An ultra fast, low power readout chain for single photon sensitivity with multi-anode photomultiplier tubes for the RICH upgrade at LHCb

C. Arnaboldi a,b, A. Giachero a,b, C. Gotti a,c,*, M. Maino a,b, G. Pessina a,b

a Sezione INFN di Milano Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy
b Dipartimento di Fisica, Università di Milano Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy
c Dipartimento di Elettronica e Telecomunicazioni, Università di Firenze, Via S. Marta 3, I-50139 Firenze, Italy

1. Introduction

Ring Imaging Cherenkov (RICH) counters are used in the LHCb [1–3] detector at CERN to detect and separate charged particles originated in the high-energy hadron interactions. In this detector system, an incoming charged particle interacts with a radiator and creates a Cherenkov ring of a few tenths of photons, which are detected with a two-dimensional array of pixel photon detectors with a spatial resolution in the millimeter range. Single photon sensitivity, along with high speed, high counting efficiency and low cross-talk, are thus the main requirements the pixel detectors must meet.

The Cherenkov rings are currently read out by Hybrid Photon Detectors (HPDs) [4]. However, other solutions are currently studied, in view of the LHCb upgrades foreseen for the next years. One of the possibilities is the use of commercial multi-anode photomultiplier tubes (MA-PMTs), whose anodes are arranged in bidimensional pixel arrays. The single photon sensitivities of two of such devices, the 16 × 16 pixel H9500 [5] and the 8 × 8 pixel R7600, both from Hamamatsu, were tested, but will not be discussed here. By recent measurements, not published yet, the R7600 proved to perform better for single photon signals, due to a lower rate of dark counts and cross-talk events.

Since the main LHCb upgrade target is to increase the luminosity, reaching an event rate of 40 MHz, the electronic readout chain must be fast and aim to the full baseline recovery before about 25 ns, in order to suppress spillover. A wideband preamplifier is thus required, with bandwidth in the GHz range, in order to reach transition times below 1 ns, followed by an equally fast discriminator. In addition to the above requirements, the readout chain must be low power, i.e. its dissipation must be limited to a few milliwatts per channel, due to the large number of channels in the system. This is a strong constraint on circuit design, since wideband amplifiers usually require a power of one or two orders of magnitude higher.

A tradeoff is to be found between a low power dissipation and slower transition times on a side, or higher power dissipation with faster transition times on the other. Transistors in Silicon–Germanium (SiGe) technology were chosen, for they can usually operate up to tenths of GHz at 10 mA or more; at lower currents, bandwidth reduces, but it still can be well over 1 GHz at 1 mA and below, allowing for fast operation while keeping power dissipation low.

2. Readout chain architecture

When properly biased, these photomultipliers give charge pulses of about and above one million electrons at the anodes in
response to single photon excitations. Charge collection time is specified to be 1 ns or below. These are fairly big signals, as opposed to other photon detectors such as HPDs; thus noise is not of main concern (unless very small jitter is required; see Section 3), and the design of the very front-end preamplifier should instead aim to minimize the input impedance to avoid such fast and big pulses to cross-talk between neighbouring pixels via stray capacitances. To reach this aim, the very front-end exploits a local feedback structure which will be explained in detail in Section 3. The response of the analog shaper is integrator-like, so that the charge pulse becomes a voltage pulse whose rise time is limited by the bandwidth of the circuit or by the charge collection time of the photomultiplier, and whose fall time is set to about 5 ns by the choice of a proper RC relaxation constant. So far a prototype of the analog shaper was built with discrete SiGe transistors, whose performances are presented in Section 4.

The analog voltage pulse is then triggering a discriminator with settable threshold and hysteresis. The discriminator should have transition times below 1 ns while keeping power dissipation of the whole chain below 5–10 mW per channel. Its output should stay triggered for as long as the subsequent digital circuitry needs to count the signal, which in any case should be less than 25 ns, and guarantee the lowest possible blind time after a pulse. A prototype of a discrete SiGe discriminator which meets these requirements is currently under development, but will not be described here.

