Contents

5 Front-end electronics and readout

5.1 Front-end electronics

5.1.1 Introduction and overview

This section reviews the requirements for the front-end electronics and discusses briefly the general architecture and the basic building blocks.

5.1.1.1 General requirements

As detailed in Sections ?? and ??, the front-end electronics has to read out the charge detected by 570 000 pads located on the readout chambers at the TPC end-caps. These chambers deliver on their pads a current signal with a fast rise time (less than 1 ns), and a long tail due to the motion of the positive ions. The amplitude, which is different for the different pad sizes, has a typical value of 7 μA. The signal is delivered on the detector impedance which, to a very good approximation, is a pure capacitance of the order of a few pF.

The main requirements for the readout electronics, listed in Table 5.1, are briefly discussed below.

- Diffusion and electron statistics limit the resolution both in the drift and azimuthal directions. Monte Carlo studies (Section ??) indicate that to reach the detector resolution limit a signal-to-noise ratio of 30 : 1 is required.

- The maximum pad and time bin for a hit corresponds typically to a charge of about 4.8 fC (3×10^4 electrons) for a minimum ionizing particle, leading to a maximum acceptable noise (r.m.s.) of about 1000 electrons.

- The ionization to be measured in a single pad can be large. The most probable proton transverse momentum is expected to be about 550 MeV/c; 280 MeV/c and 340 MeV/c protons, e.g., have

Table 5.1: Front-end electronics requirements.

Parameter	Value
Number of channels	570 132
Signal-to-noise ratio (MIP)	30 : 1
Dynamic range	900 : 1
Noise (ENC)	1000 e
Conversion gain	12 mV/fC
Crosstalk	$< 0.3\%$
Shaping time	about 200 ns
Sampling rate	5.66(–7.69) MHz
Tail correction after 1 μs	0.1%
Bandwidth to DAQ/Level-3	8.4 GByte/s
Maximum dead time	10%
Power consumption	< 100 mW/channel

about 15 and 8 times the energy loss of a MIP, respectively. Owing to Landau fluctuations, another factor 2–4 in dynamic range has to be considered. Thus, the electronics should not saturate for signals up to 30 MIP (Section ??), leading to a desired dynamic range of at least 10 bits.

- The amplifier conversion gain has to be such that the maximum output signal matches the input dynamic range of the ADC. An ADC with 2 V dynamic range, e.g., requires a conversion gain of about 12 mV/fC.

- In general, the shaping time has to be a compromise between the need for achieving a high signal-to-noise ratio (bandwidth limitation) and for avoiding overlap of successive signals. It should be noted that, owing to longitudinal diffusion and to the track inclination, the signal spreads in time. Therefore, as soon as the shaping time becomes small as compared to the width of the pulse, the current pulses due to the different primary electrons will be visible at the shaper output. The signal width varies from 124 ns (r.m.s.), for 90°-tracks, to 400 ns (full width), for 45°-tracks. It can be shown [1] that a shaping time of about 200 ns, comparable to the signal width (FWHM), is compatible with the low-noise requirement.

- The shaping time of about 200 ns makes a sampling frequency of 5–6 MHz plausible. We therefore divide the total drift time of 88 μs into about 500 time bins, leading to a sampling frequency of 5.66 MHz. Each of the 500 time bins corresponds to a drift distance of 5 mm. Simulations show that with 130 ns time bins a somewhat better momentum resolution can be achieved (see Section ??) albeit at the expense of a larger data volume. We keep the option open to run with a sampling frequency of up to 7.69 MHz, possibly just for the smaller radii.

- Owing to the high channel occupancy, in order to minimize pile-up effects, a very precise tail cancellation, at the level of 0.1% of the maximum pulse height, is required in the front-end stage. This can be done either before or after the analog-to-digital conversion.

- The large granularity or the TPC (3×10^8 pixels for 500 time bins) leads to event sizes of about 84 MByte after zero suppression (132 MByte for time bins of 130 ns). To achieve the necessary rate capability the zero suppression has to be done in the front-end before the data is transferred to the DAQ system. In Pb–Pb running, event rates will reach 10^4 minimum bias events per second, while in pp running the maximum interaction rate will be of the order of 10^5 interactions per second. A few percent of these rates correspond to central collisions that will be trigger selected. Zero suppression at the front-end will reduce the data volume by a factor 2.5, leading to a data throughput of 8.4 GByte/s (13.2 for the shorter time bins) with 100 events/s transferred to the DAQ/Level-3 processing [1]. The reduction from the Level-1 rate of up to 200 Hz to 100 Hz after Level-2 is due to pileup protection against further interactions during the entire drift time of the TPC. We imagine that this could be relaxed depending on occupancy and experiences (from STAR and also our own) to a shorter protection period. Taking this into account, at least 140 Digital Data Links (DDLs) should be foreseen for the data transfer from the front-end electronics to the DAQ (as outlined in Section 5.1.1.2 actually 180 links will be used). In pp mode the detector will produce a data volume smaller by a factor 5.

- A critical aspect in the TPC operation is the temperature stability; to ensure a constant drift velocity it has to be controlled at the level of about 0.1 °C over the whole volume. The large number of channels (570 000) requires development of a system with low power consumption. The aim is to keep the total power consumption below 60 kW (100 mW/channel). We plan to remove the heat from the readout modules with a water cooling system (see Section ??).

[1] The compression from 10-bit data to nonlinear 8-bit and according packaging could be done at the front-end level leading to 66 MByte per event. For the data format also see Section 5.1.3.5.

- The radiation load on the TPC is low, with a total dose received over 10 years of less than 300 rad and a neutron flux of less than 10^{11} neutrons/cm^2. Thus standard radiation-soft technologies are suitable for the implementation of this electronics. Nevertheless, some special care should be taken to protect the system against potential damage caused by Single Event Effects (SEEs).

- The electronics will be located in an area with limited access. High reliability is thus a concern.

The front-end electronics system has to satisfy many other constraints while meeting the required performance specifications. Mainly, the readout electronics needs to fit into the overall detector structure and, in particular, into the available space, which has important consequences for our requirements on reliability, power, and cooling.

The following sections provide more details on these requirements and show how we plan to meet these challenges.

5.1.1.2 System overview

The front-end electronics for the ALICE TPC consists of 570 000 channels. A single readout channel is comprised of three basic units (Fig. 5.1): a charge sensitive PreAmplifier/ShAper (PASA); a 10-bit 10 MHz low-power ADC; and an ASIC which contains a shortening digital filter for the tail cancellation, the baseline subtraction and zero-suppression circuits, and a multiple-event buffer.

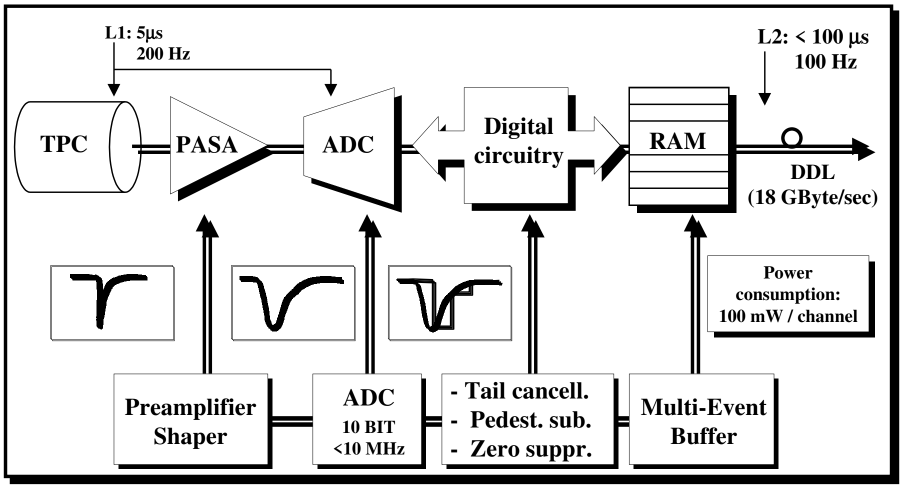

Figure 5.1: Basic components of the TPC front-end electronics.

The image charge induced on the TPC pads is amplified and integrated by a low input-impedance amplifier. It is based on a Charge-Sensitive Amplifier (CSA) followed by a semi-Gaussian pulse shaper of the second order. These analog functions are realised by a custom integrated circuit, to be implemented in CMOS technology, which will contain 16 channels with a power consumption per channel below 20 mW. The circuit has a conversion gain of 12 mV/fC and an output dynamic range of 2 V with deviations from linearity of 1%. It produces a pulse with a rise time of 120 ns and a shaping time (FWHM) of about 200 ns. The single channel has a noise value (r.m.s.) below 1000 e and a channel-to-channel crosstalk below about 60 dB.

Immediately after the PASA, a 10-bit ADC (one per channel) samples the signal at a rate of 5.66 MHz. ADCs of the required conversion time, dynamic range, and precision are commercially available. The strict limitation on the power budget (100 mW/channel) calls for low-power ADCs. Single-, double- or quadruple-channel ADCs, with power consumption below 50 mW, are available. Furthermore, most of them feature standby power consumption below 10 mW. This feature is very important since the ALICE

TPC has a duty cycle of less than 10%. A study of several commercially available ADCs has been completed, showing that the ALICE TPC requirements, in terms of performance and power consumption, are fulfilled in the frequency range of interest.

The digitized signal is processed by a set of circuits contained in an ASIC named ALTRO [2]. The ALTRO (ALice Tpc ReadOut) is a custom CMOS chip which contains the circuitry, for eight channels, to perform tail cancellation, pedestal subtraction, zero suppression, formatting, and buffering.

The circuit that removes the long tail is based on the cascade of two pole-zero shortening filters of the first order. This filter, implemented by a 16-bit fixed-point arithmetic, achieves a tail cancellation better than 0.1% of the maximum pulse height. After the tail cancellation, the ALTRO performs the subtraction of the signal baseline. This is done by a circuit, based on a look-up table, which corrects the systematic instability of the signal baseline, allowing the subtraction of (drift) time-dependent pedestal values from the input data values. Rejecting the samples of value smaller than a constant decision level (threshold) basically does the zero suppression. To reduce the noise sensitivity of this pulse detection scheme, a glitch filter checks for a consecutive programmable number of samples above the threshold. Moreover, in order to keep enough information for further feature extraction, a programmable sequence of pre-samples and post-samples can also be recorded. The zero-suppressed data are then formatted into 32-bit words according to a back-linked data structure. Trigger-related data are stored in a multiple-event buffer. The multiple-event buffer is a RAM of 1 k words, 32 bits wide, partitioned in a programmable number of up to eight fixed-length buffers. The ALTRO chip interfaces to the external world through a 40-bit control bus based on a data transfer protocol that supports a data rate of 160 MByte/s.

The complete readout chain is contained in the Front-End Cards (FEC) plugged into crates attached to the detector mechanical structure as sketched in Fig. **??** on page ??. Each FEC contains 128 channels and is connected to the readout pad plane by means of eight flexible cables with 16 signals each. 32 FECs are controlled by a Readout Control Unit (RCU), which interfaces the FECs to the DAQ, the trigger, and the Detector Control System (DCS). The RCU broadcasts the trigger information to the individual FEC modules and controls the readout procedure. Both functions are implemented via a custom bus, based on low-voltage signalling technology which provides a data bandwidth of 160 MByte/s. The interfacing of the RCU modules to the trigger and to the DAQ follows the standard data-acquisition architecture of the experiment [3].

In summary, for each of the 36 TPC sectors, the front-end electronics consists of 125 FECs, 5 RCUs, and 5 DDLs.

5.1.2 Interconnection to the readout pad plane

5.1.2.1 Electronics mapping

Several constraints define the mapping of the detector into front-end boards.

- The total length of the connecting wires has to be minimized.
- The cables used for this connection should be all of the same length. This should avoid the production of several types of cables and, especially, should guarantee the homogeneity of performance over the different regions of the detector.
- Cable crossing should be avoided to minimize the risk of error and a solution as uniform as possible is desirable.

These considerations lead to the interconnection scheme shown in Fig. ?? on page ??. The signals coming from 16 pads are grouped together and transported to the front-end cards via a cable soldered directly onto the pad plane. The high granularity in the connectivity allows to keep sufficiently short the traces by which the signals are collected from the pads and concentrated in the area where the cables are soldered (transport points). Only the pads located at the top and bottom edges of the readout chambers which, due to the mechanical frame, are not directly accessible represent an exception and require longer traces.

As each sector is segmented into three regions with three different pad sizes, the first two requirements call for at least three different types of boards. To avoid a further increase in the number of different types of boards, we opted for a radial orientation of the boards as shown in Fig. ?? on page ??. The boards are grouped in five crates of (from the inside out) 20 (C1), 25 (C2), 22 (C3), 26 (C4), and 32 (C5) boards. All FECs have the same functionality and the same number of channels (128) but slightly different geometry. The FECs of the first two crates are connected to the pads of the inner readout chamber and are the smallest ones. The boards of the next two crates are connected to the central-sector pads and the last crate is connected to the outer sector. The number of different types of boards and their dimensions are summarized in Table 5.2.

Table 5.2: Classification of the front-end electronics cards.

Group	No. of boards	Dimensions [cm^2]
C1	20	19×14
C2	25	19×14
C3	22	29×14
C4	26	29×14
C5	32	40×14

5.1.2.2 Cables and connectors

To keep the crosstalk within the specified limit all the connections inside the connectors must be low inductance. The crosstalk of the whole chain is currently under simulation, but a finite-element 3-D model of the cable and connector has already been created and simulated.

The cables used to transfer the signals from the pads to the input of the amplifier are polyimide striplines, especially designed to minimise crosstalk, pick-up noise, and heat transfer to the pad plane. Each cable contains 16 channels and has a length of about 7 cm. As is shown in Fig. 5.2, the signal

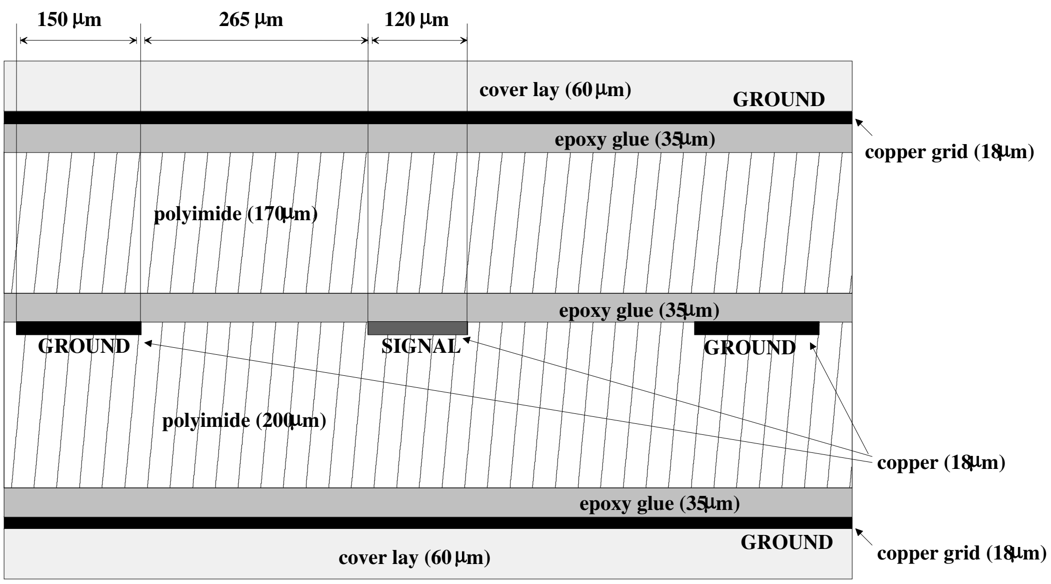

Figure 5.2: Cross-section of the front-end cable.

traces are 120 μm wide with a pitch of 0.8 mm. The ground returns, 150 μm wide, are alternated with the signal strips from which they are at a distance of 265 μm. Two solid ground planes shield completely the striplines. The crosstalk has been simulated to be less than 0.1%. The results of these simulations are shown in Table 5.3.

The proposed connector is a matched impedance connector (QSE-SAMTEC) [4] developed for high-bandwidth applications (higher than 1 GHz) and low crosstalk. The connector consists of a socket and terminal set with an integral ground plane running between two rows of pins. The signal pins have a pitch of 0.8 mm. The main characteristics are listed in Table 5.4.

Table 5.3: Electrical parameters of the front-end cable: impedance (Z_0), propagation delay (t_d), capacitance (C), mutual capacitance (C_M), inductance (L), mutual inductance(L_M), and crosstalk.

Z_0 [Ω]	t_d [ns/m]	C [pF/m]	C_M [pF/m]	L [nH/m]	L_M [nH/m]	crosstalk [%]
64	6.4	99.70	0.06	410	0.03	0.05

Table 5.4: 50 Ω connector, S-G-S configuration with center plane grounded.

0.8 mm pitch (QTE)	100 MHz	500 MHz	1 GHz
Impedance [Ω]	50.7	55.0	60
Attenuation [dB]	-0.0305	-0.2429	-0.4607
Crosstalk [dB]	-60.6	-49.0	-39.5
Propagation delay [ps]	58.4	58.4	58.4

5.1.3 Front-end basic components

5.1.3.1 Preamplifier/shaper

The basic requirements for the preamplifier/shaper chip have been outlined in Section 5.1.1.1 and in Table 5.1 on page 101. The typical number of electrons (before amplification, etc.) generating the signal on the maximum pad and time bin of a cluster is shown in Fig. 5.3 for the three pad sizes and three ranges in polar angles. As can be seen it varies systematically with pad area and for the largest pads the signals are about a factor 2 bigger than for the smallest inside pads. In comparison, the dependence on polar angle is small. In order to achieve a signal-to-noise ratio of 30 : 1 for all pad sizes a factor 2 variation in gas gain is required. However, working with the same gas gain entails only a moderate loss in resolution as detailed in Section **??**. As far as the preamplifier/shaper is concerned we envision one gain setting that puts the noise equivalent of 1000 electrons in channel 1 of an ADC with 2 V full scale. This defines the required gain as 12 mV/fC.

To achieve an overall system noise of 1000 electrons per channel for a typical input capacitance of about 4 pF for the pads plus about 10 pF for the on-board traces plus the polyimide cables requires a low-noise device.

We intend to take the design of the NA45 CERES TPC preamplifier/shaper [5] as a starting point and to optimize it for the ALICE needs, migrating from 0.8 μm CMOS technology to 0.35 μm and incorporating again 16 channels on one chip. The schematics of this circuit are shown in Fig. 5.4. For the ALICE application the tail-cancellation functionality will be removed from the preamplifier/shaper chip and be implemented in the digital chip (see below).

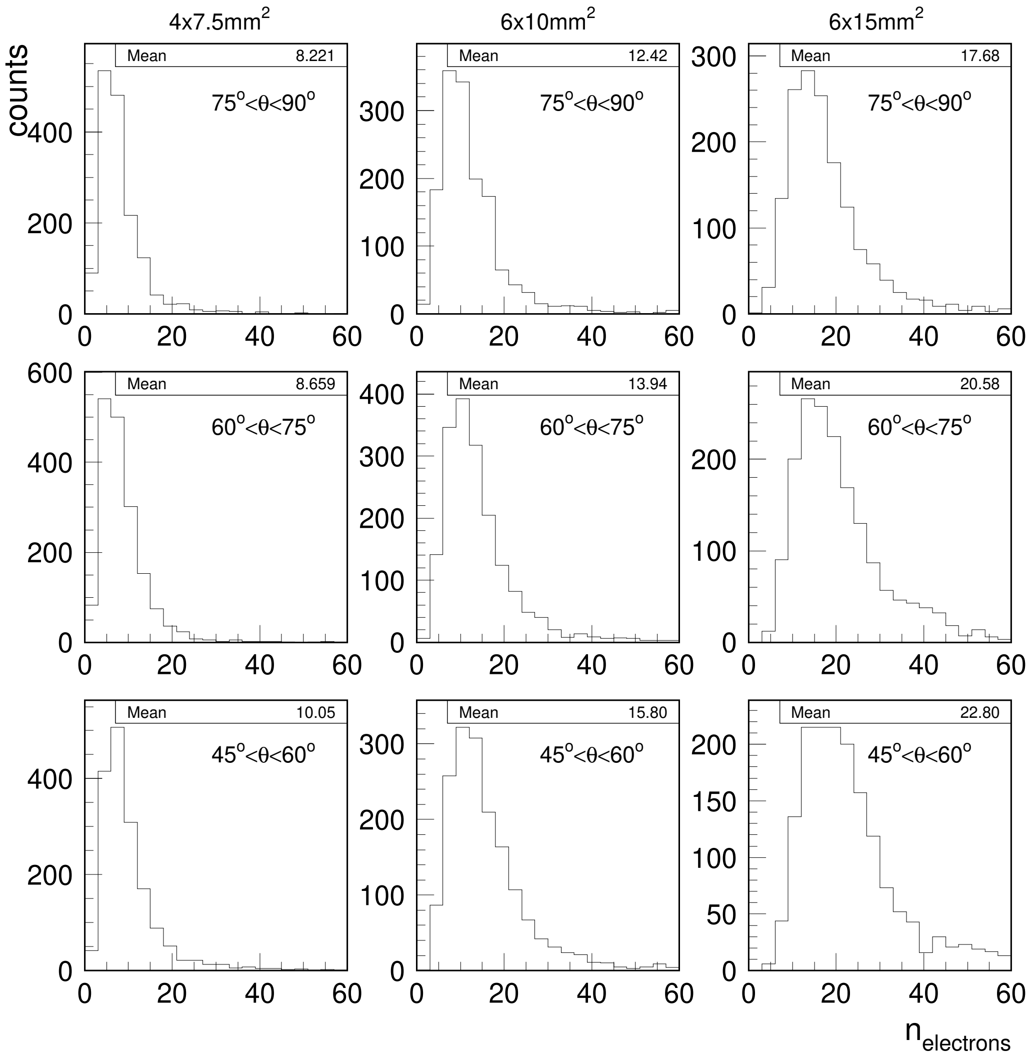

Figure 5.3: Number of electrons generating the signal in the maximum pad and time bin of a cluster for the different pad sizes and averaged over different regions in polar angle.

