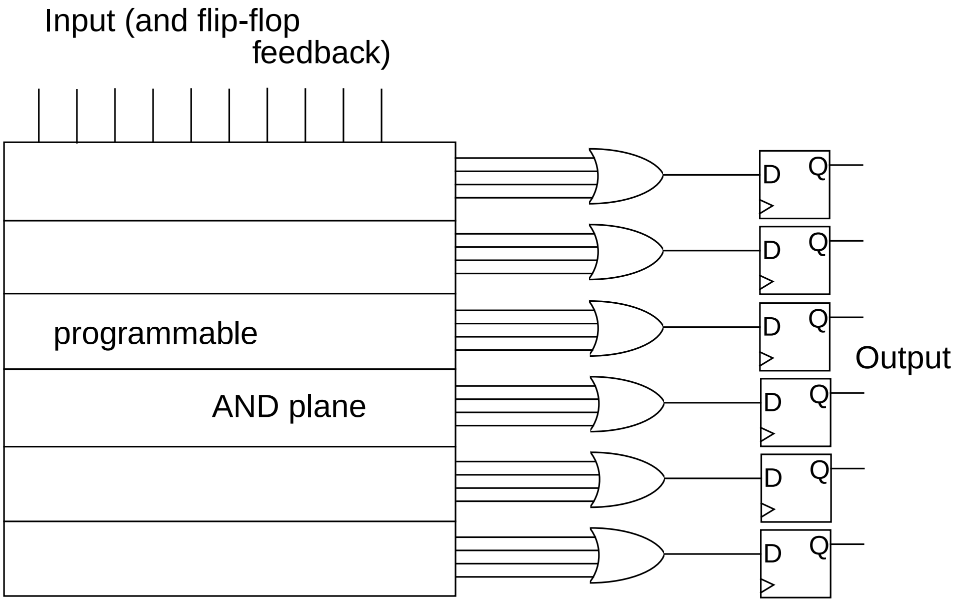
The first device developed specifically for implementing logic circuits was the field-programmable logic array, or simply PLA for short. A PLA consists of two levels of logic gates: a programmable, wired-AND plane followed by a programmable, wired OR plane. A PLA’s structure allows any of its inputs (or their complements) to be AND-ed together in the AND plane; each AND plane output can thus correspond to any product term of the inputs. Similarly, users can configure each OR plane output to produce the logical sum of any AND plane output. With this structure, PLAs are well-suited for implementing logic functions in sum-of-products form. They are also quite versatile, since both the AND and OR terms can have many inputs (product literature often calls this feature “wide AND and OR gates”).

When Philips introduced PLAs in the early 1970s, their main drawbacks were expense of manufacturing and somewhat poor speed performance. Both disadvantages arose from the two levels of configurable logic; programmable logic planes were difficult to manufacture and introduced significant propagation delays. To overcome these weaknesses, Monolithic Memories (MMI, later merged with Advanced Micro Devices) developed PAL devices. As Figure A.4 shows, PALs feature only a single level of programmability: a programmable, wired-AND plane that feeds fixed-OR gates. To compensate for the lack of generality incurred by the fixed-OR plane, PALs come in variants with different numbers of inputs and outputs and various sizes of OR gates. To implement sequential circuits, PALs usually contain flip-flops connected to the OR gate outputs. A slice of a SPLD including a flip-flop, an OR-gate and the corresponding portion of the AND plane can be referred as *macrocell*.

The introduction of PAL devices profoundly affected digital hardware design, and they are the basis of some of the newer, more sophisticated architectures that we will describe shortly. Variants of the basic PAL architecture appear in several products known by various acronyms.

The category of SPLD is the smallest and consequently the fastest and least-expensive form of programmable logic. An SPLD is typically comprised of four to 22 macrocells and can typically replace a few 7400-series TTL devices. Each of the macrocells is typically fully connected to the others in the device. Most SPLDs use either fuses or non-volatile memory cells such as EPROM, EEPROM, or FLASH to define the functionality.

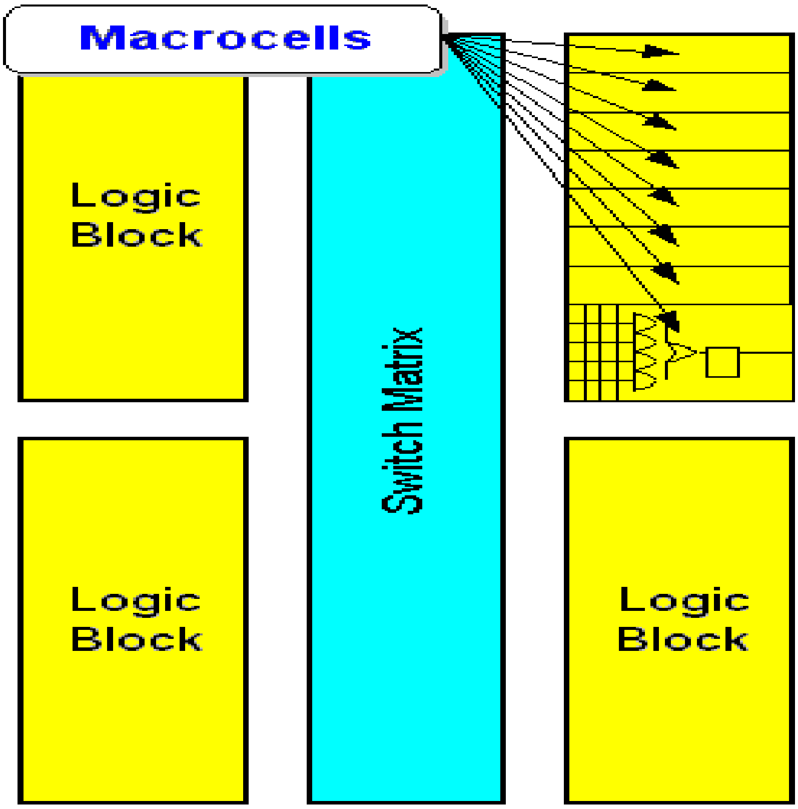
FIGURE A.4. **PAL structure**



A.2.2 CPLD - Complex Programmable Logic Device

Advances in technology have produced devices with higher logic capacities than SPLDs. The difficulty with increasing a strict SPLD architecture's capacity is that the programmable-logic plane structure grows too quickly as the number of inputs increases. The only feasible way to provide large-capacity devices based on SPLD architectures is to programmably interconnect multiple SPLDs on a single chip. Many programmable device products on the market today have this basic structure and are known as complex programmable-logic devices (CPLD, Figure A.5). However, CPLD products can be much more sophisticated than SPLDs, even at the level of their basic SPLD-like blocks (that are sometime called logic block). A typical CPLD is the equivalent of two to 64 SPLDs, where each SPLD typically contains 4 to 16 macrocells[1].

FIGURE A.5. CPLD architecture



In some architectures, when the number of product terms required exceeds the number available in the macrocell, additional product terms are borrowed from an adjoining macrocell. This

[1] A CPLD's macrocell can have a structure different from the simple block described in "SPLD - Simple Programmable Logic Device" on page 7. Typically supports four to sixteen product terms with many inputs, but the complexity of the logic function is limited. Compare this to most FPGA logic blocks where the complexity is unlimited (as the combinatorial logic is replaced by a Look-Up Table), but the logic function has just four inputs.

makes the CPLD device useful for a wider variety of applications. Borrowed product terms usually means increased propagation delay.

