electronics

MDPI

Article

Ultra-Fast Digital Silicon Photomultiplier with Timestamping Capability in a 110 nm CMOS Process

Tommaso Maria Floris [1,2,*], Marcello Campajola [3,4], Gianmaria Collazuol [5,6], Manuel Dionísio Da Rocha Rolo [7], Giuliana Fiorillo [3,4], Francesco Licciulli [8], Mario Nicola Mazziotta [8], Lucio Pancheri [9,10], Lodovico Ratti [1,2], Luigi Pio Rignanese [11], Davide Falchieri [11], Romualdo Santoro [12,13], Fatemeh Shojaei [1,2] and Carla Vacchi [1,2]

1 Università degli Studi di Pavia, Dipartimento di Ingegneria Industriale e dell'Informazione, I-27100 Pavia, Italy; lodovico.ratti@unipv.it (L.R.); fatemeh.shojaei01@universitadipavia.it (F.S.)
2 Istituto Nazionale di Fisica Nucleare, I-27100 Pavia, Italy
3 Università degli Studi di Napoli Federico II, Dipartimento di Fisica "Ettore Pancini", I-80126 Napoli, Italy; macampajola@na.infn.it (M.C.)
4 Istituto Nazionale di Fisica Nucleare, I-80126 Napoli, Italy
5 Università degli Studi di Padova, Dipartimento di Fisica e Astronomia "Galileo Galilei", I-35131 Padova, Italy
6 Istituto Nazionale di Fisica Nucleare, I-35131 Padova, Italy
7 Istituto Nazionale di Fisica Nucleare, I-10125 Torino, Italy
8 Istituto Nazionale di Fisica Nucleare, I-70125 Bari, Italy
9 Università degli Studi di Trento, Dipartimento di Ingegneria Industriale, I-38123 Trento, Italy
10 Trento Institute for Fundamental Physics and Applications, Istituto Nazionale di Fisica Nucleare, I-38123 Trento, Italy
11 Istituto Nazionale di Fisica Nucleare, I-40127 Bologna, Italy; davide.falchieri@bo.infn.it (D.F.)
12 Università degli Studi dell'Insubria, Dipartimento di Scienza e Alta Tecnologia, I-20200 Varese, Italy; romualdo.santoro@uninsubria.it
13 Istituto Nazionale di Fisica Nucleare, I-20133 Milano, Italy
* Correspondence: tommaso.floris01@universitadipavia.it

Abstract

A monolithic digital Silicon Photomultiplier (SiPM) featuring 1024 microcells with a 30-micrometer pitch and a 50% fill factor has been designed in a 110-nanometer CMOS image sensor technology. The device under consideration integrates both SPAD sensors and front-end electronics in the same substrate. It can count up to 1024 photons in less than 22 ns, while assigning timestamps to the first and last detected photons with a time resolution of less than 100 ps. A parallel counter structure combined with a fast adder tree provides photon counting in digital form with low latency, whereas a carefully balanced fast NAND tree ensures a fixed-pattern time uncertainty not exceeding 26 ps. The architecture incorporates in-pixel memory for individual cell disabling and configurable thresholding on the timing signal for noise mitigation. In order to optimize the fill factor, a part of the electronics is placed outside the array, while the most sensitive elements of the timing and counting circuits are laid out close to the sensor, in the SPAD array. A serial readout is employed to provide a single output connection per SiPM, thereby simplifying system integration.

Keywords: CMOS SPAD; digital SiPM; photon counting; fast timing; monolithic sensors

Academic Editor: Costas Psychalinos

Received: 24 February 2026
Revised: 16 March 2026
Accepted: 17 March 2026
Published: 20 March 2026

1. Introduction

Silicon Photomultipliers (SiPMs) have emerged as the detector of choice for a wide range of applications requiring single-photon sensitivity, including medical imaging, high-energy physics, LiDAR systems, and quantum optics experiments. SiPMs consist of

arrays of Single-Photon Avalanche Diodes (SPADs) operated in Geiger mode, offering several advantages over traditional photomultiplier tubes, including compact size, low operating voltage, insensitivity to magnetic fields, and compatibility with the standard semiconductor manufacturing processes. The fundamental principle of SiPM operation relies on the avalanche multiplication effect, where a single photon can trigger a self-sustaining avalanche current in a reverse-biased p–n junction, producing a detectable electrical pulse [1].

The evolution of SiPM technology has been driven by the demand for improved timing resolution, higher photon detection efficiency (PDE), extended dynamic range, and reduced power consumption. Early SiPM implementations employed analog readout schemes, where the individual SPAD signals were summed to produce an analog output proportional to the number of detected photons. While this approach is simple and effective for many applications, it suffers from several limitations, including susceptibility to electronic noise, limited dynamic range, and the need for external analog-to-digital converters (ADCs) with high sampling rates and resolution. Furthermore, analog readout architectures face challenges in preserving precise timing information, particularly in high-rate photon-counting scenarios where multiple photons arrive within short time intervals [2,3].

Typical readout solutions for analog SiPMs rely on complex external mixed-signal ASICs, such as the ALCOR chip [4] or the OMEGA family architectures (e.g., TRIROC [5] or PETIROC [6]). These systems must manage high input capacitances and significant electronic noise arising from interconnection parasitics, such as bond-wire inductances and PCB traces. Such parasitic elements often lead to a fundamental trade-off between timing resolution and power consumption per channel. Furthermore, the integration of large-area detectors with external ASICs increases system complexity and cost.

To address these limitations, the concept of digital SiPMs (dSiPMs) has gained significant attention in recent years. Digital SiPMs integrate the photon detection elements with on-chip digital processing circuitry, enabling direct digital output without the need for external ADCs. This monolithic integration approach offers several key advantages: improved noise immunity through digital signal processing, enhanced timing accuracy via on-chip Time-to-Digital Converters (TDCs), increased dynamic range through parallel counting architectures, and reduced system complexity by eliminating external analog front-end electronics [7,8]. The state-of-the-art in digital SiPM technology has seen remarkable progress, with recent implementations demonstrating sub-100 picosecond timing resolution, MHz-level frame rates, and sophisticated on-chip data compression schemes [9].

Monolithic digital SiPM solutions have been implemented in various CMOS technology nodes, leveraging the maturity and cost-effectiveness of standard semiconductor processes. Notable examples include perimeter-gated SPAD-based architectures with asynchronous address-event representation (AER) readout for positron emission tomography (PET) applications [10], segmented digital SiPMs with direct multi-bit digital outputs [11], and advanced implementations featuring per-pixel or per-quadrant TDCs with fast parallel counter architectures [12]. These developments have demonstrated that monolithic integration not only simplifies system design but also enables novel functionalities such as pixel-level masking for dark count rate management, on-chip hit mapping for spatial resolution enhancement, and configurable trigger modes for noise rejection.

For a digital SiPM to be a practical alternative to conventional analog SiPMs in applications such as medical imaging, LiDAR, or particle physics, it must first and foremost fulfill the core requirement of accurately determining the number of detected photons with minimal latency, ideally within a few nanoseconds, to keep pace with high photon fluxes. While analog SiPMs rely on external electronics for signal digitization and timing extraction,

the digital approach enables the integration of additional functionalities directly on-chip. Among these, timestamping of photon arrivals can be offered as a valuable enhancement, supporting time-of-flight or coincidence-based measurements without requiring complex off-chip circuitry. Moreover, by delivering photon-count information in native digital form, the dSiPM eliminates the need for external analog-to-digital conversion, thereby reducing system complexity, improving noise immunity, and enabling more compact and scalable readout architectures.

This paper presents an ultra-fast digital Silicon Photomultiplier implemented in 110 nm CMOS technology, featuring a square array of 1024 SPADs with integrated fast digital readout electronics. The proposed architecture employs a dual-path processing scheme, separating energy measurement (photon counting) from timing measurement to optimize both functions independently. The energy measurement path utilizes 16 parallel counters combined through an external adder tree to produce an 11-bit photon count, while the timing path implements a carefully balanced hierarchical NAND tree to generate enable signals for two TDCs with sub-100 ps resolution. The device can count up to 1024 photons, with the result available within 22 ns of the last detected photon, and provides timestamps for the first and last detected photons with sub-100 ps accuracy. Through monolithic integration and the use of fast digital electronics, the proposed dSiPM delivers direct digital output without any analog-to-digital conversion, offering significant advantages in terms of system simplification, power consumption, and overall performance.