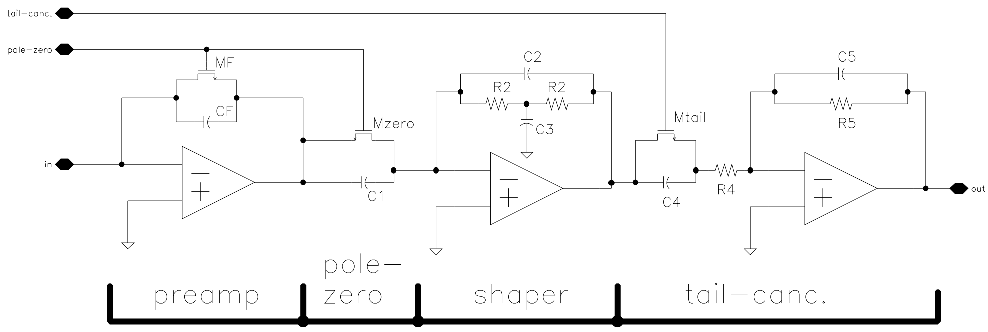

Figure 5.4: Schematics of the NA45 CERES TPC preamplifier/shaper circuit.

In contrast to e.g. the preamplifier/shaper developed for the STAR TPC which has a pulsed reset [6], the CERES design used a MOS transistor (MF) with a feedback capacitance C_F that is continuously discharged with a decay time $\tau_d = C_F \cdot R_{ds}(MF)$ (see Fig. 5.4). This continuously sensitive design is particularly suitable for a detector with high occupancy. Noise considerations dictate a large feedback resistance R_F (for e.g. a peaking time of 400 ns a value larger than 4 MΩ) and the associated drain-source resistance R_{ds} of a MOSFET transistor. This has the advantage that R_{ds} depends on the biasing condition (the self-adaptive bias technique is used [7]) and that larger signals are discharged with a faster decay time (see Fig. 5.5). This is particularly relevant in an environment where signals of minimum-ionizing particles have to be measured in the presence of heavily ionizing low-energy particles. Near mid-rapidity it is expected that the abundance of protons and antiprotons is about 8% of pions and kaons with a most probably momentum of 0.56 GeV/c, implying that 15% of these have a mean energy loss of more than 8 MIP and Landau fluctuations will lead to even much larger signals. Therefore, beyond the question of dynamic range, in the design of the chip attention has to be payed to the response of the circuit to saturating signals.

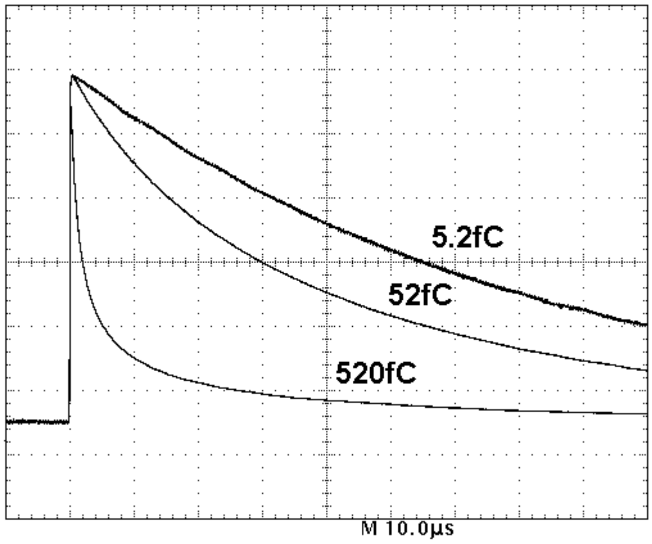

Figure 5.5: Response of the CERES charge-sensitive amplifier to input signals of increasing magnitude.

Particular attention in the design has to be given to the gain linearity in the presence of a nonlinear resistance in the feedback loop. In the CERES design this has been achieved by the self-adaptive bias technique [7] leading to a nonlinearity of the preamplifier of less than 2% for up to 350 fC input charge. For the ALICE application the maximum signal corresponds to 166 fC. In terms of noise requirements, the CERES preamplifier/shaper fulfills the requirement with an ENC of less than 600 electrons for input capacitances between 15 pF and 19 pF.

In terms of power consumption, the CERES chip requires about 60 mW per channel. For time reasons and because this was not a critical issue it has not been optimized for low power consumption. Already removing the tail cancellation and going to 0.35 μm technology should reduce the power consumption to about 1/3 and we are confident that the goal of less than 20 mW per channel can be achieved for an optimized design.

5.1.3.2 Analog-to-digital conversion

The analog signal output of the preamplifier/shaper circuit is sampled, at a rate of 5.66 MHz, by an analog-to-digital converter (ADC) with 10-bit dynamic range. Fast-conversion ADCs of the required dynamic range and precision are commercially available. Conversion times of the order of 100 ns can be achieved with flash-ADCs or with successive approximation pipelined ADCs. The requirement on the power consumption (below 100 mW per front-end channel) calls for low-power ADCs, limiting the choice to pipelined ADCs. Fortunately, due to the continuous growth in wireless communication systems and portable devices where power consumption is a major concern, 10-bit 20 MHz ADCs with low power

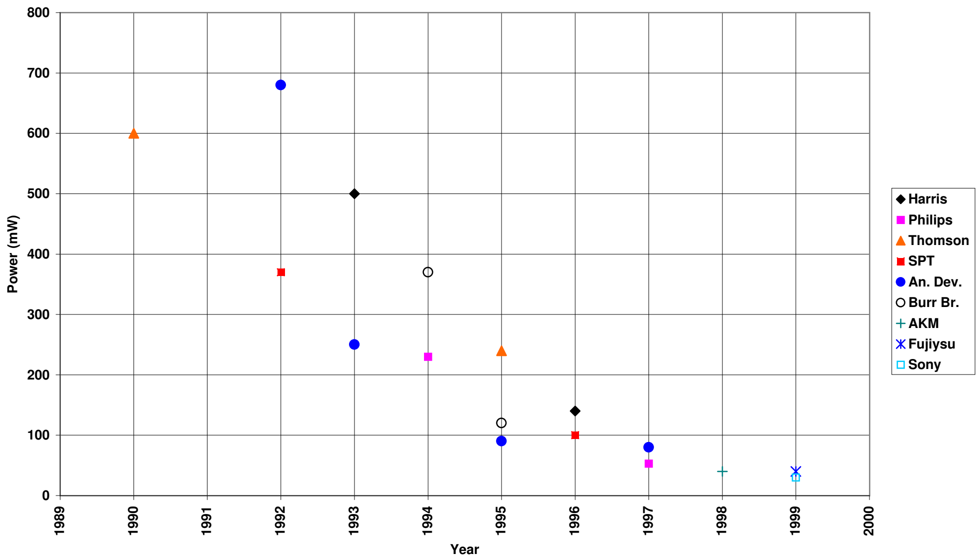

Figure 5.6: Power consumption of 10-bit 10 MHz commercial ADCs.

consumption are now widely available. Fig. 5.6 shows the reduction of power consumption, for 10-bit ADCs, over the last 10 years.

We notice that, since 1997, several commercial ADCs with power consumption below 100 mW have been available, and that in the last year a few devices with power consumption below 50 mW have appeared on the market. Moreover, the trend shown by the curve in Fig. 5.6 indicates that we can reasonably expect devices with a power consumption of 20 mW by the year 2002. In addition, most of them feature standby power consumption below 10 mW. In this respect it should be noted that the ALICE TPC has a duty cycle of about 2%, as the digitization occurs for 88 μs (TPC drift time), triggered by the Level-1 decision (200 Hz). In 'low-power mode' the clock signal is disabled, causing a drop in the device's static power. From the 'low-power mode' the device can be recovered almost instantaneously. We will come back to this point later in this section.

Table 5.5 lists the most recent devices suitable for the TPC application. Concerning the performance of these devices, the documentation provided by the manufacturers is not always exhaustive. As a matter of fact, while the d.c. accuracy of the devices is normally well documented, often the dynamic perfor-

Table 5.5: Commercial ADCs suitable for the ALICE TPC readout.

ADC	Manufacturer	Channels per chip	Max. rate [MHz]	Power at max. rate [mW/channel]
HI5710	HARRIS	1	20	140
AD9200	Analog Devices	1	20	80
TDA8766	Philips	1	20	53
ADS901	Burr-Brown	1	20	48
MByte40C360	Fujitsu	1	25	40
CXD3300R	Sony	1	20	40
AD9201	Analog Devices	2	20	100
SPT7852	SPT	2	20	80
SPT7852	SPT	3	20	130
XRD6414	EXAR	4	20	30

mance of the device is reported only for a few values of the input signal frequency and a defined sampling frequency. Besides, the characterization of the device by the manufacturer is often done in 'ideal' conditions, far from the user application. In the front-end electronics for the TPC, the ADC is close to the analog circuit PASA which provides the signal to be sampled, and to the fast digital circuit ALTRO which processes the signal. Therefore, it is important to study the effect of the ADC on the analog circuit performance, and its susceptibility to the fast digital circuits. For this purpose, an ADc Evaluation Board (ADEB) [8], containing the complete readout chain, has been developed. The ADCs are mounted on small printed-circuit boards plugged into the motherboard as daughter cards. Up to four (different) ADC chips with different analog inputs can be tested at the same time in an automated test. The study of five off-the-shelf ADCs (in bold in Table 5.5) from different manufacturers has been accomplished.

There are many reasons why an ADC with N bits does not perform like an ideal N-bit converter under dynamic operating conditions. Several parameters are universally used to define these different sources of errors: Differential NonLinearity (DNL), Integral NonLinearity (INL), offset error, gain error, Total Harmonic Distortion (THD), aperture jitters, aperture delay, etc. The values of these parameters characterize an ADC and are reported on the device's technical data sheet. To get in-depth knowledge of these devices and an independent verification of the information provided by the manufacturers, all these parameters have been measured for the five commercial ADCs under evaluation. For a detailed description of the measurements and the results we refer to [9, 10]. Here we focus only on the measurement of a parameter called 'Effective Number Of Bits' (ENOB), which in a way includes all the effects described separately by the parameters mentioned above. This number may be considered as the number of bits in a perfect ADC whose r.m.s. 'quantization' error would be equal to the total r.m.s. error from all sources in the ADC under test. Figure 5.7 shows the ENOB versus the frequency of the input signal. We conclude that all the devices under test would fulfil the ALICE TPC requirements in terms of performance in the frequency range of interest.

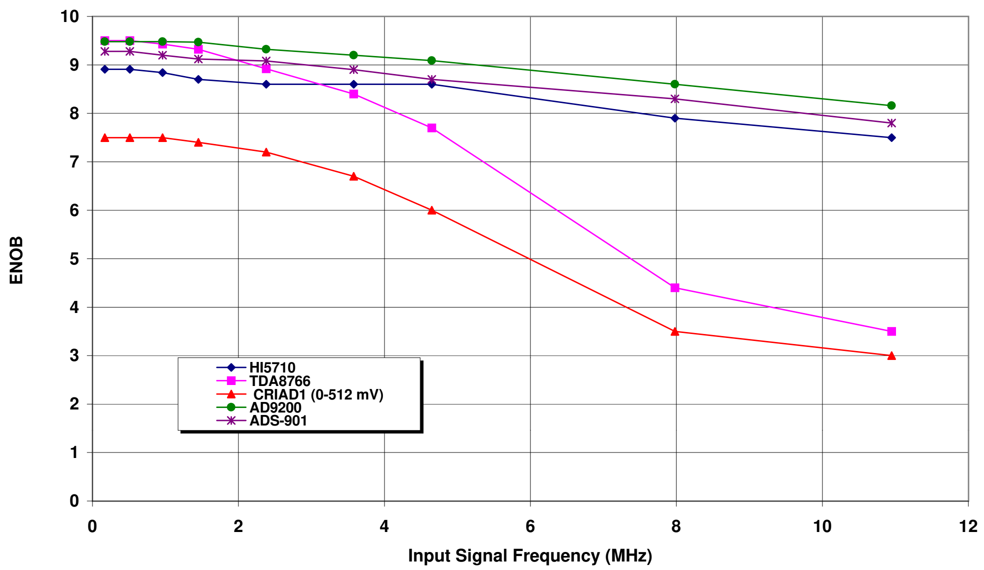

Figure 5.7: Input signal frequency versus the effective number of bits (ENOB) for five different ADCs.

One of the five ADCs (AD9200 by Analog Devices) tested has been used in the readout electronics [11] for the RCC prototype (about 1400 channels) discussed in Section ??. This readout electronics is based on a VME 48-channel Front-End Digitizer Card (FEDC) shown in Colour Fig. XI. The results achieved with this readout system, in terms of digitization resolution, are excellent. The Signal-to-Noise Ratio (SNR) measured is below 0.4 ADC counts for all channels.

In the FEDC system it is possible to put all ADCs simultaneously in 'low-power mode' to be activated only for 88 μs after the trigger signal. What matters for the recovery time is how long the device has

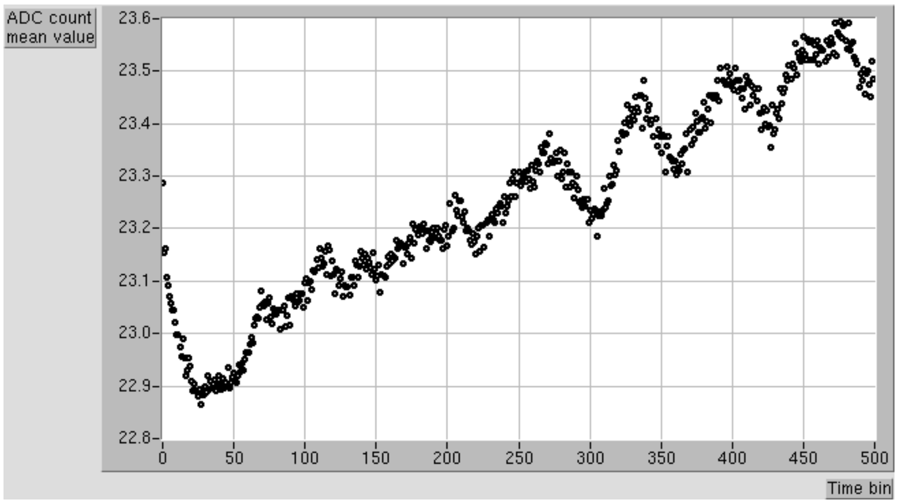

Figure 5.8: Time evolution of the signal pedestal after power-on for the ADC AD9200. Time bins are 100 ns.

been in the off-state. When the device is in 'low-power mode' the capacitors, which provide the reference voltages, start to discharge. When the device is switched back to running mode, the nominal value of the reference voltage is recovered in a time that depends on how long the device was kept in 'low-power mode'. Figure 5.8 shows the time evolution of the signal pedestal when the ADC is powered on after a 'sleeping time' of 10 ms. It can be seen that the discrepancy is not higher than 1.4 mV even for 'sleeping times' of the order of 10 ms, corresponding to the typical distance between two consecutive Level-1 triggers activating the TPC readout. It should be noted that, since this error depends only on the time from the last trigger, the digital processor circuit could apply a correction before the zero suppression is performed.

Another potentially interesting option is the development of a custom ADC. A custom design provides the possibility of integrating the ADC with the rest of the readout chain, e.g. the ALTRO chip, if the technology for the two circuits is the same. Another advantage offered by a custom design is the possibility to adapt the ADC transfer function to the accuracy required in the different regions of the dynamic range. For this purpose a special version of the CRIAD [12], a low-power ADC designed at CERN for other applications, was developed for the ALICE TPC. The CRIAD is a multiple-range linear ADC with four ranges. In each range the resolution is defined by an 8-bit linear conversion performed between two references, the lower one being ground (0 V). The range selection is automatic with the signal amplitude. For comparison with the commercial ADCs, Fig. 5.7 shows also the performance of the CRIAD in the first sub-range. Despite the very promising results, this experience has shown that it is very difficult to develop, in house, a device which is competitive, in terms of performance, power consumption and price, with the commercial devices which, due to the wide market, are significantly improved year after year.

What may be more attractive here is the possibility of using a semi-custom ADC. More and more, CMOS low-power ADCs are included in the standard libraries provided by the silicon manufacturers. As an example, we mention here the case of Austria Mikro Systeme (AMS), a European silicon foundry, which includes in its 0.35 μm CMOS library a 10-bit 20 MHz ADC that, at 5 MHz, has a power consumption of 100 mW. This device features a power consumption only slightly higher than the best off-the-shelf devices.

In conclusion, at present the baseline solution is the use of a commercial ADC, however, the system design is open to the possibility of including the ADC, bought as a cell, in the ASIC (ALTRO) which contains the rest of the readout chain.

5.1.3.3 Digital readout

After the analog-to-digital conversion the TPC signal is processed by a set of digital circuits integrated in a chip named ALTRO (ALice Tpc ReadOut) [2]. The ALTRO chip is a custom integrated circuit dedicated to the processing of a 10-bit digital signal for the readout of the trigger-related data. It contains eight processing channels operating concurrently, at the ADC sampling rate (up to 20 MHz), on the digitized signal coming from eight independent inputs. Starting from a trigger signal each channel performs, on a stream of up to 1000 samples, tail cancellation, pedestal subtraction, zero suppression, formatting, and data storage in a multiple-event memory. As sketched in Fig. 5.9, the ALTRO's processing chain is based on five main units described in some detail in the following sections. First the tail cancellation occurs, followed by baseline subtraction and zero suppression. Then the data are put into a compact format in the Data Formatting Unit (DFU) and finally stored in a multiple-event buffer.

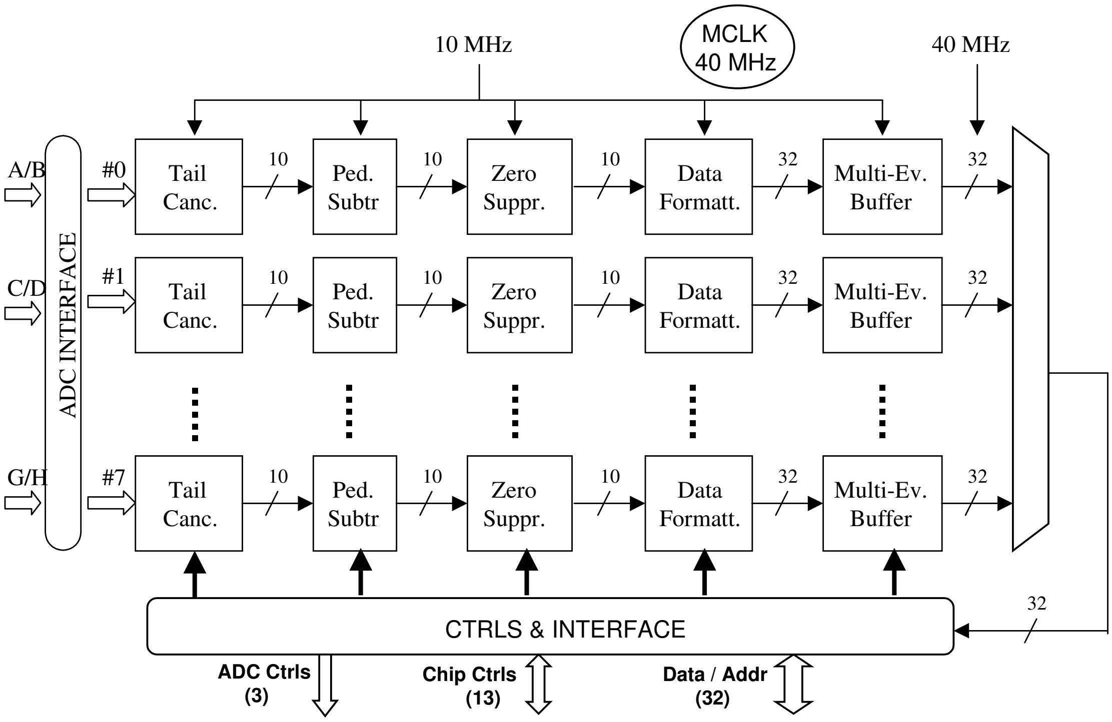

Figure 5.9: ALTRO block diagram.

