VMM1 - An ASIC for Micropattern Detectors

Gianluigi De Geronimo, Jack Fried, Shaorui Li, Jessica Metcalfe, Neena Nambiar, Emerson Vernon, and Venetios Polychronakos

Abstract – We present VMM1, the first prototype of a family of front-end ASICs designed for the ATLAS muon upgrade. The ASIC will operate with MICROMEGAS and TGC detectors providing charge and timing measurements along with a number of features including sub-hysteresis discrimination, address of the first event in real time, and digital output per channel for Time-over-Threshold measurements. The shaper, designed using the concept of Delayed Dissipative Feedback (DDF), allows analog dynamic ranges in excess of 10,000. With a capacitance of 200 pF and a nominal peaking time of 25 ns it can provide charge and timing resolution below 1 fC and 1 ns respectively, for input charges up to 2 pC. Designed in a commercial 130 nm technology it dissipates about 4.5 mW per channel.

Index Terms – ASIC, sub-hysteresis, DDF, ToT, MICROMEGS, TGC, ATLAS, muon, noise.

I. INTRODUCTION

The ATLAS muon system upgrade needs to provide accurate momentum measurement as well as participate in the Level 1 trigger [1-3]. The upgrade consists of the development of two new small wheels (NSW) based on MICROMEGAS (Micromesh Gaseous Structure) and Thin Gap Chamber (TGC) detectors [4,5]. The challenging detector requirements must be coupled with appropriate readout electronics, which must provide pulse amplitude, timing measurements, and trigger information for more than 2 million channels. The limited bandwidth of the readout link requires on-chip zero suppression. For reliable operation in the ATLAS environment, radiation tolerance is also required.

The VMM is a family of front-end ASICs being developed for the MICROMEGAS and TGC detectors. The ASICs must provide charge and timing measurements, triggers, direct timing outputs, sparse and derandomized readout, and multiplexing. We present the architecture and preliminary measurements on the first prototype of the VMM family, VMM1, and illustrate our plans for the second prototype, VMM2, currently in design. Conceived to operate (i) with input capacitances from few pF to 1 nF, (ii) at input charge ranges from 0.1 to 2 pC measured with resolutions down to 0.03 fC rms, (iii) at processing times from 100 ns to 1 µs measured with resolutions down to 200 ps rms, and (iv) integrating a number of additional amplitude, timing, trigger and readout features, the VMM family can be considered for use in a wider range of applications based on micropattern detectors.

In this work Section II describes the VMM1 architecture with circuit details and preliminary experimental results, and Section III presents the plans for the next prototype, VMM2.

II. ARCHITECTURE AND PRELIMINARY RESULTS

In Fig. 1 the architecture of VMM1 is shown. It is composed of 64 front-end channels each providing a low-noise charge amplifier (CA), a shaper with baseline stabilizer, a discriminator with trimmer, a peak detector, a time detector, some logic, and a dedicated digital output for ToT (Time-over-Threshold) or TtP (Time-to-Peak) measurements.

Shared among channels are the bias circuits, a temperature sensor, a test pulse generator, two 10-bit DACs for adjusting the threshold and test pulse amplitudes, a mixed-signal multiplexer, the control logic, and the ART (Address in Real Time) which consists of dedicated digital outputs (flag and address) for the first above-threshold event. The most relevant features of these blocks are summarized in the next sub-sections. VMM1 is designed and fabricated in a commercial 130nm CMOS process from IBM and it dissipates about 4.5 mW per channel.

Analog section - The input MOSFET is a p-channel with gate area $L \cdot W = 180 \text{ nm} \cdot 10 \text{ mm}$ (200 fingers, 50 µm each) biased at a drain current $I_D = 2 \text{ mA}$, which corresponds to an inversion coefficient $IC \approx 0.22$, a transconductance $g_m \approx 50 \text{ mS}$, and a gate capacitance $C_g \approx 11 \text{ pF}$. The MOSFET has been optimized [6] for the expected MICROMEGAS strip capacitance of 200 pF targeting an Equivalent Noise Charge (ENC) below 5,000 rms electrons at a peaking time of 25 ns. The input branch is based on the dual cascode configuration [7].