2. System Architecture Overview

The system architecture, illustrated schematically in Figure 1, comprises a square matrix of 1024 SPADs organized in 32 rows and 32 columns over a 1 mm^2 area. The SPADs employed in this design were characterized in a test chip fabricated in the same 110 nm CMOS technology used for the digital SiPM [13]. They exhibit a breakdown voltage V_{BD} of 18.5 V, with relatively small chip-to-chip dispersion, characterized by a standard deviation of the order of a few tens of mV. Their dark count rate (DCR) performance is shown in Figure 2, showing the cumulative DCR curve for a set of 832 SPADs. The measurement was obtained for an excess bias voltage V_{EX} = 1.5 V and at a temperature T = 25 °C. In the plot of Figure 2, the DCR is normalized to the SPAD active area. Although fabricated in a standard commercial CMOS process not specifically optimized for SPADs, the achieved DCR levels are reasonably good, indicating that acceptable single-photon detection performance can be obtained even with a CMOS, non-custom fabrication technology. The SPADs and their associated readout electronics are monolithically integrated in the same silicon substrate, forming a dSiPM. This device measures incident light energy and generates two timestamps corresponding to the arrival of the first and last detected photons.

As already mentioned, the digital SiPM prototype is fabricated in a 110 nm CMOS image sensor (CIS) process, a technology node optimized for low-noise optical sensing. The SPAD microcells, which a schematic cross-section i shown in Figure 3, rely upon a $p+$/low-doped n-implant junction. To prevent premature edge breakdown, a virtual guard ring is implemented using a deep n-well with a retrograde doping profile. The pixel geometry is square with smoothed corners to mitigate electric field concentrations at the edges [14].

While experimental photon detection efficiency (PDE) characterization on the test chip is currently ongoing, a peak photon detection probability (PDP) exceeding 50% at 455 nm is expected for a V_{EX} of 6 V, based on the previous characterization of individual SPAD structures fabricated in the same 110 nm CIS process [14]. Given the design fill factor (FF) of 50%, the peak PDE in the blue spectral range is estimated to be approximately 25%.

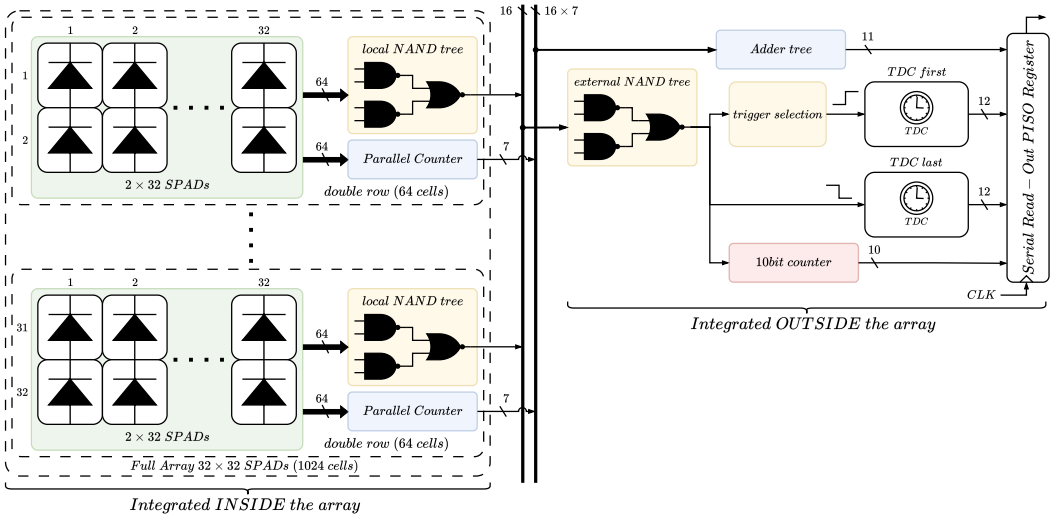

Figure 1. Block diagram of the dSiPM architecture, distinguishing components integrated inside or outside the pixel array.

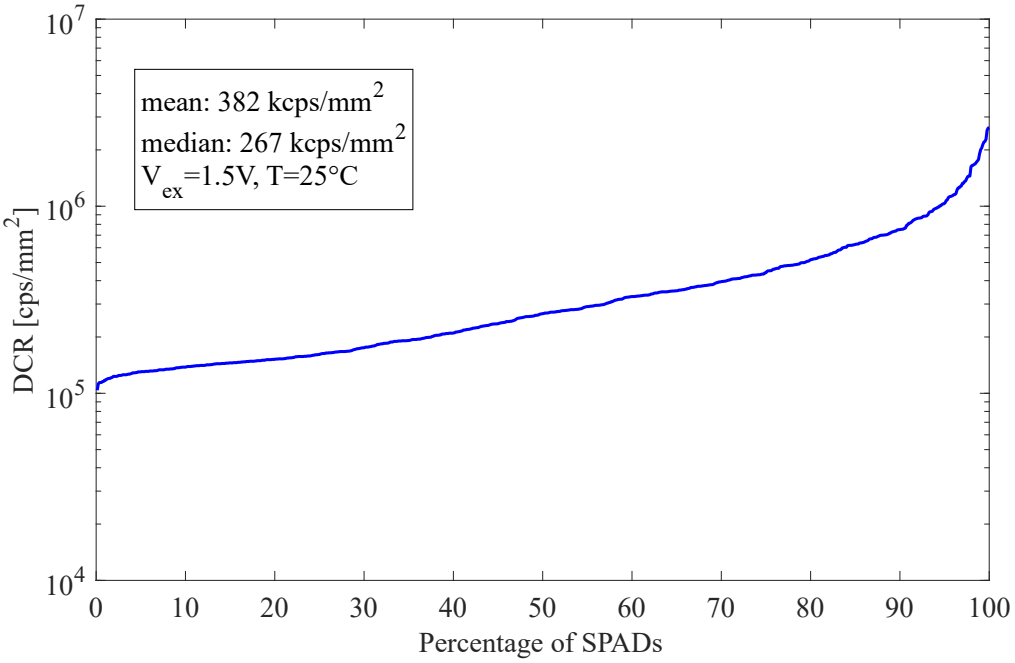

Figure 2. Cumulative distribution of the dark count rate (DCR) across the SPAD array, measured at the operating temperature and excess bias voltage specified in the figure. The mean and median DCR values are also indicated.

Signal processing occurs through two parallel pathways: one dedicated to energy measurement and one to timing measurement. Each pathway is divided into in-pixel electronics embedded within the array as shown in Figure 4 and peripheral electronics located outside the array. The in-pixel circuitry is allocated in the space between adjacent pixel rows, while the peripheral electronics occupy a region adjacent to the array.

The energy measurement path employs 16 parallel counters, each servicing a group of two rows (64 pixels) and producing a 7-bit output. The inputs to the parallel counters are provided by the pixel-level electronics (see the next section). The outputs from the parallel counters are summed externally using an adder tree to generate an 11-bit energy value representing the total number of detected photons.

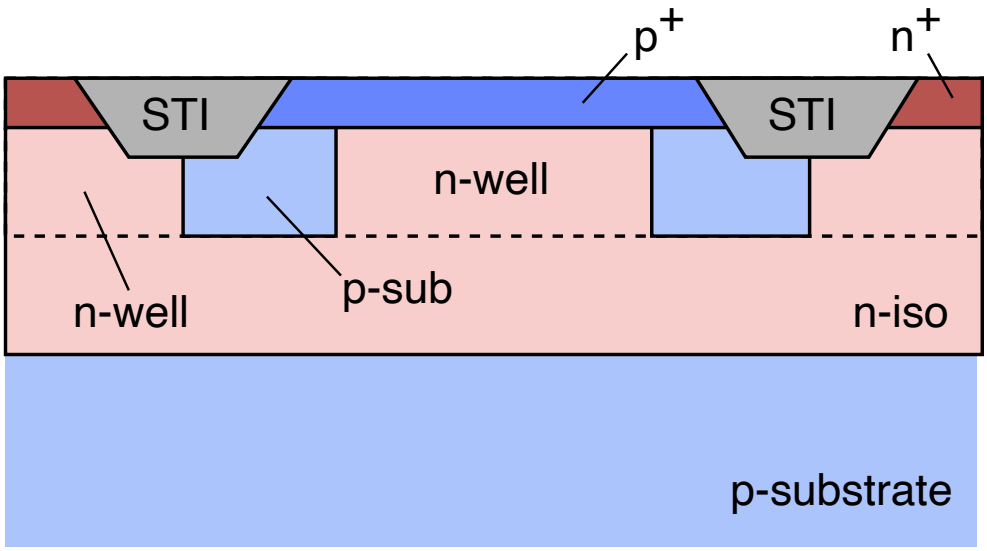

Figure 3. Schematic cross-section of the implemented Single-Photon Avalanche Diode (SPAD). The active area is defined by a p^+ anode and an n-well junction. The device is electrically isolated from the bulk p-substrate by a deep n type isolation layer. Cathode contacts are established through n^+ and n-well diffusions, while Shallow Trench Isolation (STI) is used for lateral isolation between regions.

