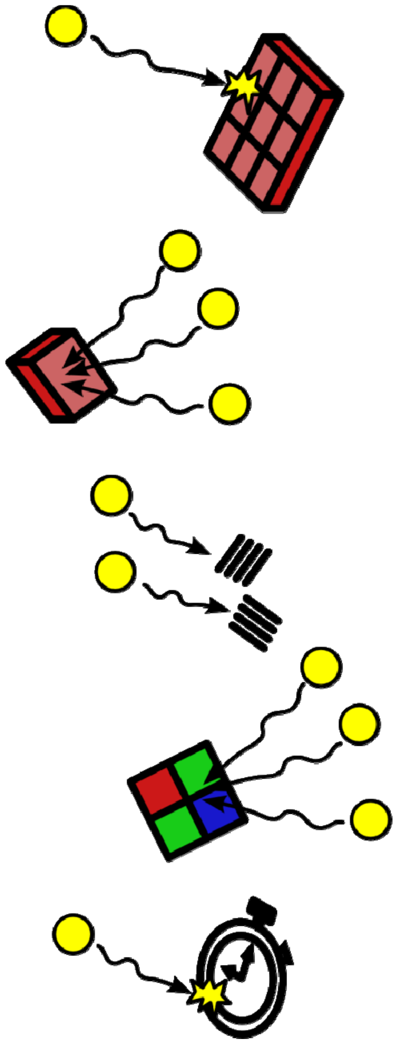
Single-Photon Detection in CMOS

Matteo Perenzoni
Integrated Radiation & Image Sensors - FBK
perenzoni@fbk.eu

February 11, 2018

Motivation: Why Single-Photon?

- Eyes are sensitive to intensity and color (wavelength)
- But there is much more!
- Think about gathering all about every single photon:
  - Position
  - Direction
  - Polarization
  - Wavelength
  - Time

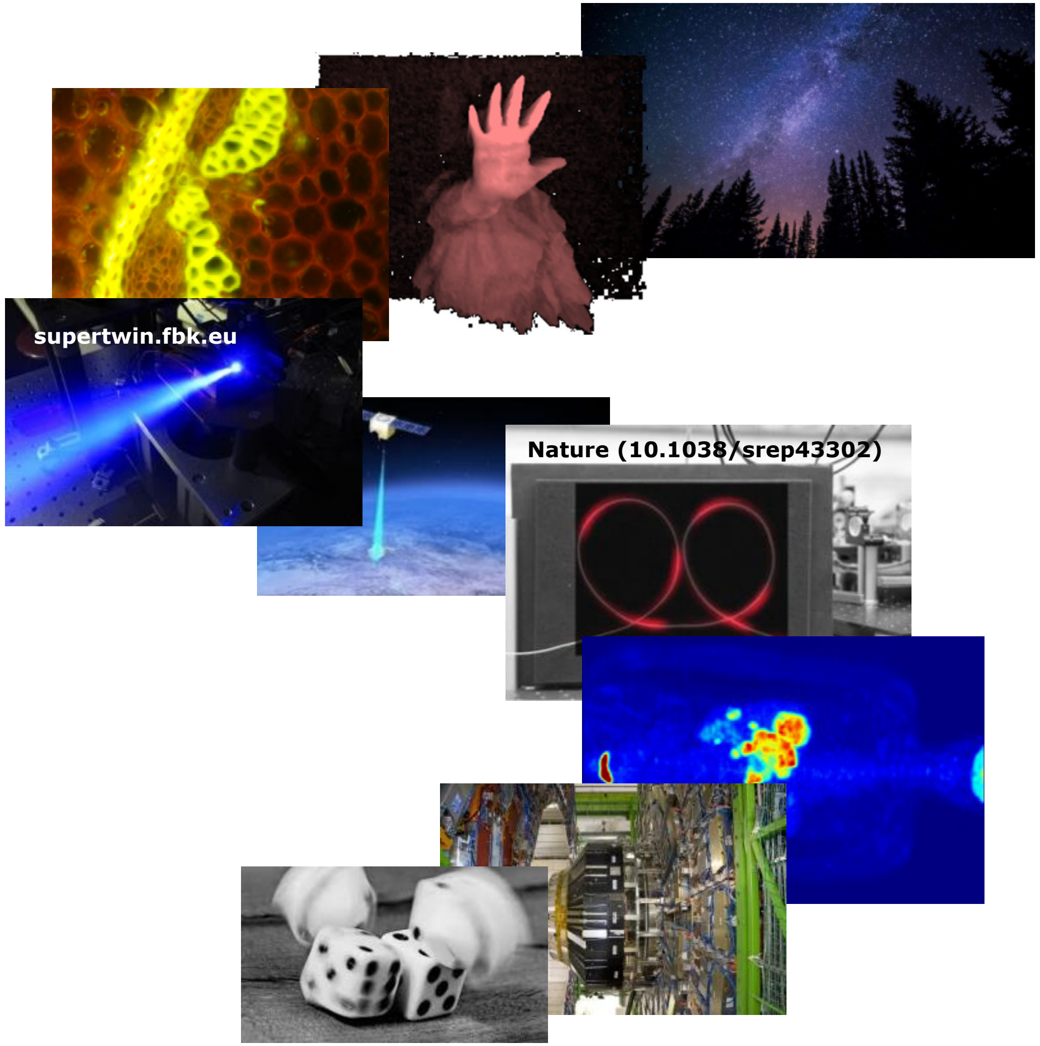


With single-photon sensing and imaging, more dimensions can be added to the measured space of information

Motivation: Examples of Single-Photon

- Low-light imaging
- Time-of-flight (single point, imaging)
- Fluorescence microscopy
- Quantum physics (entangled photons)
- Quantum communications
- Fast optical events
- Biomedical imaging (PET)
- High-energy physics
- Quantum Random Number Generation

- Many of these require imaging
  - Pixels → very small area required!



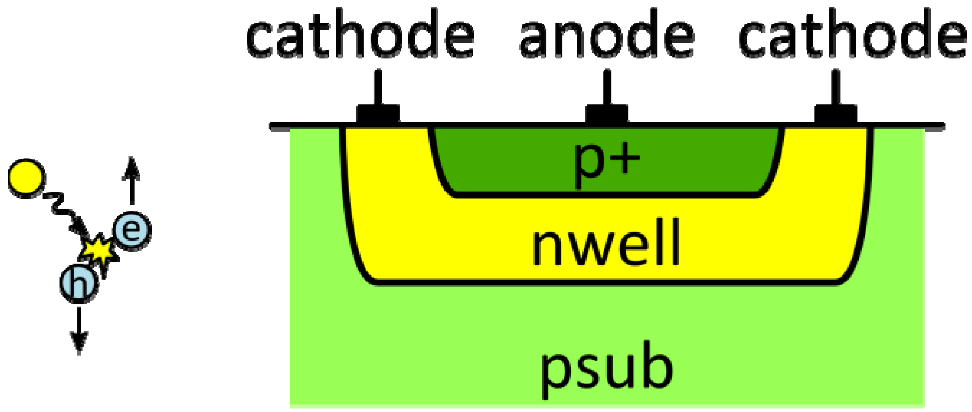
Outline

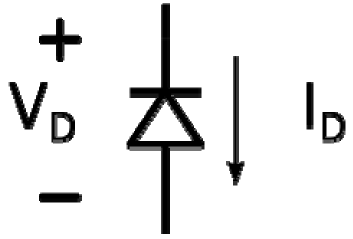
- The Device
- SPAD Frontend Circuits
- Building Blocks for SPAD Sensors
- SPAD-based Architectures
- Summary

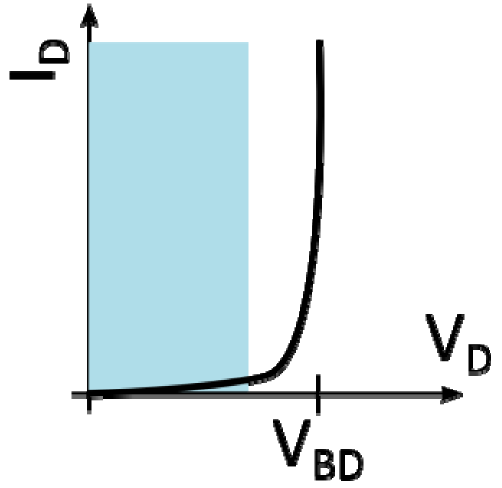
How to catch one photon at a time

THE DEVICE

Light Detection in CMOS

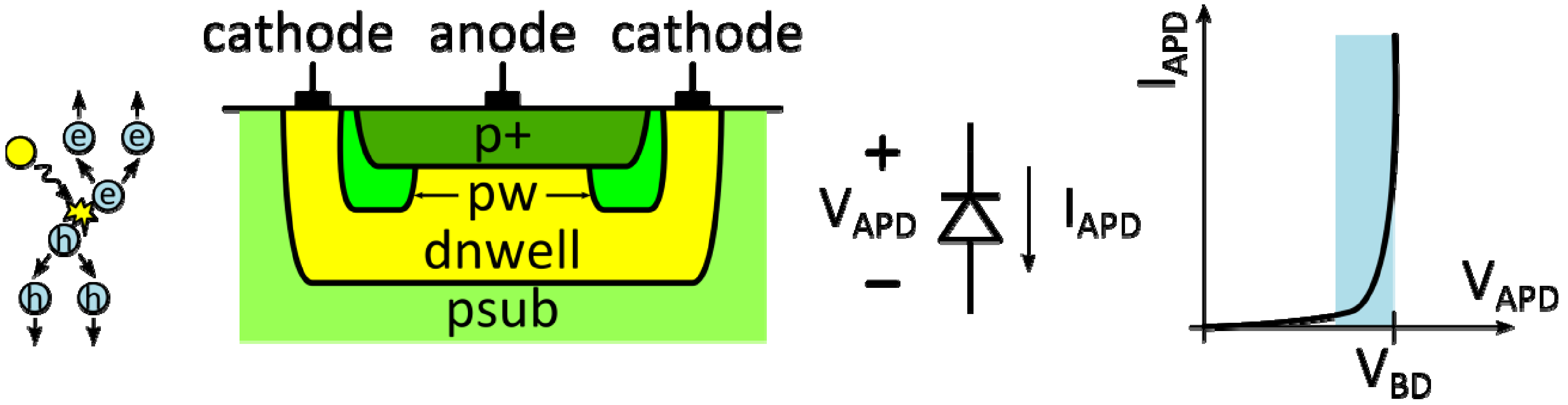






- Typical sensor is a reverse biased diode (photodiode)
- Each photon may interact with Si and generate an electron (photoelectron)
- Reverse bias field brings the generated charge to the electrodes
  - Output: photocurrent
- Typically: one absorbed photon ≈ one electron-hole pair
- Need many photons to observe a relevant signal

Increasing the Field: Avalanche PD



- Increasing the reverse bias brings the field to a critical level
- If a photon generates an electron, it is accelerated and *may* impact
- Result: one absorbed photon ≈ 10-1000 electrons
  - this is called “internal gain” ***M***
- Still an analog device
  - affected by noise

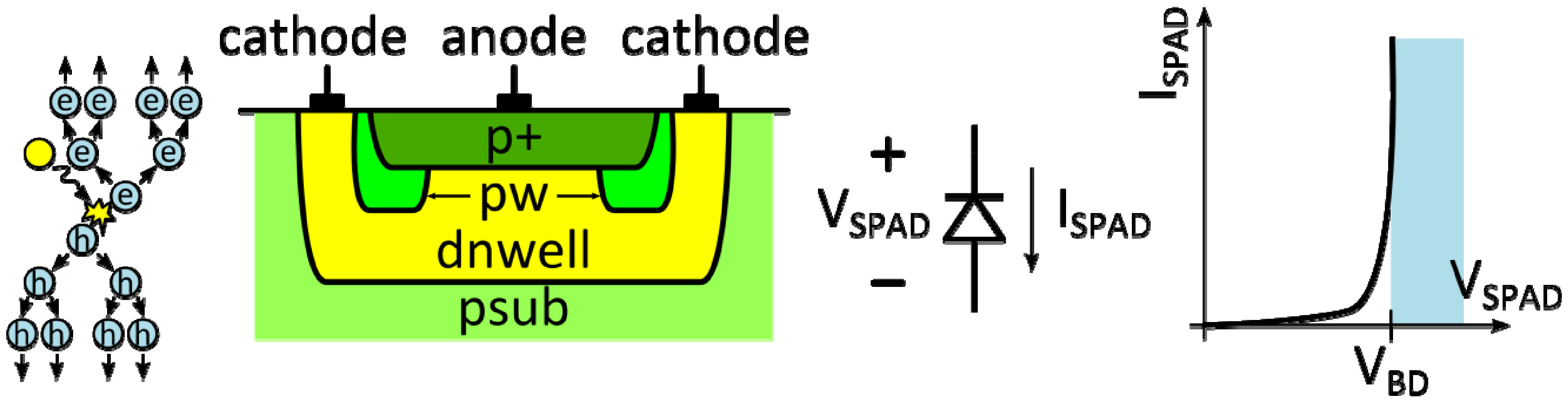
[McIntyre1972]

Excess noise factor

$$F = \frac{\beta}{\alpha} M \left(1 - \frac{\beta}{\alpha}\right)\left(2 - \frac{1}{M}\right)$$

