The laser control of the muon g-2 experiment at Fermilab

A. Anastasi\textsuperscript{a,j}, A. Anastasio\textsuperscript{b}, S. Avino\textsuperscript{b,r}, F. Bedeschi\textsuperscript{f}, A. Boiano\textsuperscript{b}, G. Cantatore\textsuperscript{c,q}, D. Cauz\textsuperscript{c,o}, S. Ceravolo\textsuperscript{a}, G. Corradi\textsuperscript{a}, S. Dabagov\textsuperscript{a,m}, P. Di Meo\textsuperscript{b}, A. Driutti\textsuperscript{c,o}, G. Di Sciascio\textsuperscript{d}, O. Escalante\textsuperscript{b,i}, G. Galgliardi\textsuperscript{b,r}, R. Di Stefano\textsuperscript{b,n}, C. Ferrari\textsuperscript{a,R}, A. T. Fienberg\textsuperscript{p}, A. Fioretti\textsuperscript{f,s}, C. Gabbanini\textsuperscript{f,s}, A. Gioiosa\textsuperscript{e}, D. Hampai\textsuperscript{i}, D. W. Hertzog\textsuperscript{p}, M. Iacovacci\textsuperscript{i,i,p}, M. Incagli\textsuperscript{f}, M. Karuza\textsuperscript{c,k}, J. Kaspar\textsuperscript{p}, A. Lusiani\textsuperscript{f,h}, F. Marignetti\textsuperscript{b,n}, S. Mastroianni\textsuperscript{b,1}, D. Moricciani\textsuperscript{d}, A. Nath\textsuperscript{b}, G. Pauletta\textsuperscript{c,o}, G.M. Piacentino\textsuperscript{e}, N. Raha\textsuperscript{d}, L. Santi\textsuperscript{c,o}, M. W. Smith\textsuperscript{i-p}, G. Venanzoni\textsuperscript{f}

\textsuperscript{a}Laboratori Nazionali Frascati dell’INFN, Frascati, Italy
\textsuperscript{b}INFN, Sez. Napoli, Napoli, Italy
\textsuperscript{c}INFN, Sez. di Trieste e G.C. di Udine, Italy
\textsuperscript{d}INFN, Sez. di Roma Tor Vergata, Roma, Italy
\textsuperscript{e}INFN, Sez. di Lecce, Lecce, Italy
\textsuperscript{f}INFN, Sez. di Pisa, Pisa, Italy
\textsuperscript{g}Istituto Nazionale di Ottica del CNR, Pisa, Italy
\textsuperscript{h}Scuola Normale Superiore, Pisa, Italy
\textsuperscript{i}Università di Napoli “Federico II”, Napoli, Italy
\textsuperscript{j}Dipartimento MIFT, Università degli studi di Messina, Messina, Italy
\textsuperscript{k}University of Rijeka, Croatia, Rijeka
\textsuperscript{l}PN Lebedev Phys Inst, Moscow, Russia
\textsuperscript{m}NR Nuclear University MEPhI, Moscow, Russia
\textsuperscript{n}Università di Cassino, Cassino, Italy
\textsuperscript{o}Università di Udine, Udine, Italy
\textsuperscript{p}University of Washington, Seattle, USA
\textsuperscript{q}Università di Trieste, Trieste, Italy
\textsuperscript{r}Istituto Nazionale di Ottica del CNR, Pozzuoli, Italy

E-mail: stefano.mastroianni@na.infn.it, michele.iacovacci@na.infn.it

\textsuperscript{1}Corresponding author.
The Muon $g-2$ Experiment at Fermilab is expected to start data taking in 2017. It will measure the muon anomalous magnetic moment, $a_\mu = (g_\mu - 2)/2$ to an unprecedented precision: the goal is 0.14 parts per million (ppm). The new experiment will require upgrades of detectors, electronics and data acquisition equipment to handle the much higher data volumes and slightly higher instantaneous rates. In particular, it will require a continuous monitoring and state-of-art calibration of the detectors, whose response may vary on both the millisecond and hour long timescale.