The design of the timing path follows a similar approach, with the array divided into 16 groups of 64 pixels each. Within each group, pixel outputs (again, coming from the pixel-level electronics, see next section) are combined by an in-pixel logic tree functionally equivalent to a 64-input NAND gate. The outputs from the 16 groups of pixels are then merged by a peripheral logic tree, equivalent to a 16-input NAND gate, producing a single timing signal. The rising edge of this signal marks the arrival of the first photon, while the falling edge indicates the last photon detection. Two Time-to-Digital Converters (TDCs) digitize these transitions relative to an external reference signal, providing precise timing measurements.

Following each acquisition cycle, the energy and timing data are loaded into a Parallel-In Serial-Out (PISO) shift register and transmitted serially off-chip. Additionally, a 10-bit counter monitors the rising edges of the timing signal to record the total number of triggering events. This count, also transmitted via the PISO register, enables the characterization of the device dark count rate (DCR).

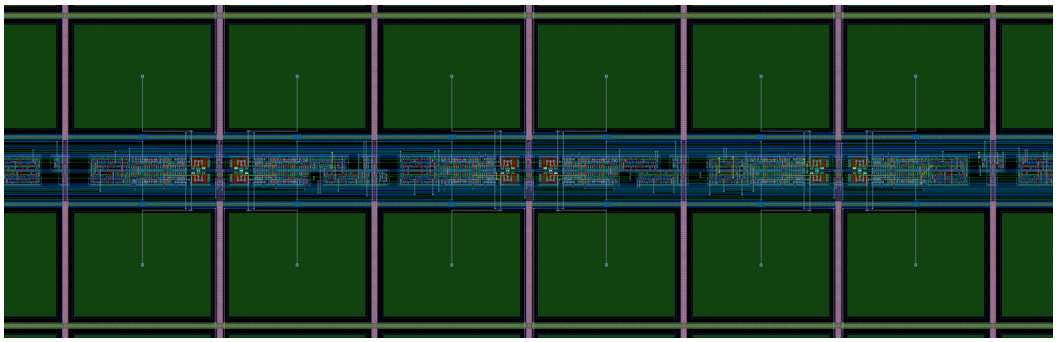

Figure 4. Layout view of the electronics located between two rows of pixels. The spacing between the electronics and the SPADs is constrained by the minimum spacing rules imposed by the fabrication technology.

3. Pixel-Level Electronics

Each pixel has a pitch of 30 µm and achieves a fill factor of 50% through an optimized layout of the SPAD biasing lines, which are shared among adjacent pixels. This design reduces the area occupied by metallization and expands the active SPAD region, as shown in the layout view of a 2 × 2 pixel cluster in Figure 5.

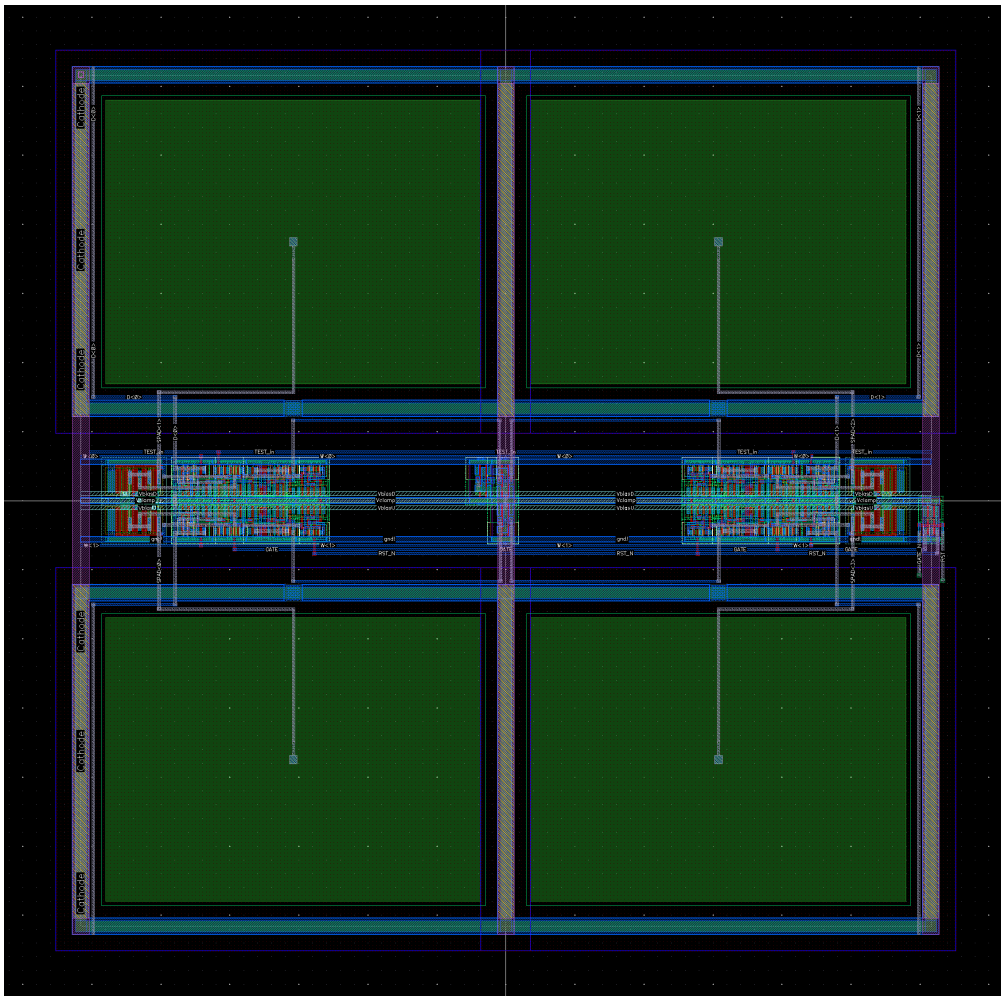

Figure 5. Layout view of a 2 × 2 pixel cluster. Each pixel contains the SPAD and a row of digital circuitry. Both the SPAD bias lines and the power supply lines for the digital cells are shared among adjacent pixels to optimize area usage and maximize the fill factor.

Each pixel comprises a front-end circuit, shown in Figure 6, consisting of a passive quenching circuit, and a digital section that generates signals for counting and timing measurements. The quenching circuit is directly connected to the SPAD anode and employs thick-oxide MOSFET transistors in a cascode configuration, capable of withstanding the high bias voltages required for operation at an excess voltage up to a nominal 6.6 V, which enhances the detector's timing and photon detection efficiency (PDE) properties (while, on the other hand, degrading the dark count rate performance). The quenching network is based on thick-oxide MOSFET transistors in a cascode configuration. This architecture, already involved in other works, like [15], allows the circuit to withstand the high bias voltages required for operation at an excess voltage up to 6.6 V while ensuring that none of the transistor terminal voltages (V_{GS}, V_{GD}, V_{DS}) exceed the safety limit of the used technology (3.3 V).