Beta for holes
Alpha for electrons
Beta<alpha

Beyond VBD: the SPAD

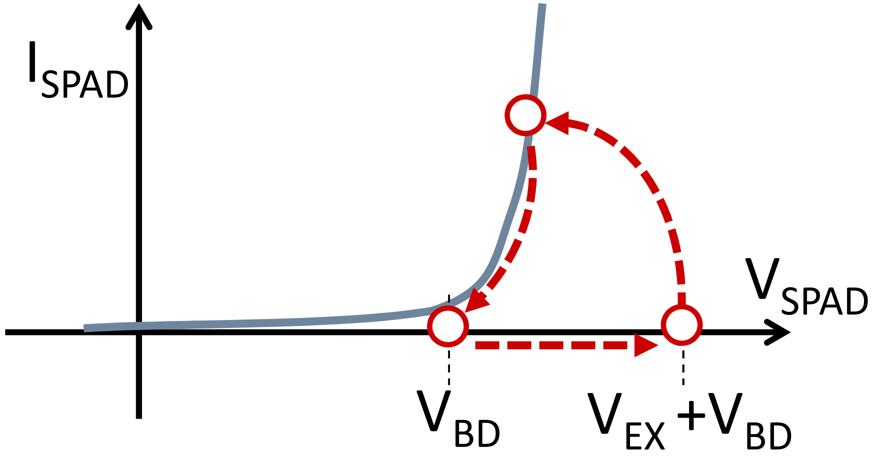


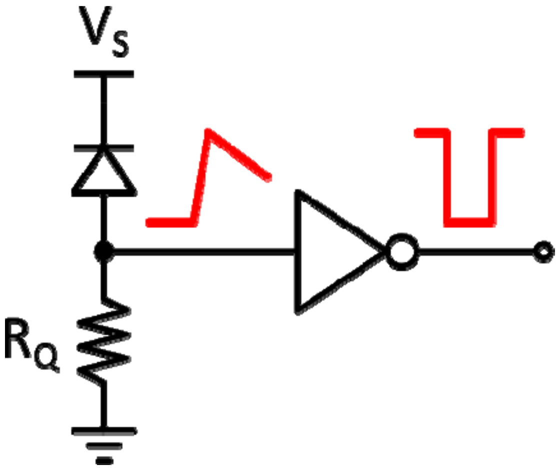
- Beyond the breakdown voltage the diode operates in Geiger Mode
- Internal gain above 10^5-10^6
- One photon ≈ a self-sustained breakdown!
- Basically, a digital device:
  - Avalanche → huge signal → not sensitive any more
- Extremely fast (10s of picoseconds)!

Excess bias

$$V_{EX} = V_{SPAD} - V_{BD}$$

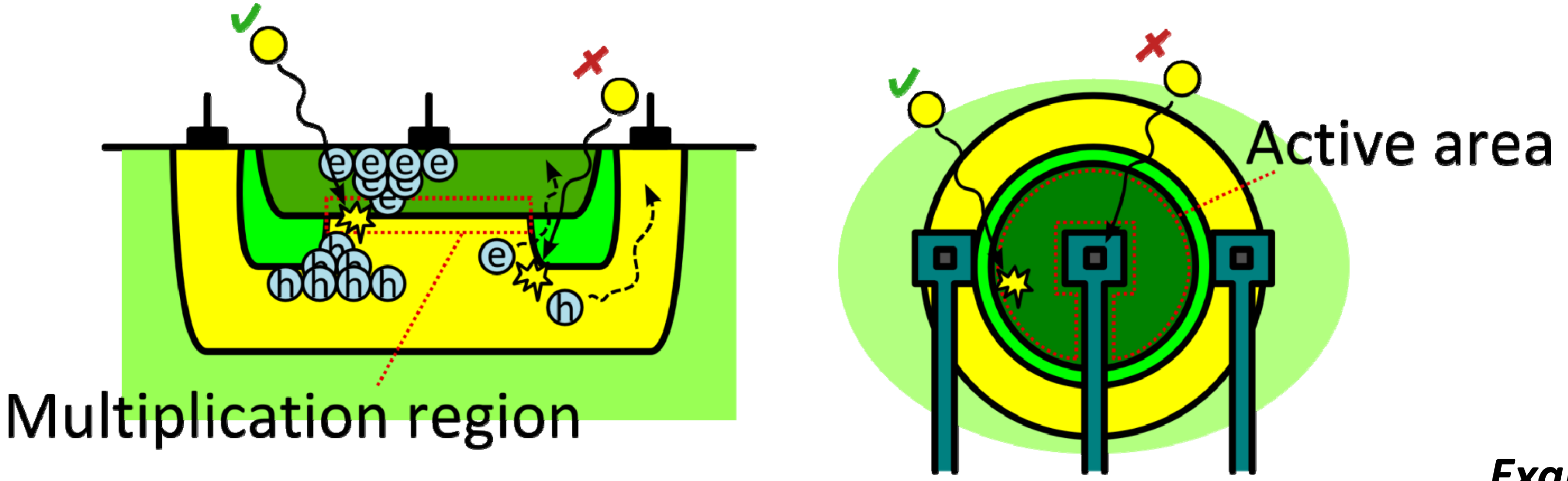
SPAD Typical Operation



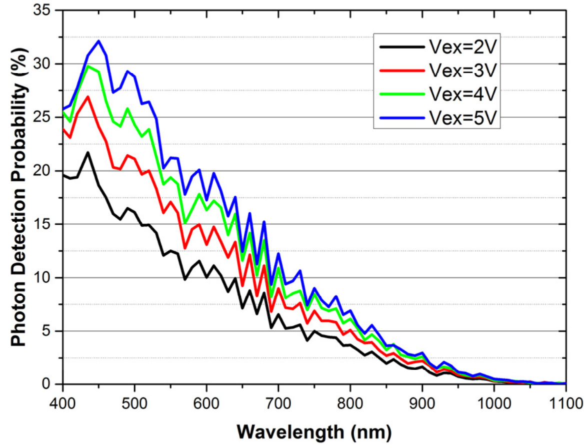


- How to operate the SPAD to make it usable?
- Three states
  - Bias beyond V_{BD}
  - Avalanche (photon, noise)
  - Recharge (also called quenching)
- Result: one photon >> one digital pulse

SPAD Characteristics (I)



Example [Xu2017]

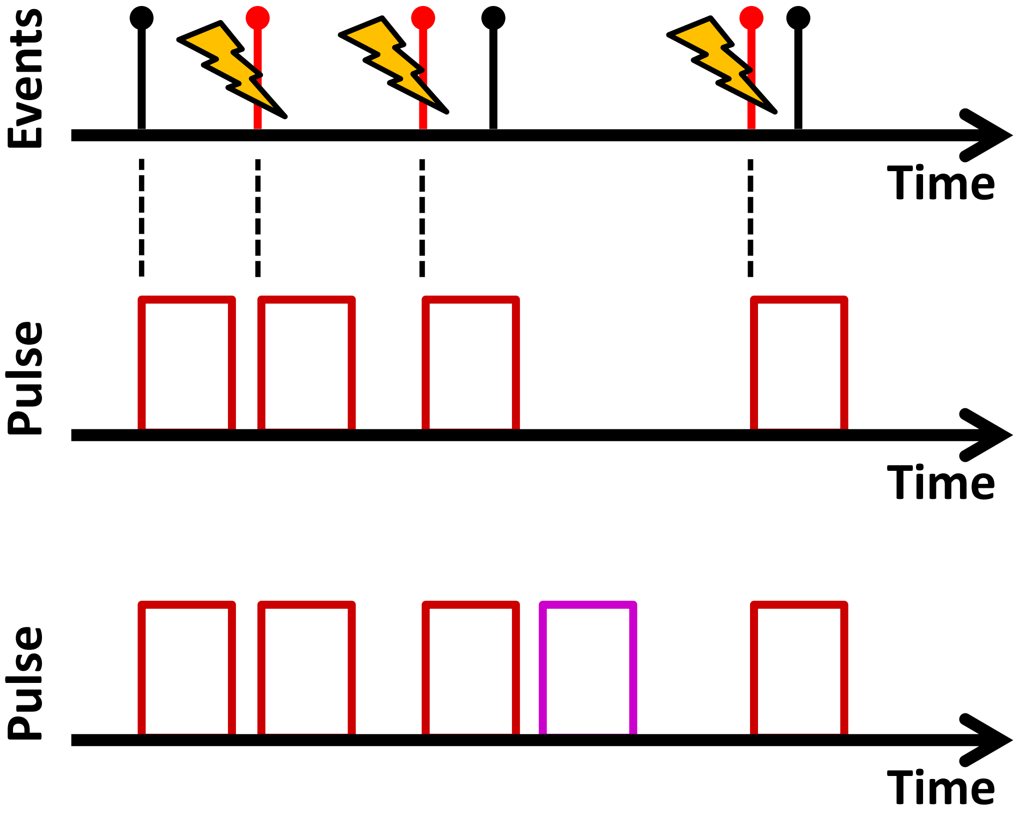


- Photon Detection Probability (PDP)
  - Avalanche probability when a photon hits the active area
- Photon Detection Efficiency (PDE)
  - Avalanche probability when a photon hits the pixel area

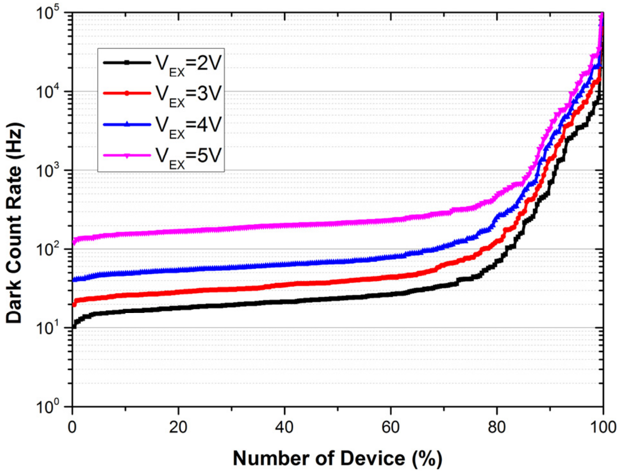
$$PDE = FF \cdot PDP$$

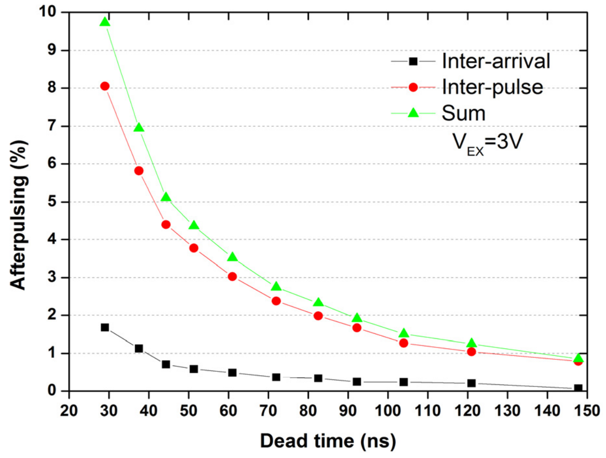
SPAD Characteristics (II)

- Dark Count Rate
  - (DCR)
- Dead Time
- Afterpulsing



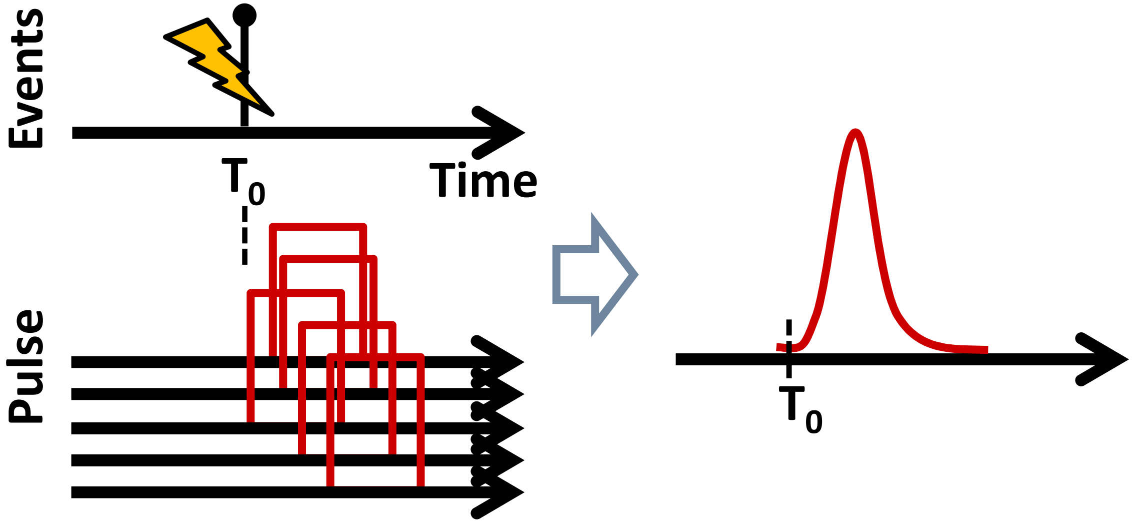
Example [Xu2017]