5.1.3.4 Tail cancellation

As mentioned above, the signal coming from the ALICE TPC presents a long tail. A technique to cancel the tail of the signal arising from the amplification region was developed for a proportional chamber and implemented by a circuit based on discrete components [13]. The same technique underlies the tail-cancellation scheme used in the preamplifier/shaper circuit developed for the readout of the CERES TPC [5]. In this method, the measured TPC signal is approximated by the sum of three exponential terms:

$$i(t) \approx e(t) = I_0 \cdot \left[A \cdot \exp \frac{t}{\alpha \cdot \tau_1} + B \cdot \exp \frac{t}{\beta \cdot \tau_2} + C \cdot \exp \frac{t}{\gamma \cdot \tau_3} \right]. \tag{5.1}$$

Hence, the cancellation of the signal tail can be realized by a deconvolution filtering technique [13]. The signal is passed through a linear network with a response that cancels perfectly one of the exponential

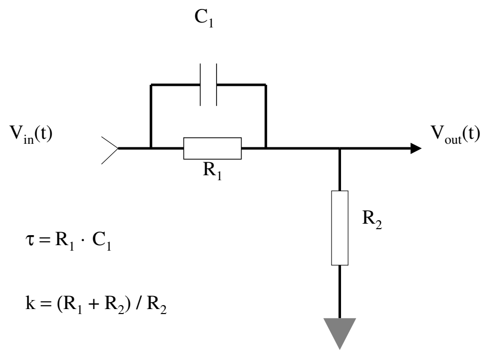

Figure 5.10: Analog pole-zero network.

terms. Let us consider a pole/zero network with transfer function

$$F_1(s) = \frac{1+s\cdot\tau}{k+s\cdot\tau}. \tag{5.2}$$

It can be shown [13] that if the time constant τ is adjusted to be $\tau = \tau_3$, k can be adjusted in a way that all the terms containing τ/k add to zero. In this way the pole τ_3, corresponding to the long tail ($\tau_3 > \tau_2 > \tau_1$), is perfectly cancelled. Similarly, a second pole/zero network with transfer function $F_2(s)$ can be added with constants adjusted so that the remaining terms only contain the time constant τ_1. As a result, the convolution of the two filters with the chamber signal is

$$i_0(t) = i(t) \otimes h_1(t) \otimes h_2(t) \cong I_0 \cdot \exp\frac{t}{\alpha\cdot\tau_1}. \tag{5.3}$$

The network with transfer function Eq. (5.2) is shown in Fig. 5.10, where $\tau = R_1 \cdot C_1$ and $k = (R_1 + R_2)/R_2$. A circuit diagram of the complete tail-cancellation filter is shown in Fig. 5.11. Figure 5.12 details the behaviour of the circuit as simulated by using SPICE. The accuracy of a shortening filter realized by an analog network is limited by the tolerances of the network components. Owing to the poor precision in the matching of the passive components, provided by the actual integrated circuit technologies, the shortening filter described above cannot reach a high accuracy ($\sim 10^{-3}$) if it has to be included in an integrated circuit with no external tunable components.

On the other hand, a digital system provides much better control of accuracy requirements. This results in specifying the accuracy requirements in terms of word length, floating-point versus fixed-point arithmetic, and similar factors. Furthermore, a digital system allows flexibility in reconfiguring the digital signal-processing operations simply by changing a number of programmable coefficients. This is indeed extremely important, considering that the exact shape of the signal which depends on the details of the chamber and pad geometry is difficult to predict and is known with high precision only when the detector is operated. Also, this allows some flexibility in the choice of the gas composition and drift field depending on first running experiences. The first-order filter expressed by the Eq. (5.2) can be expressed in the discrete time domain using the z transform. We obtain the following z transform function:

$$F(z) = \frac{L_0 + L_1\cdot z^{-1}}{1+K_1\cdot z^{-1}}, \tag{5.4}$$

where L_0, L_1 and K_1 are constants that are functions of $A, B, C, \alpha, \beta, \gamma$. This circuit is represented in a schematic way in Fig 5.13.

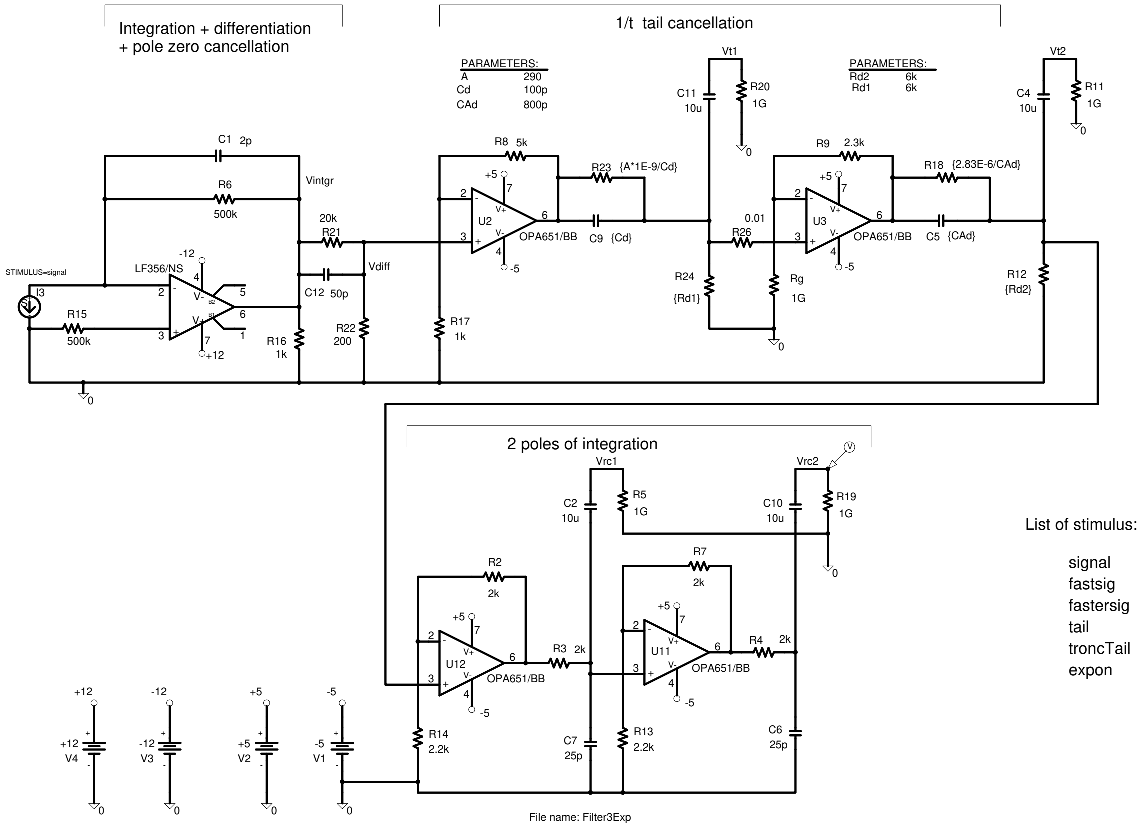

Figure 5.11: Block diagram of the analog tail-cancellation filter.

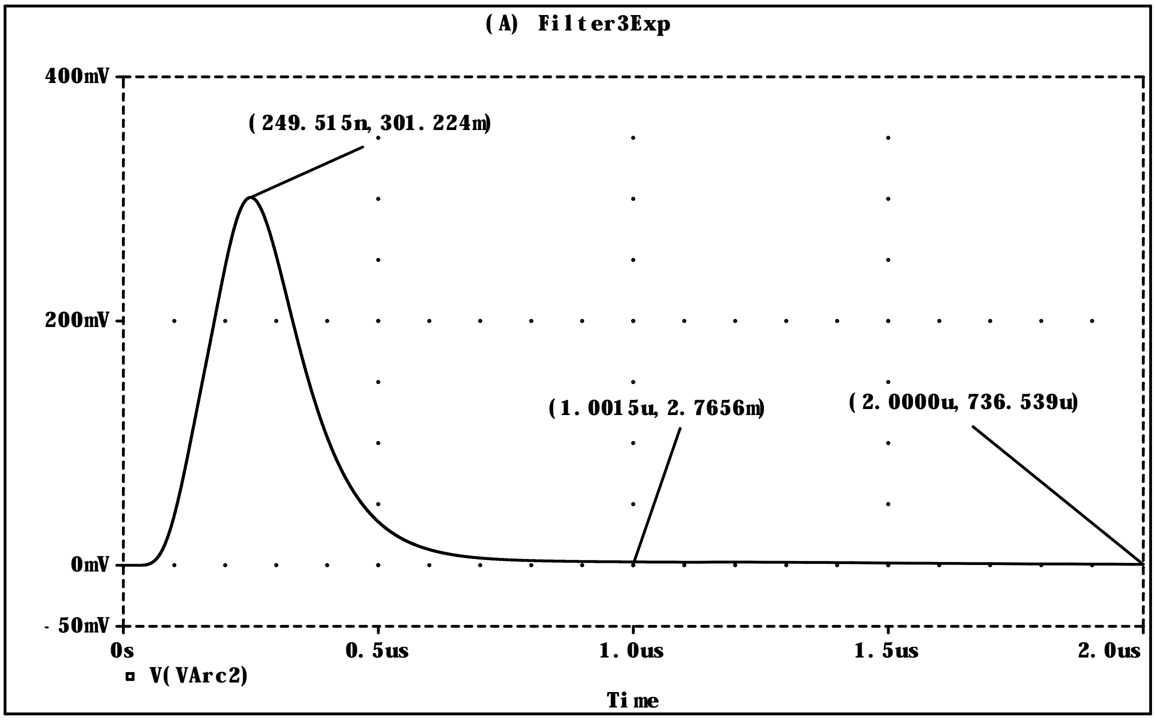

Figure 5.12: SPICE simulation of the analog tail-cancellation filter.

- IIR filter
  - 4 N-bit registers
  - 2 N-bit adders
  - 3 N-bit multipliers
- For N = 16 (0.25 µm):
  - prop. delay = 25 ns
  - area = 1 mm^2

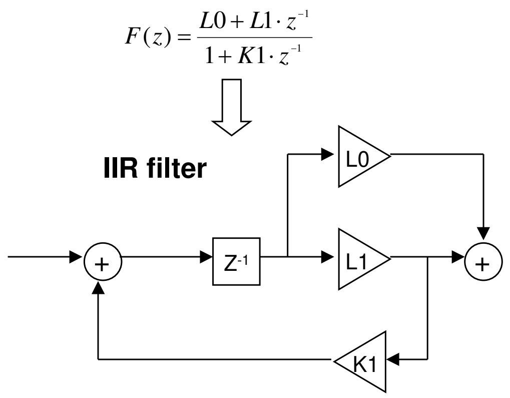

Figure 5.13: Digital pole-zero network.

It consists of three multipliers, two adders, and a register. Even though there are no negative values in the input, their use is needed for the internal calculations of the filter, therefore the two's-complement arithmetic is used. The Verilog implementation of this circuit was performed and the results are shown in this section. In order to work in the digital domain, the analog input signal according to Eq. (5.1) was sampled at 10 MHz, and 16- or 32-bit 'quantizations' were performed.

Figure 5.14 is an overlay of three signals: input, response after the first filter, and final output. These plots show that we can achieve a tail suppression with the three-exponential approximation. The final result was normalized so that the maximum value is equal to one. The error definition was to consider the ratio between a given sample and the maximum value. The filter coefficients were slightly adjusted to have a small undershoot providing a fast decline towards the baseline. Between the samples, there is a simple linear interpolation in order to improve readability of these plots. Only the 16-bit case is shown (see also Table 5.6).

To estimate the size of the digital filter circuit, the Verilog description has been synthesized and targeted to a 0.25 μm CMOS technology. The results are shown in Table 5.7.

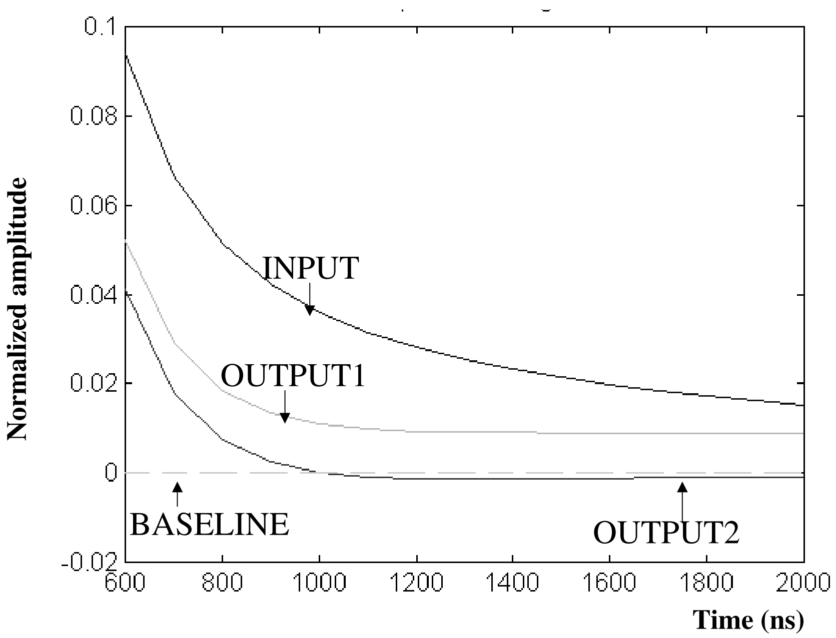

Figure 5.14: Digital filter: input signal, response after the first filter, and final output.

Table 5.6: Shortening filter performance.

	Input	Output 1	Output 2
After 1 μs	0.036	0.011	-1.83×10^{-4}
After 2 μs	0.0153	0.0088	-9.17×10^{-4}
Area [0-1 ms] AU	1.0773	0.8565	0.8270

Table 5.7: 16-bit fixed-point first-order digital filter.

	Slow	Fast
Propagation delay	25 ns	10 ns
Area	0.5 mm^2	1 mm^2

5.1.3.5 Baseline subtraction, zero suppression, data formatting and buffering

The baseline subtraction is performed by the circuit schematically depicted in Fig. 5.15. It has three modes of operation:

- *Subtraction mode*. The circuit performs the subtraction of pedestal values from the input-signal values. The subtracted value can be either fixed or time-bin dependent. In the former case — *fixed subtraction mode* — the value to be subtracted from the input signal is a constant stored in a register (DIN in Fig. 5.15). In the latter case — *variable subtraction mode* — the time-bin-dependent pedestal values to be subtracted are stored in the pedestal memory which, in this configuration, is addressed by a time counter started by the trigger signal. Clearly the *variable subtraction mode* is useful whenever the signal is affected by some systematic baseline variation correlated with the readout cycle.

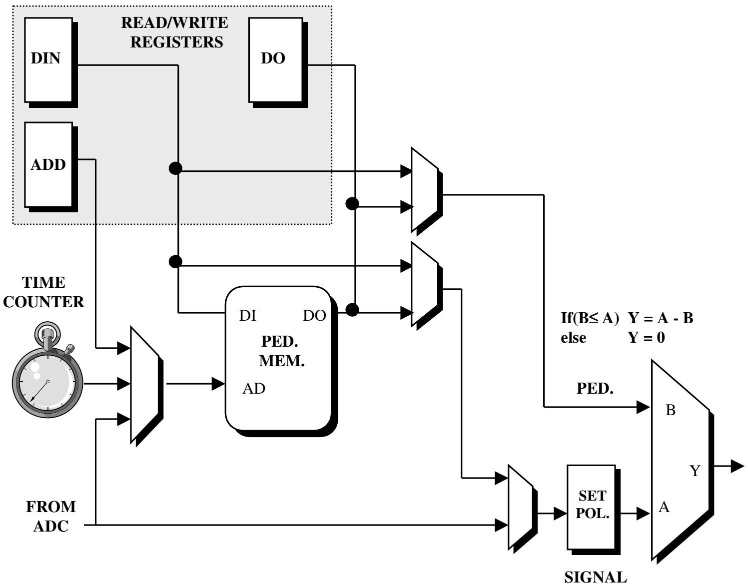

Figure 5.15: Baseline subtraction block diagram.

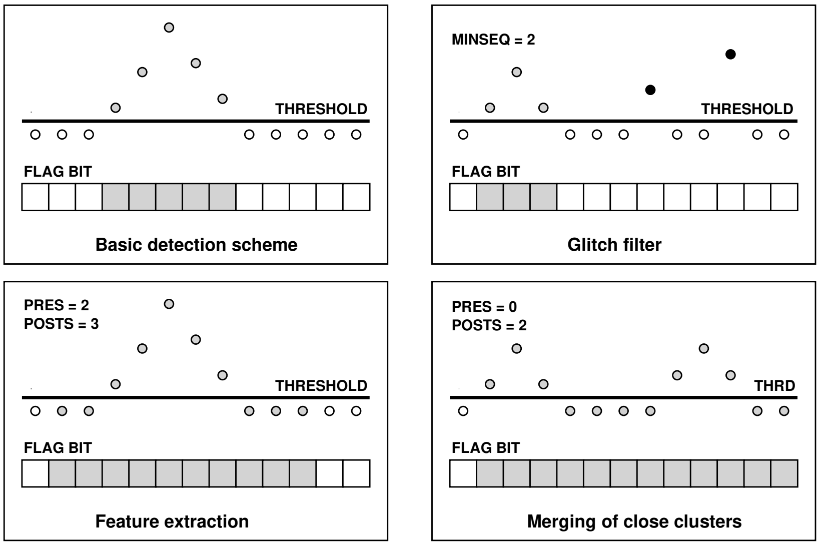

Figure 5.16: Zero-suppression scheme.

- *Conversion mode*. The circuit can perform a memoryless (static) conversion of the input signal of the type $y_n = F(x_n)$. At any instant n, the output y_n depends at most on the input sample x_n at the same time, but not on past or future samples of the input. The output values y_n are stored in the pedestal memory addressed in this case by the input x_n. Concurrently the subtraction of a fixed value from the converted signal can be performed.

- In the third mode of operation — *test mode* — the pedestal subtraction unit can be used to generate a pattern to be injected into the processing chain for test purposes. This is an important feature that allows a complete test of the overall digital readout chain.

One obvious way to compress the data stream is to discard ‘zero’ data, i.e., samples so close to the reference level (pedestal) that they are considered to contain no useful information but rather to be due to noise. The basic pulse-detection scheme is based on the rejection of samples with value smaller than a constant decision level (threshold). When a sample is found above the threshold, it is considered to be the start of a pulse (Fig. 5.16). To reduce the noise sensitivity, a glitch filter checks for a consecutive number of samples above threshold, confirming the existence of a real pulse. The minimum sequence of samples above the threshold that defines a pulse can vary from one to three. In order to keep enough information for further feature extraction, the complete pulse shape must be recorded. Therefore, a sequence of samples (pre-samples) before the signal overcomes the threshold and a sequence of samples (post-samples) after the signal returns below the threshold can be recorded. The number of pre- and post-samples can vary independently in the range between zero and four.

The data format is as illustrated in Fig. 5.17 and described in the following. The pulse thus identified and isolated must be tagged with a time stamp, in order to be synchronized with the trigger decision for validation. Otherwise the timing information would be lost by the removal of a variable number of samples between accepted pulses. This requires the addition of a time word to the set of sample data. Since 1000 is the maximum length of the data stream that can be processed by the ALTRO chip, the time information can be encoded in a 10-bit word. The principle is to label each pulse with a time stamp that defines the time distance from the trigger signal (e.g. ‘T06’ in Fig. 5.17). So the samples of the processed data stream are numbered starting from 0 to 1000.