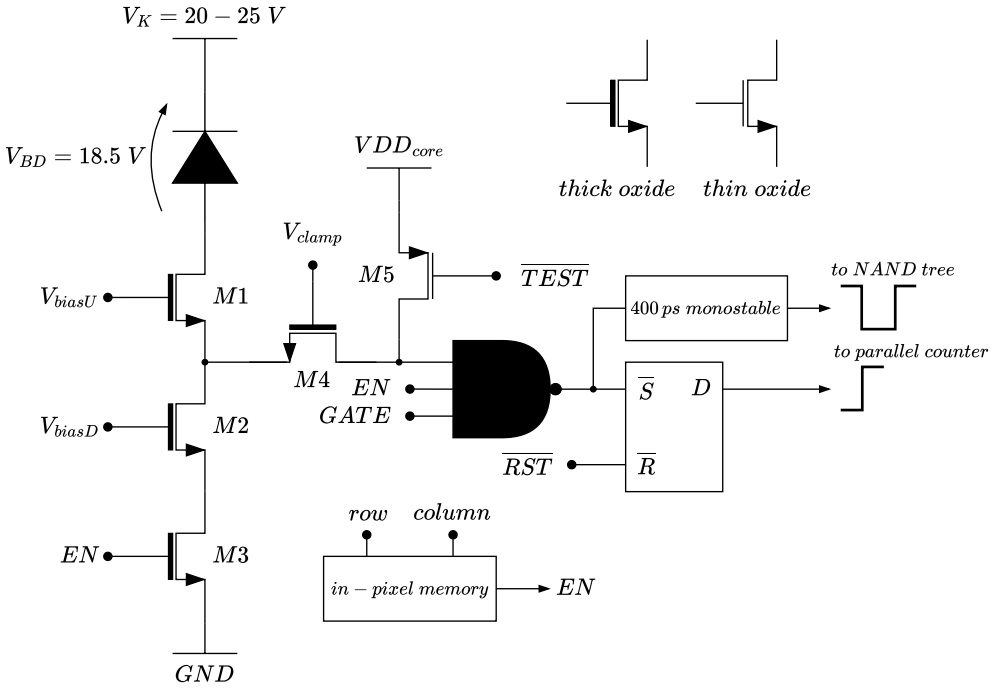

Figure 6. Schematic of the pixel-level electronics. The quenching network is represented at the transistor level, while the digital front-end is depicted at the block and logic gate level. Output signals are labeled with their destination structures and waveform shape: a negative pulse for the monostable output used in timing and a set bit (logic 1) for the pixel-hit counter.

The SPAD rise time and reset time can be adjusted through external control of the gate voltages of transistors M1 and M2. Transistor M4 serves as a voltage-domain interface between the high-voltage quenching circuit and the low-voltage digital circuits based on core, 1.2 V transistors. M4 ensures that its drain voltage never exceeds the gate voltage minus the threshold voltage, thereby protecting the thin-oxide transistors of the digital logic from dielectric stress.

The dead time of the microcell, defined as the period during which the SPAD is insensitive to new photon arrivals following an avalanche, is primarily determined by the passive quenching network. In this design, the gate bias voltages of the cascode transistors serve distinct roles: V_{biasU} is used to set the maximum allowed excess voltage to ensure device reliability, while V_{biasD} regulates the recharge phase of the SPAD. Post-layout simulations indicate that by adjusting V_{biasD}, the dead time can be continuously modulated within a range of approximately 4 ns to 16 ns. During this interval, the quenching network discharges the avalanche-generated carriers from the SPAD depletion region, during which the device remains unresponsive to further incident photons. Since the quenching topology is purely passive, both the recharging time and the resulting insensitivity period depend exclusively on the discharge current characteristics of the quenching network itself. This flexibility is crucial for adapting the sensor to different experimental conditions, allowing for a shorter dead time to maximize dynamic range under high photon flux, or a longer one to effectively suppress afterpulsing effects.

Transistor M1 is controlled by an enable signal stored in a local SRAM cell, which is programmable during device configuration. This can be leveraged for selective activation of pixels of interest or for disabling noisy pixels. Additionally, a thin-oxide transistor (M5) is integrated within the quenching network to emulate a photon events, for functional testing of the processing logic.

The digital front-end generates two distinct outputs: one for energy measurement (counting) and one for timing. The operation is controlled by three external signals: GATE,

TEST, and RESET. Upon photon detection, a logic "1" is stored in an SR latch, whose output drives a local counter to record the event. Simultaneously, a monostable circuit generates a fixed-duration pulse (400 ps) that propagates through the timing NAND tree. The temporal overlap of pulses from all triggered pixels produces an aggregate signal whose rising edge marks the first photon arrival and whose falling edge indicates the last photon arrival (provided that the duration of the signal at the output of the monostable circuit is taken into account).

The RESET signal initializes the SR latch before each acquisition, while the TEST signal enables the injection of photon-like events for verification of circuit functionality. The GATE signal, applied simultaneously to all enabled pixels, defines the acquisition time window, enabling the time-correlated operation of the detector and limiting dark count noise.

The enable state of each pixel is stored in an SRAM cell, programmable via two 32-bit peripheral shift registers, one for rows and one for columns. During configuration, these registers supply the write enable signal and data to be stored, respectively. This architecture enables simultaneous configuration of large pixel groups while maintaining the possibility to control each individual array element.

4. Photon-Counting Electronics

The electronics for energy measurement (i.e., photon counting) is organized in a two-level hierarchical architecture, as illustrated in Figure 7. The first level consists of 16 parallel counters embedded within the array, positioned in the spacing between pixel rows. The second level includes an external adder tree that aggregates the outputs of all parallel counters.

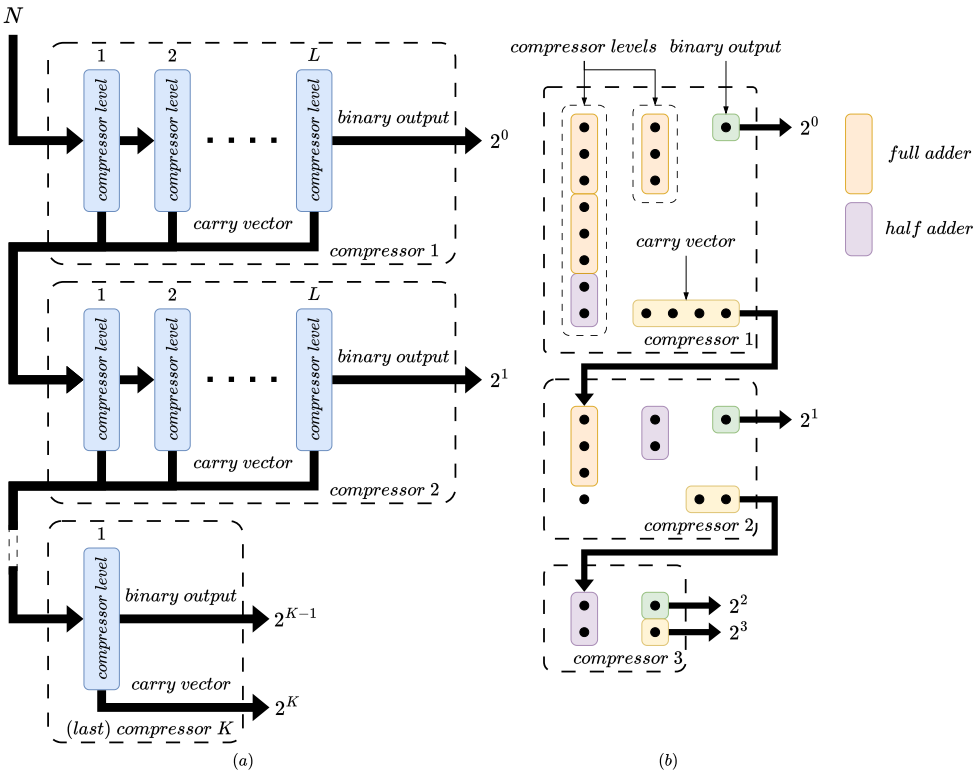

Figure 7. (**a**) Schematic representation of the parallel counter architecture, showing its decomposition into K compressors, each organized into L compression levels depending on the number of inputs, with N inputs feeding the first compressor. (**b**) Example of a parallel counter structure for an 8-input configuration.

Each parallel counter (block diagram shown in Figure 6) processes signals from 64 pixels (two adjacent rows) and produces a 7-bit binary word encoding the number of triggered pixels. All parallel counters share an identical architecture, automatically synthesized from parametric Verilog code. The internal structure of each counter is recursive, composed of $K = \lfloor \log_2(N) \rfloor$ binary compressors, where $N = 64$.

