A Comprehensive Survey of Readout Strategies for SiPMs Used in Nuclear Imaging Systems

Shahram Hatefi Hesari *, Mohammad Aminul Haque and Nicole McFarlane

Abstract: Silicon photomultipliers (SiPMs) offer advantages such as lower relative cost, smaller size, and lower operating voltages compared to photomultiplier tubes. A SiPM’s readout circuit topology can significantly affect the characteristics of an imaging array. In nuclear imaging and detection, energy, timing, and position are the primary characteristics of interest. Nuclear imaging has applications in the medical, astronomy, and high energy physics fields, making SiPMs an active research area. This work is focused on the circuit topologies required for nuclear imaging. We surveyed the readout strategies including the front end preamplification topology choices of transimpedance amplifier, charge amplifier, and voltage amplifier. In addition, a review of circuit topologies suitable for energy, timing, and position information extraction was performed along with a summary of performance limitations and current challenges.

Keywords: SiPM; readout electronics; optical detectors

1. Introduction

Silicon photomultipliers are widely used in a number of different applications in the health, environmental, and scientific discovery fields. Silicon photomultipliers, shortened to SiPMs, are semiconductor-based optical detectors that, when implemented in silicon-based semiconductors, image in primarily the visible region of light. Silicon photomultipliers are not significantly affected by magnetic fields and have lower operating voltages compared with photomultiplier tubes (PMTs). SiPM properties are also relatively consistent over time, and aging does not affect the SiPM’s performance [1–3]. In addition to this, SiPMs implemented in commercial CMOS foundries allow for the inclusion of the electronics directly on chip, leading to reduced parasitics and higher throughput and speed. Due to this, SiPMs are being widely used to replace PMTs and other photodetectors in applications such as Light Detection and Ranging (LiDAR) [4–7], optical imaging systems [8–12], medical imaging [13–17], fluorescence imaging [18–20], astrophysics and gamma detection [21–30] where they may be used in different types of experiments such as time correlated single photon counting [31–33].

A SiPM is an array of single photon avalanche diodes, or SPADs, and quenching devices. The overall structure is shown in Figure 1a. Each SPAD is biased in reverse breakdown (Geiger mode), with a quenching device in series with it. When photons are incident on the SPADs, electron–hole pairs are created and accelerated under the high electric field. They undergo impact ionization leading to avalanche multiplication and a large current. The quenching device can be implemented using passive devices (resistor) or active devices (MOSFET transistor). A simple SiPM equivalent circuit model is shown in Figure 1b, where the model is divided into avalanching vs non-avalanching cells, and each cell consists of parasitic devices [34–36]. The model shows only a single avalanching cell. To account for more cells avalanching at a given time, the model would be adjusted to have F firing microcells and M non-avalanching microcells where \( F + M = N \). It should be noted that all cells do not necessarily fire at the same time, and that there may be a small
delay that is not captured in this model. This model also does not take into account the full probabilistic nature of the photon-induced avalanche process, but is suitable for designing, simulating, and characterizing the readout electronics' performance before committing to potentially expensive devices or fabrication. The parasitic capacitances are due to the parallel and fringe capacitances of the metal interconnects and the depletion capacitances of the diodes. The use of devices integrated on the same chip offers lowered parasitics compared to off chip devices. SiPMs can be either analog or digital. In analog SiPMs, the output is measured at the common cathode or the common resistor node or through a capacitor connected at the resistor–SPAD junction (Figure 1a). The other end of all the capacitors are a common node \( V_0 \). This approach offers improved timing performance.

In digital SiPMs, the output of each SPAD is taken directly at the node between the quenching resistor and diode through its own dedicated readout circuit. In this case, the output of the SPAD is typically a digital signal. SiPMs have been implemented using custom fabrication processes and commercial foundry processes. Implementing SiPMs in commercial foundries can require an initial fabrication run to develop test devices in order to extract parameters suitable for simulating the model shown in Figure 1b with the designed electronics. However, in recent years, a few commercial foundries, such as XFAB, have begun offering single photon avalanche diodes as part of their fabrication options which will enable easier simulations of the SiPMs and the electronics on the same chip.