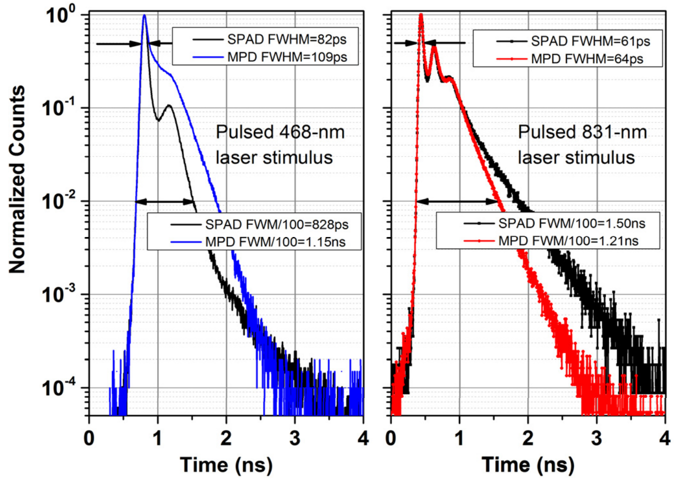


SPAD Characteristics (III)

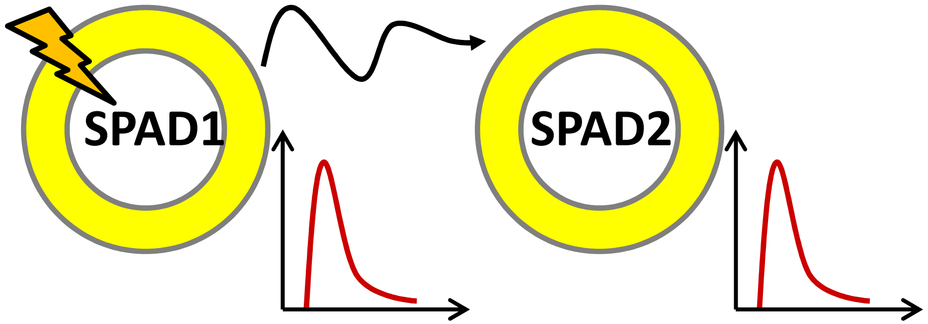
- Timing jitter

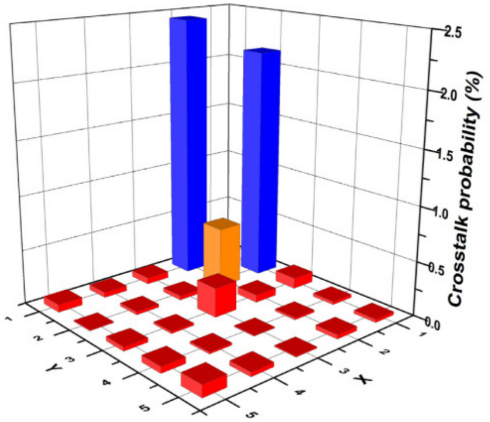


Example [Xu2017]



- Crosstalk





How to make a SPAD in CMOS

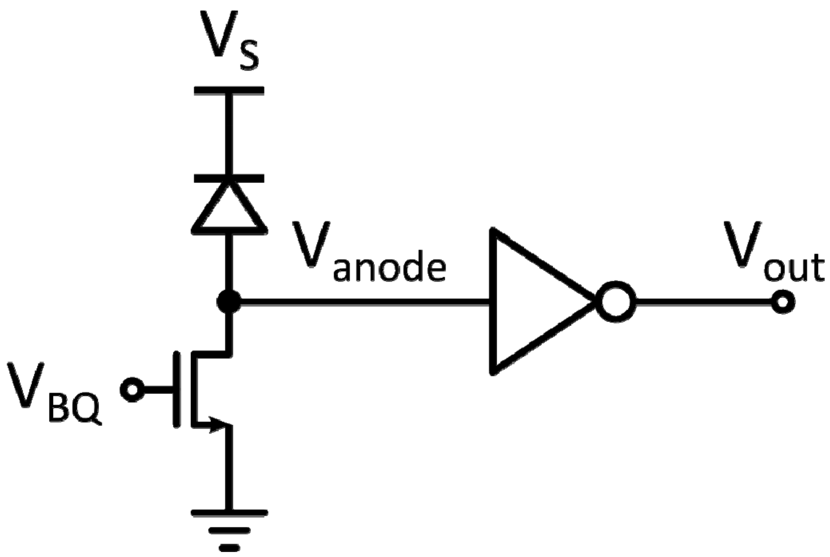
No guard-ring: early breakdown   Guard-ring: uniform avalanche area

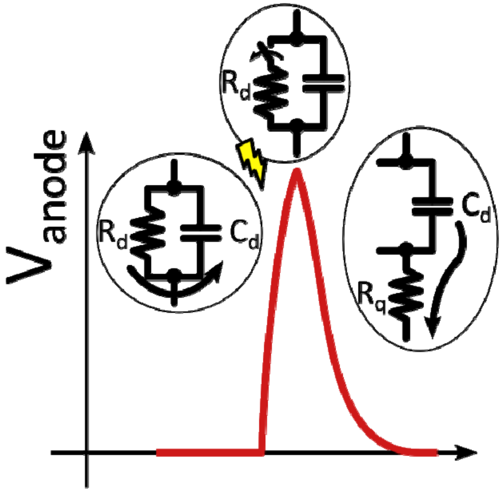
- SPAD designed in CMOS have to coexist with circuits → design rules
- First goals in designing a SPAD
  - To create a junction where to apply a high field
  - To prevent early breakdown at edges (guardring)
- Improvements
  - Increase PDP, reduce DCR, make it faster, make it smaller, reduce crosstalk, …
  - Small SPADs, [AlAbbas2017]: 8.25μm in 40nm tech

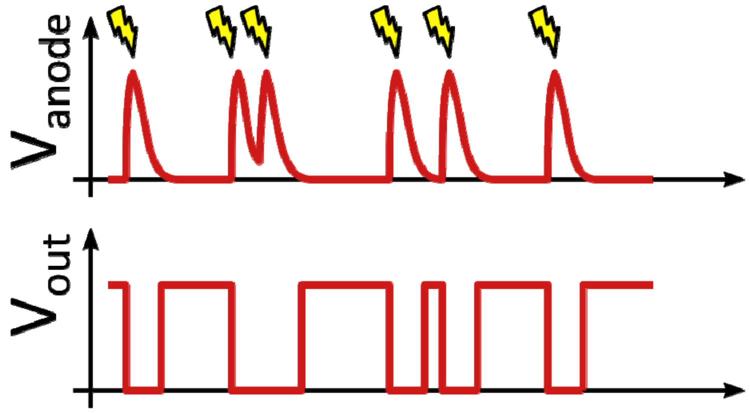
How to bring the wild SPAD under control

SPAD FRONTEND CIRCUITS

Passive Quenching





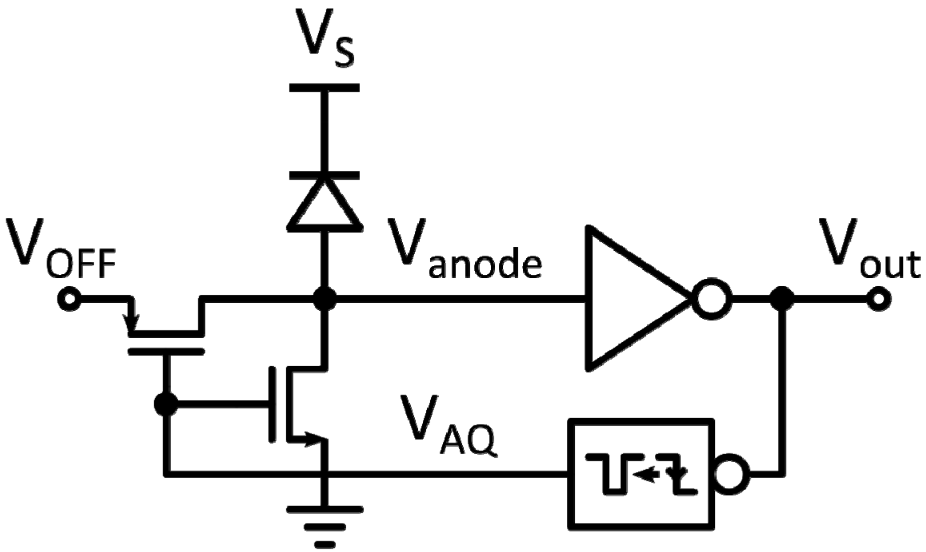


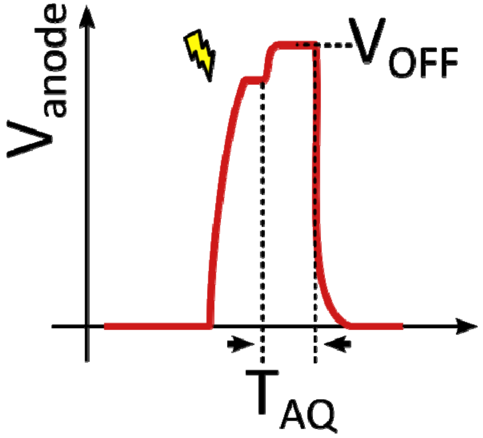
- SPAD recharge is done through a resistor (or a MOS) [Cova1996]
- MOS can be used as
  - Resistor R_Q (long MOS, high V_{BQ})
  - Current source (med-L MOS, low V_{BQ})
- Very compact in area and low power: suitable for arrays

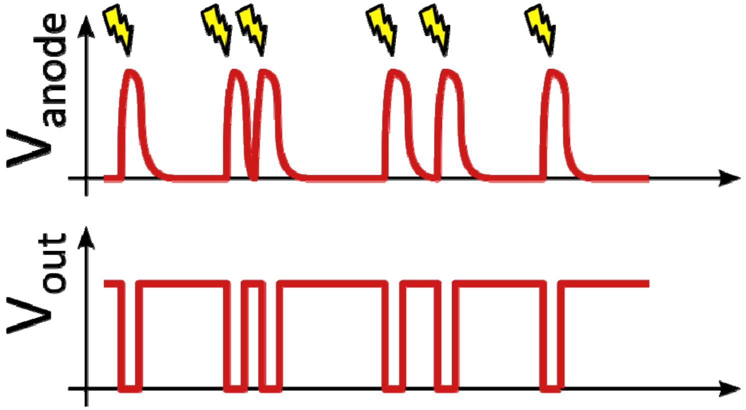
$$\frac{V_{EX}}{R_q} \ll \frac{V_{OFF}}{R_d} \qquad \text{Min } R_Q$$

$$T_{dead} \approx C_d R_q \qquad \text{Max } R_Q$$

Active Quenching

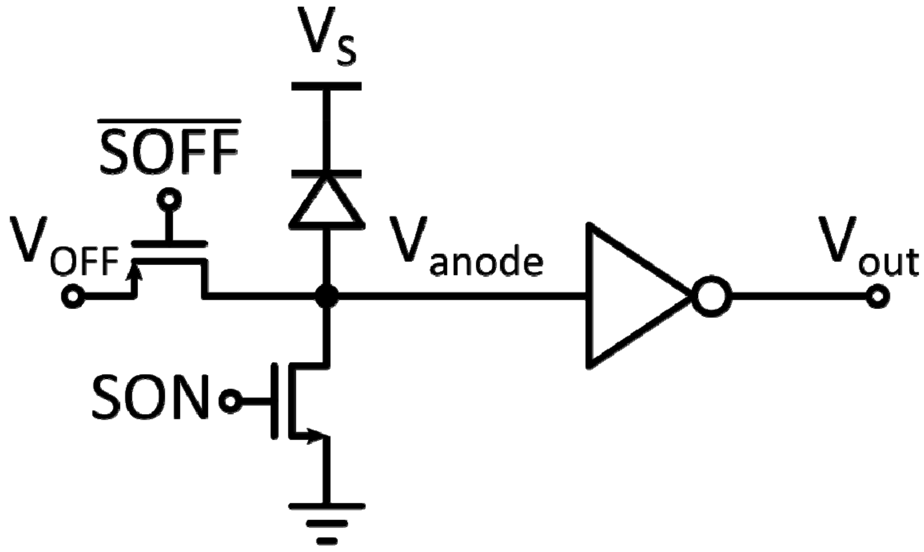


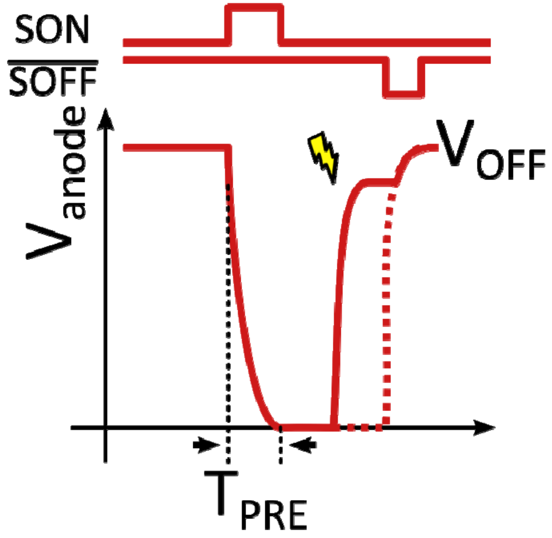


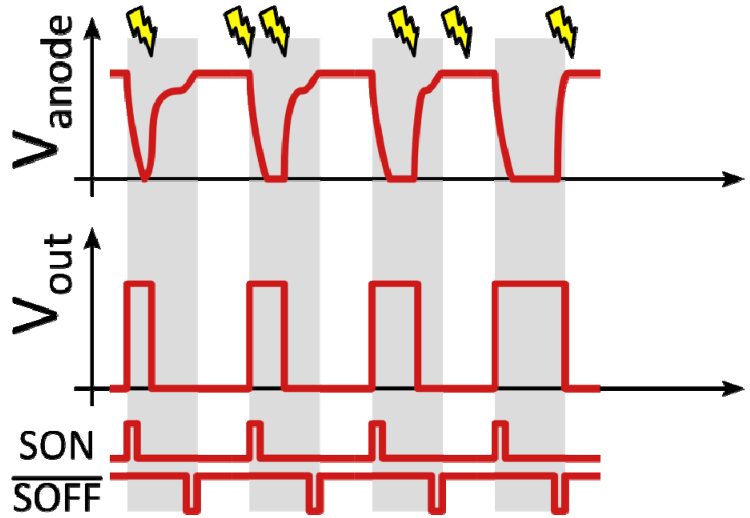


- SPAD is actively recharged [Cova1996]:
  - avalanche is detected and recharge is forced
- Very short deadtime
- Large in area and potentially power hungry → single detector or small arrays