In Fig. 2 the schematic of the front-end voltage amplifier is shown. Concerning the input section, $M_i$ is the input MOSFET, $M_{C1}$ and $M_{C2}$ are the stages of the dual cascode, and $M_{s2}$ and $C$ were included to avoid the initial saturation of the amplifier input stage when operating with positive input signals (i.e. positive charge). When the current in $M_i$ decreases, the drain of $M_{C1}$ might decrease by an amount large enough to disable $M_{C2}$, causing the amplifier’s gain to decrease to a negligible level and slowing...
down the amplifier’s response. The role of C and MS2 is to maintain control of the current at the output of the stage (drain of MS2) thus avoiding the initial saturation and resulting in a faster response for positive inputs. The MOSFETs MC2 and MS2 are biased using the matched bias circuit shown to the right of Fig. 2.

The output stage of the amplifier is inverting with a bootstrapped n-channel current source and with an adjustable lead-lag compensation for either small (<30 pF) or large input capacitance. All the MOSFETs in the charge amplifier are sized and biased so that their total noise does not exceed 10% of the noise from the input MOSFET, thus contributing no more than 10% to the equivalent input noise of the voltage amplifier.

The charge amplification is achieved using two adaptive-reset stages [8] each with adjustable gain (i.e. adjustable mirror ratio) followed by one additional stage which provides signal inversion and which is enabled only for positive input charges. Consequently the signal current entering the shaping amplifier is always of the same, negative polarity. The channel is also equipped with integrated test capacitor, connected by registers to the integrated pulse generator.

The shaping amplifier is a third order semi-Gaussian filter composed of one real pole and two complex conjugate poles. It has adjustable peaking time (25, 50, 100, and 200 ns) and it is realized using the delayed dissipative feedback architecture (DDF), which offers lower noise (i.e. higher dynamic range) at equal capacitance than other classical configurations [9]. A feedback loop based on the BLH concept [10] provides the equalization and stabilization in rate and temperature of the output baseline. Fig. 3 shows the measured channel dispersion of the output baseline and pulse amplitude (peak) measured using the integrated injection capacitor.

![Fig. 3 - Measured channel dispersion of the output baseline and pulse amplitude (peak) measured using the internal injection capacitor.](image)

The complete analog chain (charge amplification and shaping) provides a gain adjustable to 0.5, 1, 3, and 9 mV/fC in order to cover an input charge range either positive or negative of 0.11, 0.33, 1.0 and 2.0 pC respectively. The charge polarity is selectable per channel.

Fig. 4 shows (a) the theoretical and measured ENC and (b) the measured gain versus input capacitance for different settings of gain and peaking time. From Fig. 4(a) a non-negligible disagreement can be observed for the simulated and measured ENC at the lowest peaking time (25ns) and large values of input capacitance. This is due to an error in the adjustable lead-lag compensation of the front-end amplifier, which made it impossible to enable the compensation for large values of input capacitance. The consequence was a reduced speed of the charge amplifier and an increase in the peaking time resulting in an effective decrease of the ENC. For the same reason a decrease in gain larger than the expected was observed at large input capacitance visible in Fig. 4(b). Another issue was a considerable leakage current (several nA) from the ElectroStatic Discharge (ESD) protection circuit at the input of the channel due to the use of standard rather than high-threshold (low-power) devices. This resulted in a parallel noise contribution larger than expected, visible at low values of input capacitance and long peaking times. A residual disagreement at low values of input capacitance and short peaking times still needs to be understood and might be due to the modeling of the low-frequency noise amplitude and slope extrapolated from the foundry data during the design of VMM1.
however, accompanied by an increase in the slope coefficient (exponential $\alpha_f$ in $1/f^\alpha$). In Fig. 6 the extracted coefficients for $L = 120$, $180$, and $360$ nm are shown.

Both the dependencies of $K_f$ and slope with IC should be carefully taken into account during the input MOSFET optimization process. However, when extracting the equivalent $K_{feq}$ i.e. when modeling the low-frequency noise as a $1/f$ (slope $\alpha = 1$) in proximity of the white noise [11] the value for the p-channel is almost constant suggesting that this dependence has little effect in the optimization process as anticipated in [11]. It is evident the advantage of using the p-channel over the n-channel as input MOSFET in those cases where the low-frequency noise has non-negligible contribution. It was also found that both $K_f$ and $K_{feq}$ don’t have a monotonic trend with the channel length $L$. These results will be further analyzed and compared with the ENC measurements and the outcome will be used to optimize the input MOSFET in the revision VMM2.