The CPLD's logic blocks are partly or fully interconnected via a programmable switch matrix; this is vendor and family specific. With a fully populated (100% connection) switch matrix, a design will route even with a majority of the device resources used and with fixed I/O pin assignment. Generally, the delays within a fully populated switch matrix are fixed and predictable, while the delays within a partially-populated (less that 100% connection) switch matrix are not fixed and less easily predicted, similar to most FPGA devices. A device with a partially-populated switch matrix may have problems routing complex designs and there is a chance that the device will not route. Also, it may be difficult to make design changes in these devices without using a different pinout. Routing to a fixed pinout is important: it is far easier to change the program of a PLD than it is to re-layout a circuit board. Though a partially populated switch matrix has some potential limitations, it is less expensive to manufacture.

Generally, CPLDs are CMOS and use non-volatile memory cells such as EPROM, EEPROM, or FLASH to define the functionality. Many of the most-recently introduced CPLD families use a EEPROM or FLASH and have been designed so that they can be programmed in-circuit.

CPLD applications.

Their high speeds and wide range of capacities make CPLDs useful for many applications, from implementing random glue logic to prototyping small gate arrays. An important reason for the growth of the CPLD market is the conversion of designs that consist of multiple SPLDs into a smaller number of CPLDs. CPLDs can realize complex designs such as graphics, LAN, and cache controllers. As a rule of thumb, circuits that can exploit wide AND/OR gates and do not need a large number of flip-flops are good candidates for CPLD implementation. Finite state machines are an excellent example of this class of circuits. A significant advantage of CPLDs is that they allow simple design changes through reprogramming (all commercial CPLD products are reprogrammable). In-system programmable CPLDs even make it possible to reconfigure hardware (for example, change a protocol for a communications circuit) without powering down. Designs often partition naturally into the SPLD-like blocks in a CPLD, producing more predictable speed performance than a design split into many small pieces mapped into different areas of the chip. Predictability of circuit implementation is one of the strongest advantages of CPLD architectures.

Altera pioneered CPLDs, first in their Classic EPLD chips, and then in the MAX 5000, 7000, and 9000 series. Because of a rapidly growing market for large field-programmable devices, other manufacturers developed CPLD devices, and many choices are now available. CPLDs provide logic capacity up to the equivalent of about 50 typical SPLD devices, but extending these architectures to higher densities is difficult. Building programmable devices with very high logic capacity requires a different approach.

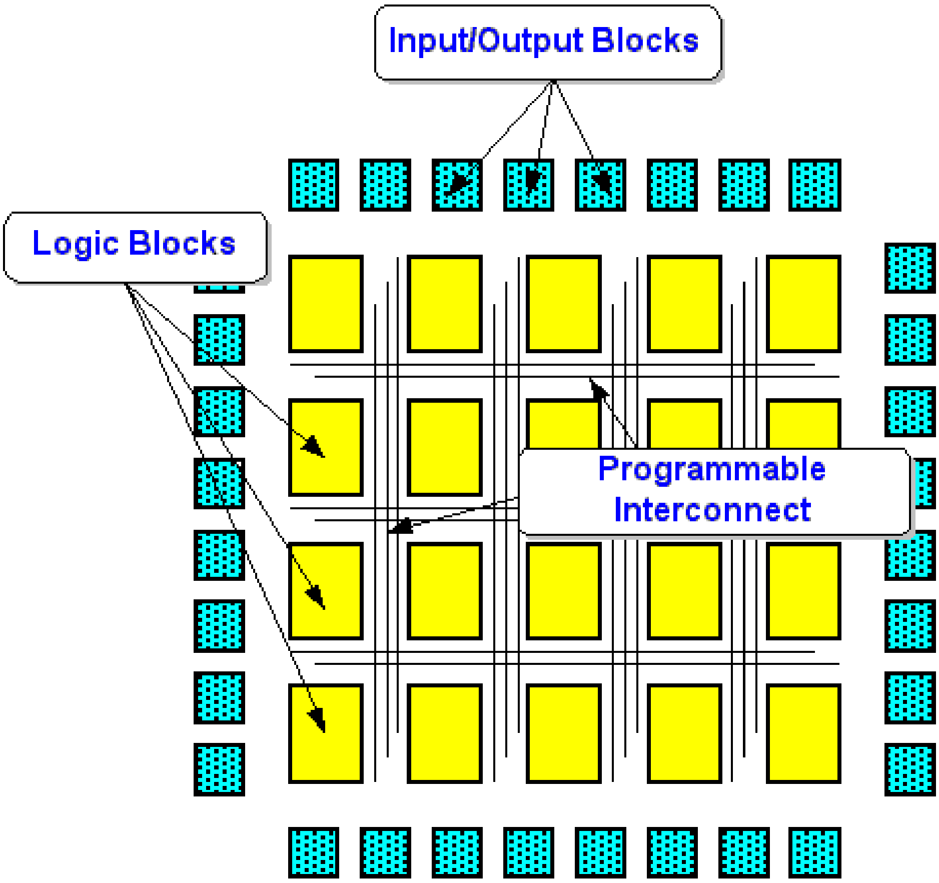
A.2.3 FPGA - Field Programmable Gate Array

The highest capacity general-purpose logic chips available today are the traditional gate arrays sometimes referred to as mask-programmable gate arrays (MPGA). An MPGA consists of an array of prefabricated transistors customized for the user's logic circuit by means of wire connections. Because the silicon foundry performs customization during chip fabrication, the manufacturing time is long, and the user's setup cost is high. Although MPGAs are clearly not field-programmable devices, we mention them here because they motivated the design of the field-programmable equivalent, FPGAs. Like MPGAs, an FPGA consists of an array of uncommitted circuit elements (logic blocks) and interconnect resources, surrounded by I/O blocks, but the end user configures the FPGA through programming. Figure A.6 shows a typical FPGA architecture. As the only type of field-programmable device that supports very high logic capacity, FPGAs have engendered a major shift in digital-circuit design.

A typical FPGA contains from 64 to tens of thousands of logic blocks and an even greater number of flip-flops. Most FPGAs do not provide 100% interconnect between logic blocks (to do so would be prohibitively expensive). Instead, sophisticated software places and routes the logic on the device much like a PCB autorouter would place and route components.

Although no technical reason prevents application of EPROM or EEPROM to FPGAs, current commercial FPGA products use either SRAM or antifuse technologies. Currently, the highest-density FPGAs are built using static memory (SRAM) technology, similar to microprocessors.

FIGURE A.6. **FPGA architecture.**



SRAM-based devices are inherently re-programmable, even in-system. However, they require some form of external configuration memory source. The configuration memory holds the program that defines how each of the logic blocks functions, which I/O blocks are inputs and outputs, and how the blocks are interconnected together. The FPGA either self-loads its configuration memory (from a PROM) or an external processor downloads the memory into the FPGA. The configuration time is typically less than 200 ms, depending on the device size and configuration method.

The other common process, the anti-fuse technology has benefits for more plentiful programmable interconnect. Anti-fuse devices are one-time programmable. Once programmed, they cannot be modified, but they also retain their program when the power is off, so no external EPROMs are required. Anti-fuse devices are programmed in a device programmer either by the end user or by the factory or distributor.