High-Z Quenching

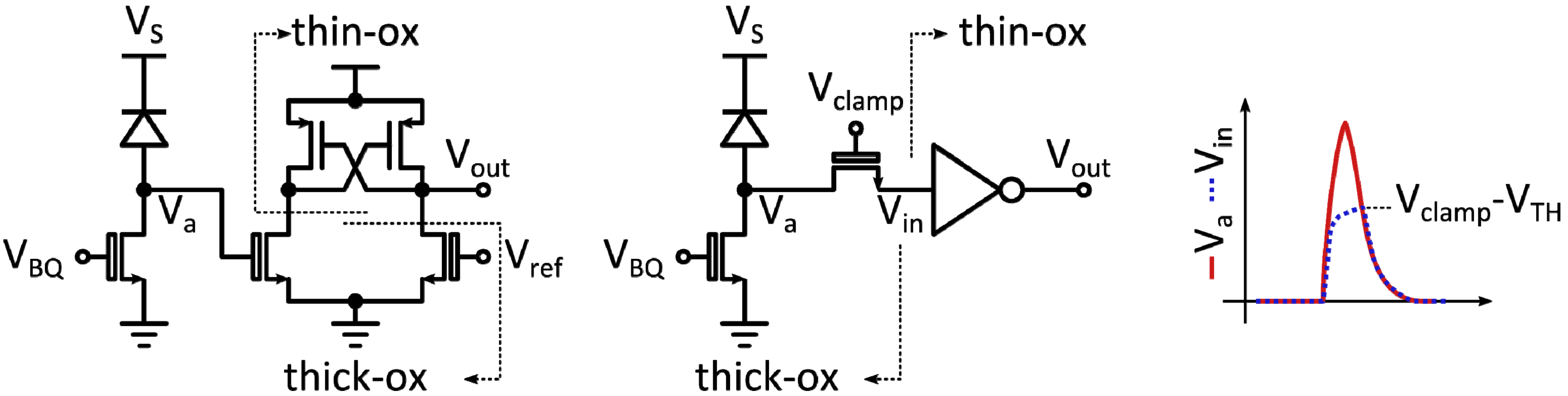






- SPAD is precharged and left in Hi-Z state
  - Avalanche stops automatically when $V_{SPAD} < V_{BD}$
- Typically, synchronous operation: precharge is cyclically repeated
- Periodical discharge can be used to reduce afterpulsing
- Very compact in area: suitable for arrays

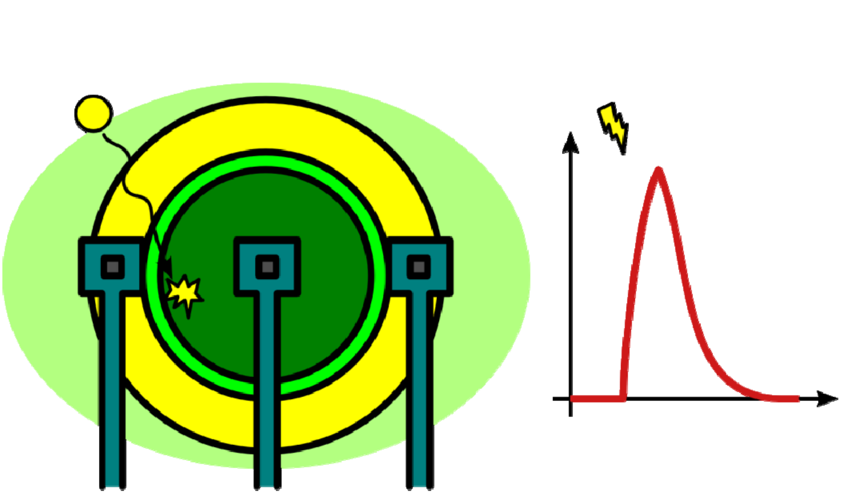
SPAD Output Voltage

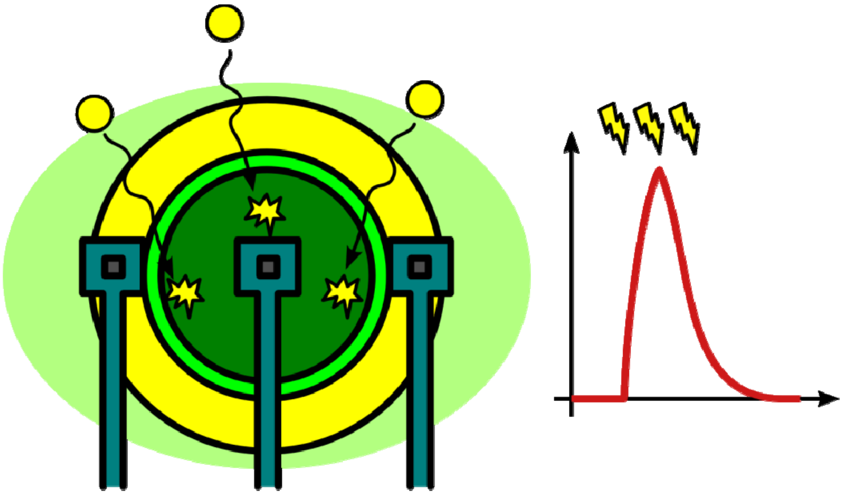


- Typical voltage domain of digital logic: 1.8V, 1.2V, …
- Typical excess bias of SPAD for optimal PDP/timing: 3-5V
- Solutions
  - Level translator
  - Clamping transistor

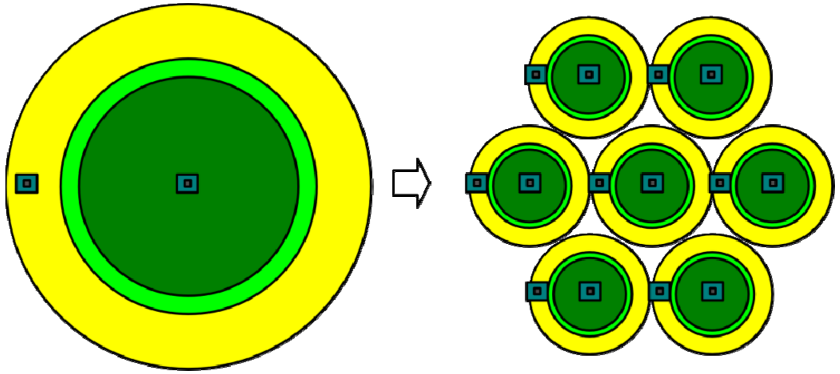
⇨ Level translators needed!

Multiple: the Silicon Photomultiplier (SiPM)

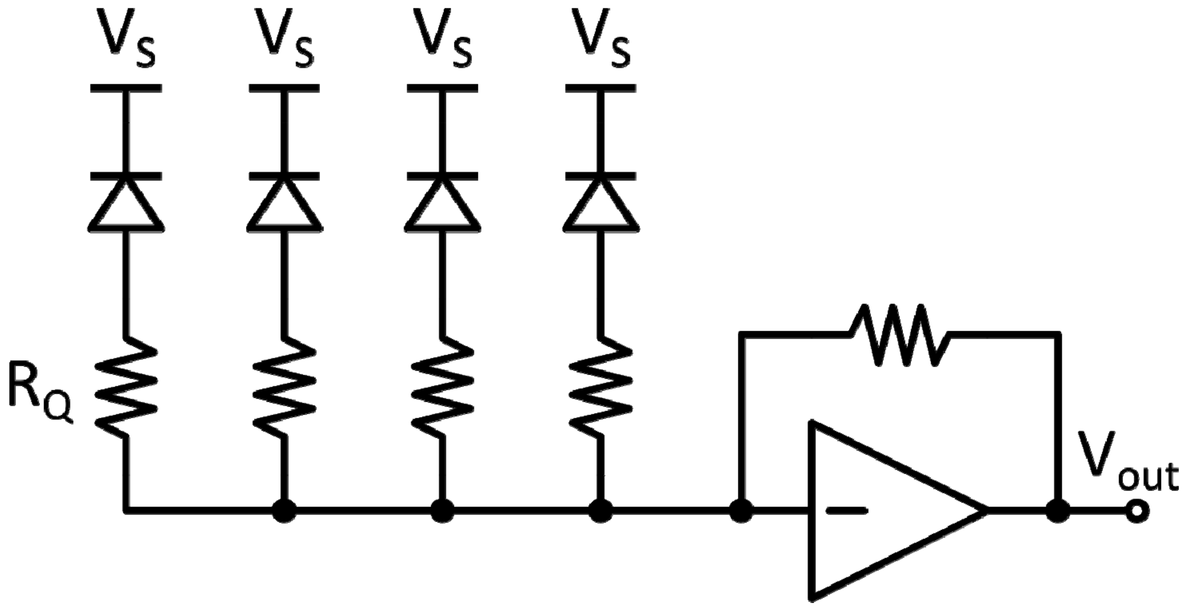


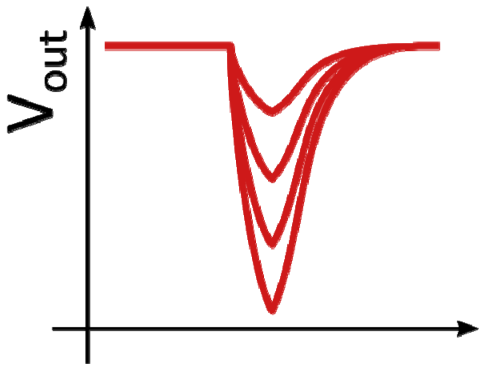


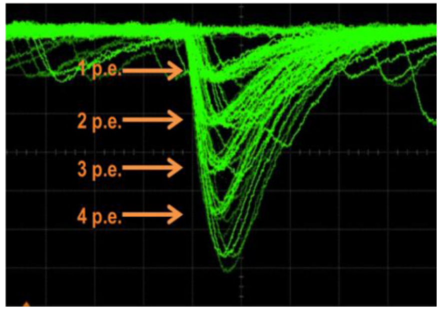
- One SPAD → one photon (at least during dead time)
- Why not putting inside your "detecting unit" more than a SPAD?
- Concept: **silicon photomultiplier**



Multiple: the analog SiPM (a-SiPM)

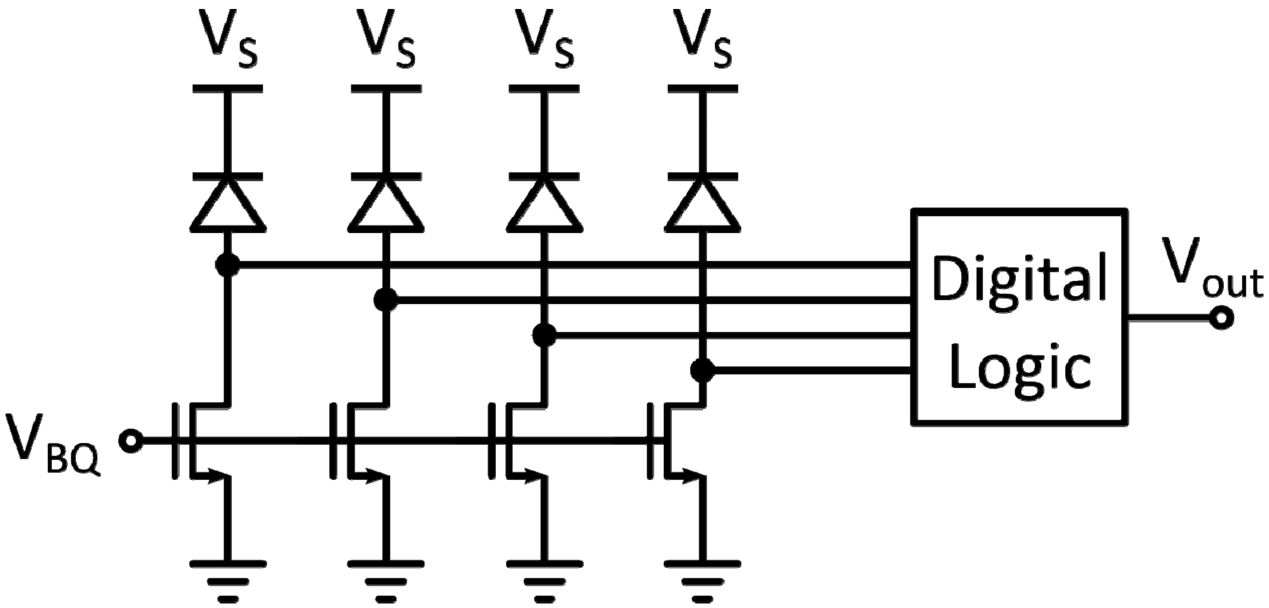


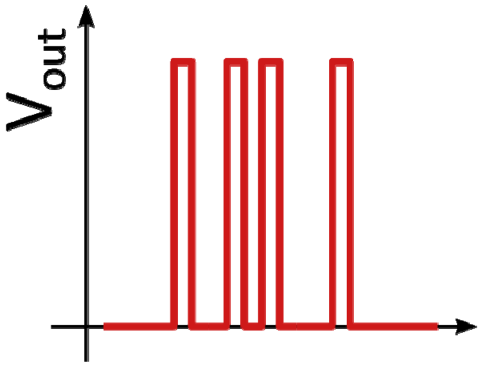




- **Analog SiPM**
- Many SPADs with quenching resistor connected in parallel
  - Measure the current
  - Count the photons with quantization
- Typical solution for custom technology SPAD [Piemonte2016]