Figure 1. (a) A SiPM is an array of SPADs each with its own quenching resistor \( R_Q \). For analog SiPMs the output is the sum of the avalanche currents taken at either the common cathode \( C \) or common anode \( A \). (b) SiPM circuit model showing 1 avalanching cell and (N-1) non-avalanching cells. \( C_Q \) is the quenching device’s parasitic capacitance, \( C_D \) is the diode parasitic capacitance, and \( C_G \) is the equivalent parasitics associated with routing. \( R_{in} \) represents the input impedance of the readout circuit [36].

SiPM-based imaging systems can consist of commercially available SiPMs with discrete off the shelf or custom electronics on a PCB board [37]. However, this approach generally has higher parasitics which may limit the system speed. It also has optical dead zones which increase the fill factor and spatial resolution. Integration of the electronics in a single chip can reduce those parasitics. However, the electronics, particularly if complex with multiple stages, could potentially use the available area for the optical detector, reducing the overall fill factor across the chip. Fill factor is typically calculated as the ratio of the optical active area, typically defined by the depletion region of the pn junction, and the area of non-optical-related devices such as resistors and transistors as well as any metal routing combined with the total pn device area. This integration of readout and optical devices on the same chip can be accomplished by implementing both the SiPM and electronics in standard CMOS processes [38,39]. In vertical integration, the diodes are in one layer, while
the other devices are in a separate layer. Thus, this offers the largest fill factor compared to a 2D implementation \[40,41\].

In this review, the primary focus is on nuclear imaging which is accomplished using a scintillator material, such as LYSO (lutetium–yttrium oxyorthosilicate), coupled to the SiPM. Scintillators emit light when particles, such as gamma rays or neutrons, are incident on them. A schematic of a typical nuclear imaging system is shown in Figure 2, where time of arrival, energy, and position are the three quantities of interest. There are a number of options for integrating the SiPM and scintillator. The exact method will depend on the material used to fabricate the scintillator. The method of integrating the two will also depend on the application, as this will drive the accuracy, sensitivity, and resolution constraints of the entire system. In general, the SiPM and scintillator can be integrated using commercially available optical coupling gels or grease. More details on the performance parameters of optical coupling materials can be found in \[42\]. A typical readout architecture, whether in integrated circuits or discrete off-the-shelf components, is composed of a number of modules, and is shown in Figure 3.

Figure 2. Scintillators convert nuclear particles into photons which are then converted by the SiPM into electrons. These can be amplified and further processed using charge-, current-, or voltage-based readout electronics.

Figure 3. The typical readout in a nuclear imaging system requires various circuit topologies to extract time, energy, or position information.

Regardless of the application, the initial SiPM output signal must be amplified before being further processed and digitized. The three typical approaches for front-end amplification that are used to readout the SiPM arrays consist of charge, transimpedance, or voltage amplification and are shown in Figure 4. The paper first gives an overview of the preamplifier structures. This is followed by a discussion on energy, timing, and position measurement along with a summary of array readout strategies. Finally our conclusion summarizes the paper and describes the challenges and possible future paths.
Figure 4. Typical preamplifier topologies (a) charge amplifier, (b) transimpedance amplifier, and (c) voltage amplifier.

2. Preamplifier Structure in Nuclear Imaging Systems

The preamplifier is the front-end circuitry between the raw SiPM readout and the signal processing electronics. Thus, its characteristics directly affect the performance of subsequent stages. The preamplifier stage can be implemented as either single ended or fully differential. When implemented as a fully differential amplifier, a common-mode feedback stage is required to maintain the common mode at a known level \[43\]. Typical common mode approaches will take the difference of the average of the amplifier outputs with the common mode voltage to adjust the tail current or the currents in the active loads to move the common mode voltage up or down as required. Implementations of common mode feedback can use either diff amp-based or switch capacitor-based methods. Reference \[44\] describes an auxiliary amplifier-based common-mode feedback circuit (CMFB) that has a phase margin of 62.1° and gain margin of 25 dB.