Discrimination - The front-end and core of the discriminator is the comparator circuit. Comparators almost always use positive feedback to guarantee a fast response when very small differential inputs occur, i.e. when signals exceed the threshold by very small amounts. In Fig. 7(a) a simplified schematic of the input stage of the comparator is shown. The two outputs are converted into a digital signal by additional circuits not shown in the Fig. 7(a). The positive feedback is achieved by using the two cross-connected MOSFETs $M_h$, which are typically equal in size and about twice the size of the diode-connected MOSFETs. The positive feedback, along with increasing the speed of the comparator, introduces a voltage hysteresis, characterized by an upper threshold responsive to the positive slopes of the input signal and a lower threshold responsive to the negative slopes of the input signal, as shown in Fig. 7(b). The hysteresis window has a typical size from several mV to a few tens of mV. During operation, the hysteresis window is placed with the lower hysteresis threshold just above the baseline in order to guarantee that after the discrimination the comparator returns to its original state, ready to process another pulse.

Fig. 5 - Equivalent input noise spectral densities for n-channel and p-channel devices with $L = 180$ nm, measures at different values of the inversion coefficient IC and drain-to-source voltage $V_{ds}$.

With regards to the ENC, once the issue with the charge amplifier compensation is addressed we expect to satisfy the requirement of $\text{ENC} < 5,000 \, \text{e}^{-}$ at 25 ns and 200 pF.

Also observable in Fig. 4(a) is an analog dynamic range (defined as ratio $Q_{max}$ to ENC) in excess of 12,000 at some low-gain setting, which is in large amount due to the low-noise design of the shaper based on the DDF architecture [9].

Fig. 6 - Coefficients $K_f$, $\alpha_f$ (exponential $\alpha_f$ in $1/f^\alpha$) and $K_{feq}$ (coefficient for the equivalent $1/f$) extracted from the measurements of devices with lengths $L = 120$, $180$, and $360$ nm.

The event is detected when the rising edge (positive slope) of the input pulse crosses the upper threshold. When the falling edge (negative slope) of the input pulse crosses the lower threshold (see the large blue pulse in Fig. 7(b)) the comparator returns to its original state. Unfortunately this setting prevents the detection of
pulses with amplitude lower than the hysteresis window (see for example the small red pulse in Fig. 7(b)). Considering the typical size of the hysteresis window, this setting introduces a severe limit to the dynamic range of the system, especially in deep sub-micron technologies where the available voltage swing is relatively small (about 1 V in our case).

In order to overcome this limitation some straightforward solutions could be adopted. One would consist of switching the threshold input of the comparator between two values. In order not to affect the rate, the switching has to occur in a relatively short time. The drawback is in the trade-off between the switching time and the stabilization and filtering (noise) of the threshold voltage. Another solution, with challenges evident to the reader, would consist of generating a small negative unipolar and low-noise pulse at the signal input.

Fig. 7 - Simplified schematic (a) and operation (b) of the input stage of the comparator. The circuits converting \( V_{\text{in}} \) and \( V_{\text{out}} \) into the output digital signal have been omitted.

The solution adopted in VMM1 allows discrimination of signals of amplitudes lower than the hysteresis window by acting directly at the core of the comparator. Implemented as a switchable option, the sub-hysteresis discrimination (SHD) concept is illustrated by the schematic in Fig. 8.

Fig. 8 - Simplified schematic (a) and operation (b) of the sub-hysteresis discrimination (SHD) concept implemented in VMM1.

As shown in Fig. 8(a), the positive feedback MOSFETs are, in part, switchable with the complementary switching signals \( s \). The switches allow changing the effective ratio between the left and the right MOSFETs. By altering the ratio it is possible to control the size and/or position of the hysteresis window. In order to achieve the sub-hysteresis discrimination, we initially set the window with the upper threshold just above the signal baseline as shown in Fig. 8(b). After a pulse is detected (i.e. once the comparator triggers) and the peak has been processed, the switches \( s \) are activated moving the hysteresis window up of an amount comparable to the hysteresis as shown in Fig. 8(b). Once the comparator is reset to the original state, the hysteresis window is moved to its original state. In our case we switch between ratios of 3:3 and 4:2, resulting in a shift of the hysteresis window without modification in size. The time required to switch the hysteresis window between the two positions is lower than 1 ns. An additional logic circuit, not shown in Fig. 8, makes sure that the comparator stabilizes to its reset state for a specific amount of time (about 10 ns) before a new discrimination can be processed.