The calibration system is composed of six laser sources and a light distribution system will provide short light pulses directly into each crystal (54) of the 24 calorimeters which measure energy and arrival time of the decay positrons.

A Laser Control board will manage the interface between the experiment and the laser source, allowing the generation of light pulses according to specific needs including detector calibration, study of detector performance in running conditions, evaluation of DAQ performance.

Here we present and discuss the main features of the Laser Control board.

**KEYWORDS:** Laser Calibration, FPGA
1 Introduction

A new measurement of the muon anomaly \( a_\mu = (g_\mu - 2)/2 \) will be performed at FNAL by the Muon g-2 Experiment (E989). This experiment[1] plans to measure the muon anomalous magnetic moment to an uncertainty of \( 1.6 \times 10^{-10} \) (0.14 ppm), derived from a 0.10 ppm statistical error and roughly equal 0.07 ppm systematic uncertainties on the precession rate and magnetic field strength measurements. The unique properties of the Fermilab beam complex enable production of the necessary flux of muons, which will be injected and stored in the muon storage ring.

The beam circulates the ring at a cyclotron frequency \( (\omega_C) \) while the muon spin vector precesses at Larmor frequency \( (\omega_S) \) induced by the ring’s magnetic field. The stored muon decays preferentially produce the highest energy positrons when the muon spin and momentum are aligned. Therefore the number of high energy decay positrons, at fixed position along the muon ring, is time modulated by the anomalous precession frequency \( \omega_a = \omega_S - \omega_C \) which is proportional to \( a_\mu \). In this way, measuring the arrival time of the high energy positrons allows determination of the anomalous precession frequency in the muon storage ring.

The detector, consisting of 24 electromagnetic calorimeter stations placed on the inside radius of the magnetic storage ring, must accurately measure the hit times and energies of the positrons, which curl to the inside of the ring following the muon decay. For maximum acceptance the calorimeters are located partly within the storage ring’s highly uniform 1.45 T magnetic field and extend inward radially to a region where the field falls to about 0.8 T. Each calorimeter station consists of 54 lead fluoride (PbF\(_2\)) crystals in a 6 × 9 wide array, with each crystal read out on the rear face by a large-area SiPM coupled directly to the crystal surface. A 12 bit waveform digitizer samples each photodetector channel at a rate of 800 MBPS and the data are transferred to a bank of GPU
The light pulses are monitored through the monitoring electronics both at the source, or at laser output, Source Monitor, and at the end of the distribution system, Local Monitor, before delivery to the calorimeters. processors for on-line data processing.

To achieve a systematic uncertainty of 0.07 ppm, a high accuracy of detector calibration and stability, on both short and long time scales, is needed. Over the time window of 0-700 $\mu$s into a muon fill, where the instantaneous positron rate drops by more than four orders of magnitude, the gain fluctuations must be limited at the sub per-mil. Over longer time scales, the gains should be stable at the sub percent level.

Consequently, a laser calibration system [2], [3], [4] is used to continuously calibrate the calorimetric stations, that is, short laser pulses are supplied directly to each calorimeter crystal through a chain of optical fibers and other optical elements. Each laser pulse should mimic the SiPM response to an electromagnetic (EM) shower for an impinging positrons. The intensity of the light source and the stability of the light distribution system are monitored at the level of precision required by the experiment, that is percentage variation in time of $10^{-4}$ and single pulse measurement precision of $10^{-3}$.

In addition the laser calibration system will allow continuous checking of the linearity of the SiPM response. Within the calibration system a key role is played by the Laser Control Board (LCB) which manages the interface between the beam cycle and the calibration system itself. It takes care of the generation of the laser pulses and distributes the time reference signals to the monitoring electronics. A key feature of the LCB is the capability of starting the lasers according to simulated positron arrival time, as in the experiment. This is accomplished by using random number generators which are implemented in two different modes on the same platform.

In Sec. II the calibration system will be presented; the LCB architecture and implementation will
be discussed in Sec. III and then in Sec. IV the test results are presented and discussed.