2.1. Charge-Sensitive Preamplifiers (CSP)

A charge sensitive preamplifier has an output voltage that is proportional to the integrated input current. A typical architecture consists of an operational amplifier, a capacitor in the feedback path, and a reset switch or circuit. The reset circuit discharges the feedback capacitor, otherwise the output voltage eventually saturates. High input impedance guarantees the current flows primarily through the feedback capacitor. Due to the Miller effect, the feedback capacitor ($C_f$), appears as large capacitance at the input ($C_{eff} = (1 + A)C_f$) which is in parallel with other parasitic input capacitances. Thus, the gain of this structure is proportional to the feedback capacitance and is relatively independent of the devices’ capacitances. The feedback network can be implemented using passive or active components. The advantage of active feedback (transistors) over passive networks (RC) lies in the active devices’ lower noise and tunability. In addition, small chip area is required to implement the circuit. An active feedback with leakage current compensation is highlighted in \[45\]. Ref \[46\] uses an active feedback network for gamma-ray tracking detectors with a low gain, noninverting amplifier between the charge amplifier and the feedback resistance. Thus, the discharge time constant can be made shorter, while leaving the RC network unchanged. In another work, an active feedback network is designed upon two different active feedbacks based on the mosfet devices and a voltage-controlled switch as the reset network. Using this technique, the device can operate at as high as 4.5 MHz and can read out 1 to $10^4$ photons with an energy of 1 and 10 keV \[47\]. In \[48\], a two amplifier feedback network replaces the large resistor. One amplifier maintains the DC offset voltage at zero and the other one constitutes the negative feedback around the main amplifier.

The Opamps used for CSPs can be based on BJT transistors \[49,50\] or MOSFET and JFET transistors \[51–53\]. JFET-and BJT-based opamps are widely available commercially. In comparison to MOSFET, JFET CSPs provide better noise performance. A noise model for
JFET CSPs can be found in [51]. The result of this work revealed that, at 10 kHz, the JFET amplifier’s input-referred inherent noise exhibited 22 dB improvement as opposed to MOS-based CSPs.

However, MOSFET-based approaches can offer reduced size and the ability to integrate the SiPM with the amplifier on the same chip [43,54–57]. BiCMOS (Bipolar and CMOS on the same die) technologies are less radiation-hard compared to CMOS [57], and in general CMOS is less radiation tolerant than SOI (Silicon on insulator) [58,59]. The irradiation tests in [57] demonstrate that, with high input gamma-rays, the CSPs output amplitude signal and the SNR reduces up to 34.3% and 11.6 dB. Moreover, the fall time also increases significantly from 201 ns to 1730 ns. This research also indicates that MOS transistors are more tolerant to radiation sources and provide better performance than BiCMOS.

A challenge for CMOS-based CSPs is the relatively large area capacitor. A typical capacitor may be 10 times as large as typical analog sized transistor. The capacitor has to discharge periodically. This can be performed using feedback resistors, transistors and other novel active devices [43,45–47]. Passive resistance is common; however, the large value has implications on the opamp DC characteristics and requires large area if implemented in integrated circuits. $R_f$ can be implemented with MOS transistors to reduce area and provide greater design flexibility [60]. The offset, bandwidth, power consumption, dynamic range, and gain of CSP, are functions of the feedback components [44,48,61,62]. Figure 5 shows a comparison of the gain and bandwidth of characteristics of a few implemented charge amplifiers.

![Figure 5. Gain and bandwidth of charge amplifiers proposed in [44,57,63–67].](image)

### 2.2. Transimpedance Preamplifier

The conventional preamplifier used in nuclear imaging systems comprises a charge sensitive and shaping amplifier [68]. An ideal nuclear imaging system needs to be fast, with a small rise time and dynamic slew rate correction and be low power and small sized for radiation detection readout [69,70]. However, even with design techniques, the CSP’s slew rate is in the microseconds range [71]. In practice, CSP and the shaping amplifier can be substituted with a transimpedance amplifier, where the current can be directly converted into an output voltage [72].