Since the PMT comes in an 8 × 8 pixel arrangement, 8-channel boards with an analog shaper and a discriminator per channel seem currently to be the best choice. A commercial FPGA on each 8-channel board could in a future read the outputs of the eight discriminators and transmit data to a remote location for storing.

3. The very front-end concept

The basic idea of the analog front end is the two-transistor structure depicted in Fig. 1. The circuit configuration is based on the so-called active cascode [6–8]. This circuit has been already exploited in high-energy physics applications [9–13], due to its natural capability to readout big charge pulses coming out of radiation detectors. This time the design was optimized for very high speed.

$Q_1$ is the input transistor, and if its base was held at constant voltage, as in a usual common base configuration, its input impedance at the emitter would be $1/gm_1$, the inverse of the transconductance of the device. A low input impedance in this case would thus require a high transconductance, which in turn requires a fairly high polarization current, increasing power dissipation; furthermore, changes in temperature would affect the input impedance via the well-known dependence of the transconductance from temperature, affecting gain also.

The role of $Q_2$ is to create a feedback loop around $Q_1$. If an input signal current pulls down the voltage at the input node $B$, then $Q_2$ amplifies this voltage change, with negative sign, and sends it back to the base of $Q_1$. This heavily counteracts the voltage change, which turns out to be negligible, and node $B$ is thus a low impedance node, i.e. a virtual ground.

To state this in a quantitative way, one needs to calculate the loop gain, which is quickly done. Calling $Z_b$ the impedance load at node $A$ (given by $R_A$ shunted by the input impedance of $Q_2$), we have that a voltage signal at node $B$ gets multiplied by common emitter amplifier $Q_2$ by a factor $-gm_2 Z_b$. Then this signal gets back to node $B$ via $Q_1$, which acts as a voltage follower, reduced by a factor $gm_1 Z_a/(1 + gm_1 Z_a)$, where $Z_a$ is the impedance load at the emitter of $Q_1$, which is $R_b$ in parallel to the detector capacitance and the base-emitter impedances of $Q_1$ and $Q_2$.

The loop gain $T$ is thus given by

$$ T = -gm_2 Z_b \frac{gm_1 Z_a}{1 + gm_1 Z_a}. \quad (1) $$

Since the emitter of $Q_1$ is held at one diode drop above ground by $Q_2$, the current in $Q_1$ is set by the DC value of $Z_a$, and so is $gm_1$; if $Z_b$ is simply a resistor to ground, then a low $Z_b$ leads to a higher $gm_1$ and vice versa, but in any case $gm_1 \gg 1/R_b$ and so the fractional term in Eq. (1) is very close to one, and thus negligible. This of course is true up to the frequency where $Z_a$ is equal to its DC value; in the case of a big detector, its capacitance to ground would cause $Z_a$ to drop with frequency, gaining a higher gain to be less effective, and possibly the whole system to oscillate because of the added pole in the loop gain. If this is the case, the value of $gm_1$ should be kept high to push this pole to higher frequency, increasing power dissipation in $Q_1$, and thus possibly making other solutions preferable. However, this circuit arrangement is intended for small pixel detectors (the anodes of the R7600) which exhibit a negligible capacitance to ground, of the order of 1 pF.

Once the loop gain $T$ is known, the effect of the feedback loop is to lower the input impedance by a factor $1 - T$ from its open loop value of $1/gm_1$, so that it becomes

$$ Z_{in} = \frac{1}{gm_1} \left( 1 + gm_2 Z_A \frac{gm_1 Z_a}{1 + gm_1 Z_a} \right)^{-1} \approx \frac{1}{gm_1 gm_2 Z_A}. \quad (2) $$

The stability of the feedback loop is not an issue, unless $gm_1$ and $gm_2$ have very small values, and capacitances to ground in $Z_A$ and $Z_B$ are kept negligible (more on this in the following). Eq. (2) shows that, provided that stability constraints are satisfied, this simple two-transistor solution allows to lower the input impedance to a negligible value while keeping the polarization currents and thus power dissipation low.