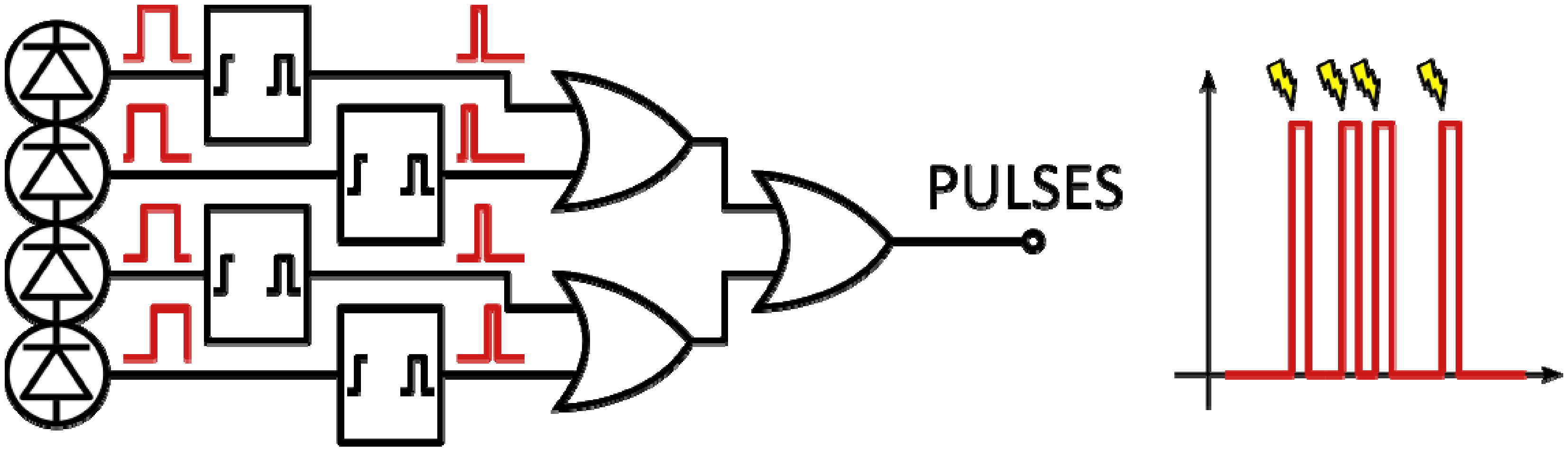
Multiple: the digital SiPM (d-SiPM)





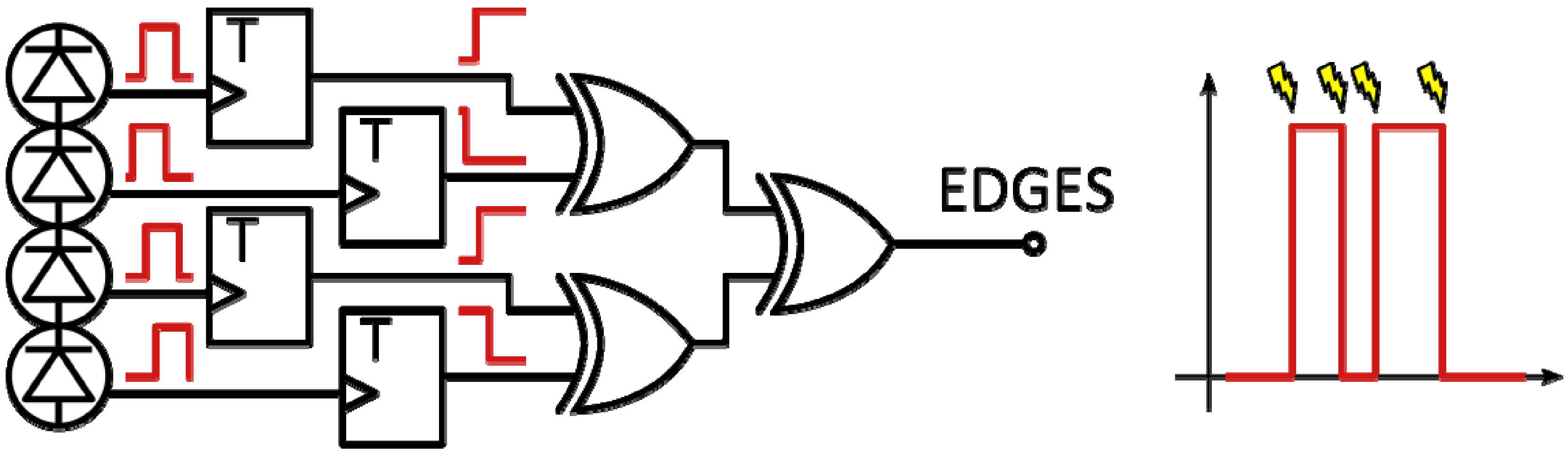
- **Digital SiPM**
- Multiple SPADs and front-ends digitally combined
  - Use digital signals
  - More electronics, but more functions
- Typical solution for CMOS SPAD [Braga2014]

Frontend for dSiPM



- Convey on one single wire the **pulses** of N SPADs
- Time-compression
  - Shorten pulse (monostable)
  - Combine (OR-tree)
- Small area, bandwidth limited by pulse width (≈100s of ps)

Frontend for dSiPM



- Convey on one single wire the SPADs avalanches as **edges**
- Edge encoding
  - Pulse to edge (toggle FF)
  - Combine (XOR-tree)
- More area but double bandwidth (≈ or below 100ps)

How to extract what we need from a SPAD output

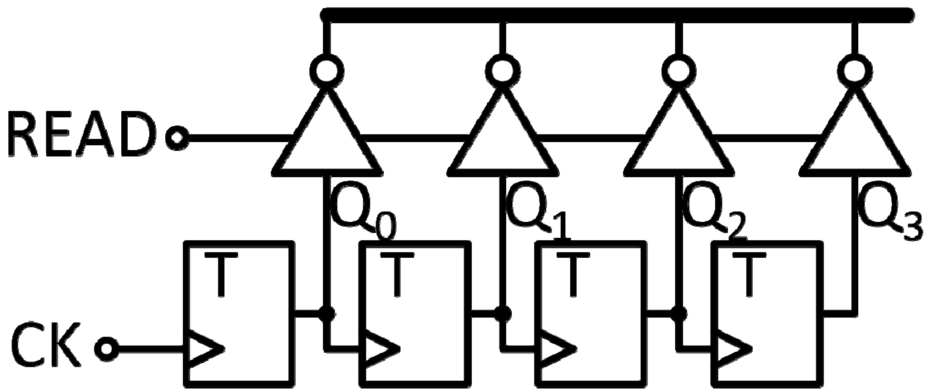
BUILDING BLOCKS FOR SPAD SENSORS

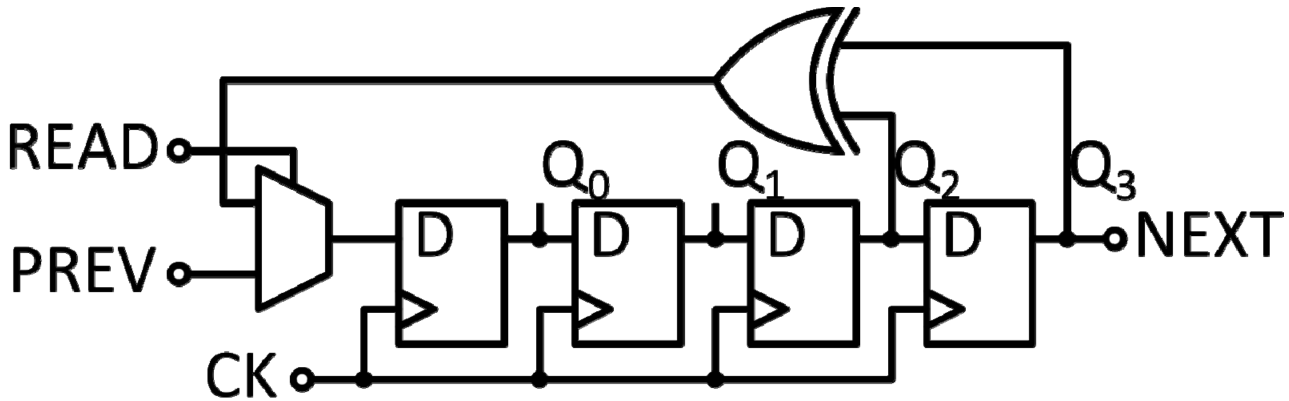
Number of Photons: Counting

- Intensity imaging
  - low-light imaging
  - time-gated observation

- Digital counter
  - Precise, reliable, robust
  - Large area (toggle flipflops, output bus, …)
- Analog counter
  - Noisy and non-uniform
  - Small area
  - Still need to convert to digital

Digital Photon Counting

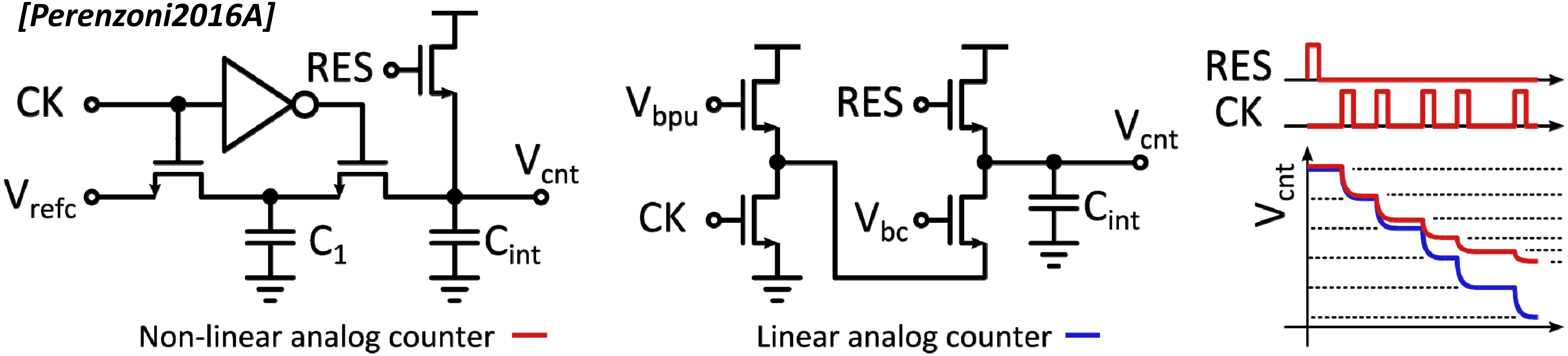




- Easiest: ripple counter
- A bit more clever: linear feedback shift register (LFSR)
  - Advantage: serial readout, only one bitline
  - Disadvantage: cannot make operations
- How to save more area? Dynamic logic, but light can erase it!

Analog Photon Counting

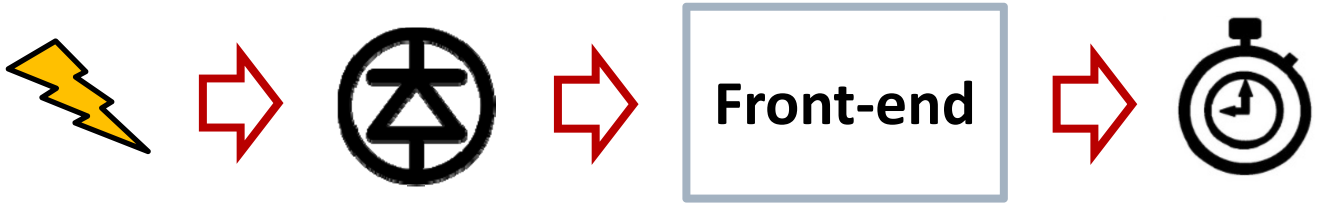
[Perenzoni2016A]



- Basic principle of analog counter: drop charge packets on a capacitor
- Challenges:
  - Uniformity
  - Linearity
  - Noise

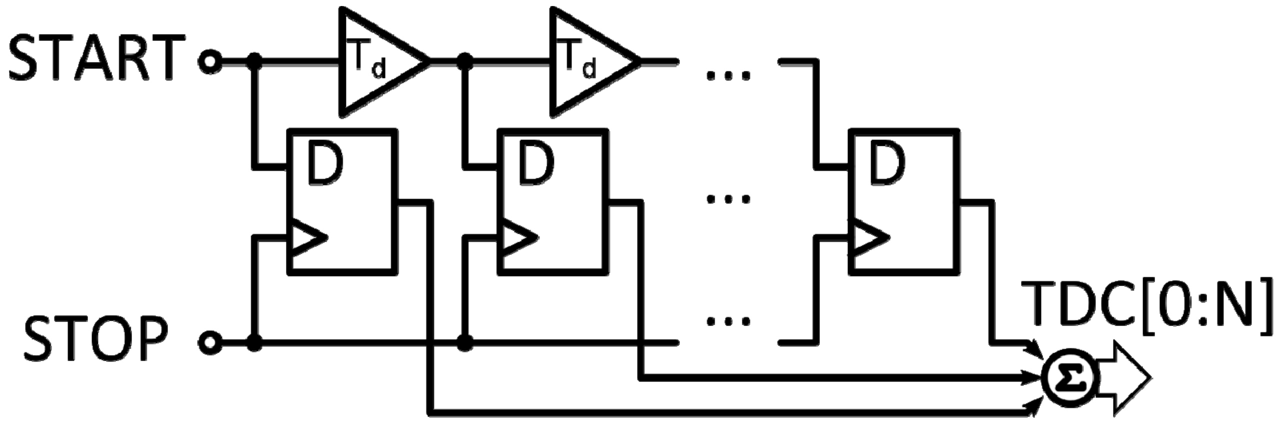
⇨ Practical limit: 6-8bit
Area can be very small (2T + 2C)