FPGA can be classified also according to the architectures:

- coarse-grained architectures
- fine-grained architectures.

Coarse-grained architectures consist of fairly large logic blocks, often containing two or more look-up tables and two or more flip-flops. In a majority of these architectures, a four-input look-up table (think of it as a 16x1 ROM) implements the actual logic. The larger logic block usually corresponds to improved performance.

TABLE A.2. **FPGA Architectures and Processes.**

Architecture	Static Memory	Anti-Fuse	Flash
Coarse-grained	Altera: FLEX, APEX Atmel: AT40K DynaChip Lucent: ORCA Vantis: VF1 Xilinx: XC3000, XC4000, Spartan, Virtex	QuickLogic: pASIC	
Fine-grained	Actel: SPGA, ProASIC Atmel: AT6000	Actel: ACT	Gatefield

The other architecture type is called fine-grained. In these devices, there are a large number of relatively simple logic blocks. The logic block usually contains either a two-input logic function or a 4-to-1 multiplexer and a flip-flop. These devices are good at systolic functions and have some benefits for designs created by logic synthesis.

Some FPGAs have system-level features built-in, like on-chip bussing, on-chip RAM for building small register files or FIFOs, and built-in JTAG boundary-scan support.

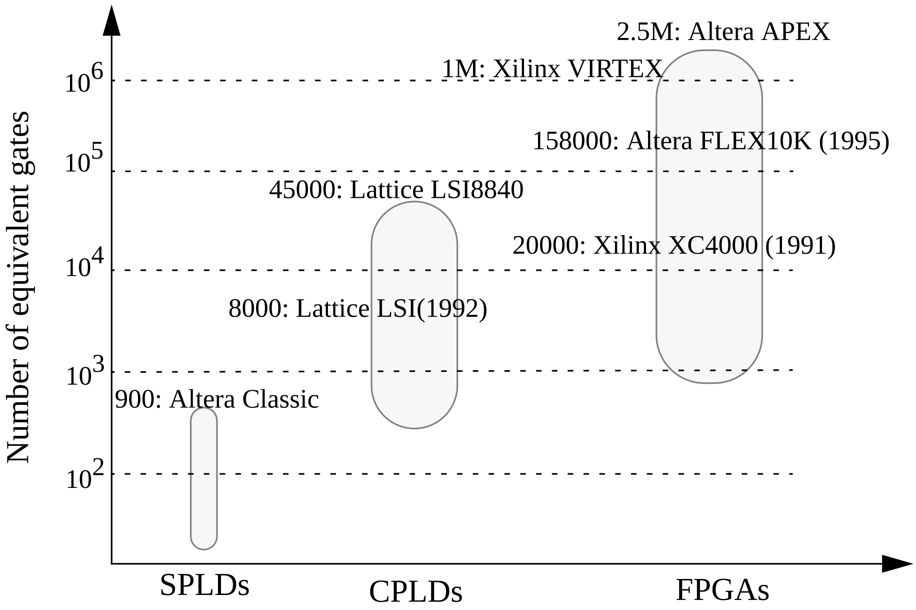
FPGA applications

FPGAs have gained rapid acceptance over the past decade because users can apply them to a wide range of applications: random logic, integrating multiple SPLDs, device controllers, communication encoding and filtering, small- to medium-size systems with SRAM blocks, and many more. Another interesting FPGA application is prototyping designs to be implemented in gate arrays by using one or more large FPGAs. (A large FPGA corresponds to a small gate array in terms of capacity). An application only beginning development is the use of FPGAs as custom computing machines. This involves using the programmable parts to execute software, rather than compiling the software for execution on a regular CPU.

As mentioned earlier, pieces of designs often map naturally to the SPLD-like blocks of CPLDs. However, designs mapped into an FPGA break up into logic-block-size pieces distributed through an area of the FPGA. Depending on the FPGA's interconnect structure, the logic block interconnections may produce delays. Thus, FPGA performance often depends more on how CAD tools map circuits into the chip than does CPLD performance.

Figure A.7 illustrates the logic capacities available in each field-programmable device category. "Equivalent gates" refers loosely to the number of two-input NAND gates. The chart serves as a guide for selecting a device for an application according to the logic capacity needed. However, as we explain later, each type of field-programmable device is inherently better suited for some applications than for others. There are also special-purpose devices optimized for specific applications (for example, state machines, analog gate arrays, large interconnection problems). Since such devices have limited use, we do not describe them here.