In general, a typical TIA implementation using off the shelf discrete components is a feedback resistor with an amplifier. For large gain, the input resistance, $R_{in}=R_f/(A + 1)$, is minimized. The TIA input slew rate is a function of the input RC time constant. Lower RC gives a higher slew rate, and is suitable particularly for high count rate applications [68,72–77]. The input impedance of a transimpedance amplifier is,

$$Z_{in} = \frac{R_f}{g_m R_0} + sR_f C_0 = \frac{R_f}{g_m} \left( \frac{1}{R_0} + sC_0 \right)$$

where $R_f$ is the resistor in the feedback path, $g_m$ is the transconductance gain, $C_0$ is the internal capacitance of amplifier, and $R_0$ is load resistance. This can cause oscillations when
coupled with the large input capacitance from a SiPM. To mitigate this, a capacitance is placed in parallel with the feedback resistor. The large feedback resistor may be implemented by an MOS device operating in triode [78,79]. In order to detect and transform nanosecond-wide signals, the TIA requires a large bandwidth. A TIA’s bandwidth can be written as,

\[
BW_{TIA} = \sqrt{\frac{GBW}{2\pi R_F C_D}}
\]

(2)

where \(BW_{TIA}\) is the bandwidth, \(R_F\) is the resistance, \(GBW\) is the gain bandwidth product, and \(C_D\) is the detector capacitance. The \(GBW\) is a constant for the open loop opamp or OTA structure. In [71], in order to increase the bandwidth, two inductors are series with the TIA. There are multiple approaches to implement a suitable TIA for nuclear imaging [80]. Reference [74] proposes a TIA based on second generation voltage conveyors (VCII). This current conveyer has been implemented using discrete components [75,76]. VCII, compared to the first generation, provides variable front-end gain and low impedance at the input of VCII that helps to decrease the effect of usually large output parasitic capacitance associated with SiPM.

Common base and common gate amplifiers provide low input impedance and do not require feedback resistances. These single transistor amplifier architectures may be modified into regulated structures or incorporate cascoding to improve performance [80,81]. In order to achieve the best performance with a regulated common gate TIA, such as that proposed in [80], the DC output voltage of the TIA needs to close to the overdrive voltage, \(V_{DSAT}\), of the input transistor and the load resistor needs to be approximately ten times larger than the drain to source resistance. Similar to CSPs, TIAs can also be made of technologies other than CMOS. The first TIA using organic thin-film transistors was proposed in [79], where the TIA was based on a voltage-controlled resistor and common gate input stage.

A multi-channel readout based on a transimpedance topology was proposed in [73] (Figure 6), where in order to prevent the possible oscillations, a series resistor (\(R_s\)) was placed between the common anode of the SiPM and the input port of the operational amplifier. In this design, the inductors represent the parasitic inductances presented by the metal lines, wires, and printed circuit board. In order to modify the offset, output signal tail and the undershoot, [82] uses trimmers. A TIA implemented in SiGe BiCMOS that used a transformer-based input stage for improved frequency response was proposed in response [83]. Figure 7 shows the gain and bandwidth of a few transimpedance amplifiers that have been implemented.

![Figure 6. A multi-channel SiPM’s readout based on transimpedance topology, proposed in [73].](image)
2.3. Voltage-Sensitive Preamplifiers

Robustness, simplicity, and ease of implementation and simulation make the voltage amplifier an ideal choice of readout circuit when dealing with unknown SiPMs. However, this is sometimes achieved at the cost of implementing a large shunt resistor, compared to the TIA design. Large shunt resistors impose higher RC time constants which correspond with wider output pulses. In addition, the gain of this structure depends on the signal size and the currents flow through the device [73].

The use of a voltage amplifier requires a resistor in series with the SiPM. The voltage developed across the resistor is then amplified. A typical implementation using commercially available off the shelf parts may use the standard inverting or non-inverting configuration where the gain is a function of the resistive feedback network (Figure 4c). The size of $R_{\text{load}}$ affects the input DC bias of the opamp [89].

The total input capacitance, formed by the parallel combination of parasitic capacitance of the SiPM and amplifier, $C_d$ and $C_{\text{in}}$ and the total input impedance, form a time constant that should be minimized such that the fastest input signal can be recognized [90].