2 Laser Calibration System

Fig. 1 displays a sketch of the laser calibration system design. Six laser heads source the pulses to the calorimeters. Each laser provides light to four calorimeters, namely each laser beam is split into four lines that are focused into long quartz optical fibers. The output of each fiber is transmitted through a diffuser that uniformly illuminates a bundle of optical fibers. Each individual fiber of the bundle delivers light to a single calorimeter crystal.

Along its path to the crystals, the light is monitored twice, the first time at the exit from the laser (Source Monitor, SM), another time to the delivery point (Local Monitor, LM) by sending a fiber back to the laser hut and comparing its output pulse with a reference pulse given by the SM. Specific photodetectors are devoted to this and specific electronics have been designed to read, process and digitize the corresponding signals. Also, these electronics provide the supply voltage to the photodetectors, read the different temperatures (environmental, on the preamplifier and on the board itself) and, eventually, stabilize the performance of the readout channel. Indeed the electronics is able to self generate pulses of known amplitude and to send them at the input of the readout channel, meaning that it has capability of self-calibration.

The LCB manages the interface between the calibration system and the experiment’s synchronous control system, the Clock and Control Center (CCC). The CCC provides the triggers to the LCB timed appropriate to delivery of the muon beam. The LCB, described in detail below, decodes the trigger mode and generates the suitable laser pulse sequence. Given that the LCB operation is driven by the beam arrival, a simplified scheme of the laser pulse sequences is shown in Fig. 2 with respect the muon beam structure of the E989 experiment. Nonetheless the same system could accommodate widely different schemes from what we have in the muon g-2 experiment at Fermilab.

3 The Laser Control Board

3.1 The requirements

The laser must be operated in two distinct modes. The first is enabled, during physics runs, to correct for systematic gain variation of the SiPMs caused by the high muon decay rate at the beginning of the muon fill. The second is devoted to the test runs, without beam, in order to exercise the detector and DAQ with specific laser pulse time sequences and to study the SiPM response to double pulses. Also, the laser is used for time alignment of the SiPMs in a calorimeter and between calorimeters. Accordingly the LCB allows:

- Calibration mode, or generation of pulse trains, at programmable frequencies, superimposed on the physics data provided in a 700 μs muon fill. The pattern is shifted by a fixed time in order to have the 700 μs sampled in 140 points. The number of samples at each point is determined by the calibration goal of a 10^{-4} relative error. Considering the number of
photon in each pulse, the muon fill repetition rate and the rate of calibration pulses within the 700 $\mu$s window, we expect that a few thousand samples at each point will be sufficient to reach the needed accuracy. This translates into a capability of calibrating the entire detector in one to two hours.

- Physics event simulation, or operation in “flight simulator” mode, entails triggering the laser according to the exponentially decreasing time function, $e^{-t/\tau}$, as expected in the experiment due to muon decay. In fact, an essential feature of the LCB is the capability of generating pulses, or triggering the laser, according to any time distribution. This mode provides flexible testing of the SiPMs to determine, for example, their response linearity and gain stability. Moreover it allows fully realistic tests of the readout electronics, DAQ and data processing.

- Synchronization of detectors and electronics by providing a reference pulse on request, or in connection to an accelerator machine signals.

![Figure 2](image)

**Figure 2.** The main cycle of the accelerator machine is represented by 16 repetitions of muon fill and decay windows (700 $\mu$s long, represented by the square signal in figure, or "in-fill phase") typically separated by 10 ms (or "Out-of-fill phase"). Actually there are two bunches of 8 filling-decay windows separated by about 200 ms and 1000 ms. In the lower part is shown the laser pulse sequence, which is structured according to many self-explaining parameters. Details of are in section 3.3

### 3.2 The architecture

The interface with the CCC system is implemented in the Clock Logic block where "beam arrival" signal coming from the CCC, starts the laser procedure. The LCB checks the status of the monitoring electronics boards (SM and LM) and if no error flag is active, the LCB initiates the laser patterns. The pulse generation is managed in the Laser Control Mode block and the details are provided in the next subsection.