FIGURE A.7. **PLD logic capacities**

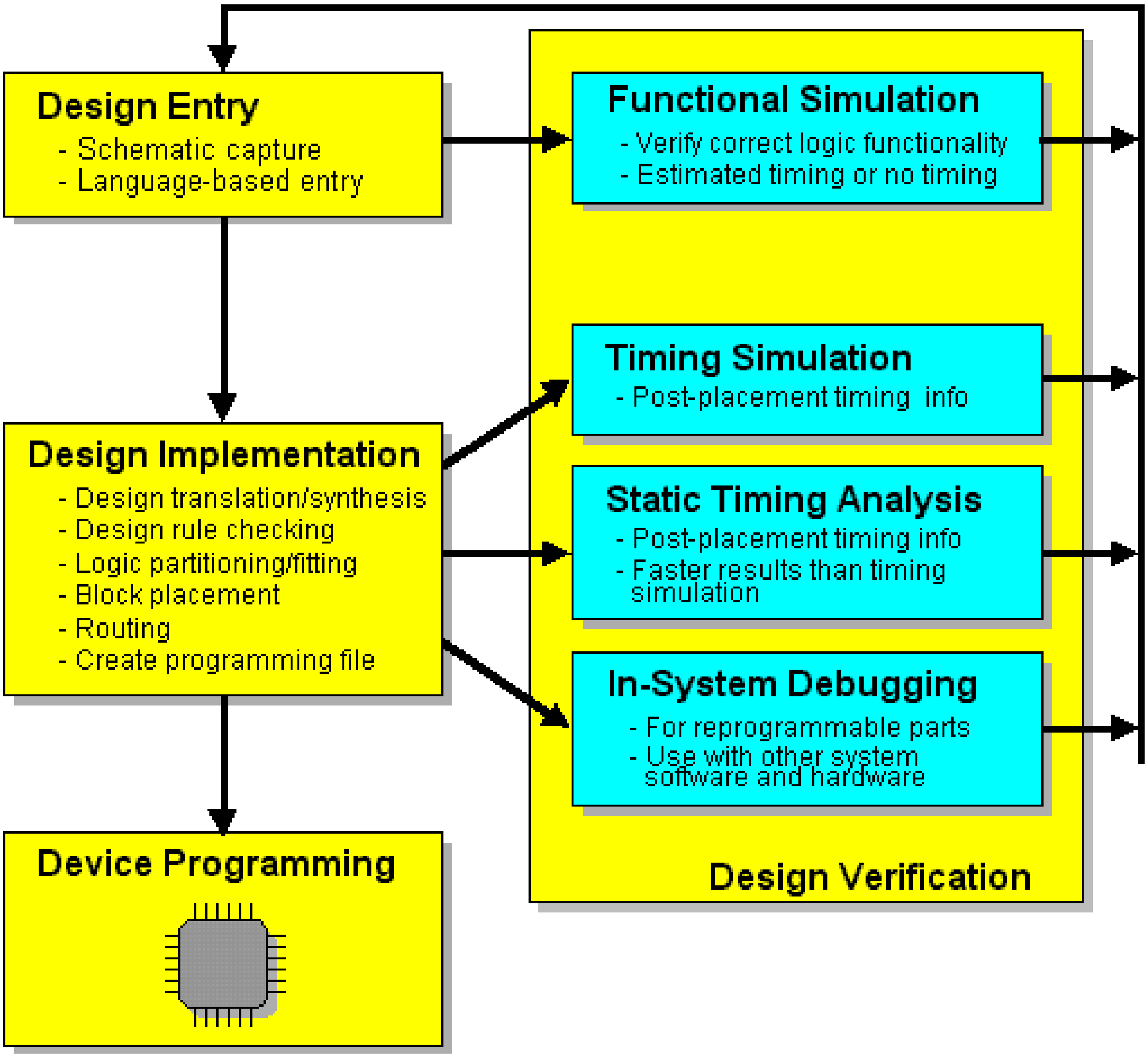


Data of year 1999, unless differently stated.

A.3 Design with programmable logic

A typical programmable logic design involves three steps showed in the left side of Figure A.8.

FIGURE A.8. **The main steps in the design of FPGAs**


Design Entry
- Schematic capture
- Language-based entry
Design Implementation
- Design translation/synthesis
- Design rule checking
- Logic partitioning/fitting
- Block placement
- Routing
- Create programming file
Device Programming
Functional Simulation
- Verify correct logic functionality
- Estimated timing or no timing
Timing Simulation
- Post-placement timing info
Static Timing Analysis
- Post-placement timing info
- Faster results than timing simulation
In-System Debugging
- For reprogrammable parts
- Use with other system software and hardware
Design Verification

A.3.1 Design Entry

A variety of available tools are available to accomplish the design entry step (as schematic, using a hardware description language or a combination of these methods). Traditionally, schematic-based tools provided experienced designers more control over the physical placement and partitioning of logic on the device. However, this extra tailoring took time. Likewise, language-based tools allowed quick design entry but often at the cost of lower performance or density. Synthesis for language-based designs has significantly improved in the last few years, especially for FPGA design. In either case, learning the architecture and the tool helps you to create a better design. Technology-ignorant design is possible, but at the expense of density and performance.

Cores, such as a PCI bus interface or a DMA controller, are a new and increasingly important addition to the programmable logic world. These cores are predefined functions specifically implemented and verified in programmable logic. Cores have been available for gate arrays for quite some time. Now that FPGA devices push beyond the 100,000 gate density level, cores should become a popular design entry tool for programmable logic users. Core help reduce design and verification time on commonly used functions.

A.3.2 Design Implementation

After the design is entered using schematic capture or synthesized, it is ready for implementation on the target device. This task is usually performed by CAD tools.

Computer-aided design programs are essential in designing circuits for implementation in field-programmable devices. Such software tools are important not only for CPLDs and FPGAs, but also for SPLDs. A typical CAD system for SPLDs or CPLDs includes software for the following tasks: initial design entry, logic optimization, device partitioning (for CPLDs), device fitting, simulation, and configuration. Since initial logic entry is not usually in an optimized form, the system applies algorithms to optimize the circuits. Then additional algorithms analyse the resulting logic equations and fit them into the SPLD. Simulation verifies correct operation, and the designer returns to the design entry step to fix errors. When a design simulates correctly, the designer loads it into a programming unit to configure an SPLD. In most CAD systems, the designer performs the original design entry step manually, and all other steps are automatic.

The FPGA design process (block “Design Implementation in Figure A.8) is similar to that of CPLDs but requires additional tools to support increased chip complexity. The major difference is in device fitting, for which FPGAs need at least three tools: a technology mapper to transform basic logic gates into the FPGA’s logic blocks, a placement tool to choose the specific logic blocks, and a router to allocate wire segments to interconnect the logic block. The last two operations are strictly connected and are often referred as Place&Route. Their primary goal is to reduce the amount of routing resources required and to maximize system performance. This is a compute intensive operation for FPGAs and larger CPLDs. The implementation software monitors the routing length and routing track congestion while placing the blocks. In some systems, the implementation software also tracks the absolute path delays in order to meet user-specified timing constraints. With this added complexity, the CAD tools take a fairly long time (often more than an hour) to complete their tasks. When the placement and routing process is complete, the software creates the binary programming file used to configure the device. In large or complex applications, the software may not be able to successfully place and route the design. Some packages allow the software to try different options or to run many iterations in an attempt to obtain a fully-routed design. A good design rule is to use less than 85% of the available device resources. This technique provides the software extra resources to help route the design. Also, some vendors supply floorplanning tools to aid in physical layout. Layout is especially important for larger FPGAs because some tools have problems recognizing design structure. A good floorplanning tool allows the designer to convey this structure to the place and route software. In order to reach high performances or density some part of a design can be placed and routed manually.

A.3.3 Verification

Design verification occurs at various levels and steps throughout the design. There are a few fundamental types of verification as applied to programmable logic. Functional simulation is performed in conjunction with design entry, but before place and route, to verify correct logic functionality. Full timing simulation must wait until after the place and route step. After place and route, the software back-annotates the logic and routing delays to the netlist for simulation. One successful technique for programmable logic design is to functionally simulate the design to guarantee proper functionality, verify the timing using a static timing calculator, and then verify complete functionality by testing the design in the system.

Some of the device vendors supply additional in-system debugging capabilities. For example, Xilinx ships a small pod called an XChecker cable that connects to your serial port and allows quick downloading of a design. With a few simple additions to your design and board, the XChecker cable is capable of stopping or single-stepping the clock and can read back the state of internal flip-flops. Likewise, Actel’s Action Probes provide access to internal nodes within their anti-fuse based FPGAs.

A.3.4 Device Programming

After creating a programming file, the programmable device is configured and ready for action. The actual programming method depends on the target technology. Most programmable logic technologies, including the PROMs for SRAM-based FPGAs, require some sort of a device programmer. In-system programmable devices, including SRAM-based FPGAs, may not require a physical programmer but do require some intelligent system resource to download the programming file into the device. This is performed with a microprocessor, micro-controller, or via a JTAG test port.

A.4 High-speed design techniques

With the development of new types of sophisticated programmable logic devices, such as Complex PLDs and FPGAs, the process of designing digital hardware has changed dramatically over the past few years. The number of applications for large PLDs has grown so rapidly that many companies have produced competing products and there is now a wide assortment of devices to choose from. A designer who is not familiar with the various products faces a daunting task in order to discover all of the different types of chips, try to understand what they can best be used for, choose a particular company’s device, and then design the hardware.

The purpose of this paragraph is to give an overview of the practical issues that face designers who wish to implement circuits in today’s sophisticated CPLDs and FPGAs. Issues facing designers who wish to use PLDs are fairly straightforward when applications are relatively small. For this reason, our focus is on the most demanding class of applications that require state-of-the-art speed-performance on large PLDs.