Since the input has negligible input impedance, a current pulse at the input is buffered to the collector of $Q_1$, where it develops a voltage across the passive components $R_F$ and $C_F$. Assuming a standard frequency response with two real poles, one due to $R_F$ and $C_F$, located at $1/(2\pi R_F C_F)$ and one at higher frequency, located at $1/(2\pi R_T C_T)$, due to the finite bandwidth of the system, a current pulse at the input carrying a charge $Q$, modeled by a delta-like signal at $t = 0$,

$$ i_{in} = Q \delta(t) \quad (3) $$

develops an output voltage at $t > 0$ given by

$$ v_{out} = \frac{Q}{C_F} \left[ \tau_F \left( 1 - \exp \left( -\frac{t}{\tau_F} \right) \right) - \tau_T \left( 1 - \exp \left( -\frac{t}{\tau_T} \right) \right) \right]. \quad (4) $$

As can be seen, $\tau_F$ mainly affects the falltime of the signal, while $\tau_T$ mainly affects its risetime. The gain of this signal is set by choosing $C_F$, while the fall time is set by the $R_F C_F$ product. In the basic schematic of Fig. 1, $R_F$ is also constrained by dynamic range considerations (the output swing must not drive $Q_1$ to saturation), and once the positive supply voltage value is set, this becomes a higher limit on the value of $R_F$; however this constraint can be bypassed with some minor modifications to the circuit.
The value of $\tau_R$ is related to the loop gain $T$ by the following relation in the complex frequency $s$:

$$\frac{1}{1+\tau_R s} \approx \frac{-T}{1-T}$$  \hspace{1cm} (5)

where the left-hand side is a definition for $\tau_R$. This relation makes sense only when the right-hand side of the equation can be approximated with a single pole. By using Eq. (1) it can be seen that the factor $-T/(1-T)$ yields two poles in $s$, which are real if $g_{m1}Z_R > 4g_{m2}Z_A$ (this is a constraint for the stability of the system, which sets an upper limit to the value of $C_D$ once the other parameters are set). Assuming $C_D$ to be negligible, and $g_{m1} \geq g_{m2}$, then the approximation in Eq. (5) is valid, and $\tau_R = C_A/g_{m2}$, $C_A$ being the capacitive load at node A, i.e. the input capacitance of $Q_1$.

As shown, this circuit solution provides amplification and shaping of the charge pulse all in a single stage. No subsequent shaping stages are required, and this helps to minimize area and power dissipation of the readout chain for each pixel. A second stage buffer can be added if the signal has to drive a low impedance, i.e. a power dissipation of the readout chain for each pixel. A second stage shaping stages are required, and this helps to minimize area and shaping of the charge pulse all in a single stage. No subsequent

4. The discrete components prototype

A prototype was built with discrete components to test the front-end concept presented in Section 3. Commercial BFP640 SiGe transistors in SOT143 package as well as resistors and capacitors in 0603 package were used, to reduce the space required for each channel; this was necessary since the pixels of the R7600 MA-PMT are squares of 2.3 mm side border included, and thus the front-end electronics for each channel is confined not to exceed 2.3 mm on the PCB. A prototype of the discriminator was placed on the PCB too, built with the same discrete components, but its performance will not be described here. For each channel, components were placed all in line, which turned out to be 2.3 mm wide and about 10 cm long (4 cm for the analog front end, 6 cm for the discriminator). Four channels per side were placed on each board, which is then able to readout a line of eight channels on the R7600 while allowing several MA-PMTs to be put side by side. A picture of the board mounted on the R7600 is shown in Fig. 2, where custom bias PCB prototypes are also visible near the PMT.

Transistor $Q_2$ of Fig. 1 was operated at 250 $\mu$A, while $Q_3$ was operated at 500 $\mu$A, and the power supply was set at 2 V, so that power dissipation in the very front end was 1.5 mW. Loop gain was set to about 20 by using a 1 k$\Omega$ resistor as collector load for $Q_2$, so that the input impedance was lowered by the feedback loop from $1/gm_1 = 100 \Omega$ to about 5 $\Omega$.