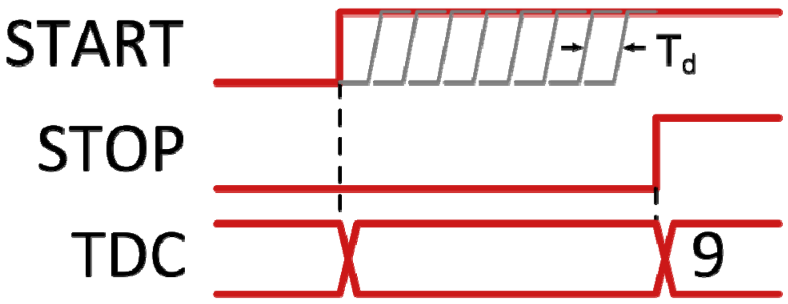
Time of Arrival: Timestamping



- SPAD + stopwatch
- Time-to-digital converter (TDC)
  - Relatively small architectures exist (still, as counters, depth matters)
  - Not exactly low power
- Time-to-analog converter (TAC)
  - Non-uniformity
  - Noise
  - Still need to convert to digital

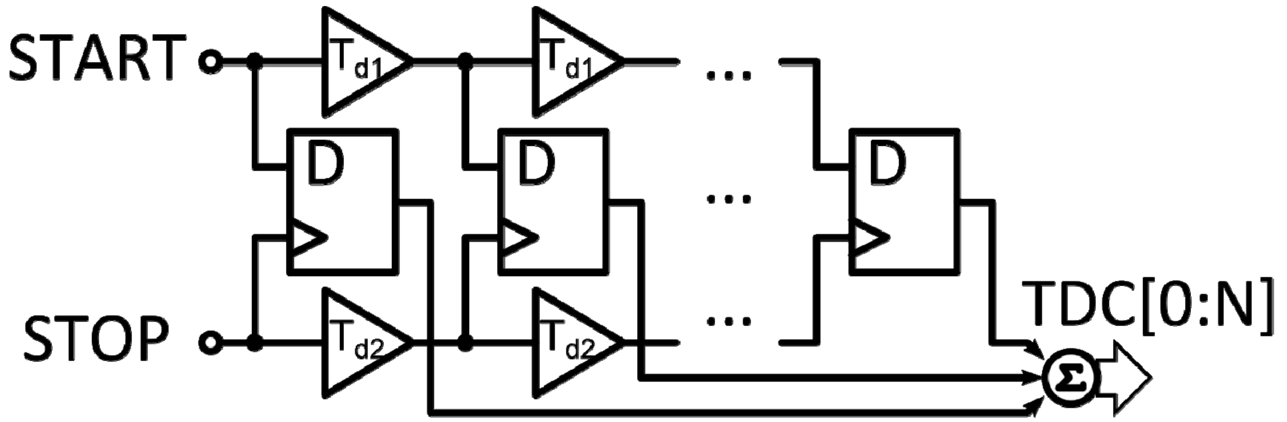
Timestamping: linear TDC

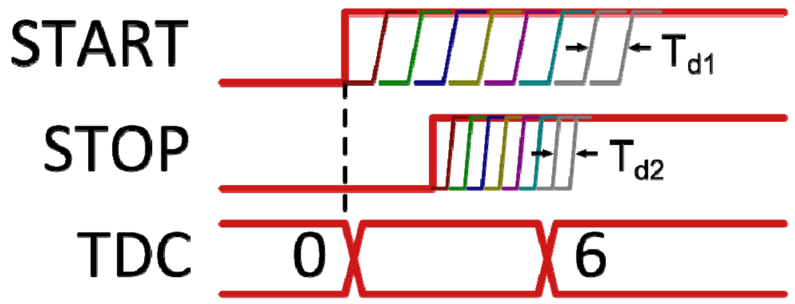




- Flash linear TDC
  - Long delay lines with sampling DFFs
  - Resolution T_d (gate delay)
  - Area consuming

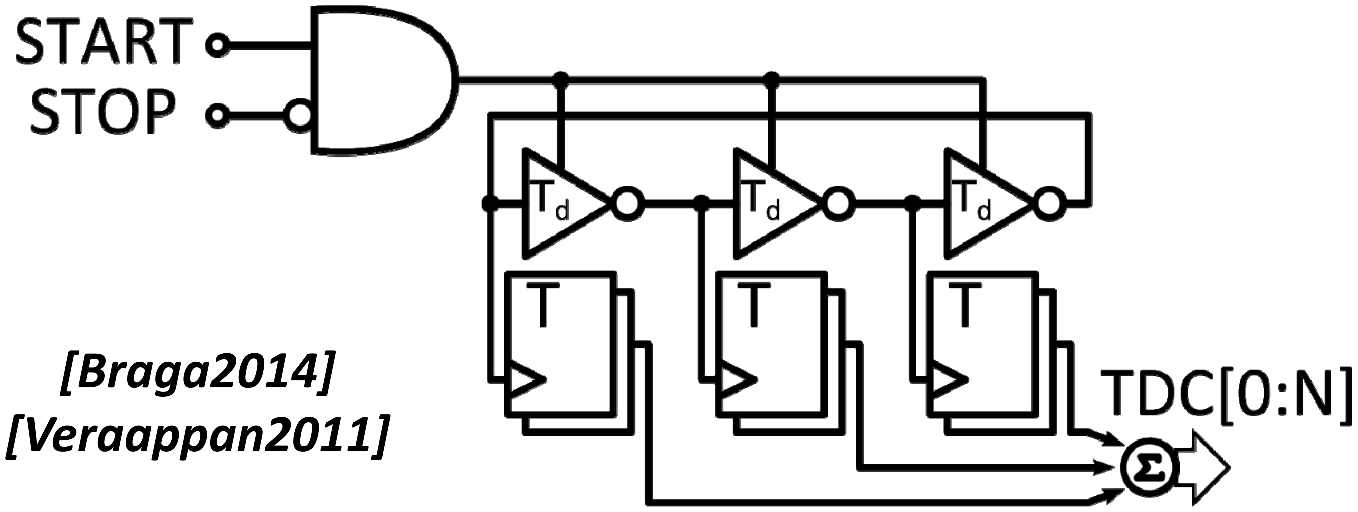
Timestamping: Vernier TDC



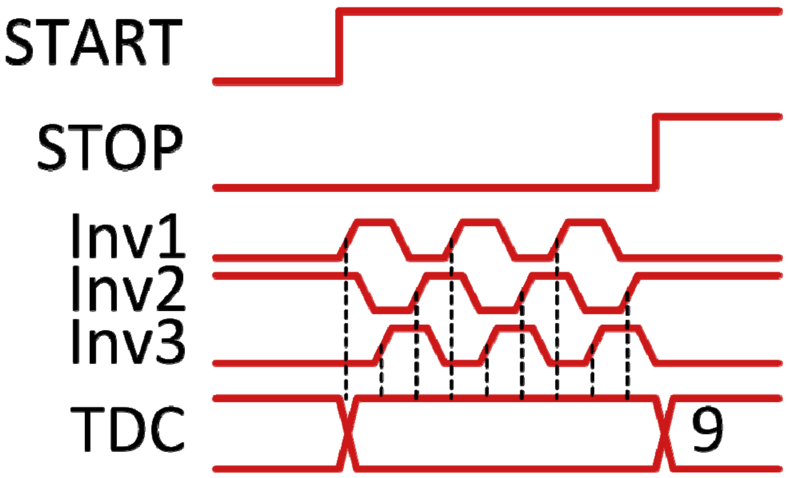


- Flash Vernier TDC
  - Propagate START and STOP on two slightly different delay lines with $T_{d2} < T_{d1}$
  - Resolution $T_{d1} – T_{d2}$ (sub- gate delay)
  - Area consuming

Timestamping: Ring-oscillator TDC

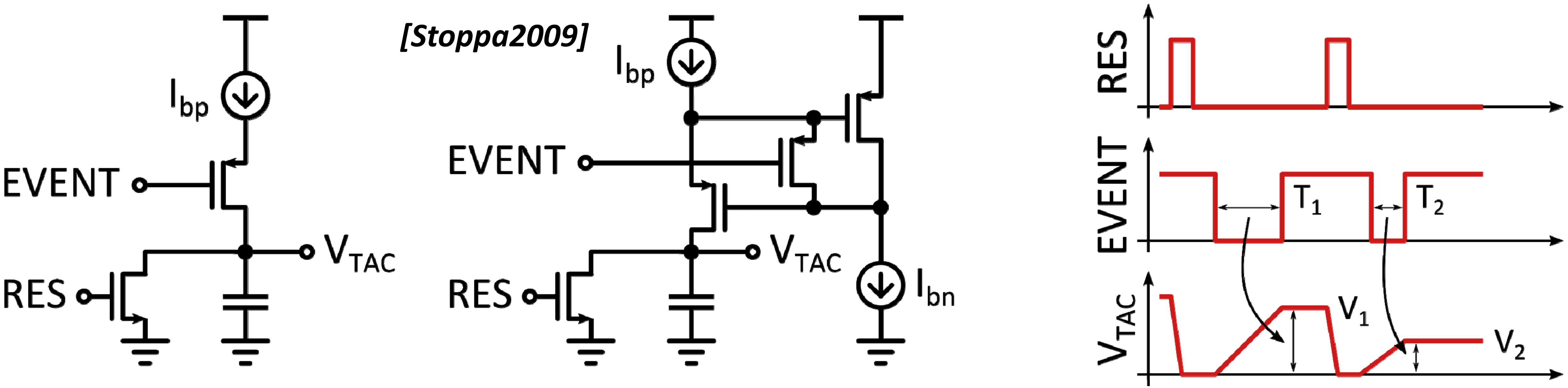


[Braga2014]
[Veraappan2011]



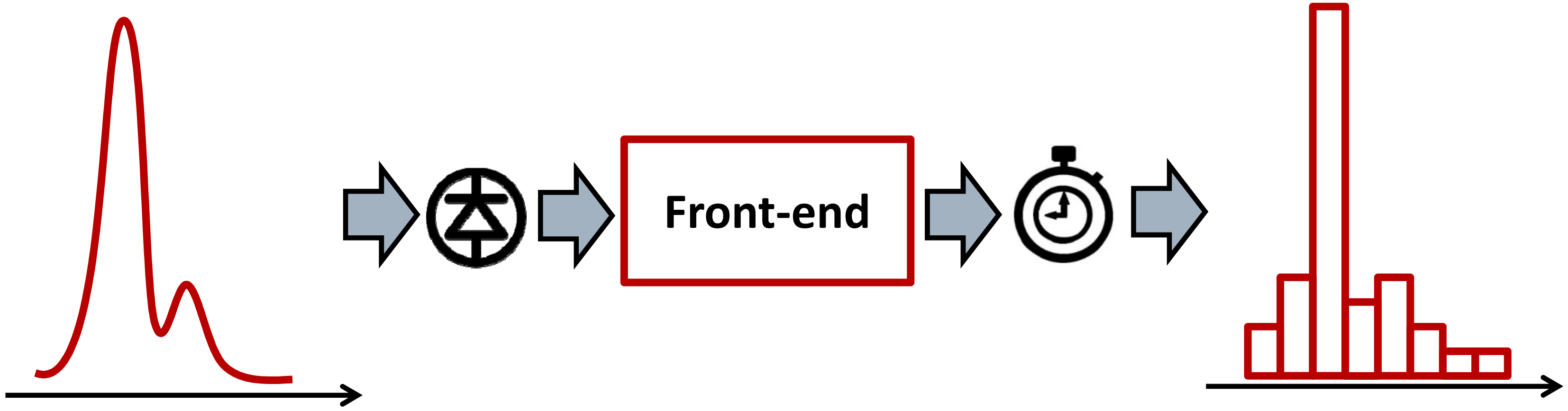
- Small area:
  - fold the delay line into a RO
  - count cycles
  - sub-period interpolation
- Fast RO → Relevant power consumption
- High depth → Relevant area consumption

Timestamping: TAC



- Very similar to analog counter, sharing pros and cons
  - Uniformity
  - Linearity
  - Noise
- Potentially very compact, main issues: robustness, need for calibration