Speed-performance achievable for a given application circuit is greatly affected by which category of chips is selected. More specifically, circuits that require fairly wide gates (such as state machines or decoders) almost always operate faster in CPLDs [70]. Even within a single category of device, products from different manufacturers (or even the same manufacturer) can result in significant differences in performance. It is important to note that such subtleties can be appreciated only through experience with the devices. PLD marketing literature often gives the impression that a certain level of performance is available for a wide range of application circuits; the reality is that maximum performance can be obtained only for applications that are well-matched to the PLD architecture. A corollary is that while today's CAD tools are sophisticated enough to map fairly abstract descriptions of a circuit into a PLD, maximum performance will only be obtained for circuits that are described in a way that provides an obvious mapping from the circuit description into the device. As an example of how PLD architecture affects speed-performance of applications, consider a generic finite state machine. If a finite state machine is to be implemented in an FPGA, then the amount of logic feeding each state machine flip-flop must be minimized. This follows because in FPGAs flip-flops are directly fed by logic blocks that have relatively few inputs (typically 4 - 8). If the state machine flip-flops are fed by more logic than will fit into a single logic block, then multiple levels of logic blocks will be needed, and speed-performance will decrease. For this reason, designers usually use "one-hot" state machine encoding when targeting FPGAs, so that the amount of logic that sets each flip-flop is minimized. Even in a CPLD architecture, speed-performance of a state machine can be significantly affected by state bit encoding; for example, in the Altera MAX 7000 CPLDs, flip-flops that are fed by five or fewer product terms will operate faster than those that require more than five terms. In general, designers who wish to obtain maximum performance for applications need to constantly consider the nuances of their PLD's architecture.

The FPGA structure limits the wiring between logic cells to a small number of lines, while other technologies do not have such constraints. The wiring in an FPGA is also peculiar as adds large delays to the logic. Even the cell architecture demands an unusual design style, and that style varies between the different FPGAs. This means that FPGA designs done using traditional styles won't perform well. In order to design a fast CPLD/FPGA circuit some rules exist:

- Minimizing logic cell fan out
- Duplicating logic in critical paths
- Pipelining the design
- Using decoded state machines
- Hand-crafting critical circuits
- Floorplanning the design
- Tailoring to the specific architecture
- Keeping the logic utilization under 85% (it eases the Place&Route)

We go back to discuss briefly the *design entry* methods. There has been an on-going battle between schematic entry and hardware description languages (HDLs). We will summarize the main point of each design entry.

HDLs allow an high-level description (and simulation) of the circuit, thus reducing the time spent for development and design changes. Moreover, HDLs are standards and are technology independent, allowing the portability of the code. On the other hand the HDL synthesis tools are still strongly evolving, and not always fully reliable.

The schematic entry describes the circuit at a lower level. This implies a longer time for the design and the correction, but it provides the designer more control on the logic implementation and especially on the P&R. Tipically it is non-portable between different technologies (as it make use of libraries), but there is an on-going effort toward its portability (Library of Parametrized Modules, schematic editors generating HDL descriptions). For high-performances and density issues, schematics done by an experienced designer are still better than an HDL description.

A.4.1 Synchronous design

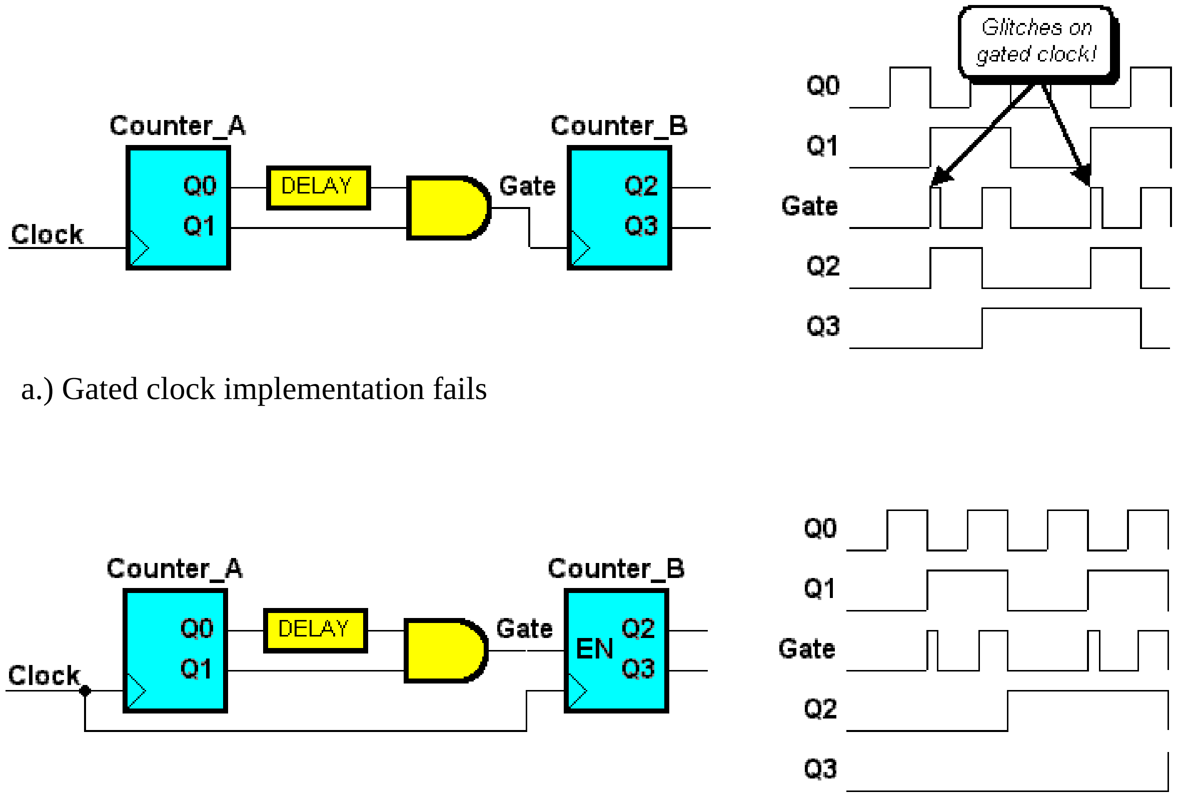
The basic rule in order to obtain a safe, reliable and easy-to-test design is to keep it strictly synchronous. This means that *every* flip-flop of the circuit is clocked by the same (low-skew) signal. The flexibility of FPGA and CPLD might tempt unwary designers into developing bad asynchronous habits, for example using gated clocks. Because these devices all use programmable interconnect, different signal arrival times coupled with asynchronous logic invite a digital disaster.

Synchronous designs are inherently easier to simulate and debug than asynchronous designs. Designers can simply predict the behaviour of synchronous systems and model them in simulation. The analysis comes down to the worst-case path between clock edges: a simple process, especially using a static timing analyzer. But the standard simulation tools are not reliable for asynchronous circuits, that require the analysis of all combinations of best- and worst-case signal paths over temperature, voltage and process variations: a far more onerous task. Consequently, synchronous designs work with a much wider variation in device timing parameters and over a broader temperature range than do most asynchronous designs. As process technology improves, circuit delays decrease. A vendor might ship faster devices that meet all datasheet specifications but behave differently than the old part. Problems with asynchronous logic usually don’t appear until the board is done: one batch of parts works fine, another batch fails the system test, while another batch seemingly fails intermittently depending on operating conditions.