A two-transistor inverting second stage with gain of $-2$, AC coupled, whose schematic is given in Fig. 3, was put thereafter to drive a 50 $\Omega$ terminated cable to the oscilloscope with a low output impedance and a positive polarity of the output signal, which is preferable because the output transistor is of NPN type. Nominal gain of $-2$ was obtained by $R_C = 300 \Omega$ and $R_H = 600 \Omega$, as with an opamp in standard inverting voltage amplifier configuration. Actual gain was set about $-1.5$ however, since its loop gain is lowered by the 100 $\Omega$ load at the output. By choosing appropriate values for bias resistors $R_P$ and $R_Q$, the second stage was dissipating 1.5 mW too, so that total power dissipation for the analog section of the front end was 3 mW.

The prototype was tested with the R7600, with single photons sent to a pixel via a commercial blue led, as described in Ref. [5]. Results are shown in Fig. 4. Biggest signals are about 50 mV peak. Risetime is 1.4 ns, maybe limited by the PMT charge collection time, as SPICE simulations suggest that the front end itself could be slightly faster (measurements with a fast pulser, as soon as it is available, will clarify this point). Falltime was set to about 3 ns by choosing $R_p//C = 200 \Omega$, $C_T = 5$ pF. With these values of risetime and falltime, baseline is restored well before 25 ns, thus completely eliminating spillover at the LHC event rate.
5. Noise and jitter performances

The noise performance of the prototype is of interest; not much for signal to noise ratio (since PMT signals are big) as for timing performance, since noise introduces jitter. For the first stage, the main contribution to white noise is given by Q2, whose input referred series noise is roughly given by $2kT/gm^2$, and with $1/gm^2 = 50\Omega$, input noise due to Q2 is at about 0.64 nV/$\sqrt{Hz}$. Noise from Q1 is divided by the loop gain, and is thus negligible in a first place. Input voltage noise due to the first stage is thus expected below 1 nV/$\sqrt{Hz}$. Input current noise due to the base current of Q2 is negligible.

The input referred voltage noise was measured by putting a 100\Omega resistor to ground at the input of the preamplifier. In this case, the series noise of the resistor, 1.3 nV/$\sqrt{Hz}$, adds in power to the contribution from the preamplifier. Also the noise of the second stage of Fig. 3 should be considered, which is mainly due to $R_s$, $R_c$ and $R_m$, about 1.3 nV/$\sqrt{Hz}$ when referred at the input. The total input series noise due to these major contributors is about 2 nV/$\sqrt{Hz}$. The result of this measurement is plotted in Fig. 5 from 10 kHz to 10 MHz. As can be seen, noise is all white, and slightly higher (due to all the second-order contributions) but compatible with the expected value.

The timing error, i.e. jitter, due to the preamplifier noise can be estimated using the following formula (the trigger threshold is assumed set at half of the signal maximum):

$$\sigma_t = \frac{\sigma_v}{V_{max} t_R}$$  \hspace{1cm} (6)

where $t_R$ is the risetime, $V_{max}$ is the output signal maximum, about 50 mV as in Fig. 4, and $\sigma_v$ is the r.m.s. noise at the output of the preamplifier. Noise power density at the output is about 5 nV/$\sqrt{Hz}$, which, assuming a 1 GHz bandwidth, gives 175 $\mu$V r.m.s.; Eq. (6) with these values gives less than 5 ps r.m.s., a very low value. Measurements with a Tektronix DPO 7254 2.5 GHz oscilloscope, with 25 ps as maximum resolution, did not show timing errors. Work is currently being done preparing the set-up to measure the actual jitter performance.

6. Conclusions and future developments

A charge amplifier was presented as the very front end to readout MA-PMT pulses in a future upgrade proposal for the LHCb RICH photon detectors. The amplifier is fast and low power, and exploits the local feedback loop of an active cascode at the input to keep the input impedance at a negligible value, while still keeping power dissipation low. A discrete components prototype was built, with risetime of about 1 ns and power consumption of about 3 mW, capable of driving a 50 \Omega terminated cable at the output. Preliminary results show very good noise and jitter performances, although a deeper characterization is to be made. Even if the performances of the discrete components prototype match the requirements, a monolithic version is going to be designed in Milano Bicocca to try push performances to a higher level.

References