In some applications further processing may require the output voltage to be converted back into current. To avoid multiple conversions, topologies such as that in Figure 8a have been proposed [91]. To achieve better performance, the current mirror that is used in [91] can be replaced with improved topologies with higher output impedance by cascoding or regulated drain structures. These alternative topologies employing feedback such as Figure 8b,c have also been implemented in 350 nm CMOS processes [92], where (b) achieves higher dynamic range and tunability and (c) is optimized for bandwidth and input impedance. In [77] the voltage amplifier is coupled with TIA. In this method, the width of the input current pulses is controlled by the circuit time constant which can be adjusted by TIA’s feedback network. Figure 9 shows a comparison of the gain and bandwidth of voltage amplifiers implemented in the literature.

Figure 8. (a) The current mode readout circuit proposed in [91]. Alternative current mode readout topologies for the current buffer (b,c) where $i$ represents the SiPM current [92].
3. Energy Measurements

The photon energy of an event (defined as photons being incident on the SiPMs), is proportional to the SiPM’s charge or current pulse, and is an important characteristic. The peak signal provides information on the energy of the photons. However, regardless of readout type, energy resolution may be affected by the scintillator material, scintillator thickness [99], and other physical mechanisms such as ballistic deficit, pulse pileup, or baseline fluctuation [25,100–104]. Table 1 shows a relative comparison of the energy resolution using different scintillator materials. The total photon absorption depends on the type and thickness of material and incident energy of photons. For example, four different scintillators, CsI(Tl), NaI(Tl), LaBr3(Ce) and GAGG(Ce), showed a direct relationship between energy resolution and material type and thickness [99]. Thus, it is important to take the scintillator material into account when characterizing the full system. In addition, since the scintillator is the front end transducer, its properties will limit the overall imaging system characteristics. Figure 10, shows readout circuit topologies and scintillators.

Table 1. SiPM readout energy resolution based on Scintillator type.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scintillator</td>
<td>LaBr3:Ce</td>
<td>CeBr3</td>
<td>LaBr3(Ce)</td>
<td>LaBr3(Ce)</td>
<td>LaBr3(Ce) co-doped with Sr</td>
</tr>
<tr>
<td>SiPM</td>
<td>Custom</td>
<td>SensL</td>
<td>FBK</td>
<td>FBK</td>
<td>FBK</td>
</tr>
<tr>
<td>Resolution</td>
<td>% 4 at 661 KeV</td>
<td>% 4.54 at 661 keV</td>
<td>% 3.19 at 661 keV</td>
<td>% 3.4 at 661 keV</td>
<td>% 2.6 at 661 keV</td>
</tr>
</tbody>
</table>

Figure 10. Readout circuits used in nuclear imaging with various scintillator materials, from 2012 to 2021 [30,108–135].

Pulse pileup occurs when the time distance between the two or more pulses is less than the pulse resolution time [136]. Resolution time is the precision of a SiPM in determining
the arrival time of a single photon. Pulse pile up can be either head pileup or tail pileup. Head pileup occurs when the rising edge of a second pulse come to pass the rising edge of the first pulse. Tail pileup occurs when the rising edge of the second pulse occurs on the falling edge of the first pulse [137]. These pileups can result in a shift of the baseline voltage [54]. Pulse pileup can be mitigated by rejection, filtering, or a transformation algorithm. Many of these are implemented in software or using off chip microprocessors and FPGAs.

The shaper stage (also known as shaper filters), which may be present in energy, timing, and position measurement [138,139], are used to tackle the problem of pulse pileup. Gaussian-based architectures are popular shaper implementations [140]. Frequency bandwidth limitations can improve the SNR [141]. However, bandwidth limitation, may cause negative overshoot for large shaping parameter [138]. Quasi-Gaussian or trapezoidal shapers have also been implemented [139]. Low-pass filters, for example using a Sallen–Key filter, can calibrate for drift and improve signal-to-noise [142]. Shapers have also been implemented using digital CR − RC [143]. The pulse pile up problem needs to be investigated at high input rate application. This problem can be addressed by implementing pole zero cancellation circuits (PZC). An example of this type of circuit can be find in [54], where the PZC has been implemented by twenty PMOS transistors and a 24 pF capacitor. In addition, to tackle the problem of pile up, Pole zero cancellation eliminates the undershoot and allows for higher rate counting [139,144–146].