One problem most vulnerable to process changes and common in programmable logic (and gate-array) designs is the gated clock. Glitches on a gated clock often result from differences in arrival times for its input signals caused by routing delays inside a device. Further, gated clocks introduce additional delay in the clock path, which increases clock-to-output times and might introduce hold-time problems.

The example in Figure A.9.a, presents a case where a gated input drives the Clock input of the second counter: a bad situation. The difference in routing delays causes glitches on the AND gate during specific states on Counter_A. These glitches cause incorrect clocking for Counter_B. A better implementation appears in Figure A.9.b, where both counters operate from the same clock source. The terminal count from Counter_A drives a Clock Enable signal on Counter_B. The clock path is clean and unfettered with asynchronous gates.

FIGURE A.9. **Gated clocks often have glitches due to differences in signal arrival times (a). Changing to a synchronous solution guarantees success (b).**



a.) Gated clock implementation fails

b.) Fully synchronous version works, despite glitches

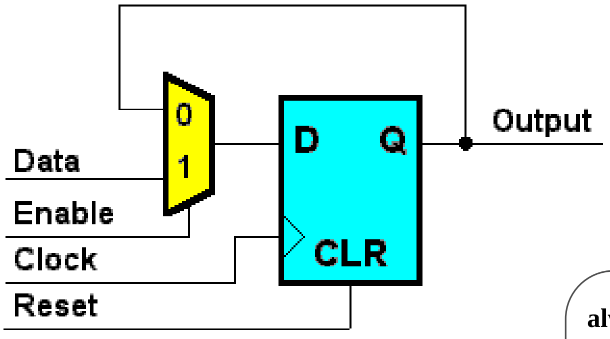
Taking into account this example, avoiding gated clocks seems a trivial advice. But all of the following situations involve a glitch-prone path through some combinatorial logic. For each circuit, there is an alternative, fully-synchronous implementation, usually involving a flip-flop with a Clock Enable input. Consider these common situations:

- Clocks derived from the Terminal Count of a counter. Most circuits generate Terminal Count signals with an AND gate. This situation is common in applications where a high-speed clock is divided down and redistributed within a device.
- Clocks derived from a decoder. Many designs commonly use this logic to load various banks of flip-flops.
- Clocks derived from a multiplexer output. This situation is common in applications that select various clock frequencies.

The solution in example of Figure A.9, and many others, includes a flip-flop with a Clock Enable input. In some FPGA devices, such as those from Xilinx and Lucent, the internal flip-flops have a built-in Clock Enable input. In other devices, you can easily build a clock input by including a 2:1 multiplexer in front of the data input (Figure A.10). If designing circuits with VHDL or Verilog, note that not all synthesis packages automatically create a flip-flop using the built-in enable, even if one is available in the target technology.

FIGURE A.10. **Some FPGAs have flip-flops with built-in clock enables, some don’t. Building an equivalent solution is simple as implemented in schematic (a), VHDL (b) and Verilog (c).**

a.) Schematic implementation



b.)VHDL code fragment

```
architecture RTL of FlipFlop is
          begin
            process (Reset, Clock)
            begin
              if (Reset = ’1’) then
                Output <= ’0’;
              elsif Clock’event and Clock = ’1’ then
                if (Enable = ’1’) then
                  Output <= Data;
                end if;
              end if;
            end process;
          end RTL;
```

c.)Verilog code fragment

```
always @ (posedge Clock or posedge Reset)
          begin
            if (Reset)
              begin
                Output = 0;
              end
            else
              begin
                if (Enable)
                  Output = Data;
              end
          end
```

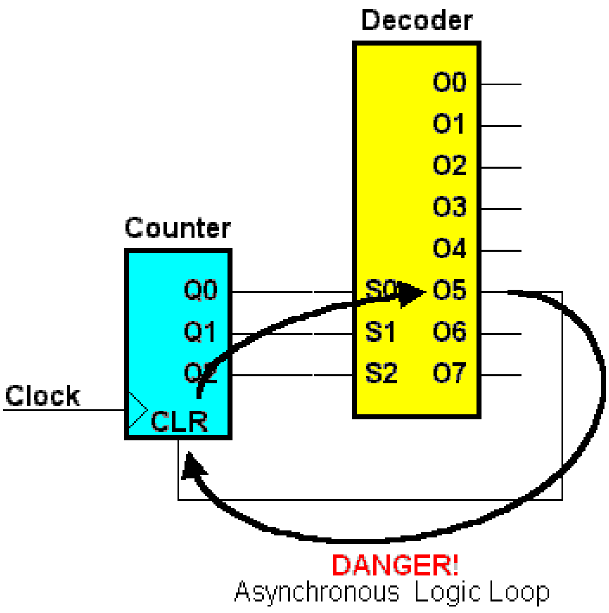
In large circuits a single global clock can be not enough for the system requirements. In this case the system should be partitioned in a few sub-systems each one of them is strictly synchronous with its own clock Then the interfaces between sub-systems should be simple and clear, in order to guarantee a correct behaviour of the overall system.

A designer using CPLDs might wonder about the wisdom in using a product-term clock, which essentially is a gated clock. In fact, internal delays on a CPLD product term are more closely matched than routing delays inside an FPGA or gate array. However, gated-clock problems can still arise and need to be studied carefully.

Another clocking aid is a *global clock buffer*, a high-speed, low-skew, high-fanout clock-distribution network built into most CPLDs and FPGAs. These devices come with two or more global clock buffers, and some have as many as eight. If a design does not use them, these valuable resources are wasted. Ideally, a design should have one or two clock inputs. These clocks, typically with high fanout, should use the global clock buffers to simplify overall device design. With most vendors tools, you must specifically request a global buffer either using a special symbol in the schematic or by instantiating the buffer through VHDL or Verilog, but some tools automatically infer the use of a global buffer. A drawback of global clock buffers is that they are among the largest power consumers in an FPGA or CPLD with about 2-10 mW/MHz. For most designs, this level is second only to I/O switching in overall power consumption. However, the benefit of a clock buffer typically outweighs the extra power consumption and engineering required to design without them. Luckily, in most devices it is possible to connect as many flip-flops as desired to a buffer without consuming additional power. If a design requires the absolute minimum power and the designer decides not to use any global clock buffers, special care is required on clock-skew problems.

Many synchronous flip-flops, even with global clocks, use asynchronous Set and Reset inputs. This situation is again risky. With modern fast devices, even a momentary glitch is enough to inadvertently change a flip-flop state. Figure A.11 shows a common circuit that decodes the output of a counter, and the decoded output asynchronously resets the counter. The problem is that a decoded output might be too short to reliably reset all the flip-flops in the counter. A better approach is to use a counter with a synchronous reset. Asynchronous flip-flop inputs should originate from either device inputs or from other flip-flop outputs.