Since the ALTRO data format does not make use of extra flag bits to distinguish the samples’ data from the time data, we introduce a further word for each accepted cluster to represent the number of words in the cluster (e.g. ‘07’ in Fig. 5.17). Since a cluster can be as long as the complete data stream

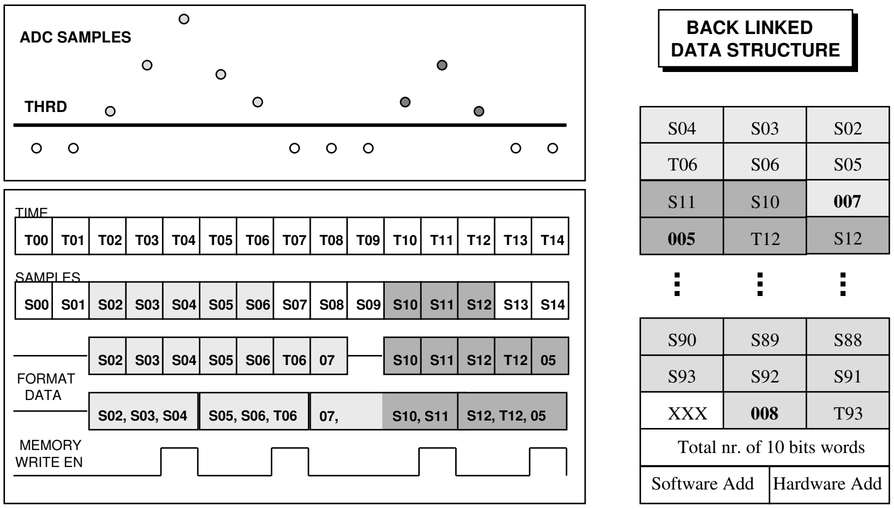

Figure 5.17: The ALTRO data format.

the new word is also a 10-bit word.

The DFU formats the stream of zero-suppressed data by adding, to each set of samples, two extra 32-bit words, and encoding the 10-bit words into 30-bit words. The two 32-bit words form the trailer of each data packet. The first word is fundamental for the decoding of the data packet and expresses the total number of 10-bit words in the packet. Indeed this word provides the position of the last 10-bit word in the data packet. Starting form the last 10-bit word the clusters are back-linked, as each set contains the pointer to the first word of the previous cluster (Fig. 5.17). The second trailer word contains in the 16 LSB a software identifier, which is read from one of the configuration/status registers, while the byte after that contains the hardware address. This address represents a geographical address and is used in the data packet to identify unambiguously to which channel the data packet is associated. The software identifier is a programmable number that can be used to specify to which data source the processing channel is connected.

Finally the data are stored in a multiple-event buffer, a 512×32 RAM that can hold up to eight events. The ALTRO chip interfaces to the external world through a 40-bit control bus based on a data transfer protocol that supports a data rate of 160 MByte/s.

5.1.3.6 ALTRO prototype

In its final version the ALTRO will be implemented in a deep sub-micron CMOS process: at the moment, 0.25 μm seems the process feature size suitable for this application. At the present phase of the project the chip functions are not frozen yet. Thus in order to have an estimate of the size and power consumption, this circuit has been studied using the standard cell library for a 0.25 μm CMOS process, leading to a die size of 8 mm^2/channel and a power consumption of 8 mW/channel. With the exception of the shortening filter, all the ALTRO's functions have been implemented in a FPGA, as used in the FEDC cards [11] for the readout of the RCC (Section ??). Finally, for the purpose of studying the system aspects of the FEC, a four-channel prototype has been developed as a 0.6 μm double metal CMOS process featured by AMS (Austria Mikro Systeme). It has a silicon die of 40 mm^2 and is packaged into a plastic case with 100 pins (PQFP100).

The layout of the chip is shown in Fig. 5.18 where one can recognize the four channels and the common control logic. The supply voltages are distributed to the internal cells starting from two external 80 μm wide rings that surround the circuit body completely.

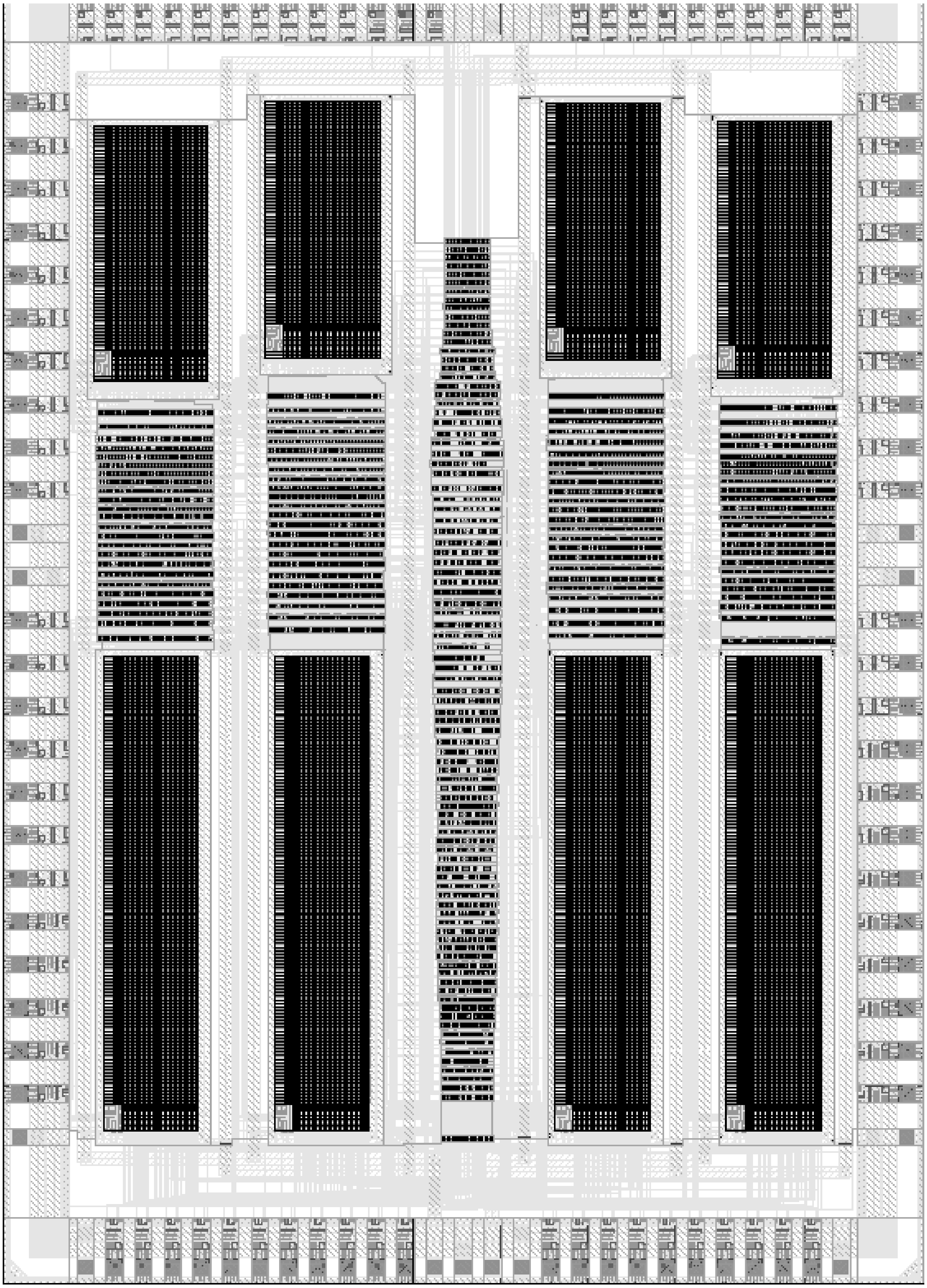

Figure 5.18: Layout of the ALTRO chip prototype.

5.1.4 Front-End Card

The Front-End Card (FEC) contains the complete readout chain for amplifying, shaping, digitizing, processing, and buffering the TPC signals. The FEC must handle the signal dynamic range of about 10 bits with minimal degradation of precision, store the signals during the Level-2 trigger latency, and provide a dead time below 10% at the maximum Level-2 trigger rate of 100 Hz. The design provides 128 channels per FEC, with an estimated maximum power consumption of approximately 12 W.

The layout of the FEC is shown in Fig. 5.19 and has the following flow. The FEC receives 128 analog signals through 8 flexible cables and the corresponding connectors as shown in the figure. The input signals are very fast, with a rise time of less than 1 ns. Therefore, to minimize the channel-to-channel crosstalk, the PASA circuits have to be very close to the input connectors. The signals are first amplified and then processed by a band-pass filter. The latter operation is done to limit the bandwidth of the ADC input signal and to reduce the signal-to-noise ratio. A 16-channel chip (PASA) contains the circuits to implement these analog functions. Each FEC contains therefore eight PASA chips. Commercial ADCs convert the PASA output signals into digital signals. ADCs with two or four converters per chip are considered suitable for this application leading to 64 or 32 ADCs per FEC, respectively.

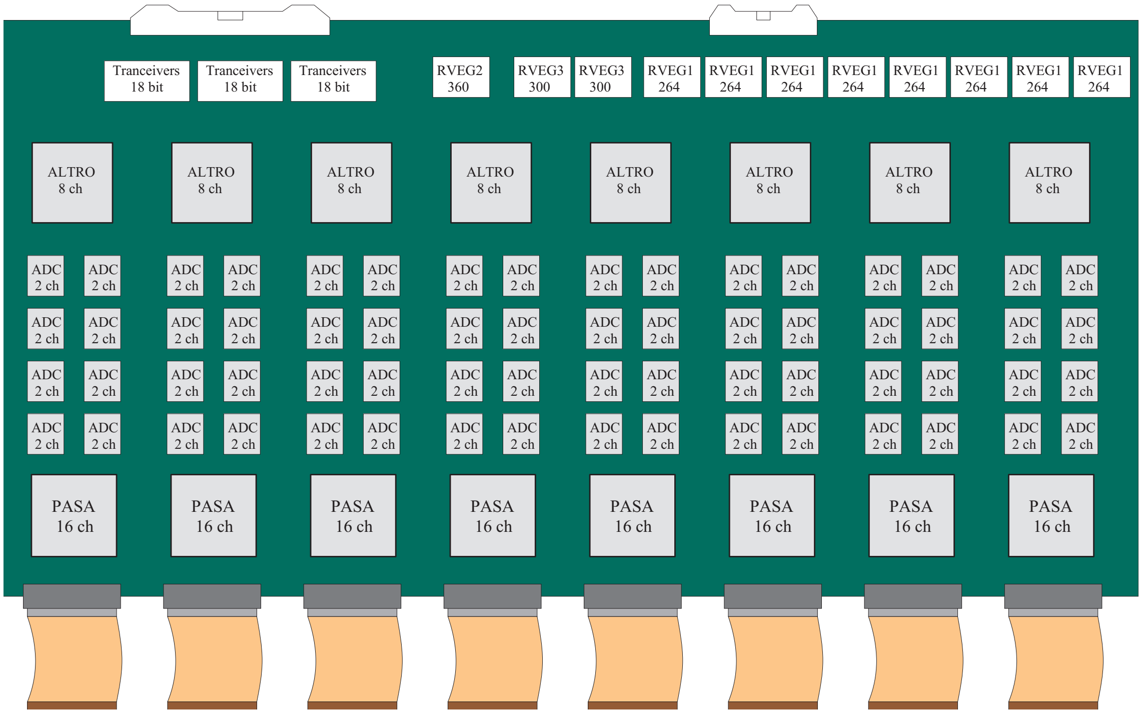

Figure 5.19: Sketch of the Front-End Card (FEC).

The digitized signals are processed and stored in a memory, where they wait for a Level-2 trigger, to be either read out or discarded. In the former case, the buffer will be frozen and protected against overwriting until the data are transferred to the Readout Control Unit (RCU). The details of the data transmission protocol are discussed in Section 5.1.5.2. In the latter case, the buffer is made available to store a new data set. All the digital circuits are implemented in the ALTRO chip that contains eight channels and the output multiplexing circuitry.

The FEC channels are multiplexed, at the board level, via a LVTTL (Low-Voltage TTL) bus. It features an asynchronous VME-like protocol, which is enhanced by a Clocked Block Transfer (CBT) that provides a bandwidth of 160 MByte/s. The FEC is interfaced to the RCU through a 40-bit bus that is based on the GTL (Gunning Transceiver Logic) standard. At the board output the bus signals are translated from LVTTL level to GTL level by bidirectional transceivers. The configuration, readout and test of the board are done via the GTL bus. However, the FEC contains a circuit, named Board Controller (BC), implemented in a FPGA, which provides the RCU with an independent access to the FEC via a field-bus. Although this secondary access is normally used to monitor the board activity, power supplies and temperature, it can also be used for the configuration and readout of the FEC at a slower rate.

The board offers a number of test facilities. As an example, a data pattern can be written into the ALTRO chip and read back exercising the complete readout chain. It is also possible to perform a boundary scan test of the ALTRO chips under the control of the BC. The BC allows the bus activities, the presence of the clock, and the number of triggers received to be verified.

The ALTRO chips and the BC work synchronously under a master clock frequency of 40 MHz. The ALTRO circuits usually perform the same operations simultaneously under the control of the RCU which can also control a single channel at a time. This is performed in the configuration phase and for test purposes. The RCU broadcasts the trigger information to the individual FEC modules and controls the readout procedure. Both functions are implemented via the GTL bus.

The board layout of the FEC can be seen as partitioned in nine different sections. The first eight sections are identical, each one containing the readout chain for a group of 16 channels. Each chain is comprised of one PASA, eight (or four) ADCs, two ALTROs and four voltage regulators. The last section contains the LVTTL–GTL translators and the BCs. As detailed in Section 5.1.8 the FEC is powered by means of four power cables carrying three supply voltages (+4.3 V, $\pm$2.5 V) and a common ground. From the three main supply voltages, for each of the eight sections, four supply voltages are derived and distributed (+3.3 V for the ADCs, +3.3 V for the digital circuits and $\pm$1.5 V for the PASA). The last section is powered independently with +3.3 V. The voltage regulators feature a power mode that is remotely controllable. The ON/OFF pin of the voltage regulators related to the same section is controlled via the BC which monitors the current supplied by each voltage regulator. It can be programmed to power down the sections where the monitored quantities exceed an upper limit.

The actual layout of the FECs has just been completed. This is an important step because it demonstrates that even for the smallest boards on the inner chamber with dimensions of only 19×14 cm^2 all components for 128 channels can fit on the board. The actual layout for the two sides of the board is shown in Fig. 5.20 and Fig. 5.21.

5.1.5 Readout bus

5.1.5.1 Requirements

The communication between the RCU and the FECs is implemented via a custom bus (front-end bus) based on a shielded ribbon cable and a custom protocol. The front-end bus is essentially an extension of the FEC's internal bus that allows the RCU to access the FEC's internal components. It is a multiple-drop single-master bus where the RCU is the master unit and the FECs are the slaves. In order to minimize the length of the bus cables the bus controller supports two branches running in opposite directions as shown in Fig. 5.22.

The front-end bus has to satisfy the following requirements.

- From the functional point of view the front-end bus is the means by which the RCU broadcasts the Level-1 trigger information to all the FECs in a crate and steers the readout procedure. Besides these basic functions, the RCU will also download the FEC's configuration parameters (pedestals, thresholds, tail-cancellation coefficients, etc.) and run test procedures via the same bus.

- To match the performance of the DDL it has to provide a bandwidth of at least 100 MByte/s.

- It must contain a number of slots that ranges from 21, for the innermost crate, to 33 for the outermost one. Physically the front-end bus is composed of two branches, each one with half the number of slots.

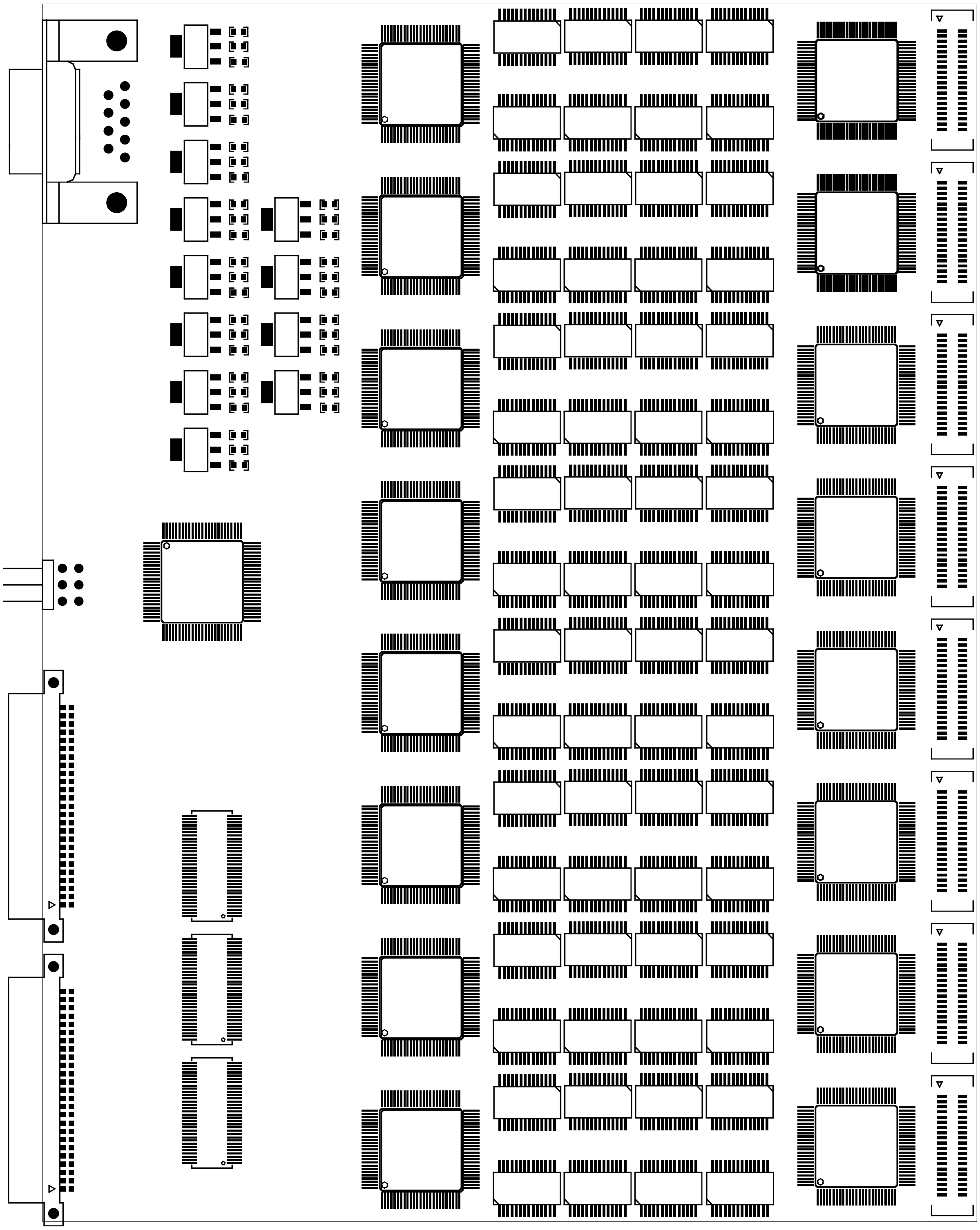

Figure 5.20: Top view of the layout of the small FEC.

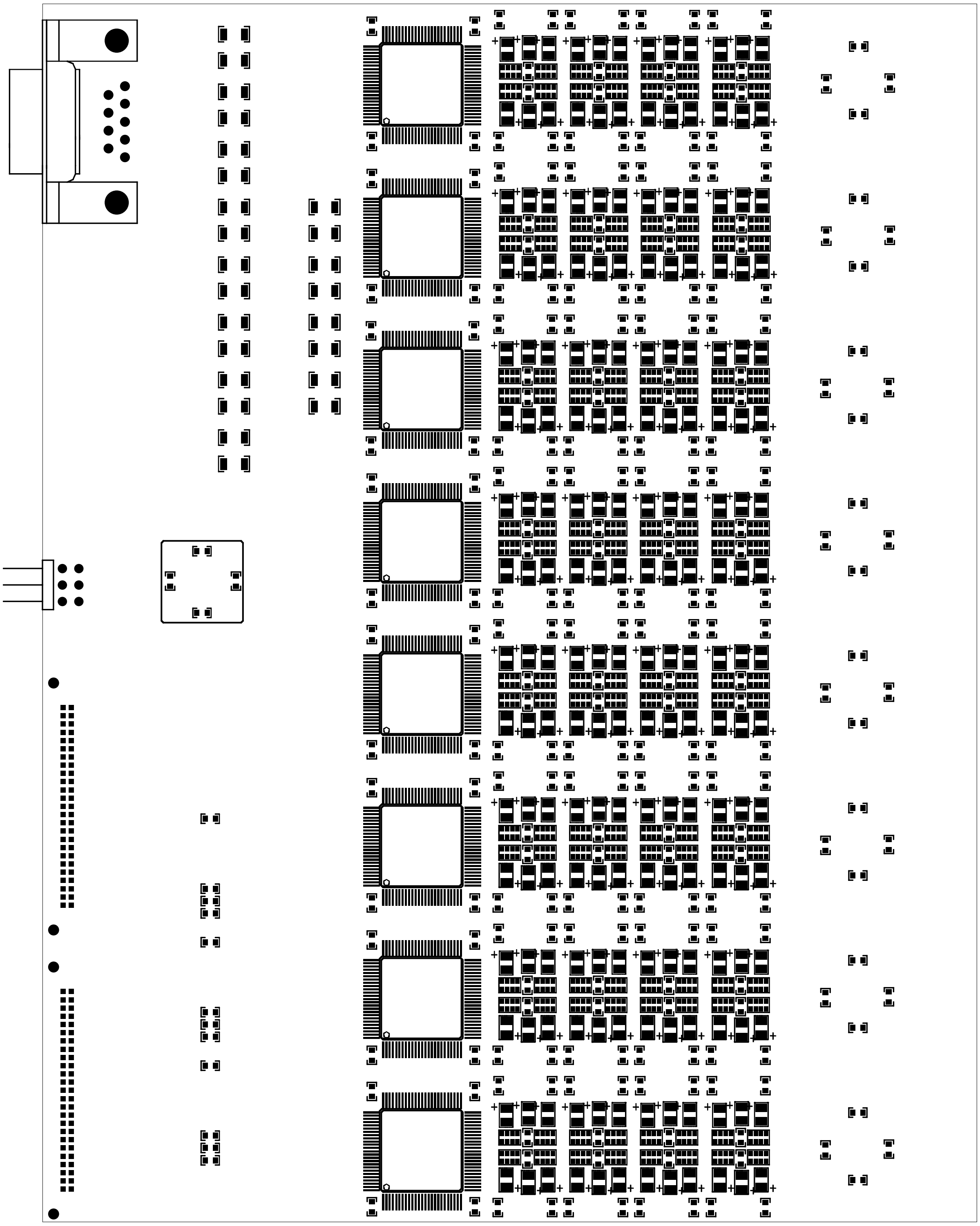

Figure 5.21: Bottom view of the layout of the small FEC.