The shaper’s output signal is passed into a sample and hold circuit (also called a peak stretcher) to store the peak value of the signal before digitizing [147]. Analog to digital converters (ADCs) are also used in energy measurement systems. They range from successive approximation (SAR), delta-sigma, dual slope, pipelined and flash architectures [25,148–150]. These architectures offer various advantages in speed, resolution, power, and area on chip. ADCs in general are power hungry and can be area intensive on chip. Thus, other digitization and quasi-digitization strategies such as time-over-threshold (ToT) circuits may be employed [151]. In a ToT circuit, the difference between two time stamps where the signal crosses a pre-defined threshold is stored. This can be implemented with time to digital conversion [152,153]. Integration of readout electronics and digital SiPM, within a single chip can make them noisier than analog SiPMs. This is because the digital SiPMs will typically be in standard CMOS where no special efforts have been made to optimize the process for optical detection [154]. Many commercially available off the shelf SiPMs are analog and are optimized for optical response. For analog and digital SiPMs both implemented in CMOS, the noise of the optical detecting device remains the same and there are distinct advantages and disadvantages for the digital or analog approach. Customized analog SiPMs have better noise performance for energy measurements as the energy resolution depends on the optical performance, i.e., the statistical properties of photon-induced charge generation. [155].

4. Timing Measurements

In addition to energy, the time of arrival of pulses, corresponding to events, is also of interest [156]. Figure 11 shows the response of various types of events. Time of flight measurements determine the distribution of the time difference between events. Uncorrelated signals will have a flat time difference distribution. Correlated signals, such as those which occur in prompt coincidence, delayed coincidence, or time of flight experiments, show a delta or spike-shaped distribution [157–159]. Noise in time of flight is controlled by the coincidence resolving time [160]. Time domain information can always be converted to frequency domain. This can be done either in software or in hardware using either analog, digital, or a combination of analog and digital techniques.
Figure 11. Time measurements due to a source or sources, (a) uncorrelated events (b) prompt coincidence (c) delayed coincidence (d) time of flight [157].

Figure 12a shows the use of a discriminator (comparator) to determine if an event occurred [161,162]. Delay units to enable proper comparison of the time stamps [163,164], time to amplitude converters (TAC) [165], and multi-channel analyzers [166,167] are required to accurately report the timing. Figure 12b filters the events within some range before the discriminator and improves the signal to noise [168]. Each of these approaches can be implemented using discrete off the shelf components or implemented using custom integrated circuits fabricated in commercial foundries. In all cases, the analog modules contain amplifiers and filters, along with the digital and quasi-digital blocks, such as comparators, and encoders [139,169–173]. Depending on the scintillator type and material, such as LGSO, LYSO, GAGG, LuAG, YAG, LSO, and LFS, various types of SiPMs may have different time resolutions [174]. The required resolution depends on application, where, for example, in PET, sub 10 ps is a research target. The maximum achievable coincidence time resolution (CTR), based on scintillator statistics, can be estimated using [175],

$$CTR = \alpha \sqrt{\frac{\tau_r \tau_f}{n}}$$

where $\tau_r$ and $\tau_f$ are the scintillation rise and fall time, $n$ is the number of photons, and $\alpha$ is a constant. For example, for an output light of between 27.9 and 49.5 kphoton/MeV measured upon gamma Cs173 excitation, the best achievable CTR with Cerium-doped GAGG:Ce scintillator was 87 ± 2 ps [161]. The other factor that has an influence on time resolution is the readout electronic noise. [156] proposed a passive compensation circuit for device capacitance to reduce the effect of electronic noise on time resolution. This technique provided no injection of noise from the SiPM to the readout circuit and improved the single photon time resolution. In some other work, different time resolution, ranging from less than 80 ps [174] up to more than 330 ps FWHM [176] can be found. In [177,178], improved time resolution is achieved by connecting multiple SiPMs in series creating multiple scintillation counters. For a readout circuit, consisting of an N counter, the resolution improves by a factor of $1/\sqrt{N}$. 
Figure 12. (a) Time to analog conversion [179], (b) time spectrometer with a differential amplifier [168], and (c) definitions of the shaping and peaking times [180].