Each compressor stage generates two types of outputs: a weighted binary output with weight 2^{K-1} (where K is the compressor index, starting from $K = 1$) that contributes directly to the final count, and a carry vector formed by the carry bits from the elementary adders (full-adders and half-adders) within the compressor, which feeds the subsequent compressor stage and have a weight equal to 2^K.

Internally, each compressor is organized into $L = \lceil \log_3(M) \rceil$ compression levels, where M represents the number of inputs to that compressor. At each level, adders operate in parallel: sum outputs propagate to the next level, while carry outputs are concatenated to form the carry vector for the next compressor. The final compressor produces both a binary output and a single carry bit, which together complete the 7-bit output word of the parallel counter.

The external adder tree combines the sixteen 7-bit words from the parallel counters to generate the final 11-bit photon count. This structure was synthesized from optimized Verilog code with netlist complexity reduction and logic gate minimization as primary objectives, ensuring both area efficiency and high-speed performance.

5. NAND Tree

The NAND tree generates the enable signal for the Time-to-Digital Converters (TDCs) from the 400 ps pulses produced by the monostable circuits in each pixel. This structure implements the logical function of a 1024-input NAND gate: the output remains at logic level 1 as long as at least one input is at logic level 0. The physical implementation leverages a hierarchical structure of alternating NAND and NOR gate levels, with a total depth of $\lceil \log_2(I) \rceil$ levels, where I is the number of inputs and $\lceil x \rceil$ is the smallest integer larger than x.

The architecture is partitioned into two sections: an in-array portion and a peripheral portion. The in-array section is comprised of 16 local trees, each processing 64 inputs from two adjacent pixel rows through six logic levels. The outputs of these local trees feed the peripheral section, which implements a 16-input tree with four logic levels.

A critical design challenge was balancing the propagation delay across all signal paths. Since the output signal precisely marks photon arrival times, with the rising edge indicating the first photon arrival and the falling edge the arrival of the last one, uniform delay from any input to the output is essential. The delay sensitivity to interconnect characteristics and the input-dependent propagation times of standard cell gates necessitated careful delay balancing during physical layout. This was achieved by adjusting the lengths of interconnection lines between logic stages to equalize path delays.

This approach ensures temporally uniform response across the entire array, preserving the timing integrity of the first and last photon detection with minimal skew (see simulation results in Section 7).

6. Timestamping Circuitry

The system employs two Time-to-Digital Converters (TDCs) to timestamp the first and last photon events. Because the monostable pulses have a fixed duration $t_{mono} = 400$ ps, the falling edge of the global signal does not strictly coincide with the arrival of the last photon. The actual timestamp of the last detected event must be mathematically corrected in post-processing as follows: $T_{last} = T_{TDC\_last} - t_{mono}$. While the monostable

pulse width is not actively tunable at the pixel level, its effective duration can be precisely characterized through timing measurements leveraging the on-chip TDCs, ensuring an accurate calibration of the correction factor.

Both TDCs are based on the current-controlled architecture described in [16], which achieves a time resolution of 100 ps. Each TDC starts its measurement upon receiving a trigger signal and stops upon reception of an external STOP signal. The arrival times of the first and last photons are reconstructed from the measured intervals and the known timestamp of the STOP signal.

Three trigger modes are available for the first-photon TDC, two of which incorporate noise rejection mechanisms to prevent spurious dark counts or uncorrelated background events from initiating timing measurements.

The first mode, without filtering, triggers the TDC directly on the rising edge of the global NAND tree output.

The second mode implements Time-over-Threshold (ToT) filtering (see Figure 8), which generates a trigger only when the NAND tree output remains high for a duration exceeding a programmable threshold. This condition requires temporal overlap of pulses from multiple pixels (produced by the relevant monostables), which can occur only when at least two photons arrive in close temporal proximity. Consequently, only coherent multi-photon events with pulse widths exceeding the threshold can enable the TDC, effectively rejecting isolated noise events.

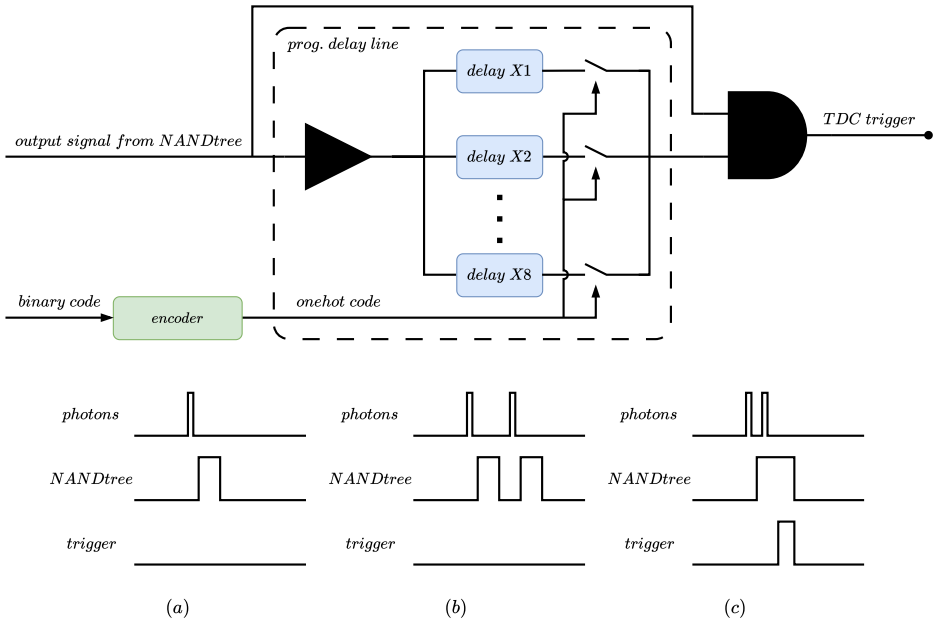

Figure 8. Schematic of the Time-over-Threshold (ToT) trigger system for the first TDC. The delay is implemented as an integer multiple of 100 ps and is selected via 3 configuration bits in the control register. The total programmable delay ranges from 600 ps to 1.4 ns. (**a**) An isolated event generates a NAND tree output pulse insufficient to trigger the TDCs; (**b**) multiple events separated by more than the monostable duration fail to generate a trigger; (**c**) a TDC trigger is produced only when multiple events on the array extend the NAND tree output beyond the selected delay.

The third mode exploits spatial coincidence by monitoring the outputs of the 16 local NAND trees, each associated with two adjacent pixel rows (Figure 9). A trigger is generated only when simultaneous events occur in at least two or three adjacent row pairs within the 400 ps monostable pulse window. This spatial correlation criterion further suppresses spurious triggers from uncorrelated noise.

The trigger mode selection and associated parameters, such as the ToT threshold or the minimum number of coincident row pairs, are configured through a programmable register, enabling flexible and dynamic control of the timing system.

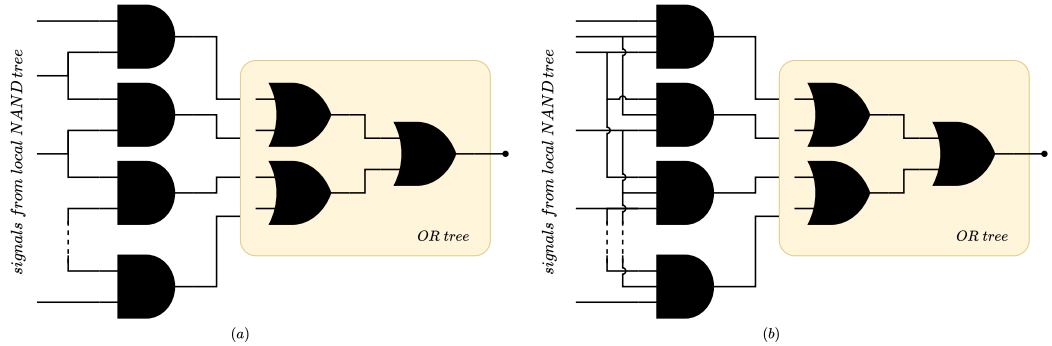

Figure 9. Schematic of the system for detecting event coincidence across adjacent double rows. The circuit (**a**) detects coincidence across two adjacent double rows. The circuit (**b**) detects coincidence across three adjacent double rows. Both configurations employ an OR tree to generate the trigger signal for the first TDC.