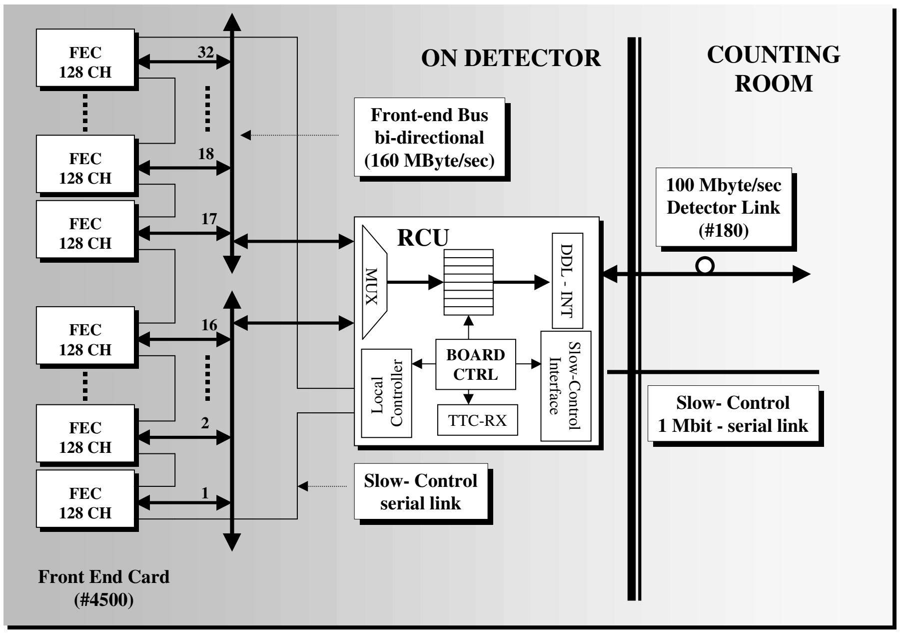

Figure 5.22: Front-end electronics system architecture.

- The front-end bus cables are very close to the analog front-end electronics; therefore low signal noise and ElectroMagnetic Interference (EMI) are an issue.

5.1.5.2 Readout protocol

The readout of one event is performed in two separate phases which are consecutive for a given event, but can otherwise be activated concurrently. In a first phase, the trigger information is received by the RCU and broadcast to all modules in the subsystem, starting the digitization of each channel, which lasts for 88 μs (the TPC drift time). During this phase tail cancellation and zero suppression are performed. In the second phase, information is moved from the multiple-event buffers to the RCU. The time needed to complete the second phase depends on the size of the event, but other triggers can be processed during the readout of the previous event, as long as the multiple-event buffers in the FEC are not full. Dead time can be generated only when this condition occurs. The transfer of data words is synchronous and modules containing valid unread data are enabled to assert data on the bus by individual addressing. To be able to perform online hardware data compression, the order in which the data are presented for optical transmission is not arbitrary. The best sequence is to send all time bins of one channel in a channel-by-channel fashion with neighbouring channels being also transmitted adjacently. However, the detailed data format may have to be revised when more detailed studies of the possible data-compression scenarios have concluded.

The front-end bus implements a VME-like protocol with an asynchronous handshake between the RCU and the ALTRO chips in the FECs. A synchronous block transfer enhances the protocol whereas words are transferred without any acknowledgement. The bus consists of 32 bidirectional lines for the transmission of data and addresses, and six control lines. Special lines are used to distribute the trigger (Level-1) and clock signals. The ALTRO chip recognizes a set of eight instructions. By means of these instructions the RCU can access each ALTRO chip in the system to

- write/read a Configuration/Status Register (CSR),
- issue a COMMAND,
- read the multiple-event buffer.

The ALTRO chip acknowledges the execution of any instruction. A special case is represented by the data readout procedure where the ALTRO acts as MASTER and strobes data on the RCU FIFO (SLAVE) synchronously without any acknowledgement. The signals used for the implementation of the protocol are BD[31:0], WRITE*, CSTB*, ACKN*, TRSF*, DSTB*, GRST*, TRG*, MCLK.

In the CSR access or COMMAND instructions, the most significant 16 bits of the BD bus specify the chip and channel addresses, the CSR address or the COMMAND code. The least significant 16 bits are used to input data in the CSR WRITE and COMMAND instructions, and to output data in the CSR READ instructions. The write/read access to a CSR is defined by the WRITE* signal. A rising edge on CSTB* (Command STroBe) informs the ALTRO chip that the addressing and data information in the BD[31:0] lines is stable and can be captured. On a WRITE* or COMMAND cycle, the ALTRO drives the ACKN* signal to indicate that it has successfully received all the data called for by the instruction type. On a read cycle, the ALTRO drives the ACKN* signal to indicate that it has placed data on the data bus. During the readout of the multiple-event buffer the chip announces the data transfer by means of the signal TRSF* (TRanSFer) which is kept low until the data packet has been completely transferred. The data transfer is not necessarily continuous and for this reason each single word being transferred is validated by the signal DSTB* (Data STroBe). The GRST* (Global ReSeT) initializes all the internal registers, counters, and state machines. The TRG* (Level-1 TRiGger) signal starts the processing of the digitized signals. The MCLK (Master CLocK) is a clock signal generated by the RCU and distributed to all FECs, which is used only by the readout circuits.

A second clock signal SCLOCK (Sampling CLOCK), with much tighter constraints on the maximum skew and jitters, is distributed via a dedicated point-to-point connection. This is the clock signal distributed to the ADCs and it has a frequency between 5 MHz and 10 MHz.

The front-end bus provides, with a clock frequency of 40 MHz, a bandwidth of 160 MByte/s on each of the two branches, leading to an aggregate bandwidth of 320 MByte/s. It is clear that a bus master-clock frequency of 20 MHz which corresponds to a bandwidth of 160 MByte/s would be adequate to match the bandwidth currently foreseen for the DDL. However a front-end bus with a higher bandwidth will allow to support possible future upgrades of the DDL performance.

5.1.5.3 Electrical and physical specifications

For the electrical implementation of the bus transceivers we plan to use GTL+. GTL, invented by W. Gunning at Xerox Corporation and standardized by JEDEC, is a low-swing input/output driver technology. It has been further modified by Intel and TI (Texas Instruments) who have increased the voltage swing to create GTL+. The typical swing for GTL+ is from 0.6 V low (VOL) to 1.5 V high (VOH) at most. TI uses tighter threshold regions to provide better signal integrity in its stand-alone devices.

It has been shown [14] that the newest TI GTL+ devices (GTL161612A) can operate up to 100 MHz, providing bit rates of 3.2 Gbit/s in a 32-bit wide backplane. They are designed with a slow rising and falling edge to offer significant system frequency improvement in heavily loaded backplanes. The GTL output consists of an edge-control circuit that provides optimized rise and fall times, typically 2.6 ns (20% to 80%), for backplanes under various loading conditions. Device power can be switched off without having to remove the device from the system because the inputs and outputs are at high impedance and, therefore, are able to tolerate active bus signals.

The basic GTL output structure is an open-drain transistor, whereas the input is a differential driver. Besides, the GTL I/Os have been designed to minimize their capacitance, an extremely important factor for distributed-load high-performance backplanes. Figure 5.23 shows the backplane electrical model for a 17-slot system located in the outermost crate (C5).

For the TPC front-end electronics a rigid backplane is not a pratical solution as the FECs are mounted on the detector first, and the backplane should be plugged onto 17 boards simultaneously. Therefore we plan to implement the front-end bus in a polyimide flexible cable. The 40 bus lines are alternated with ground lines, with a pitch of 1.27 mm, and lie above a solid ground plane. The electrical performance of

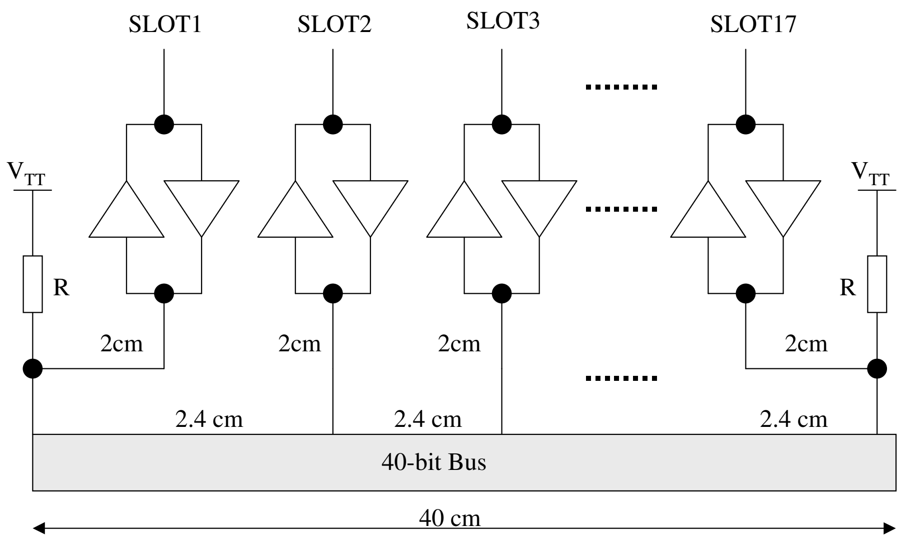

Figure 5.23: Backplane model for one branch of the front-end bus for the outermost crate.

this system is equivalent to a standard rigid backplane.

5.1.6 Readout control and services

The overall TPC electronics architecture is sketched in Fig. 5.22 on page 124. The electronics for 128 channels is mounted on a Front-End Card (FEC) as described in Section 5.1.4. The functionality performed here includes, aside from amplification and digitization, digital filtering, zero suppression, and buffering of up to four black events. For a motivation why four events are adequate refer to Section 5.1.7. Basically the FEC hosts all electronics necessary for one channel. In addition, it also performs necessary power management, which has to be as close as possible to the analog circuitry.

There will be a total of about 4500 FECs. Assuming black events and the maximum 200 Hz TPC readout each card will produce a data stream of about 11 MByte/s if the data not falling into the nominal acceptance (32%) are not read out (otherwise 16 MByte/s). The accepted ALICE data transmission standard off the detector is DDL [15, 16] supporting 1 Gbit/s. Therefore multiple TPC FECs can be merged into one DDL. This functionality is performed by the readout control unit sketched in Fig. 5.22 on page 124. However, depending on the radial position, not the total depth of the chamber will be read out and also position-dependent compression factors are expected. Therefore the number of FECs per readout controller varies from 20 to 32 such that the average throughput per link is constant. There are going to be five readout controllers per sector.

The readout controller has the following functional requirements

- merge digitized data from FECs into the DDL data stream in ordered form (time bin, pad number, pad row);
- distribute trigger and clock signals to FECs;
- interface detector control system (DCS) to FECs;

The following paragraphs discuss the functionality and interfaces of this device.

5.1.6.1 Clock and trigger interface

The accepted ALICE standard how to distribute trigger and clock information is the RD12 TTC system which will be used here [17]. Every readout controller implements an appropriate receiver (TTCRX) which produces the phase-corrected LHC clock, Level-1 trigger and Level-2 accept/reject; these have to be forwarded to all FECs via the readout bus. Due to the multiple-drop nature of the readout bus there will be some clock skew between the different FECs. With a cable length of less than 30 cm per readout bus link the clock skew between any FEC relative to the TTCRX chip which can be individually phase corrected, can be kept below 5 ns with less than 200 ps electronic jitter. This corresponds to a systematic position error of less than 200 μm with a contribution to the position resolution of less than 10 μm.

In addition, a 4-bit trigger type needs to be provided to the FECs which can be used to issue pulser or test readout triggers, such as the readout of a pre-loaded event for test purposes of the data path. These triggers can be issued as broadcast commands through the TTC. Normal physics Level-1 triggers have the default readout type zero.

ALICE implements a multiple-deadtime system allowing triggers to some subdetectors while others, such as the TPC, are busy. Therefore, an event number is not an unambiguous means to identify an event and to ensure data integrity. The TTC receiver provides a bunch-crossing number which, however, is only unique for one LHC turn. Therefore it was decided to use the combination of the bunch-crossing number and an orbit counter. Here 24 bits correspond to 20 minutes of real time, which is longer than any event is expected to remain within the DAQ and therefore adequate to ensure uniqueness of the event ID. The orbit counter has to be implemented externally to the TTCRX and be able to run at 12 kHz.

5.1.6.2 Interface to the digital optical links

The ALICE standard to transmit data off the detector is the DDL technology developed by the DAQ group [15, 16]. It implements a gigabit optical link which is driven by a Source Interface Unit (SIU) on the readout controller. Since this device is being used by multiple subdetectors this SIU will be built by the DAQ group and be implemented as a mezzanine card on the readout controller. The protocol between the readout controller and the SIU is a synchronous 40 MHz 32-bit bus.

The readout controller will start sending an event as soon as it has received a Level-2 accept for the event guaranteeing that the event will not be aborted. Since the order of events is maintained by the Level-2 accept or reject commands no special event identification logic is required but rather the queued events can be handled in the order they were triggered. The whole architecture can be implemented as a simple push architecture. Any Level-2 decision received by the TTC will be forwarded immediately to the FECs which then will start sending the next event upon an accept or will free the appropriate buffer in their round-robin buffer scheme. Therefore the readout controller acts mainly as a multiplexer for the optical transmitter (SIU).

The data format to be shipped is variable length. Every event will be encapsulated by at least three command/status words, the begin-of-event, the begin-of-data and the end-of-event commands. The begin-of-event command is followed by a fixed-length header that further defines the event type and size. Following standard practice, the first bytes of the subevent header define the total subevent size. The second 32-bit word defines the subevent type and subtype including some version number. In addition the event header has to contain the trigger type, event ID (the bunch-crossing number and orbit number), and the readout controller ID. Following the header are the zero-suppressed raw data. The exact header data format is not critical and will be defined jointly by the TPC and DAQ groups.

5.1.6.3 Detector control related interfaces

The readout controller has to have two detector control related interfaces. One is the connection to the FECs, which is required in order to configure and monitor the devices, and the other is the DCS interface itself. The functional requirements for the DCS interface are the following

- configure power state on all FECs;
- monitor power and temperature on all FECs;

- read/write any configuration parameter in the front-end, including zero-suppression thresholds, and configuration of the digital filter;
- asynchronously fire exceptions or warnings out of the front-end;
- read status parameters, such as error conditions;
- upload an event into the front-end for testing;
- download an event from the front-end as redundant fall-back.

The largest data items for up/download are complete events. However, since this is done for testing purposes only, there are no particular performance requirements here and it would be acceptable if the event data transmission took several seconds. All other parameter sets are comparatively small. The slow control interface has to be full duplex supporting the ability to address individual registers in the front-end. It is the goal to implement any front-end configuration register in read/write mode so that the as is status can be read back and verified. Further, the front-end needs to be able to assert asynchronous events such as alarms like over-power, current or temperature.

One accepted ALICE standard for slow control is the fieldbus CAN (ISO 11898–2). The maximum CAN data transmission rate is 1 Mbit/s at a maximum distance of 40 m. This would correspond in case of a central Pb–Pb event to less than 20 seconds for the readout of a complete sector. CAN is a multi-master multiple-drop serial bus implementing the CSMA/CD arbitration. There are low-cost, single-chip devices available today, such as the Siemens C167 16-bit microcontroller, which have an integrated CAN interface. This or a similar microcontroller will be hosted on the readout controller to perform all necessary detector-control functionality. The particular protocol to address the various control and status registers is not yet defined in detail as is the case for the CSR registers. This will be done in collaboration with the detector control systems group. There are no particular technological challenges. In order to reduce the single point of failure risks there will be one CAN bus per TPC sector.

The readout controller needs to be able to communicate commands to and receive status from the FECs. Most of the control and status registers will be part of an FPGA on the FEC. Therefore a simple synchronous serial interface would be adequate and would allow to keep the number of signals and cables to a minimum. This simple interface will be part of the programmable logic on the FEC and also be implemented on the readout controller as interface to the CAN microcontroller. The details of this serial interface are not defined yet and will be finalized as the prototyping of the FEC progresses. Given the speed of less than 10 MHz, there are no particular technological challenges here.

5.1.7 Front-end electronic readout efficiency

This section deals with the readout efficiency of the front-end electronics. In particular, it will be shown how the system dead time depends on the dimensions of the front-end multiple-event buffer.

We assume here that the TPC will be operated at a rate (Level-1 trigger) of about 200 Hz for Pb–Pb central collisions. At the receipt of a Level-1 trigger, the front-end electronics processes the TPC signals for 88 μs, storing the zero-suppressed data into a multiple-event buffer. We assume that the Level-2 trigger (accept/reject) has a fixed latency which corresponds to the TPC drift time. It should be noted that, owing to the trigger past–future protection, for the TPC a new Level-1 trigger only comes after the Level-2 trigger related to the previous event.

In the case of a Level-2 accept, the memory region where the event data has been stored is protected against overwriting until the data has been shipped to the RCU. Otherwise, in the case of a Level-2 reject, this region is immediately freed to store a new data set. Therefore the multiple-event buffer is not a pipelined memory but rather a memory subdivided in regions (buffers) to be accessed randomly. To avoid fragmentation each buffer is deep enough to contain a black (non zero-suppressed) event. When all memory regions are occupied the front-end electronics generates dead time.

The dead time generated by the TPC has therefore two contributions: detector dead time, i.e. the drift time, and front-end electronics dead time (readout dead time). While the multiple-event buffer scheme can reduce the second contribution, it cannot eliminate or reduce the first. This obvious fact implies that the total dead time has a lower limit determined by the detector drift time.

In order to study the readout performance of the front-end electronics and to dimension properly the multiple-event buffer, a Verilog model of the front-end chain has been created. Verilog is a hardware description language suitable for the stochastic analysis of digital systems. The results are displayed in Figs. 5.24 and 5.25 which show how the system dead time depends on the number of buffers for two different Level-1 rates: 1 kHz and 200 Hz, respectively. The simulations assume a Level-2 rate of 100 Hz and an average occupancy of 25%.

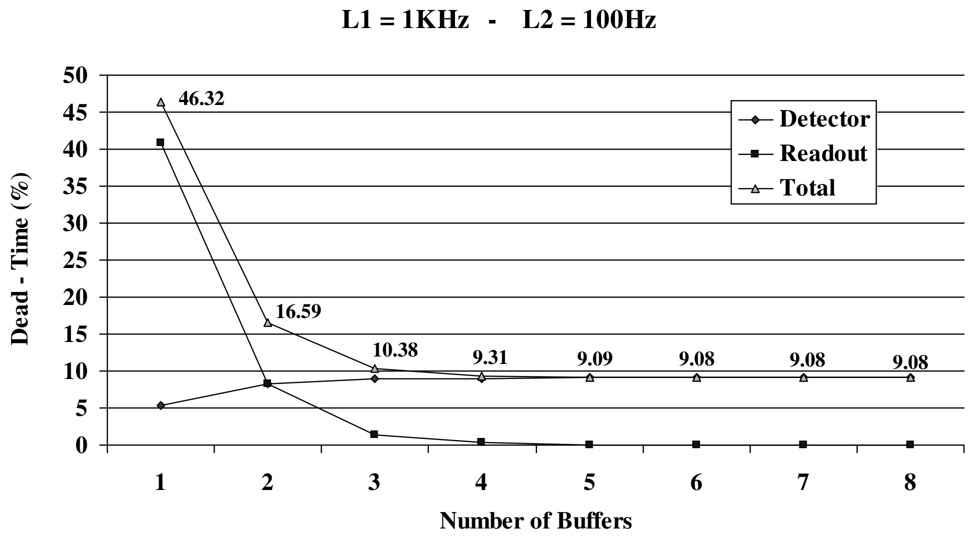

Figure 5.24: TPC dead time versus the number of buffers in the front-end memory. The simulation is made under the following assumptions: Level-1 rate = 1 kHz, Level-2 rate = 100 Hz, average occupancy = 25%.