**Peak detection** – Once a pulse exceeds the threshold it is processed by the peak and time detectors, and the corresponding analog values are stored in analog memories, queued to be read out. The peak detector is based on the 3-phase (track, peak-detect, readout) configuration [12, 13] and it has been optimized to process pulses with peaking time down to 25 ns.

Peak detection – Once a pulse exceeds the threshold it is processed by the peak and time detectors, and the corresponding analog values are stored in analog memories, queued to be read out. The peak detector is based on the 3-phase (track, peak-detect, readout) configuration [12, 13] and it has been optimized to process pulses with peaking time down to 25 ns.

![Fig. 9 - Examples of peak measurements for a large amplitude (a) and a small sub-hysteresis amplitude (b), along with a detail (c) for the large amplitude. In both cases a ~ 3 mV offsets can be observed, due to large extent to the external buffers used in the measurements.](image-url)
As discussed in the next sub-section, the peak-found signal is used as time detector and as an indicator (flag) that a pulse has been processed and is ready to be read out. While waiting for an event, the peak detectors are held in the track phase where the voltage in the hold capacitor tracks the baseline and its noise. A pulse exceeding the threshold switches the peak detector into the peak-detect phase, where the peak amplitude is detected and stored in the hold capacitor. As discussed in a following subsection, the discriminator can also enable the peak detection in channels neighboring to the ones exceeding the threshold. Once the peak is detected, the circuit is switched in the read-out phase, where the amplifier is re-used as buffer, thus canceling offset and common-mode errors [13].

Figs. 9(a) and 9(b) show two cases of complete measurements for a large amplitude and a sub-hysteresis amplitude respectively, where a signal on the order of 1.5 mV is being detected in front of a comparator hysteresis of about 20 mV. The flag (in correspondence with the peak) and the enable signals are also shown. At enable-low, the readout process starts and the stored peak is multiplexed (peak out) to a dedicated analog output. The ~3.2 mV error between the pulse peak and the peak out is to large extent due to offsets from the external buffers used in the measurement. A similar error ~3 mV is observed in the large amplitude measurement as shown in Fig. 9(c) where the vertical scale has been expanded around the peak value. The measurement in Fig. 9(b) shows evidence of pick-up from digital signals along with a settling time of the peak detector output longer than desired (about 1 µs). Both issues will be addressed in the revision VMM2.

The performance of the peak detection and sub-hysteresis circuits can also be observed from the measurements in Fig. 10 where the peak value versus the input charge is shown for a peaking time of 50 ns and a gain of 1 mV/fC. Fig. 10(a) shows the whole charge range from 0 to 1000 fC, which corresponds to a peak amplitude, including the baseline, up to about 1.13 V, i.e. about 70 mV from the 1.2 V supply rail. Fig. 10(b) shows the same measurement, but with the horizontal scale set from 0 to 20 fC. The effective threshold without sub-hysteresis discrimination is shown (horizontal dashed line, at about 204 mV, i.e. about 17 mV above the baseline) along with the measurements below that threshold made possible by to the sub-hysteresis discrimination (SHD) circuit. The linearity error as shown in Fig. 10(c) is within +/- 1 % through the whole measurement.

**Time detection** - The time detector makes use of the peak-found signal from the peak detector. Compared to a threshold crossing, this signal can offer a better resolution and a significantly smaller time walk [12, 14]. The peak-found signal feeds a Time-to-Amplitude-Converter (TAC) consisting of a voltage ramp that starts at the peak-found and stops at a signal controlled by the data acquisition system. The ramp value is sampled, stored, and read out using a circuit similar to the two-phase peak detector [13] where in the first phase the value is stored in the hold capacitor and in the second phase it is read out by re-configuring the amplifier as buffer, thus canceling errors introduced by offset and common mode gain. The full voltage range of the ramp is about 1 V and the slope can be adjusted to cover 1 V in 125, 250, 500, and 1000 ns. In the revision VMM2 a counter controlled by an external clock operating in the 10 to 50 MHz range will be added in order to provide a coarse timing measurement, thus increasing the dynamic range of the timing measurement, which will result from the superposition of the coarse timing (counter) and the fine timing (TAC).