The interface with the laser drivers and the timing characteristics of the fan-out channels are managed in the Laser Logic block. The laser status and the acknowledge signal from the driver is
monitored. Any malfunctioning of the light distributions or an anomalous delay of the feedback signals results in an error condition that can be latched. An error condition ends pulse generation. The system configuration and its monitoring are controlled by an embedded processor whose main

![Laser Control Block Diagram](image)

**Figure 3.** The Laser Control Block diagram; the architecture is based on a hybrid system which hosts an FPGA and an ARM processor (dashed line). The Laser Control Modes block, outlined by the solid line, contains the logic for all the firing modes.

components (RAM block, Ethernet and I/O peripherals) are depicted in Fig. 3. The use of an embedded processor has the advantage of flexibility regarding the I/O interface. The main elements between busses and peripheral are I2C, SPI, ADC and GIPO.

### 3.3 Implementation of the LCB

In calibration mode the LCB acts as a pulse generator with a programmable frequency and a programmable offset for the pulse train. The generation parameters for both “in-fill” and “out-of-fill” time gaps can be separately configured. As shown in Fig. 2, the “in-fill” mode is characterized by a Sync/RST pulse with a delay ($\Delta t_{\text{Sync/RST}}$) from the CCC trigger and, then, the laser pulse sequence can start within $t_{\text{seq}}$ with respect the same CCC trigger. The pulse train of sequence is defined by frequency ($1/T$), number of pulses ($N_{\text{pulse}}$) and time offset ($T_{\text{shift}}$). A typical pulsing frequency is 10 kHz. The time offset $T_{\text{shift}}$, calculated as a fraction of pulse period (T), is typically $T/20$ or 5 $\mu$s.

The “out-of-fill” sequence is started by an appropriate CCC trigger and, as in the “in-fill” mode, is defined by frequency ($1/T$), number of pulses and time delay with respect the CCC trigger. The pulse generator module has been realized in VHDL language. The logic of the pulse generator is designed synchronously with a 100 MHz clock.
In simulation mode the LCB is able to repeatedly provide a time sequence and a mean number of pulses according to any time distribution. As the arrival times of decay positron in the experiment are exponentially distributed with a decay time of 64.4 s, the simulation mode envisages a time generation according to an exponentially decreasing function.

Generally to solve such a problem one would think of a custom solution, capable of high performances thanks to a careful design of the hardware components. Such solution, though, would typically result optimized for a well defined function and then scarcely transportable from one architecture to another. The increasing of power and flexibility of the embedded processor make a general solution possible for such a demanding application by computing resources.

The LCB contains two different implementations of the simulation mode. The first (HW) is fully realized in hardware (FPGA) while the second (SW&HW) consists in a hardware generator of pulse sequences whose patterns are provided by software modules (CPU).

Related to the latter approach, the generation of random patterns applied to more general functions and implemented by means of hardware controlled by an ARM processor has been described in [6]. The details of the two implementations, along with results on performance test, are reported in the two following sub-sections.

### 3.3.1 The HW approach

The time distribution of the hits along the 700 µs window has a statistical density of events that can be represented by the exponential function $D(t) = h e^{-t/\tau}$. While $\tau$ is 64.4 µs, the parameter $h$ must be chosen in order to have a faithful reproduction of the mean event number in the 700 µs. The pulse generator has been implemented on an FPGA device, its architecture is shown in Fig. 4. The time counter is fed by the device’s system clock and it is started by an external strobe signal. The exp function is applied to the counter output $t$ and the result is multiplied by a random number $r$ in order to get the proper delay $d$ between two consecutive pulses. To generate a uniform distribution of random numbers, a 32 bit linear feedback shift register (LFSR) has been used [7], [8], [9], which is based on a registered shift register whose input bit is a linear function of its previous state. The feedback is provided by an xor combination of some bits tapped along the sequence as shown in the zoomed box of Fig. 4. The tap weights are 1 for taps “connected” and 0 for those “not connected”. The sets of feedback taps, which define the feedback polynomials, are fundamental to allow the maximum length sequence covering a uniform distribution. There are many valid feedback polynomials[10], in this paper we used the taps (1,25,26,31) so that the corresponding polynomial is $1 + x + x^{25} + x^{26} + x^{31}$.