7. Post-Layout Simulations

Simulations were conducted after parasitic extraction, following completion of the prototype physical layout (Figure 10). The analyzed structures were the pixel counting circuitry and the NAND tree, both critical to the system's timing and functional performance.

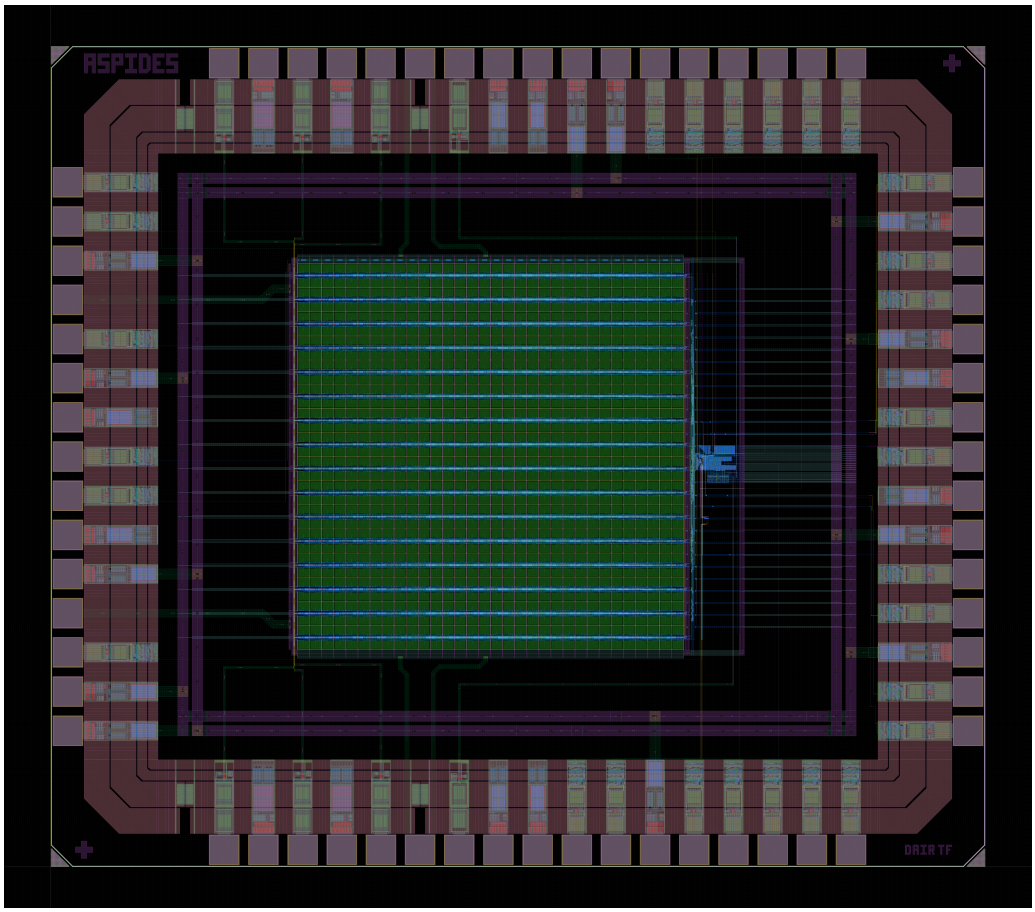

Figure 10. Layout view of the dSiPM prototype developed for the ASPIDES collaboration. It features a total of 66 pads and occupies an area of 5 mm^2.

To validate the front-end and the digital processing logic, the SPAD was emulated using a discrete-component circuit model. The electrical parameters of this model, such as the junction capacitance and the series resistance, were extracted from dedicated TCAD physical simulations to ensure a realistic representation of the avalanche current pulse. It should be noted, however, that the detailed physical modeling of the SPAD was not the primary focus of the research presented in this work, which instead concentrates on the integrated readout architecture and its timestamping performance.

7.1. Counting Circuitry Simulations

Two distinct methodologies were adopted to evaluate the counting logic behavior:

Maximum delay analysis across process corners: In this configuration, all 1024 pixels were stimulated randomly in a defined time window, simulating the worst-case scenario in terms of logic circuit loading. The delay was calculated as the interval between the arrival of the last photon at the array and the final transition at the output of the external adder tree. This scenario involves activation of all binary compressors in the local parallel counters and requires complete propagation of all carry signals through the logic stages. The results show that in the worst case scenario, occurring in the Slow–Slow corner, the delay is 22 ns. An example of the output signals is shown in Figure 11.

Simulations with partial array stimulation (10, 50 or 100 pixels): In this case, groups of 10, 50 or 100 pixels positioned in specific array regions were stimulated to evaluate the influence of spatial location on processing delay and to quantify the energy dissipated per conversion. The results, shown in Figure 12 indicate that the delay depends primarily on the number of activated pixels rather than their position. This phenomenon is attributed to the generation of an increasing number of carry bits as the number of events increases, requiring more extensive propagation through the parallel counter levels and the adder tree outside the array. Consequently, the dissipated energy rises with the number of triggered pixels, as larger counts activate deeper logic stages in both the parallel counters and the external adder tree, resulting in higher switching activity and capacitive loading.

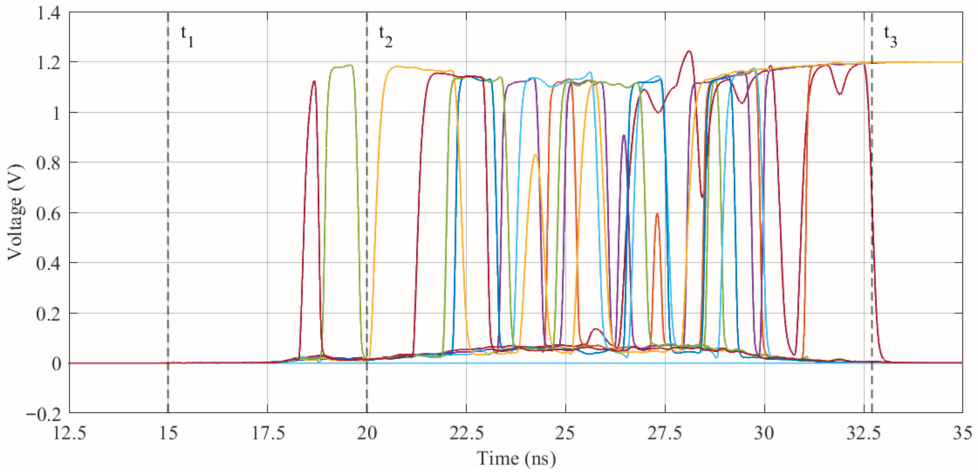

Figure 11. The 11-bit digital output of the photon-counting logic. This result was obtained by simulating the arrival of 100 photons within the time interval from $t_1 = 15$ ns to $t_2 = 20$ ns (i.e., an instantaneous burst 15 ns after the simulation has started). The timestamp t_3 marks the final transition at the output of the counting logic. The minimum settling time, defined as the shortest interval after the arrival of the last photon during which a valid count can be read out, is given by $t_3 - t_2$.

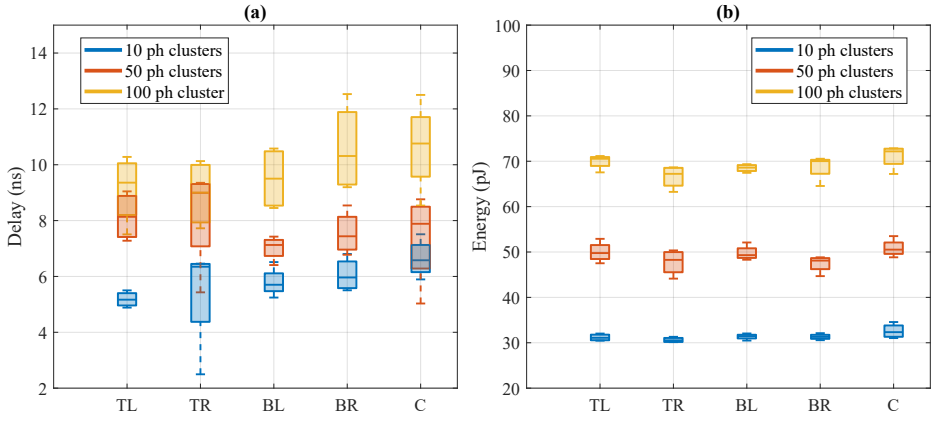

Figure 12. Results of post-layout simulations showing (**a**) propagation delay and (**b**) energy consumption. For each region of the pixel array, multiple hit patterns (with 10, 50 or 100 activated pixels) were simulated to evaluate the dependence of timing and power on both location and number of triggered pixels.