FIGURE A.11. **A common circuit demonstrates unreliable operation. The output from the decoder depends upon and also resets the counter flip-flops. The asynchronous reset signal might be too fast to reliably reset all of the counter flip-flops.**

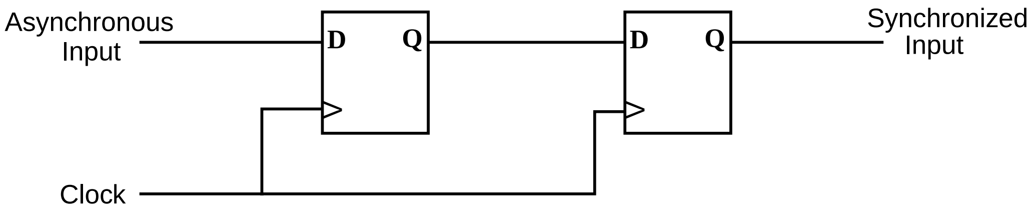


A few simple guidelines summarize the synchronous approach:

- Use as few clocks as possible. The ideal synchronous system has a single clock input.
- Avoid gated signals on any asynchronous flip-flop input including Clock inputs as well as asynchronous Set or Reset inputs.
- If available, use global clock buffers to distribute any high-fanout or skew-critical clock signals. Otherwise, is required an evaluation of each separate clock path for clock-skew problems.
- Synchronize inputs coming from non-coherent systems.

The last point is worth to be commented. Also if a system is fully synchronous, its inputs can be asynchronous. This means that when such inputs (or other signals derived from them) go into a FF, they can violate the set-up or hold time. In this case the transition on the FF output can be delayed beyond the specific clock-to-output delay of the FF. This situation is called *metastabilit*y (see for example [71]), and can induce an undefined behaviour in the rest of the circuit. A common design technique to avoid metastability is based on a cascade of two (or more) FFs, acting as a synchronization circuit (Figure A.12). This method does not guarantee that the second FF while not clock an undefined value, but it dramatically increases the probability that the data will go to a valid state before it reaches the rest of the circuit. One drawback of this circuit is that it takes longer for the system to respond to an asynchronous input.

FIGURE A.12. **Basic synchronization circuit**



A.5 Market overview

Ranked by 1997 sales volume, the major high-density programmable logic suppliers are:

1.Altera
2.Xilinx
3.Vantis (formerly AMD’s programmable logic division, now part of Lattice)
4.Lattice Semiconductor
5.Actel
6.Lucent Technologies
7.Cypress Semiconductor
8.Atmel
9.QuickLogic

Table A.3 summarizes the main vendors and their products.

TABLE A.3. **Vendors, architecture, technologies and products.**

Vendor	**FPGA**	**CPLD**	**SPLD**
Altera	**SRAM** APEX 20K, FLEX 10K, FLEX 6000, FLEX 8000	**EEPROM** MAX 9000, MAX 7000 **EPROM**: MAX 5000 **FLASH**: FLASHlogic	**EPROM** Classic
Xilinx	**SRAM** Virtex, Spartan, XC4000, XC5200, XC3x00A, XC6200, XC2000	**FLASH** XC9500 **EPROM** XC7000	
Vantis (AMD)	**SRAM** VF1	**EEPROM** MACH5 / 5A, MACH4 / 4A, MACH1 & 2	**EEPROM** PAL
Lattice		**EEPROM** pLSI 8000 (SuperBIG), pLSI 5000 (SuperWIDE), pLSI 6000, pLSI 3000, pLSI 2000, pLSI 1000/E	**EEPROM** GAL
Actel	**Anti-fuse** SX Series, MX Series, 3200DX, 1200XL, ACT3, ACT3PCI, ACT2, ACT1 **Flash/CMOS:** ProASIC		
Lucent	**SRAM** ORCA3+, ORCA3, ORCA2C / ORCA2T, ATT 3000		
Cypress		**EEPROM** Ultra 37000, FLASH, Delta39K, Flash 370/370i	**EPROM**, **PAL**, **PLD**
Atmel	**SRAM** AT40K, AT6000	**FLASH**: ATF **EPROM**: ATV	**FLASH**: ATF
QuickLogic	**Anti-fuse** QuickRAM, QuickPCI, pASIC3, pASIC2, pASIC1		
Gatefield	**EEPROM** GF260F, GF250F		
DynaChip	**SRAM (CMOS):** DL6000 **SRAM (ECL I/O):**DL5000		

Conclusions

After the investigation of several approaches and architectures, a first prototype of the digital system for the read-out and control of the ALICE Pixel Detector has been built and tested.

The implementation on a VME board is based on four programmable logic devices. The design, both at board level and programmable circuit level, was aimed to guarantee a very high flexibility, in order to match future changes in the specifications and in the interfaces with the other parts of the detector, that are still under study at the time of the design of the Pilot prototype. Special care has been put also on design for testability issues, introducing a pixel emulator on the board, and allowing the fast access to the data flow in a intermediate step of the processing. A dedicated set-up for testing has been built. It includes the development of new hardware and of a software based on the LabView environment. This set-up will be also useful in the following stages of the pixel detector assembly.

In summary I consider that the design, simulations, construction and measurements presented in this thesis have demonstrated the viability of the Pilot system and possibility to integrate it on an ASIC without any major problem.

The Pilot logic can be implemented in a synchronous way. My recommendation is to avoid any asynchronous circuitry, in order to improve the reliability of the full Pixel detector.

The design of the Alice detector started in 1996. When LHC will become finally operational (the expected date is 2005), the detector will be installed and a decade of intensive preparations will be rewarded by the exploration of a new frontier of elementary particle physics.

-

Bibliography

[1]. The LHC Conceptual Design Report - The Yellow Book, CERN/AC/95-05 (LHC).

[2]. D. Denegri, “Standard model physics at the LHC (pp collision)”, CERN-PPE/90-181 (1990) 56-117

[3]. Alice Technical Proposal CERN/LHCC 95-71

[4]. Alice ITS Technical Design Report CERN/LHCC 99-12

[5]. A. Badala et al., Internal Note ALICE 99-01.

[6]. B.G. Taylor,"TTC Distribution for LHC Detectors", IEEE Trans. Nuclear Science, Vol. 45, No. 3, June 1998, pp. 821-828

[7]. R.K. Bock and W. Krischer, The Data Analysis BriefBook, Springer 1998; an Internet version is available at URL http://www.cern.ch/Physics/DataAnalysis/BriefBook/.

[8]. O. Villalobos et al.,Internal Note ALICE 99-39.

[9]. F. Anghinolfi et al., IEEE Trans. Nucl. Sci. 39 (1992) 654.

[10]. M. Campbell et al., Nucl. Instr. and Meth. A290 (1990) 149.

[11]. M.G. Catanesi et al., Nucl. Physics B (Proc. Suppl.) 32 (1993) 260 (Como 1992).

[12]. M. Campbell et al., Nucl. Instr. and Meth. A342 (1994) 52.

[13]. E.H.M. Heijne et al., Nucl. Instr. and Meth. A349 (1994) 138.

[14]. F. Antinori et al., Nucl. Instr. and Meth. A360 (1995) 91.

[15]. P. Middelkamp et al., Nucl. Instr. and Meth. A377 (1996) 532.

[16]. F. Antinori et al., Nucl. Phys. A590 (1995) 139c.

[17]. V. Manzari et al., J. Phys. G25 (1999) 473.

[18]. E.A. Vittoz, Proc Int. Workshop on Silicon Pixel Detectors, Leuven 1988, Nucl. Instr. and Meth. A275 (1989) 472.

[19]. F. Krummenacher et al., Nucl. Instr. and Meth. A305 (1991) 527-532.

[20]. E. Andersen et al., Phys.Lett. B433 (1998) 209.