The exponential function should be implemented avoiding sequential statements because they are timing consuming and do not allow an efficient implementation. For this reason the IEEE library functions have been excluded. Other algorithms such as CORDIC [11] or the one based on parabolic synthesis [11], that boast higher performance, are not fast enough. Because our application does not require a high precision calculation, we have opted to implement the exponential using a look-up table of 4096 unsigned long.
3.3.2 The SW&HW method

The main components of the SW&HW implementation, depicted in Fig. 5, can be divided in two groups: those operating on the CPU for event simulation and the others hosted in the FPGA for pulse generation control.

A FIFO on the FPGA holds a list of time records during which the trigger should be fired, where the time binning granularity is a user-specified parameter $\Delta t$. When a new muon fill signal arrives from the CCC system a Finite State Machine (FSM) in the Laser Control Mode block commences a firing pattern according to the sequence stored in the FIFO and pulls the first firing time from the FIFO. A local time counter is incremented appropriately given $\Delta t$. When the local counter matches the firing time from the FIFO, the laser is triggered and the next firing time is pulled from the FIFO.

The comparison and generation processes go ahead until the end of the time sequence is detected. The software processes inside the CPU keep the FIFO countainig the time sequences almost full by pushing asynchronously the frames of time records in it.

The Rnd Gen process manages the extraction of the sequences following the probability distribution function defined in the Init block at the setup and configuration phase. The method used for the random number generation is based on the standard library function of C/C++ software packages. The Formatter process assembles the sorted list of times according a defined granularity ($\Delta t$). The Sender controls the transfer of the pattern from the Formatter to FPGA FIFO, with handshaking accomplished using a “Progammed Full Flag”. This flag becomes active when the FIFO counter reaches a predetermined value or 90% of depth. When it is not active, the Sender transfers the next prepared frame. The status of the FIFO is continuously monitored by the Spy&Mon process.
The initialization process (Init) manages the setup for both hardware and software components; the user can configure the parameters of the pattern, namely the number of laser pulse in the 700 μs window extracted according to Gaussian distribution with given Mean and RMS values, the time binning granularity Δt and the timing parameters represented in Fig. 2.

The next section presents the implementation of the LCB and the results of the major tests performed on it.

4 Test Results

A hybrid platform with FPGA board and ARM-based processor has been used for the LCB implementation, namely of both methods, the HW and the SW&W. Different FPGA devices (ML507, Kintex7 and Nexys3)[13] and ARM processor (BeagleBone[14] and Aria[15]) have been tested. Figure 6 shows a prototype system based on a Nexys3 board containing a Spartan6 XC6LX16-CS324 component and a BeagleBone board containing a Sitara ARM Cortex-A8 processor running at 1 GHz. The two devices are interfaced by means of a custom board that handles the input/output signals.

The ARM and FPGA devices must communicate both to allow configuration / monitoring and to transfer data. The system utilizes the synchronous serial communication protocol SPI operated in full duplex mode at a 5 MHz clock frequency. The platform provides several GPIO pins to implement the interrupt service routines and some ADC channels to sample analog signals (such as temperatures, voltages etc.).

The test bench illustrated in Fig. 7 was assembled to verify the LCB functionalities. A signal generator provides the time reference signals that in the g-2 experiment come from DAQ and accel-
Figure 6. The LC implemented by a Spartan6 FPGA board, a BeagleBone Black CPU and a custom board which plays as interface for signal exchange and component communication.

Figure 7. Block diagram of the test bench setup.
erator clock. As in the experiment, a PicoQuant LDH-P-C 405M pulsed diode laser was used. The laser receives the trigger signal from the LCB and drives the sync-out signal synchronously with the light pulse. The trigger pulse and the sync-out signals are acquired by a LeCroy oscilloscope and/or by a companion board specifically designed for monitoring and time measurements. The system configuration is managed by a remote PC connected via an Ethernet link or locally over a USB connection.