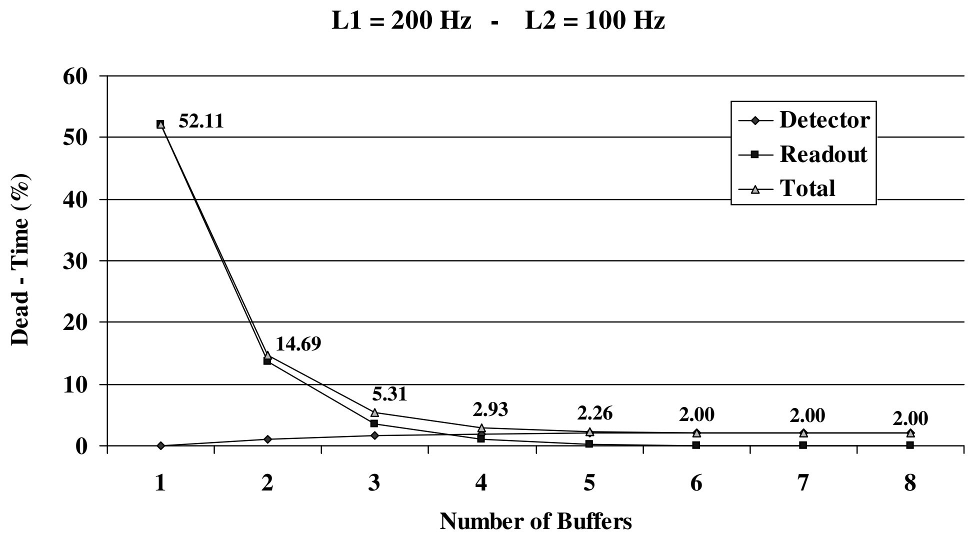

Figure 5.25: TPC dead time versus the number of buffers in the front-end memory. The simulation is made under the following assumptions: Level-1 rate = 200 Hz, Level-2 rate = 100 Hz, average occupancy = 25%.

5.1.8 Low-voltage power supplies and distribution

For the calculation of the number of low-voltage supply lines for the FECs and their cross section, it is assumed that each card dissipates at most 100 mW/channel of power (in total 12.8 W), where 20% are consumed by the preamplifier/shaper at ± 1.5 V and 80% are consumed by the digital part at $+3.3$ V. In total the supply system has to supply about 21 kA of current. It is assumed that the power supplies are located outside the L3 magnet on both sides of the TPC in the area not accessible during LHC operation. The cable length required in this scenario is 40 m.

To avoid cooling of the cables but keep the cost of cables moderate a voltage drop of 0.5 V has been assumed as a design value leading to acceptable levels of heat dissipation in the L3 cavern. In Table 5.8 the characteristics of the cables are listed.

Table 5.8: Characteristics of the low-voltage supply cables. The voltages are nominal voltages at the load. The power dissipation quoted refers to the power dissipated in 40 m of cables.

Parameter	Analog supply	Digital supply
Nominal voltage	$+1.5$ V/-1.5 V	$+3.3$ V/GND
Cross section	2×4 mm^2	2×10 mm^2
Current	2×0.85 A	2×3.1 A
Power	1700 W	6200 W

Two scenarios are considered. In the first the power for each FEC is provided individually via a separate cable with four wires and a common shield (in total 2232 cables per end-plate). This assumes that the cables have to pass through the existing gaps at the L3 magnet doors with its chicane and thus have to be flexible enough to accommodate a bending radius of less or about 5 cm. If the openings can be enlarged the need for flexible cables is alleviated. In the second scenario, therefore, the supplies for about five cards are combined into a thicker, less flexible but more economic cable. There would be a transition from the thick cable to individual cables at a patch panel mounted at the service support wheel (see below).

To allow access to the inner tracking system the TPC and its supply lines have to accommodate movements of about 4.5 m along the beam direction. Due to space limitations no distribution scheme allowing such a movement while at the same time keeping the supply lines permanently connected could be found. It is therefore foreseen to connect the low-voltage cables via patch panels mounted on the service support wheel. For service interventions all cables have to be disconnected at the patch panel.

The cross sections of all low-voltage cables connecting the patch panel with the individual FECs can be reduced easily by a factor of two compared to the 40 m long main sections. The estimated power dissipation close to the end-plate is then 200 W per end-plate.

To avoid ground loops it is foreseen that the power supplies are left floating on the supply side and connected to ground only at the detector side. A schematic layout of the low-voltage system with its grounding scheme is shown in Fig. 5.26. Since each FEC is connected to the pad plane with eight cables, each of which carries the ground for each of the 16 channels plus a common shield, a relatively large cross-section is available in case of an interruption of the ground return line.

The DCS will monitor the voltages on the FECs before regulation. If a voltage drop is detected the power supplies can be powered down sufficiently fast (on a time-scale of ms) to avoid damage due to excessive currents.

With regard to the stability of the power supplies no special requirements are necessary since the final regulation is done on the FECs themselves. Quite strict limits on noise have to be adhered to the given sensitivity of the preamplifier/shaper part. The cables are therefore shielded between the power supplies and the patch panel. At this point no analysis has been performed regarding the best grouping of power-supply lines and therefore the maximum output currents required per supply module.

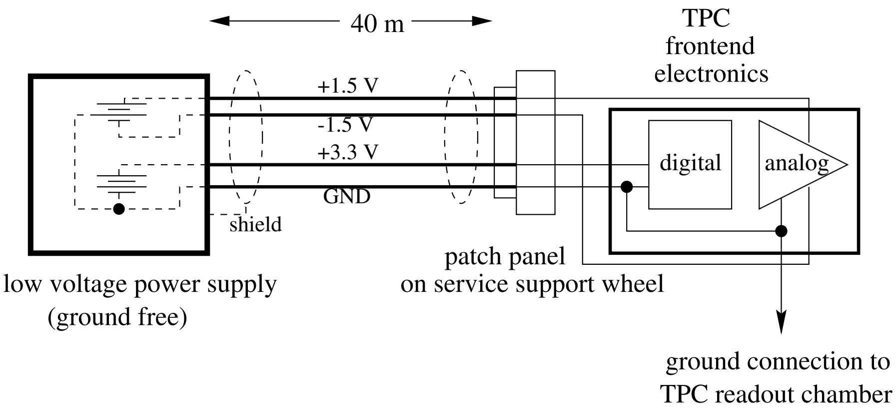

Figure 5.26: Connection of the low-voltage power supplies to the front-end electronics of the TPC.

5.1.9 System implementation

5.1.9.1 Quality assurance

The inaccessibility of the ALICE experiment during the entire year of LHC running makes stringent quality tests of the readout electronics mandatory before installation. To minimize the failure rate all FECs will be tested in a dedicated test set-up. A schematic diagram of this tester is shown in Fig. 5.27. An important part of the test procedure is a burn-in process, which will be made in an oven at a temperature of 50 °C for a period of approximately 10 hours. All crucial functions of the circuits will be cycled according to a preselected scheme. During the 10 hours of burn-in typically thousands of cycles will be completed. The results of the tests will be automatically logged to file for later investigation of the errors occurring. A malfunctioning component will be replaced with a new one and the full test will be

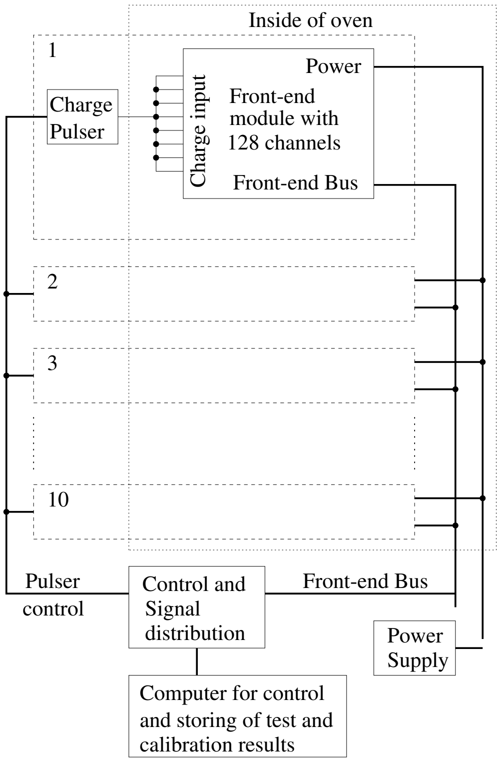

Figure 5.27: Functional diagram of the FC tester with burn-in feature.

repeated before accepting the board for installation. Special care has to be taken to control the external noise influence on the analog circuits.

The test set-up foreseen will hold ten units to be tested and two test rounds could be completed per 24 hours. A total number of 570 000 readout channels grouped in about 5000 FECs (including 10% spares) then requires about 240 days to complete the burn-in acceptance test using one test station. It is foreseen to duplicate the test set-up in order to speed up the burn-in process.

Once the boards are accepted and installed on the TPC an *in situ* testing procedure has to be implemented. All the digital parts can be tested through the DAQ and/or the DCS. To test the analog part (preamplifier/shaper and ADC) it is necessary to include the TPC itself or use the calibration pulser. During operation the full system has to be monitored for malfunctioning channels.

5.2 Readout

The ALICE Trigger/DAQ architecture was initially designed for a set of requirements described in the Technical Proposal [3]. These requirements and the corresponding Trigger/DAQ architecture have been relatively stable since then and have been refined for the preparation of the first TDR. This architecture is now facing two important modifications of the original requirements. First, the occupancy of the TPC is expected to be higher than originally assumed. Second, a new trigger detector (the TRD) has been introduced into the ALICE detector system. Both modifications require revisiting the baseline architecture in order to ensure that the changing requirements can be met by the Trigger/DAQ architecture. In the new scenario the TPC and TRD detector data are sent directly into the host processors of a Level-3 Trigger/DAQ processor farm. The system is scalable in terms of processing power and bandwidth in order to meet the computing requirements of the various physics programmes. It is absolutely mandatory that zero-suppressed raw data are recorded without any further processing, especially in the first year of running in order to ensure a complete understanding of the TPC response. The intelligent readout system (Level-3) guarantees the readout of unprocessed full TPC events at a rate of 10 Hz. Having understood the TPC response to central Pb–Pb collisions, the Level-3 system will take over more functionalities. Almost lossless data compression and selective readout for TRD events can be implemented easily, since these techniques do not need large computing power. Finally, by adding more computing power, online track finding in the whole TPC and therefore very effective data compression methods and TPC-based selective readout schemes become feasible. Should there be a need for a higher rate of unprocessed full TPC events, the bandwidth of the system can be increased. This system is referred to as Level-3 trigger, Level-3 filter and Level-3 processors in various places, depending on the functionality addressed in the context.

Extensive simulations of track-finding methods and data flow have been performed for the proposed Level-3 Trigger System of the STAR [19] experiment at RHIC. A large scale prototype of an Level-3 system is currently being installed. There is a large overlap in the physics objective, detector design, readout electronics and personell between the STAR and the ALICE Level-3 systems.

5.2.1 Physics requirements

5.2.1.1 Event size

The latest simulations for central Pb–Pb collisions indicate that the average TPC occupancy will be about 25% (see Table 5.13 on page 142 in Section 5.2.3). The readout chambers have a granularity of 570 000 pads, the number of time buckets per channel is assumed to be 512. The exact number of time buckets will be determined after the first Pb–Pb collisions have been recorded and the cluster shapes have been analysed. Depending on occupancy it may increase to 768 at the most. The 10-bit amplitude can be compressed into a nonlinear 8-bit number. This corresponds to a raw event size of 292 Mbyte. The event size can be reduced by roughly 30%–50% by removing the two inner 45° cones of the TPC volume during readout. Different solutions for coding the zero-suppressed data are currently being investigated. Assuming a 15%–40% overhead for encoding the zero-suppressed data yields an event size of about 66 Mbyte $\pm$15% per central collision. Minimum-bias events are assumed to be smaller by roughly a factor of two (in NA49: central Pb–Pb 100%, medium bias 85%, p–A 30% and pp 20%).

5.2.1.2 Event rates

In general, the required central event samples for various physics topics vary between $< 10^6$ for most inclusive hadronic observables to $> 10^7$ for rare processes such as open charm and open bottom production, quarkonia production, and the detection of direct real and virtual photons [?, 3, 22, 23]. The detailed event rates to be taken undergo continuous discussion within the collaboration.

In the following we present, for a quantitative evaluation of the proposed readout schemes, an illustrative set of event rates expected for hadron and electron physics in the TPC. However, the proposed readout scheme is sufficiently general that physics dictated variations in the rates should be accomodated without changes in the architecture.

An event rate of about 1 Hz leading to more than 10^6 events per ALICE year is currently considered sufficient for most hadronic observables. An example for an observable that would require higher event rates ($\geq$ 10 Hz) is the reconstruction of D mesons via their hadronic decay products. The ALICE DAQ was originally dimensioned to handle up to 50 Hz of events with full readout and in addition up to 1 kHz of muon triggers, with a total bandwidth to mass storage of up to 1.3 Gbyte/s. With the introduction of the TRD and the increase of occupancy in the TPC, the event size has grown substantially since the publication of the Technical Proposal and would currently limit the data rate to less than 30 Hz within the foreseen bandwidth.

As discussed above, the final bandwidth of the system can be increased beyond 10 Hz if the corresponding physics requirements are there. However, by introduction of the Level-3 system, online tracking will make it feasible to perform partial readout, event rejection and sophisticated data compression. It can significantly reduce the event size for full TPC readout (potentially by more than one order of magnitude) and/or reduce the event rate (by online filtering) and therefore restore the originally foreseen rates within or below the originally foreseen bandwidth to mass storage. Typical event rates and the TPC event size for diffent compression schemes (discussed in detail below) and different trigger samples are given in Table 5.9.

The required event rates for the full dielectron programme and an improved open-charm programme with a Level-3 trigger are as follows.

Full dielectron programme:

- TRD triggers TPC readout at $\simeq$ 60 Hz;
- Selective inspection of a few TPC sectors;
- Online processing of electron track candidates, limited processing compared to full TPC tracking;
- Output rate is $\simeq$ 30 Hz, event size 0.5 Mbyte.

Improved open charm (hadronic decays) programme:

- TPC readout at 10 Hz–100 Hz;
- Online processing of full TPC data, large processing power needed for cluster and track finding;
- Selecting high-transverse-momentum tracks;
- Output rate is 10 Hz–100 Hz, event size is 4 Mbyte.

Table 5.9: Typical TPC event rates and data rates. It has to be noted that all other detectors contribute 10 Mbyte to the size of an event for a central collision. The muon programme will increase the bandwidth by approximately 250 Mbyte/s.

	No. of events	Event rate [Hz]	Event size [Mbyte]	Bandwidth [Mbyte/s]
1 — no Level-3, full TPC readout				
Hadronic observables	10^6	1	66	66
Open charm (hadronic decays)	10^7	10	66	660
2a — Level-3, lossless data compression				
Hadronic observables	10^6	1	47	47
Open charm (hadronic decays)	10^7	10	47	470
2b — Level-3, slightly lossy data compression				
Hadronic observables	10^6	1	19	19
Open charm (hadronic decays)	10^7	10	19	190
3 — Level-3, TRD trigger and selective TPC subsector readout				
Dielectrons	6×10^7	60	1 – 2	60 – 120
4 — Level-3, TRD trigger, selective TPC tracking and event rejection				
Dielectrons	3×10^7	30	0.5 – 2	15 – 60
5 — Level-3, TPC tracking and high-p_t track selection ($>$ 0.5 GeV/c)				
Open charm (hadronic decays)	10^7	10	20	200
6a — Level-3, TPC track/cluster modeling and full data compression				
Open charm (hadronic decays)	10^7	10	4	40
6b — Level-3, TPC tracking and data compression, DE tracks, TPC summary event for correlations				
Dielectrons	3×10^7	30	5	150

5.2.2 Data rate reduction

From a trigger point of view the detectors in ALICE can be divided into two categories: fast and slow. Fast detectors provide information for the trigger system at every LHC bunch crossing. Decisions at trigger Levels-0, -1 and -2 are made using information from these detectors. Fast detectors are the Forward Multiplicity Detectors (FMD), the Zero Degree Calorimeter (ZDC), the Transition Radiation Detector (TRD) and the muon trigger chambers.

The slow detectors are tracking drift detectors and need a longer time span after the collision to deliver their data. The Time Projection Chamber (TPC) and the Inner Tracking System (ITS) are slow detectors. Their slowness is compensated for by the detailed information they provide. The Level-3 system is intended to take advantage of TPC information (up to 66 Mbyte/event at rates of up to 200 Hz) in order to reduce the data rate as far as possible to have reasonable taping cost. The data are then recorded onto an archival-quality medium for subsequent offline analysis.

A key component of the proposed system is the ability to process the raw data — performing, for example, track finding in real time — besides its baseline requirement of forwarding zero suppressed events into the DAQ data stream. Level-3 is designed to utilize the information from the TPC and the fast detectors, e.g. the TRD. The system will be flexible enough to be expanded to include other tracking devices.

Data reduction can be achieved in different ways:

- Generation and application of a software trigger capable of reducing the input data stream.

- Reduction in the size of the event data by selecting subevents. By analysing the tracking information or by utilizing TRD information, Regions Of Interests (ROI) can be defined. The data volume is significantly reduced by recording only summary information and raw data of the ROIs (e.g. electron tracks).

- Reduction in the size of the event data can also be achieved by compression techniques. General lossless or slightly lossy methods can compress TPC data by factors 2–3 [21]. By online tracking and a compressed data representation an event size reduction by a factor of 15 can be achieved [24]. This data compression method is based on a cluster model and a local track model. It is assumed that all TPC information can be reduced to a local track with its helix parameters and a mean charge, the corresponding cluster information (charge per cluster, cluster position) is stored as the (quantized) deviation from the local model. The result may be further compressed by a Huffman coding scheme (see e.g. [25]). Since all necessary information is preserved, TPC data can later be once more reconstructed by the offline chain. This method may be useful in the later phase of the experiment when the detector and the reconstruction performance are well understood.

5.2.2.1 Volume reduction

The raw data volume per event can be reduced by compression techniques or by selective or partial readout. These subevents contain only the raw data of a few tracks. Zero suppression is performed in the front-end electronics by pedestal subtraction (a threshold operation) and one-dimensional (time direction) hit-finding. The resulting long zero sequences between hits can then be compressed by, for example, run-length encoding, which is equivalent to storing only the hits and their positions. This technique achieves compression ratios which are roughly inversely proportional to the TPC occupancy. The thresholding and hit-finding operations are lossy techniques which could lead to a loss of small clusters or tails of clusters. The following encoding methods of hit sequences are lossless. All these operations are performed on the front-end electronics board (see Section 5.1.3.5).

Selective or partial readout

Subevents or ROIs can be defined on the basis of rough tracking information, including a PID supplied by the TRD. All raw data inside these regions are written on tape and all other data are dropped. For the lepton measurements the data volume can be reduced to candidate e^+e^- tracks, which would yield a few tracks per event.

Data compression

General data compression techniques have been applied to TPC data, i.e. both the NA49 raw data [21,26] and the STAR simulation data [19,20,27]. Zero suppression reduces the TPC data by a large margin (depending on occupancy). Lossless transformations like variable length codes (e.g. Huffman coding) or lossy compression methods like vector quantization can compress the ADC data by factors of up to 3. The data volume can be reduced even further by using data modelling techniques, and storing only quantized differences to a data model, i.e. a cluster and a local track model. This results in reduction factors of about 15.

Lossless compression – variable length coding The crucial observation to reduce the size of the data is that the ADC values are not equally probable. Small ADC values occur very often in the data stream but larger ones are rare (Fig. 5.28). The distribution is approximately exponential. The expected size of the data can be reduced if short words are used for frequent values and longer ones for rare values. Another advantage of this technique is that one is no longer restricted to 256 symbols — as one is with bytes — or powers of two. One can choose as many codes as necessary. This means that if one enhances

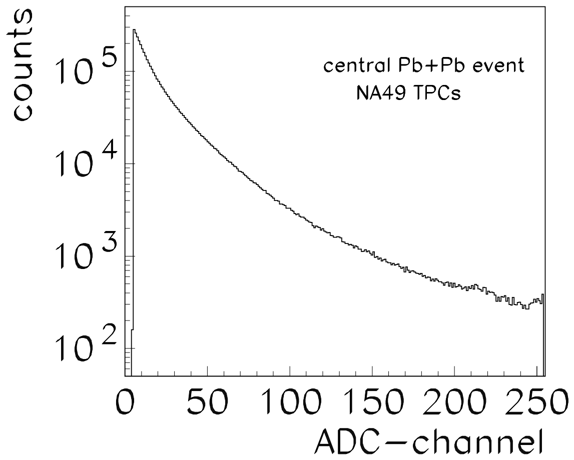

Figure 5.28: Distribution of ADC values for a central Pb–Pb collision.

the ADC resolution from 8 to 9 bits the size of the data is no longer doubled, but increased by 1/8 in the worst case.