[21]. E. Andersen et al., Strangeness enhancement at mid-rapidity in Pb-Pb collisions at 158 GeV/c, CERN preprint CERN-EP/99-29, to be published in Phys.Lett. B.

[22]. R. Caliandro et al., J.Phys. G: Nucl.Part.Phys. 25 (1998) 171.

[23]. R. Lietava et al., J.Phys. G: Nucl.Part.Phys. 25 (1998) 181.

[24]. T. Virgili et al., J.Phys. G: Nucl.Part.Phys. 25 (1998) 345.

[25]. A. Jacholkowski et al., J.Phys. G: Nucl.Part.Phys. 25 (1998) 423.

[26]. E.H.M. Heijne et al., Nucl. Instr. and Meth. A383 (1996) 55.

[27]. S. Di Liberto et al., Internal Note ALICE 98-43.

[28]. F. Riggi et al., Irradiation tests of the Omega3 pixel readout chip by 15 MeV electrons, Internal Note ALICE 99-31 (in preparation).

[29]. N. S. Saks, M.G. Ancona, J.A. Modolo, IEEE Trans. Nucl. Sci. 31 (1984) 1249.

[30]. N. S. Saks, M.G. Ancona, J.A. Modolo, IEEE Trans. Nucl. Sci. 33 (1986) 1185.

[31]. R.C. Lacoe et al., Total dose hardness of CMOS commercial microelectronics, presented at 1997 RADECS conference, to be published in the proceedings.

[32]. D.R. Alexander, Design issues for radiation tolerant microcircuits for space, short course presented at the 1996 NSREC conference in Indian Wells, Ca.

[33]. G. Anelli et al., Proceedings of the Third Workshop on Electronics for LHC Experiments, London, September 22-26, 1997, CERN/LHCC/97.

[34]. F. Faccio et al., A quarter micron CMOS technology for the radiation environment of LHC, in Proc. 6th International Conference on Advanced Technology and Particle Physics, Villa Olmo, Como, Italy. October 5-9, 1998, submitted to Nucl. Instr. and Meth A.

[35]. J.L. Pakuti, J.J.LePage, IEEE Trans. Nucl. Sci. 29 (1982) 1832.

[36]. T. Aoki, IEEE Trans. El. Dev. ED-35 (1988) 1885.

[37]. R. Menozzi et al., IEEE Trans. El. Dev. ED-35 (1988) 1988.

[38]. J.W. Gambles, A path toward low cost rad-tolerant digital CMOS, in Proc. 6th NASA Symposium on VLSI design, (1997).

[39]. A. Giraldo, Evaluation of deep submicron technologies with radiation tolerant layout for electronics in the LHC environments, Ph.D. Thesis, University of Padova, Italy, Dec. 1998.

[40]. GEANT Detector Description and Simulation tool, CERN Program Library Long Writeup W5013.

[41]. J.F. Ziegler,.Stopping cross-section for energetic ions in all elements, Volume 5, Pergamon Press.

[42]. W. Snoeys et al., Layout techniques to enhance the radiation tolerance of standard CMOS technologies demonstrated on a pixel detector readout chip, presented at the 8th European Symposium on Semicondictor Detectors, Schloss Elmau, Germany, June 1998, and to be published in Nucl. Instr. and Meth. A, 1999.

[43]. F. Faccio et al., in Proc. of the Fourth Workshop on Electronics for the LHC Experiments, Rome, Sept. 1998, CERN/LHCC/98-36.

[44]. K. Wyllie et al., "A pixel readout chip for tracking at ALICE and particle identification at LHCb", Proceedings of the Fifth Workshop on Electronics for the LHC Experiments (LEB99), 20-24 September 1999, Snowmass, Colorado, USA.

[45]. B. Gunning et al., A CMOS Low-Voltage-Swing Transmission-Line Transceiver, ISSCC Dig. of Tech. Papers, Feb 1992, 58.

[46]. Canberra Semiconductors N.V., Lammerdries 25, 2259 Olen, Belgium.

[47]. L.H.H. Scharfetter, Active Pixel Detectors for the Large Hadron Coolider, Ph.D. Thesis, Leopold Franzens University, Innsbruck, Feb. 97.

[48]. P.A. Totta and R.P. Sopher, IBM Res. Develop. 13 No.3, 226, 1969.

[49]. J.H. Lau, Flip-chip Technologies, McGraw-Hill (1995).

[50]. D.J. Pedder, Plessey Research Review (1989) 69.

[51]. P. Middelkamp, Tracking with active pixel detectors, Ph.D. Thesis, University of Wuppertal, Germany, Dec. 96, WUB-DIS 96-23.

[52]. E.H.M. Heijne et al., CERN DRDC/93-54.

[53]. E. Cantatore, Caratterizzazione Statistica di Circuiti Integrati in Tecnologia CMOS per la Lettura di Rivelatori a Pixel, Ph.D. Thesis, Politecnico di Bari, Italy, Feb. 1997.

[54]. E. Cantatore, Construction and Performance of the WA97/NA57 Silicon Pixel Telescope, in Proc. of PIXEL98 - International

[55]. Pixel detector Workshop, Fermi National Accelerator Laboratory, Batavia, Illinois, USA, May 1998, Fermilab-CONF-98/196.

[56]. M. Bosteels, http://nicewww.cern.ch/pbonneau/CoolingSystemWeb/Leakless1.htm and private communication.

[57]. IEEE Standard Test Access Port and Boundary-Scan Architecture, IEEE Std 1149.1 - 1990 (Includes IEEE Std 1149.1a - 1993).

[58]. M. Bosteels, http://nicewww.cern.ch/pbonneau/CoolingSystemWeb/Leakless1.htm and private communication.

[59]. ALICE Technical Proposal, CERN/LHCC/95-71, 1995

[60]. VMEbus Specification Manual, IEEE P1014/D1.0

[61]. Vikram S. Sundaram, Applications Engineer, Altera Corporation (email: vsundara@altera.com): private communication.

[62]. S. Ramo, J.R. Whinnery, T. Van Duzer, Fields and waves in communication electronics, II ed. , John Wiley & Sons, 1984

[63]. R.F.Tinder, Digital Engineering Design (Chapter 4), Prentice-Hall, 1991

[64]. P. Horowitz, W. Hill, The Art of Electronics, II ed. Cambridge University Press,1989

[65]. H.W. Ott, Noise Reduction techniques in electronic systems, II ed., John Wiley & Sons, 1988

[66]. XC3000 Series FPGAs Data Book, Xilinx, 1998

[67]. S. Brown, J. Rose, FPGA and CPLD Architectures: A Tutorial, IEEE Design & Test of Computers (1996) 42

[68]. E.Hamdy et al., Tech. Digest IEEE Int’l Electron Device Meeting. IEEE, Piscataway, N.J., 1988, pp. 786-789.

[69]. D. Marple and L. Cooke, "Programming Antifuses in CrossPoint’s FPGA," Proc. IEEE Int’l Custom Integrated Circuits Conf., IEEE, Piscataway, N.J., 1994, pp. 185-188.

[70]. Z. Zilic et al.," Designing for High Speed-Performance in CPLDs and FPGAs," 3rd Canadian Workshop on FPDs (FPD’95), Montreal, May 1995.

[71]. App.Note 42 “Metastability in Altera Devices”, May 1999, Altera