Several tests have been carried out to validate the operation modes. In particular, using the programmable fixed-frequency mode, measurements have been done in a large range of frequencies (from hundreds of Hz up to MHz) and with different time shifts. Fig. 8 shows the distribution of frequency measurements at 10 kHz carried out with the oscilloscope. The relative spread of the distribution ($RMS/\text{Mean}$) is about $3.3 \times 10^{-6}$. The frequency stability of the pulse generator is also monitored long term.

The second test session was dedicated to the “flight simulator” mode. The $SW\&HW$ implementation was first verified by use of reference probability functions; namely a pulse train at a fixed frequency, a uniform distribution and a Gaussian distribution. The exponential distribution function was then tested by varying the average number of laser hits in the muon fill window from 10 to 100. In each configuration about ten thousand muon fill were recorded and the effective number of pulses measured. We expect the observed number of pulses to be lower than the requested number because there is a finite probability for two trigger times to occur within a single time bin. Fig. 9 reports the average deficit (%) in the number of triggers observed versus the number of expected in (0-700) $\mu$s time interval. Up to 100 requested pulses, the deficit of generated triggers is lower than 0.4%. The effective number of hits determines the data size of the frame to be transferred from the ARM processor to the FPGA device. With this system architecture each time hit is defined by a 16 bit word.

![Frequency distribution obtained at 10 kHz and with $t_S = 5\mu s$, measured with a LeCroy oscilloscope.](image)

**Figure 8.** Frequency distribution obtained at 10 kHz and with a $t_S = 5\mu s$, measured with a LeCroy oscilloscope.
Figure 9. Deficit in the number of pulses due to a 10 ns time granularity for an exponential distribution in the time interval of (0-700) \(\mu s\). \(N_{hit_g}\) is the number of generated pulses in the time interval and \(N_{hit_m}\) is the number of detected pulses in the same time interval. The measurements have been collected with the \(SW&HW\) method. The dotted line shows the expected behavior.

This version has been used to do several tests on the SiPMs [5] and was recently used in different test beams in particular at LNF [12] and at SLAC. The possibility to modify the number of pulses in a such large range was an important benefit allowing measurements of fundamental quantities like the SiPM gain versus time and versus the number of laser hits in the 700 \(\mu s\), moreover studies of the sustainable data rate of the DAQ system and test of DAQ performances.

The last section of tests is dedicated to the comparison between the two implementations (\(SW&HW\) and \(HW\)). The \(HW\) design has been optimized with respect to the parameter \(h\) in order to collect about 64 pulses on average per fill. Fig. 10 shows a good agreement between the distributions of pulses obtained using the two implementations.

In the test bench we fully exploited the LCB capability of self-generating the signals coming from the CCC system, what allows operation of the calibration system and calorimeters in absence of beam.

5 Conclusions

The LCB handles the laser calibration system of the new Muon g-2 Experiment at Fermilab. It allows the laser to be triggered during physics runs, within and outside the 700 \(\mu s\) muon fill. In such a way we can correct systematic gain variation of the SiPMs. Also, it is capable of managing the laser pulsing according to specific time distribution, as required in calibration runs. It is able to operate without any external control signal and to simulate the beam structure, so that detector and DAQ can be exercised in any condition.
The core of the LCB is a pulse generator with two different implementations (the first fully realized in an FPGA and the second utilizing an ARM processor to control the final generation of pulses) which allow both a high level of flexibility, due to software benefits, and a high level of performance typical of hardware solutions. The time resolution of the pulse generator is 10 ns. The LCB has been installed at Fermilab on February 2017 in time for the first engineering run.

Acknowledgements
This research was supported by Istituto Nazionale di Fisica Nucleare (Italy) and by the EU Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Grant Agreement No. 690835. SD would like to thank the support by the Competitiveness Program of National Research Nuclear University MEPhI. The authors thank Lawrence Gibbons for a careful reading of the manuscript and useful suggestions.

References

Figure 10. Time distribution of pulses generated according to an exponentially decreasing function “flight simulator” for the two implemented methods. The measurements have been realized by a LeCroy oscilloscope.