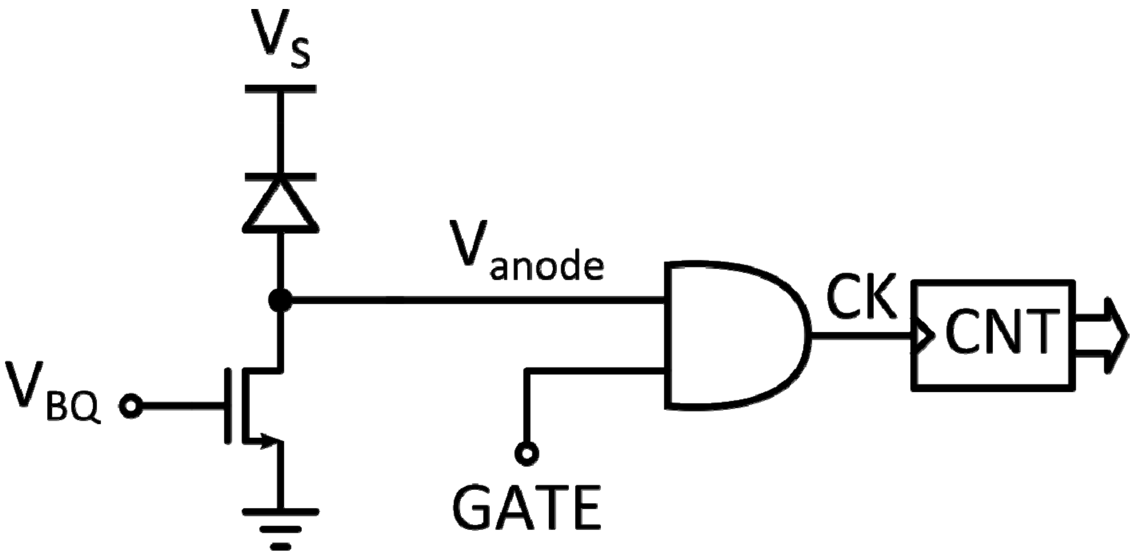
Counting and Timestamping

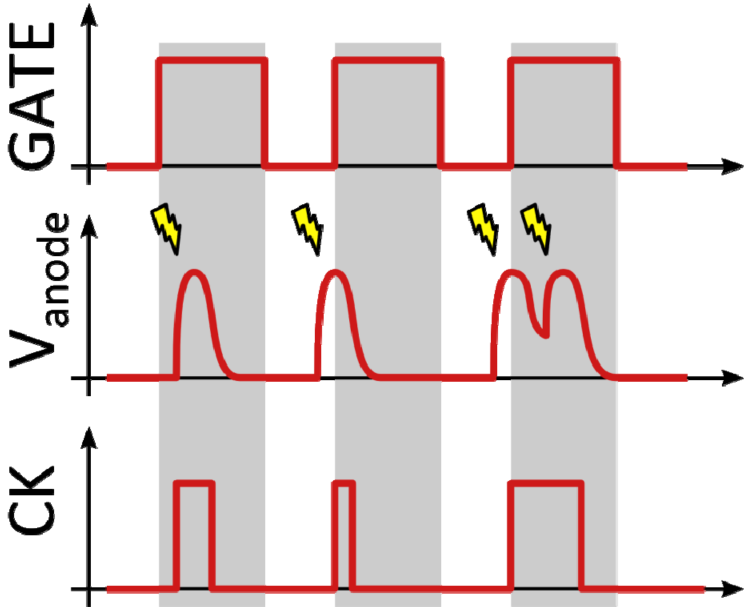


- ☐ Time-correlated single-photon counting (TCSPC)
  - Compiling an histogram of photon arrival times
- ☐ Problem: amount of data produced
  - External accumulation → high framerate, small chip/pixels
  - Local accumulation → limited memory, large chip size

⇨ Alternative solution: Time-Gating

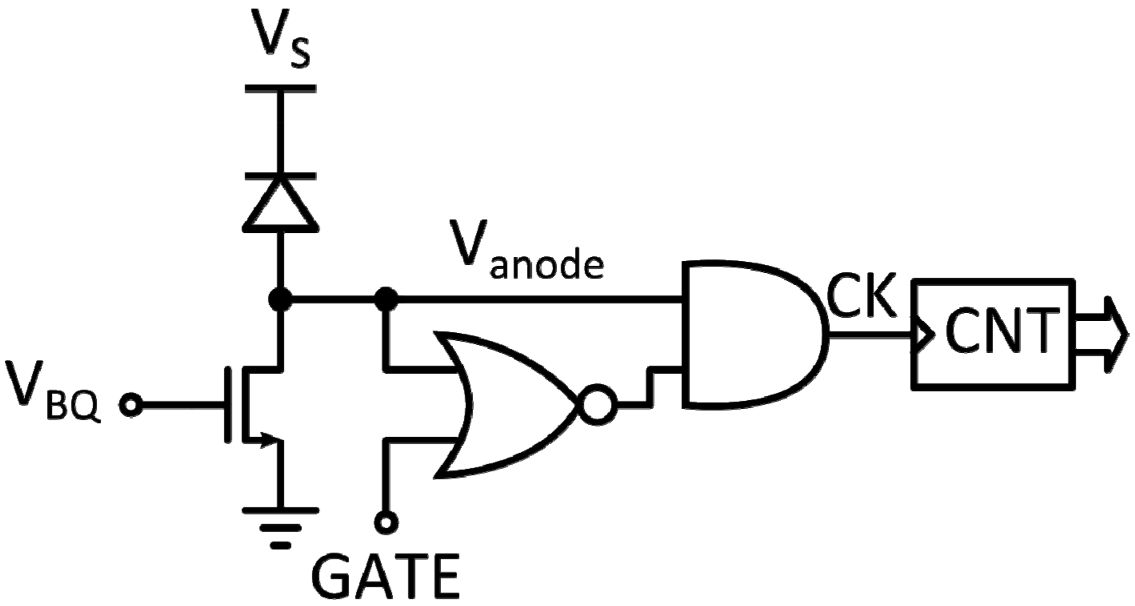
SPAD Time-Gating

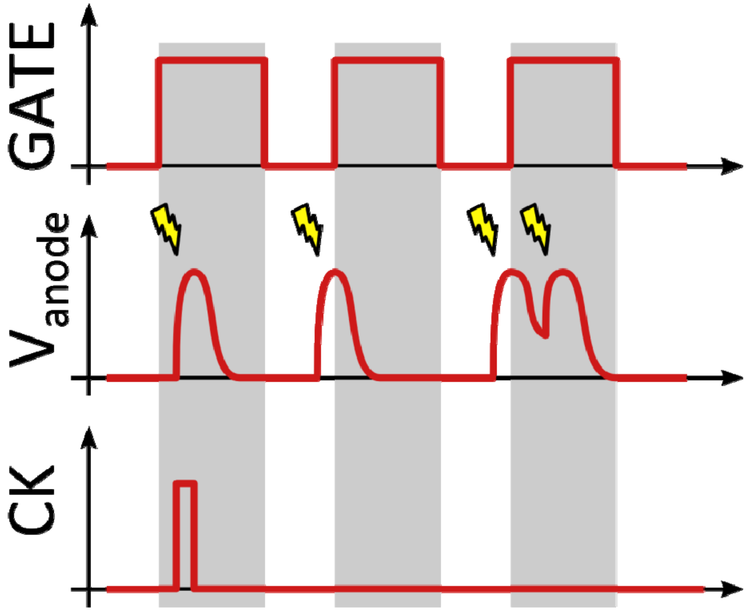




- **Standard digital gating**
- Pulses from SPAD are enabled by digital logic
  - Photons arriving just before are pushed inside the window
  - Potentially varying pulse width
  - Very simple to implement

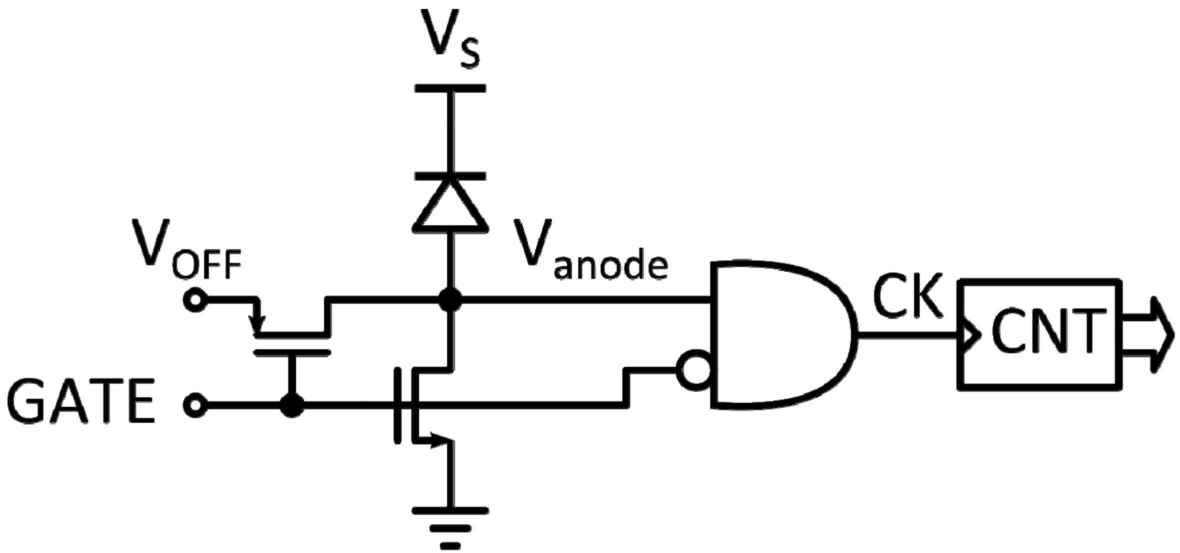
SPAD Time-Gating

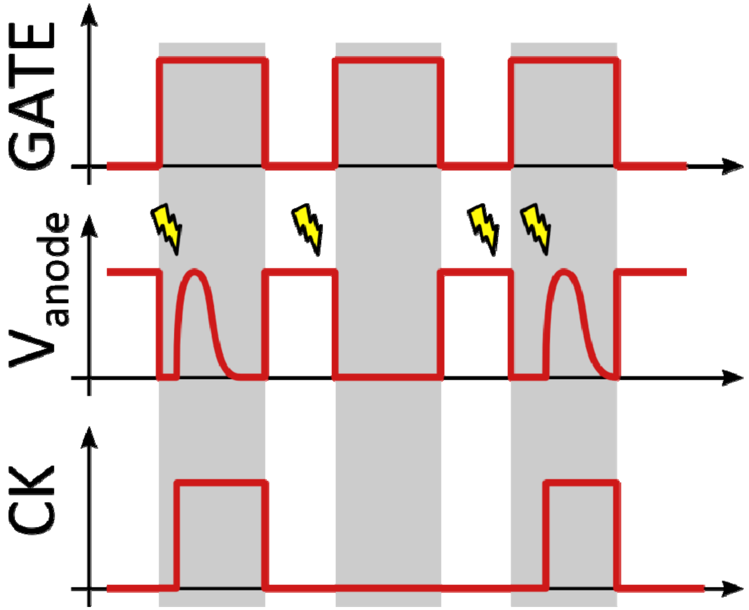




- ☐ **Edge-sensitive gating**
- ☐ Only rising edge of SPAD pulse is used
  - ■ Photons arriving just before are blinding the SPAD
  - ■ Need careful tuning → good knowledge of SPAD pulse shape, or buffer
  - ■ More area

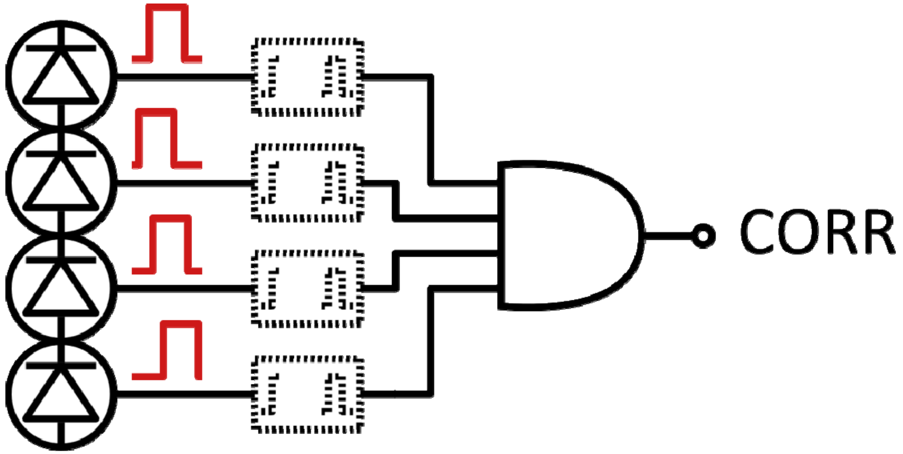
SPAD Time-Gating

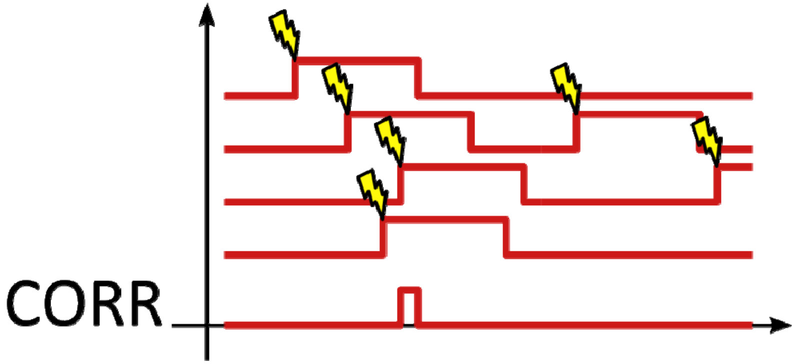




- **Turn-on gating**
- SPAD is turned on *only* when gate is enabled
  - Only photons within the window are counted
  - No blinding for photons arriving before
  - Challenges in array precharge distribution