There is a theoretical lower bound on the average word size that can be achieved by this strategy. This lower bound is called entropy E of the data source and can be computed as

$$E = \sum_{x \in \Sigma} p(x) \log p(x), \tag{5.5}$$

where Σ is the set of all possible words that are output by the data source, p is the probability of the occurrence of an ADC value x. It can be shown that this bound is tight for stochastic data sources [28]. There are compression techniques that approach this bound arbitrarily close. An example is arithmetic encoding, which is rather complex and also patented. An alternative to arithmetic encoding are Huffman codes. They are easier to implement, need less processing power and achieve good results unless the distribution of input values is too extreme. It is important to note that the average size of an event will be reduced by this approach, but one has to consider that sometimes events can become larger than uncompressed.

A TPC is obviously not a stochastic data source, as adjacent ADC values are highly correlated. Therefore it is possible to compress the data to a lower bit rate than the entropy of the ADC values. In other words, there are representations of the TPC data that have a lower entropy than the formats described above. Various methods like differentiation, prediction, etc. have been evaluated, but none of these approaches yields results that are much better than the plain Huffman coding [25, 29].

Lossy compression There is a some gain in using the lossless techniques such as simple Huffman encoding. Better results can be achieved when very small, noise-like changes of the data are tolerated. This leads to the so-called lossy compression schemes. One kind of lossy compression, the zero suppression together with a run-length encoder, is already in use at the front-end. The following section focuses on the further reduction of the remaining data.

- **Vector quantization**
  Vector quantization [30] is a sophisticated type of quantization. Here, statistical dependencies between successive data samples are exploited: Instead of quantizing data samples independently, several samples are grouped together to form a vector of data samples. Then this vector is compared to entries in a codebook and the number of the best matching vector from this codebook is

output, where best matching means lowest distance using Euclidean or any other metric, as best suited for the application. If the distribution of the input data is known in advance, the vector quantizer can produce data with a very low distortion relative to the bitrate. To be optimal, the codebook has to be trained on the statistic properties of typical input data. This is usually done before the codebook is taken into use with an algorithm known as the modified LBG algorithm [31]. The adapted codebook then remains unmodified throughout the actual quantization.

- **Vector quantization with residual encoding**
  Though this approach offers a very low bitrate, it is obvious that there is almost no possibility for the vector quantizer to change behaviour: Since the codebook is preproduced, only the given vectors are available to represent the output data. Even if the codebook were well adapted to the data, sometimes large quantization errors could occur (Fig. 5.29). To prevent this, we need to store the differences between the input data and the selected codebook entry, the so-called residuals. These residuals are then quantized and entropy encoded to achieve an even lower bitrate. A quantization of the residuals is especially effective, since the distribution of errors between codebook entry and input data is very steep: If the codebook is sufficiently trained, the vector quantizer will already absorb much of the input signal energy, so the residual encoder will mostly get values around zero. When these values are quantized, even more values are mapped to zero. Here, arithmetic compression [25] is the method of choice.

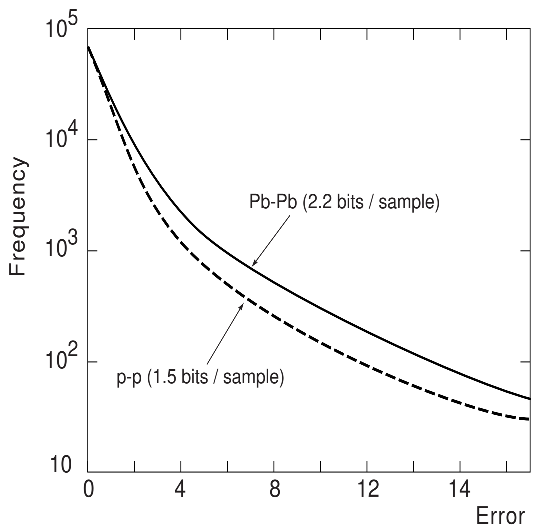

Figure 5.29: Error distribution of vector quantization.

- **Cluster and track Modelling**
  The best compression method is to find a good model for the raw data and to transform the data into an efficient representation. Information is stored as model parameters and (small) deviations from the model. The relevant information given by a tracking detector is the local track parameters and the clusters belonging to this track segment. The local track model is a helix; the knowledge of the track parameters helps to describe the shape of the clusters in a simple model [33, 34]. The track finder reconstructs clusters and associates them with local track segments. Note that track recognition at this stage can be redundant, i.e. clusters can belong to more than one track and track segments can overlap. Once the track recognition is completed, the track can be represented by helix parameters. These are curvature r, starting point (x,y,z), dip angle λ, azimuthal angle φ, track length, average charge, χ^2 of the helix fit and the number of clusters belonging to this track segment (see Table 5.10).

Table 5.10: Track parameters

Parameter	Size [byte]	Type
Curvature r	4	float
Begin x	4	float
Begin y	4	float
Begin z	4	float
Dip angle λ	4	float
Azimuthal angle φ	4	float
Track length	2	integer
Cluster charge (average)	2	fixed point
χ^2	2	fixed point
Number of clusters	1	integer
Total	31	

Table 5.11: Cluster parameters

Parameter	Size [bit]
Flag empty cluster	1
Δ time	6
Δ pad	6
Δ cluster charge	7
Δ shape	4
Total	24

In a second step, the deviation of the cluster centroid position from the track model (residuals), the deviation from the average charge and deviations from the expected shape (based on the track parameters) a re calculated for each cluster. These numbers are then quantized by a nonlinear transfer function adapted to the detector noise and detector resolution (see Table 5.11). Remaining clusters can be optionally kept as raw data arrays.

The compression method discussed above allows for a later second pass of calibration and distortion corrections, track and vertex finding and fitting, and dE/dx analysis. Since the aim of the track finding is not to extract physics information but merely to build a data model, which will be used to collect clusters and to code cluster information efficiently, any inefficiencies in track finding e.g. due to an unprecise track model will result in an inefficient compression, but not in a loss of clusters; no relevant data are lost. By fast and redundant online track recognition, reduction factors of about 15 can be achieved [24].

Experimental results The described algorithms were applied to a set of TPC events from the NA49 Pb–Pb collions as well as to pp data. The results for central heavy-ion collisions are shown in Table 5.12. Compression factors for pp data with a much lower occupancy are higher by 30%–40%. In our experiments we used a vector quantizer of length three, so that two vectors model the majority of hits of lengths up to six time bins. The size of the codebook is 256 entries (VQ3). This leads to a data rate of 2 bit/sample for the vector quantizer alone. The algorithm RVQ3 is a vector quantizer of length three with quantized residuals. Allowing an absolute error of one in the residual quantization, the change of the number of clusters was less then 10^{-4} and no change in the number of tracks was observed. First preliminary test with ALICE simulated data are in accordance with the above results [16].

Table 5.12: Performance of different compression algorithms on NA49 TPC data.

Type of encoder	Entropy bit/sample	Relative event size Pb–Pb (NA49) [%]
Zero suppressed raw data	8	100
Huffman	5.8	72
Transformations e.g. differentation (gif)	5.3 – 6.1	66 –76
RVQ3 lossless VQ Residual value	4.8 2.3 2.5	60
RVQ3 lossy (error 1 ADC value) VQ Residual value	3.8 2.3 1.5	48
VQ3 lossy	2.3	29
Cluster and track Modelling		7

5.2.2.2 Event rate reduction — Level-3 trigger

Based on the results from the TRD analysis and the online tracking information, the (sub)event building can be aborted. The Level-3 system is responsible for deriving such a trigger decision. The relevant tracks are identified by the TRD, and a processing command is sent to the appropriate sector processors of the TPC. The Level-3 processor receives this information via its Level-3 network from the TRD global trigger. This can easily be implemented at a rate of a few hundred Hz without presenting any particular technological challenges.

5.2.2.3 Modes of operation

The Level-3 system will incorporate all data compression methods mentioned above. The TPC raw data are shipped directly to commercial off-the-shelf computers, which have FEDC (Front-End Digital Crate) and LDC (Local Data Concentrator) functionalities (see Section 5.32). The number of machines is mapped to the natural granularity of the TPC (36 sectors, 180 fibre connections). This local layer of the Level-3 system is part of a hierarchical system. Data compression, subevent selection and track recognition are performed on the local level, the global layer is responsible for track merging across detector boundaries and different detectors and for deriving software trigger decisions. The mode of operation for a combined dielectron/open charm programme is shown in Fig. 5.30. The Level-3 system utilizes the full luminosity, while producing manageable data rates.

No Level-3 activity — full TPC zero-suppressed data readout

Zero-suppressed raw TPC data is passed on to DAQ without any processing beyond standard data compression. This is the baseline assumed in the Technical Proposal.

Full readout and almost-lossless data compression

In this model, the receiver processors function exactly the same way as the GDC processors in the Technical Proposal. The compression for each sector is made in parallel and the reduced data is then transmitted. The event building would take place after the compression. Thereby, the event headers would not be compressed, but the additional overhead for all the headers can be neglected.

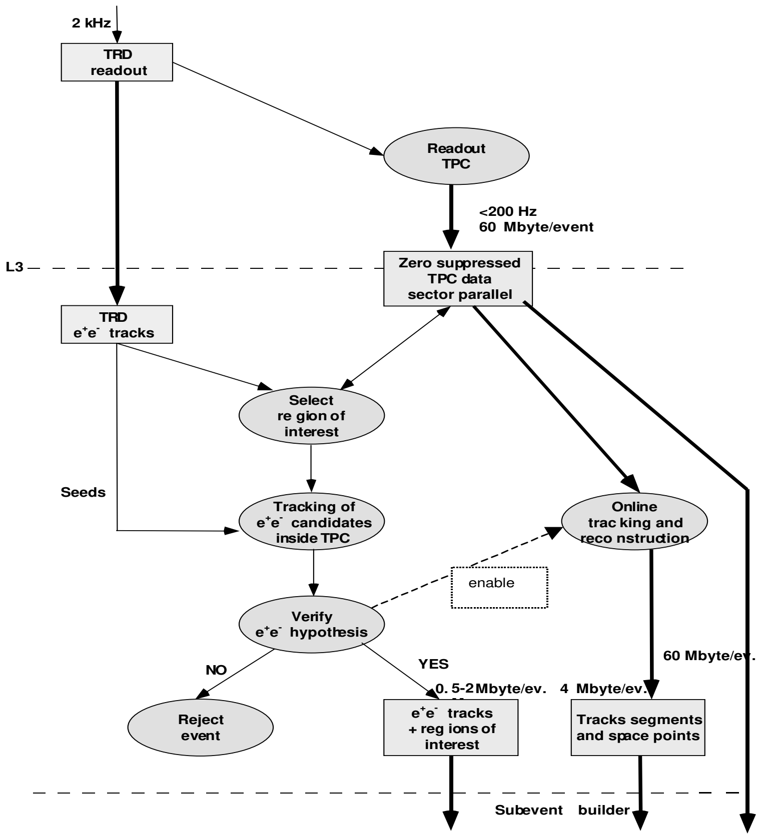

Figure 5.30: Level-3 data flow.

Full readout — data exchange with other detectors

In principle, the new architecture does not need to have an influence on other subdetectors. The raw data of other detectors can easily be distributed to the relevant processors within the Level-3 farm. It is also possible to combine the data from the receiver processors into one overall data flow after the trigger has been received and to send it to the DAQ. This ensures that, at any time, zero-suppressed and even non zero-suppressed raw data can simply be read out, though limited by the tape bandwidth.

TRD Level-3 trigger and selective readout (without tracking)

The electron tracks are identified by the TRD, and a processing command is sent to the appropriate sector processors of the TPC. The coordinates of the track candidates are distributed through the Level-3 network. After having received the event in memory, the sector processor calculates roads in the raw-data space (pad number and time bin) based on the track parameters given by the TRD trigger. If an event is to be read out, the processor receives an appropriate network message and the ROIs are transferred. If not, the internal data are marked available again in the FEDC tables and overwritten later by another event.

TRD Level-3 trigger and selective Level-3 processing

Even if the TRD should perform as designed, it will produce, for example, 40 Hz triggers but only 20 Hz of them will be good events. The intention here is to achieve this reduction by tracking the high-p_t candidates in the TPC. If the TRD selectivity is not as high as planned, the same method could be used for increasing the selectivity by tracking the track candidates in the TPC so that the resulting trigger rate would not exceed 10 Hz–20 Hz. However, this does not require the full event. The relevant tracks are identified by the TRD, and a processing command is sent to the appropriate sector processors of the TPC.

After having received the event in memory, the processor starts tracking the defined track candidate only together with the processing command. Which track candidate to use has to be distributed through the Level-3 network. During these calculations, the processor can already receive another event to make maximum use of the CPU. The Level-3 processor receives the information concerning the candidate via its network from the TRD global trigger. This can easily be implemented at a rate of a few hundred Hz without presenting any particular technological challenges. The tracking is done relatively fast.

Level-3 processing for ROI selection and partial readout

Owing to the data, each receiver already knows what needs to be done with the data record. Up to 16 TTC commands accompanying the data can be defined. In case of a full readout and Level-3 processing, the cluster finder algorithm has to be executed first in order to determine space points, a task that typically requires the longest time. This can be done in parallel for each sector. The receiver processors would take on this task. The architecture offers enough leeway to be able to provide adequate computing power. For example, four- to eight-fold processor modules could be used instead of simple processors. After the space points are determined, they could then be sent to a second layer for tracking. This is only one scenario, many other solutions are possible. The receiver board, for example, can be configured in such a way that a certain part of the events (every second, three out of four, etc.) are sent directly to the network to obtain a higher degree of parallelism. All these scenarios can be put into practice or even be changed long after the installation.

Level-3 processing for data compression

This is similar to the above, but tracking has to done only locally, though redundantly, i.e. with very open parameters, since track parameters are only used for compression purposes. The aim of the track finding is not to extract physical information but merely to build a data model, which will be used to collect clusters and to code cluster information efficiently. Therefore, the pattern recognition algorithms are differenly optimised or even different methods may be used, as compared to the offline tracking.

5.2.3 Architecture

The basic idea is to add an intelligent layer to the TPC readout and to move the data only if necessary. Data moving costs money and data storage is location independent. Custom designs should be avoided wherever possible and mass-produced electronics shall be used wherever possible. The updated architecture will allow a very high degree of flexibility by reducing the requirement for custom hardware to a minimum and using standard interfaces instead. In the new scenario the TPC and TRD detector data are sent directly into the host processors of a Level-3 Trigger/DAQ processor farm, thus avoiding the necessity for custom front-end systems. By using a standard bus, which is common in any commercial off-the-shelf computer system today, a very large degree of freedom is achieved with regard to the processor architecture.

The intelligent Level-3 layer will make use of the inherent granularity of the TPC and TRD readout. Each sector is processed in parallel, results are then merged on a higher, global level. 36 TPC sectors are processed in parallel by a single Level-3 processing farm per sector. For a Level-3 trigger (e^+e^- inspection) only a few tracks in a few sectors have to be analysed. A complete event reconstruction (Level-3 processing) results in 12 000 trackable tracks total, 333 tracks and 125 000 clusters per sector (see Table 5.13 and Section ??).

The TPC granularity is adapted to the granularity of the TRD. Therefore there will be 2×18 sectors, which are composed of two segments each (inner and outer). The analog data is preamplified, shaped, and digitized (10 bits). A digital filter will be used to achieve a better tail correction. After this, zeroes will be suppressed, and the data will be shipped to the counting room. In order to optimize fibre utilization the front-end electronics implements a small elasticity buffer.

Table 5.13: Number of tracks and clusters and size of raw data of a central Pb–Pb event ($dN_{ch}/dy = 8000$, 570 000 channels, 160 pad rows, approximately 66 Mbyte/event).

	Total	Per sector	Per optical link
TPC raw data	292 Mbyte		
Without inner cone	146 – 208 Mbyte		
After zero suppression	37 – 52 Mbyte		
Including coding overhead	≈ 66 Mbyte	1.8 Mbyte	0.37 Mbyte
Total trackable tracks	12 000	333	
Total clusters	4 500 000	125 000	25 000
Track summary (total)	0.4 Mbyte	11 kbyte	
Compressed data (total)	4.2 Mbyte	120 kbyte	

The TPC raw data are shipped optically to the counting room. The protocol proposed here is the ALICE development DDL [16]. Appropriate mezzanine cards are available in the PMC form factor. Assuming a total of 180 optical links or five links per sector, the data rate per link corresponds to about 67 Mbyte/s (at 200 Hz) leaving plenty of head room.

In the following, we will show that the FEDC functionality presented in the ALICE TP can be achieved by using an architecture, where a new RORC plugs directly into the backplane of a commercial off-the-shelf computer, such as a PC. The corresponding overall architecture of a sector is shown in Fig. 5.31. The raw data are directly transmitted to a processor's main memory via the widely used standard PCI bus (assuming the use of a fast 66 MHz/64-bit version, which is available today), while the LDC task runs on the processor's CPU. The elasticity buffers on the PCI FEDC and the main memory of the computer functions as the event buffer in this case.

The processors that hold the PCI RORC are called receiver processors in Fig. 5.31. They are generic standard processing units, which could be operated as such, ignoring or disabling the RORC, or be part

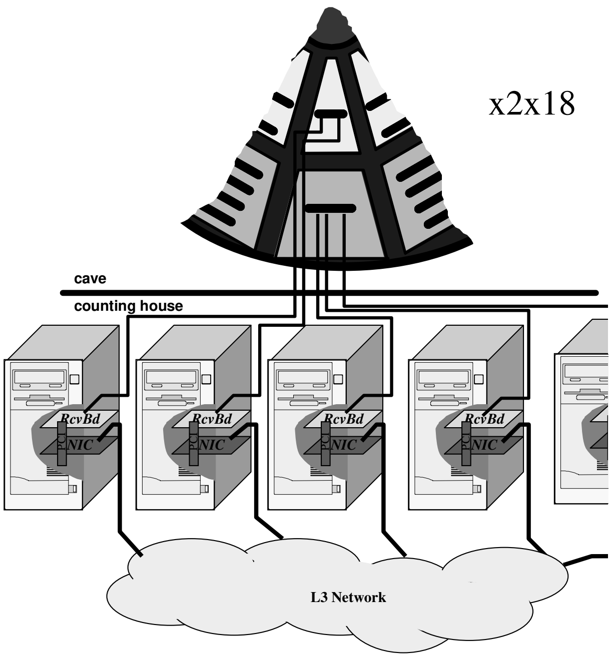

Figure 5.31: View of the readout electronics installed on a sector.