To determine the time of an event, comparators and digital gates may be used [181,182]. Comparators type (or discriminators) can be constant fraction (CFD) or leading edge (LED) [183]. However at a lower threshold, LED provide better time resolution [184,185]. Comparators can have errors due to jitter, long term and short term drift, and variation in the shape and amplitude of the input pulse [186]. This can be compensated for by 1-D techniques such as linear or logarithmic compensation or 2-D compensation techniques like artificial neural networks and polynomials [187,188]. Time resolution is determined with the full width half maximum of the time spectrum [189]. To achieve the best time resolution, we suggest setting the shaping time at 61% of the pulse amplitude. where the peaking time needs to be approximately of three to five times greater than the shaping time (Figure 12c).

Delays can be generated using coaxial cables, active circuits or lumped elements [148,190,191]. However, the use of coaxial cables may also attenuate, frequency shift or reflect the signal. Time to amplitude converters (TAC) generate an analog output whose amplitude is proportional to the measured time interval. They may be implemented using a purely analog, digital or mixed signal approach [179,192,193]. The time resolution is limited by the accuracy of the timing discriminators [145,180]. As with the energy measurements, ADCs may be replaced by time to digital converters (TDC) [194,195].

Both analog and digital readout approaches can have similar time resolution [196–199]. Due to increased power consumption, large numbers of readout channels, and challenges in digital realization [200], analog SiPMs are more prevalent in nuclear imaging applications. Figure 13 shows CTR and energy resolution for different front-ends.

Figure 13. Time and energy resolution of different readout topologies charge amplifier (CSP), transimpedance amplifier (TIA) and voltage amplifier [111–113,116–118,120,121,135,201].
5. Position Sensing

When implemented in integrated circuits, the inherently asynchronous and event-based nature of the signal makes them suitable for employing the asynchronous readout such as in address event representation (AER) protocol [202–204]. This has been implemented in a standard CMOS technology with a digital SiPM architecture [205]. By comparison of the arrival time of two or more pixels, from the AER communication protocol, position information also can be extracted [206,207] (Figure 14).

Figure 14. (a) SPAD pixel within SiPM and (b) an actual AER protocol [204].

For conventional analog SiPM arrays, typical position sensing may require a large number of readout channels and may show limited position resolution due to the SiPM’s pixel size [208]. To achieve higher spatial resolution, artificial neural networks [131,209] and novel position-sensitive SiPM (PS-SiPM) fabrication processes [210,211] can increase the position sensing performance. Multiplexing [25,212] and multi-channel integrated readouts [164,213,214] simplify the readout and improve performance characteristics (Figure 15). In this figure, each diode symbol represents a SiPM. Position detection has also been performed using multiple cameras, algorithms to determine the center of gravity of signals spread across the pixels, Compton scattering, and time of flight measurements, [215–220]. In many cases these approaches involve significant use of signal processing software and thus approaches that utilize microprocessors on the same chip as the SiPM or with 3D integration can potentially offer improved readout speed and throughput.
6. Array Readout

For small arrays, each channel may be read independently. Each can have its own amplifier and data acquisition channel [221–224]. Larger arrays require multiplexing and processing to reduce complexity and cost [225–227]. Multiplexing may reduce the capability to discriminate several events or scatter interactions and correct for anode/pixel non-uniformities [228]. Multiplexing of multiple SiPM sensors using resistor networks can lead to reduced timing performance [25]. Common multiplexing schemes include Anger, discretized positioning (DPC) [229], row/column or crosswire readout [230], and symmetric charge division multiplexing (SCD) [231]. Row/column readout can be standard X-Y or scrambled X-Y, where the devices are placed in series-parallel connections. In standard X-Y readout, all cathodes in each row and all anodes in each column are shorted. For a case with nine SiPM arrays, with each of those consisting of 16 pixels, the readout electronics require 12 threshold comparators to detect when an event has occurred. However, using a scrambled approach, where the cathodes of all the devices in each array are connected together, nine threshold detector channels can be used. Since the average energy, time, and position resolution is degraded with noise [212,232], reduction in the number of SiPMs connected to a single readout channel results in less noise and better system performance [233].