Fig. 10 shows the measured timing resolution at 200 pF as function of the output signal amplitude for different values of peaking time. The theoretical resolution [12, 14], the time-walk at 200 pF, and the timing resolution at 2 pF are also shown for a nominal peaking time of 25 ns. The difference between the theoretical and the measured resolution at 25 ns is largely due to the actual peaking time which, as already discussed, was larger than the expected one. A timing resolution of less than 200 ps can also be observed for 2 pF and 25 ns.
The data acquisition system enters the readout mode by de-

narchs (ch63 threshold, the peak and time detectors of the two neighbor chan-

nels are processed for peak and time detection. The flags of both chips (flag a

and flag b) are properly asserted. The data acquisition system enters the readout mode by de-

asserting the enable signal. At the first two clock cycles the two peak amplitudes in chip a (from ch64a and ch63a) are routed to the output for analog-to-digital conversion. The timing measurements are read out in parallel at another dedicated analog output, not shown. The 6-bit addresses of those two channels are also made available sequentially at six dedicated digital outputs, not shown as well. With one more clock cycle the token is automatically passed to the neighbor chip b, and the peak amplitude (from ch1b) is read out. In the revision VM2 the analog outputs will be converted directly on chip using integrated ADCs and stored in a deep digital memory (FIFO), thus allowing measurements and derandomization while the memories are simultaneously being read out.

Address of first event in real time - When a first signal exceeds

the threshold or, optionally, when a first peak is found, a flag is asserted (see again Fig. 12, both cases, first two signals from top) and the address of that first signal is made available at six dedicated digital outputs, not shown. The circuit self-resets in 40 ns (adjustable in the revision VM2), ready to process the first signal from the next group of events.

Neighbor processing logic and sparse readout - In normal op-

eration only the signals exceeding the threshold are processed for peak and time measurements. When the neighbor processing circuit is enabled, the channels neighboring to those exceeding the thresholds are also processed for peak and time measurements and queued for read out even if they did not exceed the threshold. The edge channels (i.e. channels 1 and 64) of a chip can communicate to corresponding edge channels of a neighbor chip through bi-directional low-voltage differential signals. If an edge channel exceeds the threshold, the corresponding edge channel of the neighbor chip is processed and read out. This is illustrated in the measurement in Fig. 12. In this case, only the pulse in channel 64 of chip “a” (ch64a) exceeds the threshold while the pulses in channel 63 of chip “a” (ch63a) and in channel 1 of the neighbor chip “b” (ch1b) do not exceed the threshold.

With the neighboring circuit enabled, once ch64a exceeds the threshold, the peak and time detectors of the two neighbor channels (ch63a and ch1b) are enabled and the corresponding pulses, both below threshold, are processed for peak and time detection. The flags of both chips (flag a and flag b) are properly asserted. The data acquisition system enters the readout mode by de-

III. PLANS FOR VMM2

Preliminary tests of VMM1 were also conducted at CERN (European Organization for Nuclear Research, Geneva, Switzerland) with both MICROMEGAS and TGC detectors. The results are encouraging and will be reported elsewhere.

Although the preliminary results on VMM1 are in close agree-

ment with the expectations, several issues have also been found. The first is the failure of the switchable compensation network of
the charge amplifier resulting in a reduced speed when operating at large input capacitive loads. The second consists of an error in the design of the ESD protection network at the channel input resulting in a leakage current of several nA. Several parasitic effects were observed in the mixed-signal circuits affecting the performance at small signal amplitudes.

All these issues will be addressed during the design of the second prototype, VMM2. A number of improvements were requested and will be implemented, among which is a wider range of gains to cover up to 5 pC on the low-gain end, and up to 20 mV/fC on the high-gain end. Finally, ADCs and FIFO are being designed to allow simultaneous measurement and readout and to provide a fully digital interface. VMM1 will also be tested for radiation tolerance and the results will be used to harden VMM2.

ACKNOWLEDGMENT

We are grateful to Ken A. Johns and Sarah L. Jones (University of Arizona, USA) and to George Iakovidis (BNL and NTU, Athens, Greece) for remarkable help with developing the test system and performing the measurements at CERN, to Nachman Lupu (Technion Haifa, Israel) for valuable discussions during the design and characterization with the Thin Gap Chamber prototype, to Howard Gordon (BNL, USA) and Craig Woody (BNL, USA) for the effective encouragement and support, to the VMM1 ATLAS review team (J. Oliver, M. Newcomer, R. Richter, P. Farthouat) for the trust and valuable feedback during the review, and to the ATLAS Collaboration.

This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

REFERENCES