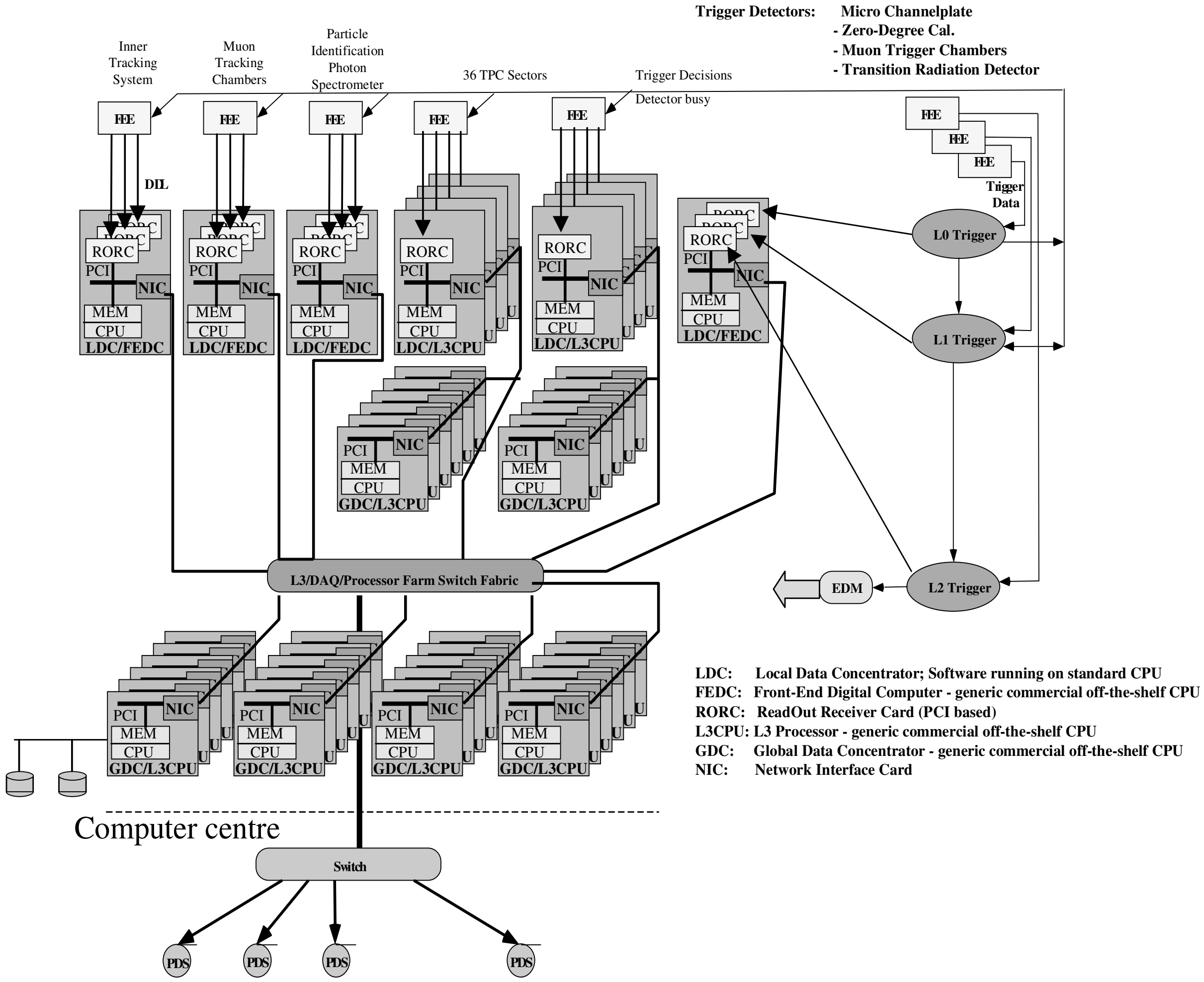

Figure 5.32: Level-3 architecture.

of the whole system or even further operate as a stand-alone single-sector DAQ system. These different configurations are only software switches.

Figure 5.32 shows a sketch of the Level-3 architecture. The receiver processors are all interconnected by a hierarchical network. The bus standard used to interface to these Network (Interface) Cards (NIC) is labelled as PCI as this is the standard that would be used if the system were to be built today. However, in the unlikely event of another standard becoming as widely accepted by the time the design has to be frozen, it would be used. At present, neither the precise network topology nor the networking technology are defined. However, it is obvious that the TPC sector granularity is reflected in the architecture processing the TPC data. There are various promising candidates and as in the case of the computing nodes the best standard network will be chosen at the latest time possible. This requires the use of portable standard network APIs. Following this Level-3 network, there are only further processors. Note that the GDC and LDC functionality can be performed on identical processing nodes. LDC and GDC functionality can even be performed on the same CPU running different tasks. Note that the permanent data storage is performed by the computer centre through an appropriate link. This very same link can be used to feed this general-purpose processing cluster when ALICE is not running in order to allow the utilization of this system as an analysis processing farm.

The system architecture provides an overall unit that can also be operated offline. In this mode, only some processors have an additional card, which may or may not be used (RORC). If the TPC is to be tested offline, an appropriate readout process could then run simultaneously in the background (similar to a computer in a network which, today, can receive and send data in the background) or a corresponding group of receiver PCs could be taken out of the Level-3 cluster. All of this can be controlled by software, without any need to make changes to the hardware. It is possible that the online and offline operation will be using different operating systems, something which could be done easily. Each computer should be equipped with a small hard disk for paging and swapping in the offline mode.

5.2.3.1 Estimated computing power

Track recognition in tracking detectors like the TPC is usually done sequentially. After finding clusters, the position and charge of the clusters are corrected for gain differences, distortions, time offsets, etc. These space points are then passed on to a track finder which builds track pieces out of clusters. Track pieces are finally merged to global vertex and nonvertex tracks.

A fast cluster finder has been optimized for speed [36]. Preliminary timing tests in the STAR environment [20, 37] showed that cluster finding for 600 clusters including deconvolution can be done in 10 ms on a 66 MHz version of a i960 microprocessor [38]. The second step in the sequential track recognition scheme is a track finder. The tracker combines a number of space points to form track segments. Track segments are then merged to form vertex and nonvertex tracks. The track finder currently used in STAR is based on an algorithm described in Refs. [39, 40]. Its main features are an optimized data organization and a conformal mapping to speed up fitting procedures. It takes about 90 ms on an ALPHA XP1000 (500 MHz) to find 400 tracks.

The processing power needed could be estimated on the basis of these benchmarks for the STAR Level-3 system [37]. A farm of the order of about 500 PCs (year 2005 PCs, assuming scalability according to Moore's law) would suffice for full tracking. Partial readout needs even less computing power. In view of the high occupancy a sequential track recognition approach may not work. Track finding on raw TPC data by transformation techniques followed by a cluster fitter is a possible alternative [41, 42]. Should there arise a need for more computing power for the track recognition, more nodes could be added to the sector Level-3 system, or PCI boards with DSPs or FPGAs could be added to perform special functions at very high speed on the data stream.

5.2.3.2 Network bandwidth and latency

The receiver processors and all other Level-3 processors are interconnected by a hierarchical network. At present, neither the precise network topology nor the networking technology are defined. The network has to have a low latency, a high bandwidth, and should allow remote access to data with low processing overhead.

5.2.3.3 Simulations and Prototyping

Extensive simulations of track recognition methods and data flow have been performed for the proposed Level-3 Trigger System of the STAR [19] experiment at RHIC. A large scale prototype of an Level-3 system is currently being installed. There is a large overlap in the physics objective, detector design, readout electronics and personell between the STAR and the ALICE Level-3 systems.

The RHIC accelerator at Brookhaven National Laboratory, USA, will start to investigate Au–Au collisions with $\sqrt{s} \leq 200\ A{\cdot}\mathrm{GeV}$ and pp collisions with $\sqrt{s} \leq 500$ GeV in 2000. The STAR experiment is a large scale, cylindrical, symmetric 4π-detector at one of the RHIC interaction points. Data taking will start in 2000 with a full size TPC (Time Projection Chamber, $r_{in} = 0.6$ m, $r_{out} = 2$ m) with 24 TPC sectors, 6912 pads each. The STAR Level-3 trigger is a distributed ALPHA processor farm, performing online tracking of $N_{\mathrm{track}} \geq 8000$ particles with a design input rate of $R = 100$ Hz. The components (cluster finder, track finder) have been benchmarked, a large scale prototype system (1/3 of the final design, $R = 20$ Hz) is envisaged for STAR data taking in 2000.

The STAR trigger system is subdivided into 4 hierarchic levels. The Level-0 input rate is 10^5 Hz, the first three levels reduce the rate by one order of magnitude each. The Level-3 trigger is supposed to reduce an input rate of 10^2 Hz to the final DAQ rate of R_{tape}=1 Hz at an expected TPC event size of $\simeq$ 15 Mbyte. The tasks of the Level-3 trigger are event selections based upon the online reconstructed track parameters of each particle. Two examples are: for Au–Au collisions the online invariant mass reconstruction of $J/\psi \rightarrow e^+e^-$, and for pp collisions the filtering of 700 pile-up events in the TPC per one Level-0 trigger.

The Level-3 trigger design is embedded into the STAR DAQ system [20]. Each physical TPC sector is mapped onto one VME crate, containing a Sector Broker, i.e. Motorola MVME-2306 VME board, carrying a PowerPC 604 (300 MHz, VxWorks), as the TPC sector master controller and a low latency, high bandwidth network interface for the raw-data transfer to the main STAR event builder and the connection to the Level-3 track finder CPU. Moreover, each DAQ crate also contains six VME receiver boards, each carrying three mezzanine cards with one Intel i960 CPUs (66 MHz, VxWorks) for data formatting, initiating the VME raw-data transfer and running the Level-3 cluster finder, and 4 Mbyte of dual-ported VRAM for buffering and pipelining of raw data for 12 events.

The Level-3 trigger scheme consists of two main parts: The sector Level-3 part is mapped onto one physical TPC sector. It contains the Level-3 cluster finder and the Level-3 track finder. The global Level-3 part consists of one master CPU for the whole STAR TPC, collecting all track data and issuing the Level-3 decision. The development of the Level-3 trigger can be subdivided into two main stages. In the first stage, envisaged for 2000, eight TPC sectors are connected. The Level-3 trigger will employ TPC deta only, and the input trigger rate is estimated to be 20 Hz. In the second stage, all 24 TPC sectors will be connected. The Level-3 trigger will employ additional information from the SVT (Silicon Vertex Tracker) and the EMC (ElectroMagnetic Calorimeter), and the final design value for the input trigger rate is 100 Hz. Probably more than one CPU per TPC sector will have to be used, implying programming parallelization techniques.

The cluster-finder algorithm runs on the Intel i960 CPUs, implemented on the DAQ receiver boards. The number of i960s is 18 per TPC sector, 432 for the whole TPC. Input to the cluster finder are zero-suppressed TPC raw data, stored in the VRAM. The output cluster data, i.e. cluster center-of-gravity and cluster total charge (ADC sum), are sent via VME to the Sector Broker, which itself ships the data via the network to the Level-3 track finder CPU (expected data transfer rate of $\simeq$ 3 Mbyte/s per TPC sector). The time constraint is $\tau_{\mathrm{cluster}} \leq 10$ ms (input rate 100 Hz). Benchmarks on the i960 were performed for 600 clusters (realistic Au–Au scenario) on the TPC's most inner pad row. The position resolution (difference between reconstructed and Monte-Carlo generated cluster positions) of $\Delta(r\varphi) \simeq 37\ \mu$m and $\Delta z \simeq 13\ \mu$m could be obtained with an algorithm within $\tau_{\mathrm{cluster}} = 7.5$ ms. If two clusters are merged, an additional deconvolution subroutine must be started, consuming 6.0% more CPU time than in the case of two separated clusters.

In case of an Au–Au collision, the track finder algorithm must be able to fit at least $N \simeq 400$ tracks per event per TPC sector, each consisting of $N_{\mathrm{point}} \leq 45$ points (given by the number of pad rows). The fast track finder algorithm has specifically been developed for the Level-3 trigger project [40]. It employs conformal mapping (transformation of a circle [2] into a straight line), followed by a fit with a follow-your-nose method. Cluster data are the input to the track finder. The track finder time constraint of $\tau_{\mathrm{track}} \simeq 110$ ms is given by the buffer time of 12 pipelined events (12×10 ms), minus the time being necessary for cluster finding $\tau_{\mathrm{cluster}} \leq 10$ ms. The track finder code was benchmarked on several Linux CPUs. According to the benchmark results, the ALPHA XP1000 500 MHz workstation was the only CPU to be able to fulfill the time requirement, if one restricts the number of CPUs per TPC sector to one. Based on these results, the ALPHA 21264 was chosen for the first Level-3 track finder implementation. Other architectures (e.g. Pentium III 733 MHz) are candidates for future Level-3 extensions.

One global Level-3 CPU receives the track data from all sector Level-3 CPUs. It is connected to

[2] In the STAR solenoid magnetic field of $B = 0.5$ T charged particle tracks can be parametrised as helices, being visible as circles in xy-projection.

the network. The global Level-3 CPU performs track merging for tracks of different sectors, a Level-3 decision algorithm based on the track data (e.g. invariant mass reconstruction) and issues the Level-3 yes/no decision.

5.2.4 Interfaces

This paragraph outlines the peripheral interfaces of the ALICE Level-3 intelligent readout system. The overall architecture has been outlined in the previous section.

5.2.4.1 Interface between TPC readout electronics and Level-3.

By far the largest amount of data is produced by the TPC. The shipping of the data has to be done optically in order to avoid any ground potential problems. In order to minimize risk, all unnecessary electronics on the detector should be avoided. 180 optical links (five per sector) will ship the raw data off the detector into the counting room. In order to utilize the optical links optimally a small eight-event elasticity buffer will be implemented on the front-end digital chip in order to support a very streaming data transfer across the optical fibre. The estimated aggregate data rate is expected to be about 73 Mbyte/s (at 200 Hz). For a detailed description refer to Section 5.1.

The decision about the Level-3 processor type and architecture will be delayed as long as possible in order to benefit from the rapidly evolving markets. On the other hand an interface is required to allow these computers, which are currently not yet defined, to receive the TPC data for processing. The most obvious approach is to use the most widely accepted peripheral bus standard in industry, PCI [43], as interface or whatever becomes the PCI replacement at the time the design has to be frozen. Basically all computers today use this bus standard as their internal peripheral bus. Given the availability of a PCI receiver card, which connects to the optical fibres being fed by the TCP front-end, it is conceivable to mount this card into a Level-3 CPU and to directly receive the TPC data into the processor's main memory. This scenario avoids the requirement for any front-end bus systems, such as VME, and uses the inexpensive host memory as elasticity buffer.

General description of the data transfer components

The only data being transmitted from the counting room to the TPC are slow controls, calibration and monitoring data. These data, however, are to be provided by the Detector Control System (DCS). It is an essential system requirement that the DCS be independent of any other system such as DAQ or trigger in order to have the necessary redundancy to debug such a complex system. Therefore an additional path to the front-end will have to be implemented in any case in order to enable downloading of test or calibration data, and to read out typical slow-control data such as temperature, voltage and current within the system. This existing data path is an ideal method for redundant but slow readout of any TPC sector, allowing simple offline detector testing without the requirement of a DAQ system. Given these arguments there is no requirement for a full duplex TPC data up-link. For a detailed description of the TPC front-end system please refer to Section 5.1.

The TPC counting room data link consists of three major components: the network feed, which is part of the front-end readout cards on the detector, the actual optical link itself and the optical receiver card in the Level-3 host. The network feed is part of the TPC front-end electronics. However, in order to achieve maximum flexibility while permitting continued progress on the TPC front-end electronics, the fibre transmitters in the front-end will be mounted on appropriate mezzanine cards using a standard bus. Prototype optical transmitters have already been designed, as has an appropriate prototype of a PCI optical receiver.

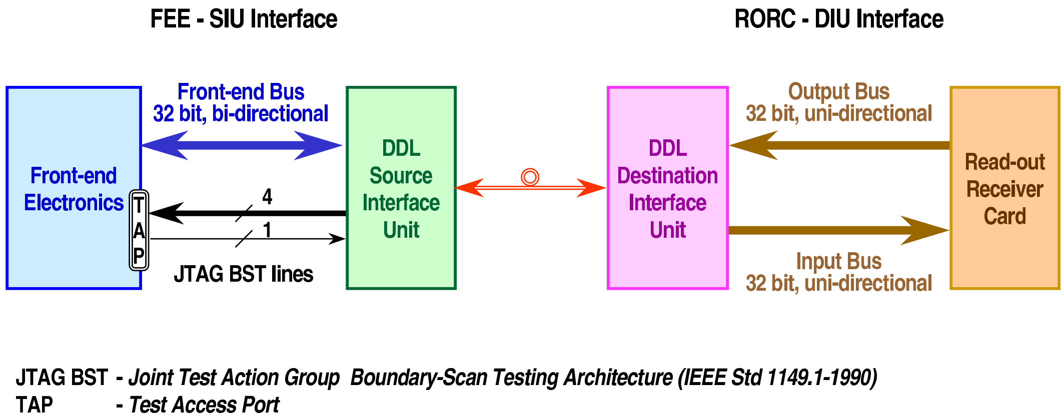

Figure 5.33: DDL logical interface. Note that this interface can be operated in full-duplex mode with handshake as sketched above, which, however, is a feature, which is not being required by the ALICE TPC.

Detector Data Link

The Detector Data Link (DDL) [16] is the proposed standard for optically transmitting data within ALICE. The logical interface of the DDL can be seen in Fig. 5.33. The DDL is composed of three hardware items: the Source Interface Unit (SIU), the fibre, and the Destination Interface Unit (DIU). It can be operated in full-duplex and half-duplex modes. The operation mode for the TPC is half duplex. The SIU is designed to be plugged into the front-end electronics of most of the detectors. The fibre will be the only medium used to ship the data between the ALICE detectors in the experimental area and the computing room located in the access pit. The DIU will be inserted on a mother card (the RORC) inside a computer located in this computing room.

A prototype of the DDL DIU and of a VME-based receiver card have been developed. A prototype of the DDL SIU is currently under design. Its main characteristics correspond to the needs of the TPC. A test of integration of the TPC readout system and of the DDL will be performed in 2000. The specifications of the DDL will be frozen after a successful integration test with the TPC and the other major ALICE detectors. Two modules have been produced to allow an easy stand-alone test of the front-end electronics. These two modules simulate a complete DDL and DAQ system by simple and cheap electronics boards [44, 45].

Read-Out Receiver Card (RORC)

A prototype Front-End Digital Crate (FEDC) housing several RORCs has been develepod and tested in the NA49 Forward TPC detector system. The current prototype of the FEDC consists of a VME crate. In this prototype, the data transfer system is performed through the VME backplane by the Local Data Concentrator (LDC). The LDC is currently implemented as a single-board computer running the UNIX operating system. The physics data transferred by the DDL are buffered in a prototype RORC, built as a VME module and carrying two DDL DIUs [46].

PCI is the proposed standard interface between the detector electronics and the Level-3 system. Almost every modern computer uses PCI as a peripheral interface bus. The processor choice for the Level-3 system is not restricted by this requirement. Should another bus standard become as widely used

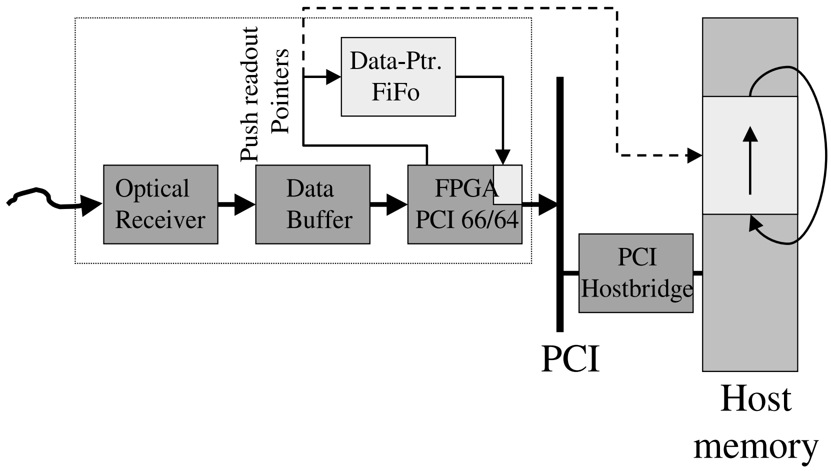

Figure 5.34: The functional blocks of the TPC receiver card (PCI-RORC).

by the time the ALICE electronics has to be frozen it would be chosen instead.

The PCI receiver card basically consists of a data converter which writes the optically arriving data to the main memory of the host computer (receiver processor) after some additional buffering. The interface is PCI 66/64 (theoretical bandwidth of 528 Mbyte/s). At present, PCI is the baseline bus interface.

Figure 5.34 shows the functional blocks of this card as it is integrated in the computer. The data which are received optically are converted into electronic signals. A small elasticity buffer follows behind, which is only necessary in order to cover PCI latencies. The arriving data is then copied directly into the main memory of the receiver processor via the PCI bus using a DMA engine on this card. The host's main memory, which now also acts as data buffer, can easily be considerably enlarged without producing significant additional costs. The target memory area is marked light grey and works like a cyclic memory buffer. The start and end long-term addresses are configurable. The receiver card copies the arriving data sequentially into the cyclic memory buffer of the host. These memory models are supported by all common operating systems, including Windows NT and Linux. The buffer principle of an Ethernet interface, for example, is very similar. For every newly started event, the start address of this event is written into a second buffer memory, which is here also implemented as a FIFO for the sake of simplicity. The host processor(s) only need(s) to read the start address of an event from the FIFO. An empty FIFO defines the absence of further data. If an event has been either transmitted to the DAQ or has been rejected, the corresponding read pointer in the CSR area of the receiver card is updated, t hus freeing the corresponding memory area.

This architecture does not require any flow control between the detector front-end and the receiver modules. At every point, there are large enough elasticity buffers to absorb maximum reaction times. The data buffer in the receiver processor can easily be enlarged to ensure that there will be no dead-time under normal operating conditions. In case of extremely long trigger bursts or false trigger thresholds, a simple model, such as high/low-water marking, would allow to generate Level-3 dead-time through software. Alternatively, a busy signal can be generated on the PCI-RORCs and daisy-chained into one TPC counter box, which sends a busy signal to the trigger. This ensures that no data can be overwritten, while the requirement for flow control hardware is nevertheless minimal. The resulting system can be debugged very easily.

The PCI interface itself can be integrated into the FPGA. There are commercially available FPGAs