7.2. NAND Tree Simulations

The simulations performed on the NAND tree aimed to verify propagation delay uniformity along all input paths, which is essential for minimizing the dependence of the signal used to trigger the TDCs on the hit position in the array. To perform this evaluation, a stimulus block written in Verilog-A was employed, programmed to activate one pixel at a time, each pixel being identified by an index. A total of 1024 simulations were executed, each with a synchronous stimulus on a different pixel, recording the signal arrival time at the NAND tree output. The results, shown by the histogram in Figure 13, reveal a delay distribution with a standard deviation of approximately 25 ps, which is negligible when combined in quadrature with the TDC resolution. This relatively narrow distribution confirms the effectiveness of the delay balancing techniques implemented during layout, ensuring temporally coherent behavior across the entire array area.

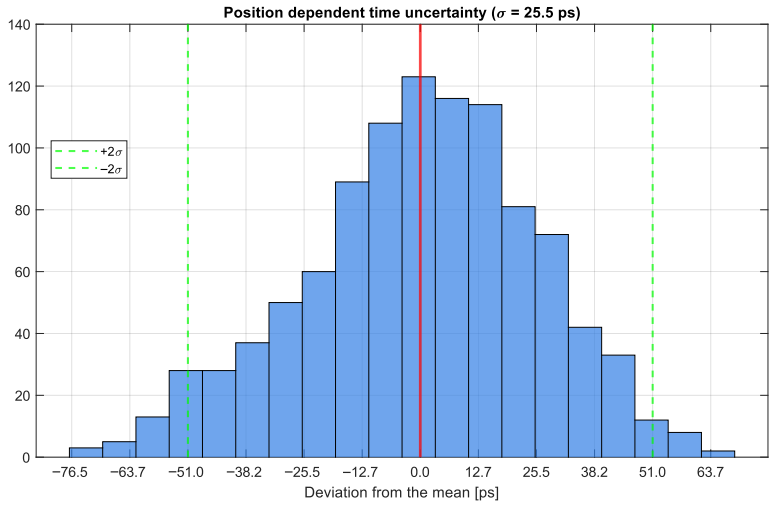

Figure 13. Results of post-layout simulations showing propagation delay and energy consumption. For each region of the pixel array, multiple hit patterns (with 10 or 50 activated pixels) were simulated to evaluate the dependence of timing and power on both location and number of triggered pixels.

8. Power Consumption

The power consumption of the system was estimated by taking into account the results from post-layout simulations of the counter and the NAND network and the energy

consumption data of the TDCs reported in [16]. As a rough estimate of the energy consumed for a single event by the two TDCs integrated in the SiPM, the energy needed to convert a time interval of 200 ns (i.e., half of the full TDC dynamic range) was used. Regarding power dissipation, the reported estimates are derived from a conservative approach to account for the lack of reliable statistical data on the temporal distribution of events across the sensor array. Without a definitive model for the photon arrival statistics in a real-world scenario, a precise power estimation is challenging. Therefore, a worst-case assumption has been made by considering an average event arrival time at the midpoint of the TDC measurement range for all microcells. This condition maximizes the dynamic switching activity of the digital logic and the TDC conversion stages. Combined with verified post-layout data for individual TDC channels and extracted parasitics for the adder tree, the upper-bound power consumption for the full 1024-pixel array is estimated to be less than 100 µW under a workload of 100 photons arriving within a 5 ns interval at a 1 MHz repetition frequency, ensuring the system remains within safe thermal and operational limits even under high-count-rate conditions.

Figure 14 presents the overall power consumption results for a single dSiPM, emphasizing its dependence on both the number of incident photons and the rate of events to be processed. The results indicate an extremely low power dissipation, even at relatively high event rates. Such a power dissipation is significantly lower than that typically reported for analog SiPM readout architectures [17–19], and is well below the threshold above which an active cooling system would become necessary, impacting the amount of material around the detector and the operation of the detector itself [20]. Furthermore, given the density of incident photons per unit area, which depends on the specific application, it is possible to estimate the power dissipated per unit area of the detector. This facilitates thermal design and integration of the device into complex systems.

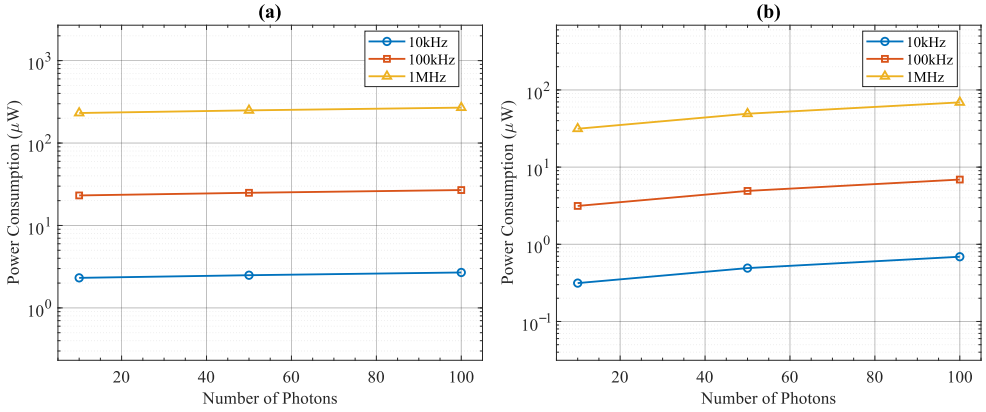

Figure 14. (**a**) Total power consumption of the entire array, including both photon-counting circuitry and timestamp generation via the two TDCs. (**b**) Power consumption associated exclusively with pixel-hit counting operations. In both cases different event rates are considered.

9. Conclusions

This paper has presented an ultra-fast digital Silicon Photomultiplier designed in 110 nm CMOS technology that demonstrates significant advances in photon-counting and timing-measurement capabilities. The proposed architecture integrates 1024 SPADs in a 32×32 array with on-chip digital processing electronics, achieving a 50% fill factor while maintaining high performance across both energy and timing-measurement paths.

The results obtained highlight the advantages of the monolithic design over traditional SiPM readout chains. The elimination of analog front-ends (TIAs and discriminators) reduces vulnerability to electronic crosstalk, a common issue in mixed-signal ASICs. The sub-100 ps timing resolution and 1024-photon dynamic range make this dSiPM suitable

also for high-performance applications such as Clinical TOF-PET, where it can enhance diagnostic image sharpness, and Flash LiDAR for space applications, where integrated spatial coincidence helps in rejecting solar background noise.

The key advantages of the proposed architecture stem from the use of fast digital electronics for photon counting, which enable a direct digital output without any analog-to-digital conversion. This approach offers several benefits. First, the parallel counter architecture with 16 independent counters allows for ultra-fast photon counting providing remarkable throughput for high-rate applications. Second, the carefully balanced hierarchical NAND tree achieves noteworthy timing uniformity across all 1024 signal paths, ensuring consistent sub-100 ps timing resolution for both first and last photon timestamps. Third, the monolithic integration of SPADs and digital processing circuitry eliminates the need for external ADCs and analog front-end electronics, significantly reducing system complexity, power consumption, and susceptibility to analog noise and signal degradation.

The direct digital output architecture provides inherent advantages over traditional analog readout schemes. By avoiding the bandwidth limitations and quantization errors associated with analog-to-digital conversion, the proposed dSiPM delivers more accurate photon counting and preserves precise timing information even at high event rates. The digital nature of the output also simplifies system integration, as the device can interface directly with digital signal processing units and FPGAs without requiring high-speed, high-resolution ADCs. Furthermore, the availability of configurable noise rejection mechanisms, including Time-over-Threshold filtering, spatial coincidence detection, and pixel-level enable/disable control through in-pixel SRAM, provides an operational flexibility that would be difficult to achieve in analog architectures.