7. Conclusions

The readout architecture of a SiPM-based array will be a function of application. Applications range from those which have weak light sources to those with stronger incident light sources. Depending on the application, either digital or analog SiPMs can be used. Digital SiPMs digitize at the microcell level, after which any digital system such as an FPGA, complex logic, or microprocessor can be used to process the signal. For many applications, an analog SiPM is used instead due to the increased optical active area. The readout of analog SiPMs typically targets energy, time, or position of interaction. The readout electronics can be implemented using commercial off the shelf components or custom integrated circuits. Both options add constraints that must be taken into account for a given application and SPICE models may be used to predict system performance. Discrete electronics will typically have larger parasitics and thus operate at slower speeds than systems where the devices and electronics are integrated on the same chip.

The main challenges in SiPM readout design can be broken down into three categories, the readout circuit, packaging and interfaces, and signal processing. In general, an imaging system consists of hundreds of SiPMs with many readout circuits that are typically multiplexed to reduce the power and complexity. However, multiplexing can impair the SiPM’s performance. Strategies may involve sharing circuits between pixels or adding in circuits for calibration. At the packaging and interface level misalignment and edge effects may affect energy, time, and position resolution. Post fabrication of a...
scintillator material directly on top of the SiPM is a possible solution to this. However, wafer dicing will be subject to inaccuracies and new materials and fabrication protocols are germane to any integration of a scintillator and optical detector. Processing of the signal can be performed after digitizing using analog to digital converters, microprocessors, and software. These are all very power-hungry and space-intensive. Readout using in-pixel digitized approaches along with asynchronous readout can help reduce complexity. Finally, the use of 3D integration can allow for increased pixel fill factor (i.e., increased optical active area) while still allowing integrated circuit sized readout electronics.

Author Contributions: Conceptualization, S.H.H., M.A.H. and N.M.; methodology, S.H.H., M.A.H. and N.M.; investigation, S.H.H., M.A.H. and N.M.; resources, S.H.H., M.A.H. and N.M.; data curation, S.H.H., M.A.H. and N.M.; writing—original draft preparation, S.H.H., M.A.H. and N.M.; writing—review and editing, S.H.H., M.A.H. and N.M.; visualization, S.H.H., M.A.H. and N.M.; supervision, N.M.; project administration, N.M.; funding acquisition, N.M. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0017414.

Conflicts of Interest: The authors declare no conflict of interest. The funder 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
3. Sun, Y.; Maricic, J. SiPMs characterization and selection for the DUNE far detector photon detection system. *J. Instrum.* 2016, 11, C01078. [CrossRef]


52. Scandurra, G.; Giusi, G.; Ciofi, C. Single JFET front-end amplifier for low frequency noise measurements with cross correlation-based gain calibration. * Electronics 2019, 8, 1197. [CrossRef]


140. Kantor, M.Y.; Sidorov, A. Shaping pulses of radiation detectors into a true Gaussian form. *J. Instrum.* **2019**, *14*, P01004. [CrossRef]


148. Krishnan, S.; Webster, C.; Duffy, A.; Brooks, G.; Clay, R.; Mould, J. Improving the energy resolution while mitigating the effects of dark-noise, for a microcontroller based SiPM sensor. *J. Instrum.* **2020**, *15*, P09028. [CrossRef]


164. Parsakordasiabi, M.; Vornicu, I.; Rodríguez-Vázquez, Á.; Carmona-Galán, R. A Low-Resources TDC for Multi-Channel Direct ToF Readout Based on a 28-nm FPGA. *Sensors* 2021, 21, 308. [CrossRef]

165. Lobanov, A. Precision timing calorimetry with the CMS HGCal. *J. Instrum.* 2020, 15, C07003. [CrossRef]


221. WB Series Silicon Photomultipliers (SiPM)—KETEK GmbH. Available online: https://www.ketek.net/sipm/sipm-modules/tia-modules/ (accessed on 2 April 2021).