The performance characteristics demonstrated through post-layout simulations confirm the viability of the proposed approach, with counting circuitry delays of 15 ns in worst-case conditions and timing path delays carefully balanced to achieve a fixed-pattern uncertainty not exceeding 26 ps.

Future work will focus on the experimental characterization of the device, including measurements of photon detection efficiency, dark count rate, timing resolution, and dynamic range under various operating conditions. Additionally, integration of the proposed dSiPM into complete detector systems will demonstrate its practical advantages in real-world applications and validate the benefits of the direct digital output architecture in terms of overall system performance and complexity reduction.

Author Contributions: Conceptualization, T.M.F., L.R. and R.S.; Methodology, T.M.F. and C.V.; Software, T.M.F. and F.L.; Validation, T.M.F.; Formal analysis, T.M.F.; Investigation, T.M.F.; Resources, T.M.F. and L.P.R.; Data curation, T.M.F.; Writing—original draft, T.M.F.; Writing—review & editing, T.M.F., M.C., G.C., M.D.D.R.R., G.F., F.L., M.N.M., L.P., L.R., L.P.R., D.F., R.S., F.S. and C.V.; Visualization, M.C., G.C., M.D.D.R.R., G.F., F.L., M.N.M., L.P., L.R., L.P.R., D.F., R.S., F.S. and C.V.; Supervision, F.L., L.P., L.R., L.P.R., D.F., R.S. and C.V.; Project administration, M.N.M. and L.R.; Funding acquisition, M.N.M. and L.R. All authors have read and agreed to the published version of the manuscript.

Funding: The activity leading to the results presented in this paper was carried out in the framework of the ASPIDES project, funded by the National Institute for Nuclear Physics (INFN), Italy.

Data Availability Statement: The experimental data obtained from device simulation can be made available upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gnecchi, S.; Dutton, N.A.W.; Parmesan, L.; Rae, B.; Pellegrini, S.; McLeod, S.J.; Grant, L.; Henderson, R.K. Analysis of photon detection efficiency and dynamic range in SPAD-based visible light receivers. *J. Light. Technol.* **2016**, *34*, 2774–2781. [CrossRef]
2. Diehl, I.; Hansen, K.; Vanat, T.; Vignola, G.; Feindt, F.; Rastorguev, D.; Spannagel, S. Monolithic MHz-frame rate digital SiPM-IC with sub-100 ps precision and 70 µm pixel pitch. *J. Instrum.* **2023**, *19*, P01020. [CrossRef]
3. Berretta, D.P.; Incoronato, A.; Sesta, V.; Severini, F.; Madonini, F.; Villa, F.A. Towards low-timing jitter photon number-resolved digital silicon photomultipliers. In *Optics + Optoelectronics*; SPIE: Bellingham, WA, USA, 2021; p. 117200I.
4. Cossio, F.; Preghenella, R.; Achari, B.R.; Agrawal, N.; Alexeev, M.; Alice, C.; Ammendola, R.; Antonioli, P.; Baldanza, C.; Barion, L.; et al. ALCOR: A SiPM readout chip for the ePIC-dRICH detector at the EIC. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2026**, *1087*, 171468. [CrossRef]
5. Ahmad, S.; Fleury, J.; de la Taille, C.; Seguin-Moreau, N.; Dulucq, F.; Martin-Chassard, G.; Callier, S.; Thienpont, D.; Raux, L. Triroc: A multi-channel SiPM read-out ASIC for PET/PET-ToF application. *IEEE Trans. Nucl. Sci.* **2015**, *62*, 664–668. [CrossRef]
6. Fleury, J.; Callier, S.; de La Taille, C.; Seguin, N.; Thienpont, D.; Dulucq, F.; Ahmad, S.; Martin, G. Petiroc, a new front-end ASIC for time of flight application. In Proceedings of the 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC), Seoul, Republic of Korea, 26 October–2 November 2013.
7. Diehl, I.; Feindt, F.; Hansen, K.; Lachnit, S.; Poblotzki, F.; Rastorguev, D.; Spannagel, S.; Vanat, T.; Vignola, G. The DESY digital silicon photomultiplier: Device characteristics and first test-beam results. *Nucl. Instrum. Methods Phys. Res. A* **2024**, *1063*, 169321. [CrossRef]
8. Diehl, I.; Feindt, F.; Gregor, I.M.; Hansen, K.; Lachnit, S.; Rastorguev, D.; Spannagel, S.; Vanat, T.; Vignola, G. 4D-tracking with digital SiPMs. *Nucl. Instrum. Methods Phys. Res. A* **2024**, *1065*, 169450. [CrossRef]
9. Shawkat, M.S.A.; McFarlane, N. A digital CMOS silicon photomultiplier using perimeter gated single photon avalanche diodes with asynchronous AER readout. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2020**, *67*, 4818–4828. [CrossRef]
10. Shawkat, M.S.A.; McFarlane, N. A CMOS perimeter gated SPAD based digital silicon photomultiplier with asynchronous AER readout for PET applications. In Proceedings of the 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), Cleveland, OH, USA, 17–19 October 2018; pp. 1–4.
11. Vinayaka, V.; Namboodiri, S.P.; Roy, A.; Baker, R.J. Segmented digital SiPM. In Proceedings of the 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), Dallas, TX, USA, 4–7 August 2019; pp. 1118–1121.
12. Torilla, G.; Giroletti, S.; Brogi, P.; Collazuol, G.; Dalla Betta, G.F.; Marrocchesi, P.S.; Minga, J.; Morsani, F.; Pancheri, L.; Ratti, L.; et al. Digital SiPMs in a 110 nm CMOS technology with fast parallel counter architecture. In Proceedings of the 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Milan, Italy, 5–12 November 2022; pp. 1–5.
13. Torilla, G. A Test Platform for CMOS SPADs and Digital SiPMs in a 110 nm Technology. Ph.D. Thesis, University of Pavia, Pavia, Italy, 2024.
14. Moreno-García, M.; Xu, H.; Gasparini, L.; Perenzoni, M. Low-noise single photon avalanche diodes in a 110 nm CIS technology. In Proceedings of the 2018 48th European Solid-State Device Research Conference (ESSDERC), Dresden, Germany, 3–6 September 2018.
15. Lindner, S.; Pellegrini, S.; Henrion, Y.; Rae, B.; Wolf, M.; Charbon, E. A high-PDE, backside-illuminated SPAD in 65/40-nm 3D IC CMOS pixel with cascoded passive quenching and active recharge. *IEEE Electron Device Lett.* **2017**, *38*, 1547–1550. [CrossRef]
16. Floris, T.; Ratti, L.; Gasparini, L.; Manuzzato, E.; Vacchi, C. Sub-100 ps LSB time-to-digital converter with voltage- or current-based resolution control. In Proceedings of the 2025 20th International Conference on PhD Research in Microelectronics and Electronics (PRIME), Taormina, Italy, 16–19 June 2025; pp. 1–4.
17. Giangrande, L.; Heller, M.; Montaruli, T.; Favre, Y. FANSIC: A fast analog SiPM interface circuit for the readout of large silicon photomultipliers. *Nucl. Instrum. Methods Phys. Res. A* **2025**, *1077*, 170523. [CrossRef]
18. Haemisch, Y.; Frach, T.; Degenhardt, C.; Thon, A. Fully digital arrays of silicon photomultipliers (dSiPM)—A scalable alternative to vacuum photomultiplier tubes (PMT). *Phys. Procedia* **2012**, *37*, 1546–1560. [CrossRef]
19. Dorosz, P.; Baszczyk, M.; Kucewicz, W. Low-power front-end ASIC for silicon photomultiplier. *IEEE Trans. Nucl. Sci.* **2018**, *65*, 1070–1078. [CrossRef]
20. Sobotka, P.; Bolek, K.; Pawłowska, Z.; Kliś, B.; Przychodzki, M.; Fornalski, K.W.; Rutkowska, K.A. Sensitivity and performance of uncooled avalanche photodiode for thermoluminescent dosimetry applications. *Sensors* **2024**, *24*, 6207. [CrossRef] [PubMed]