Photon Time Correlation



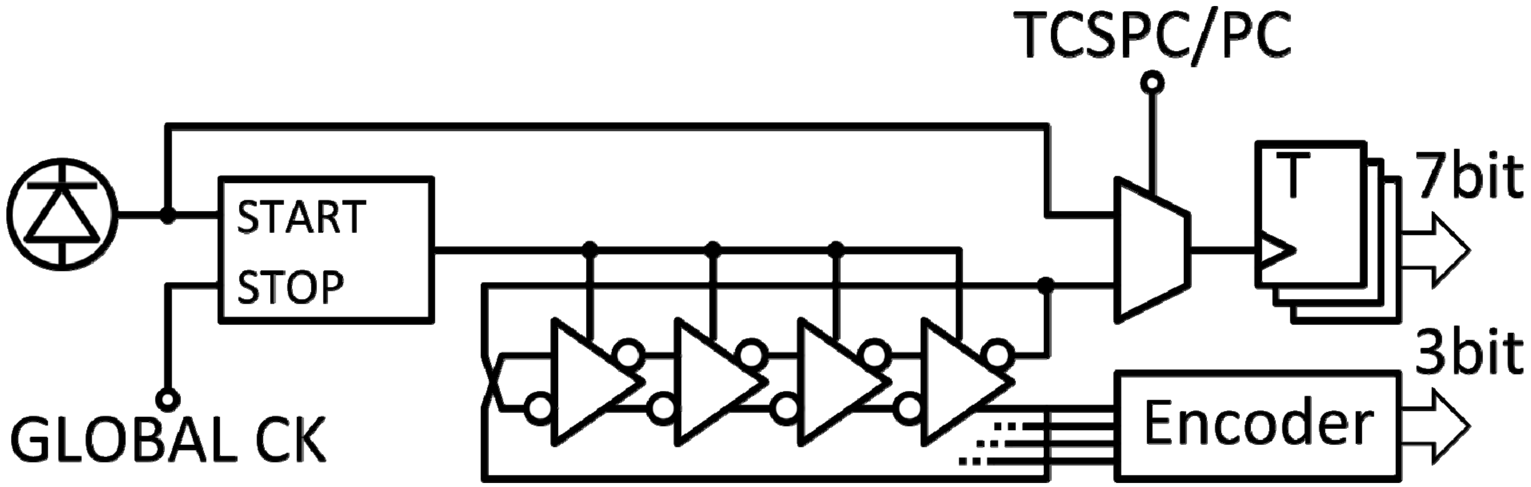


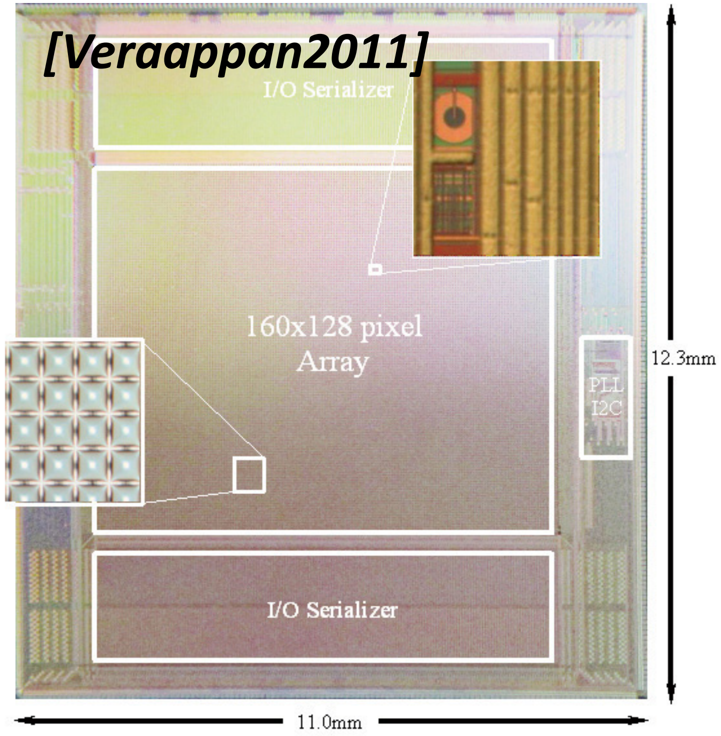
- Some applications require to do sth when **>N photons** are detected
- Exploit dead-time as correlation window
  - Not all Twin are usable, afterpulsing changes
- Add monostables
  - More area required but more flexible

How to create your perfect SPAD-based recipe

SPAD-BASED ARCHITECTURES

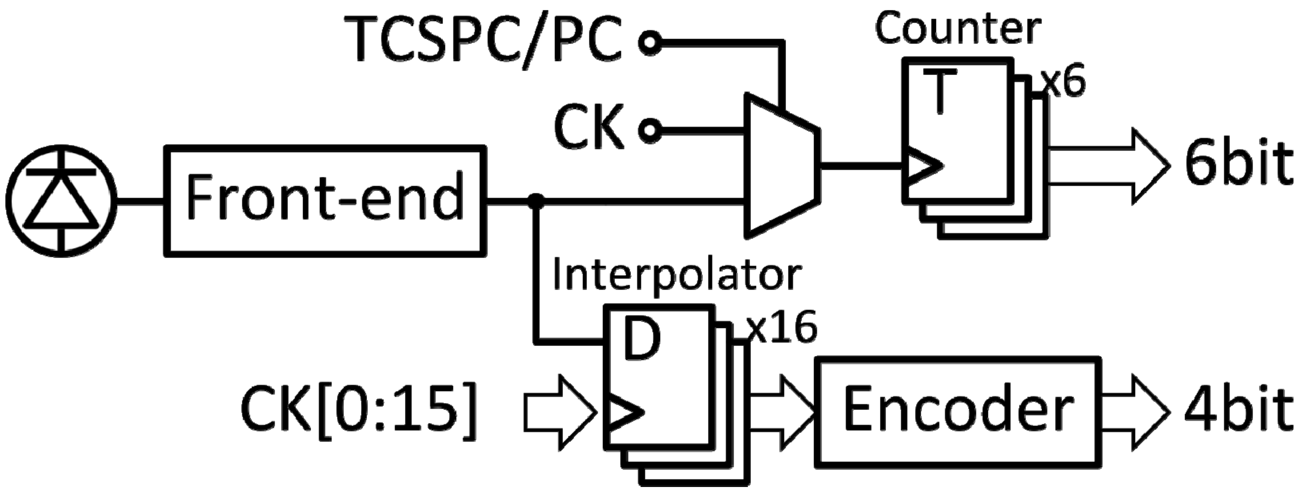
SPAD TCSPC Imaging **[Veraappan2011]**

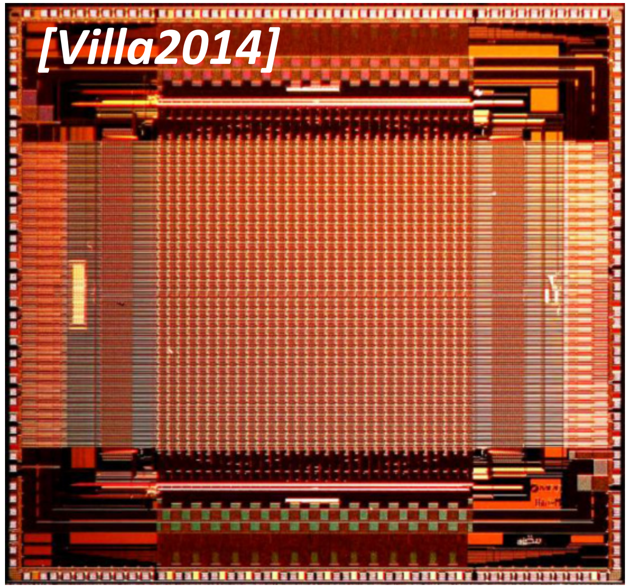




- 160x128-pixel array for TCSPC Imaging
  - 50µm pixel pitch, 1% fill-factor, 0.13µm tech
  - In-pixel 50ps TDC or 7bit counter
  - High-speed readout (1Mfps)
  - Applications: FLIM, ToF, high-speed

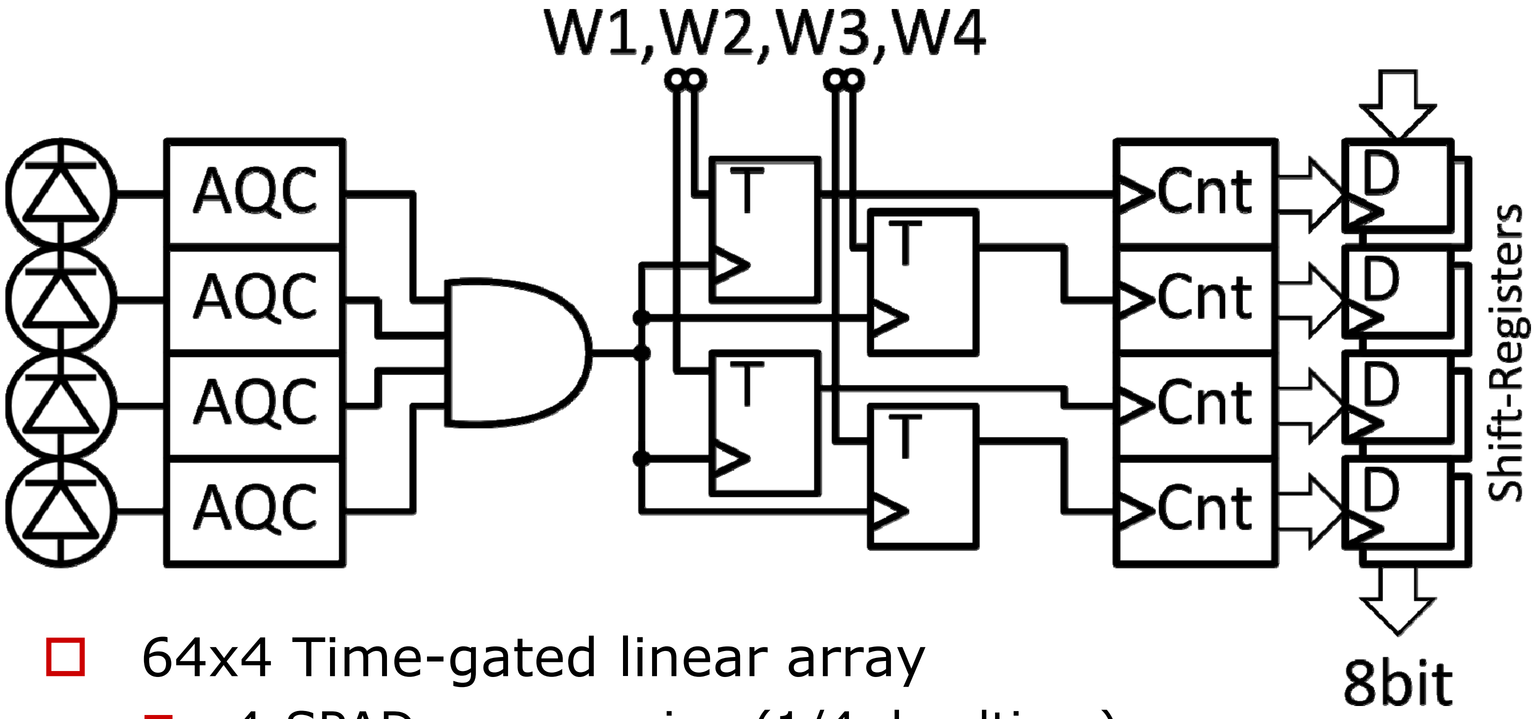
SPAD TCSPC Imaging **[Villa2014]**

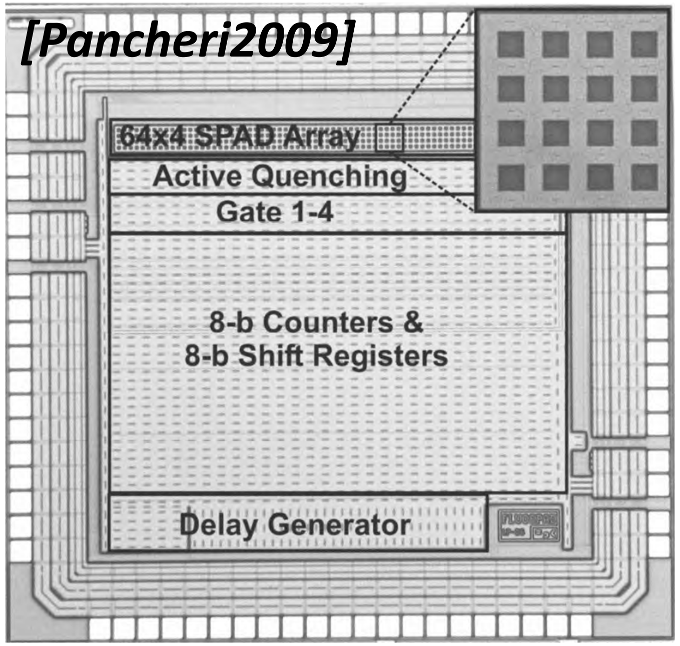




- 32x32-pixel array for TCSPC imaging
  - Extremely low-noise 30μm SPAD (customized process)
  - 150μm pixel pitch, 3.14% fill-factor, 0.35μm tech
  - In-pixel 312ps TDC or 6bit counter

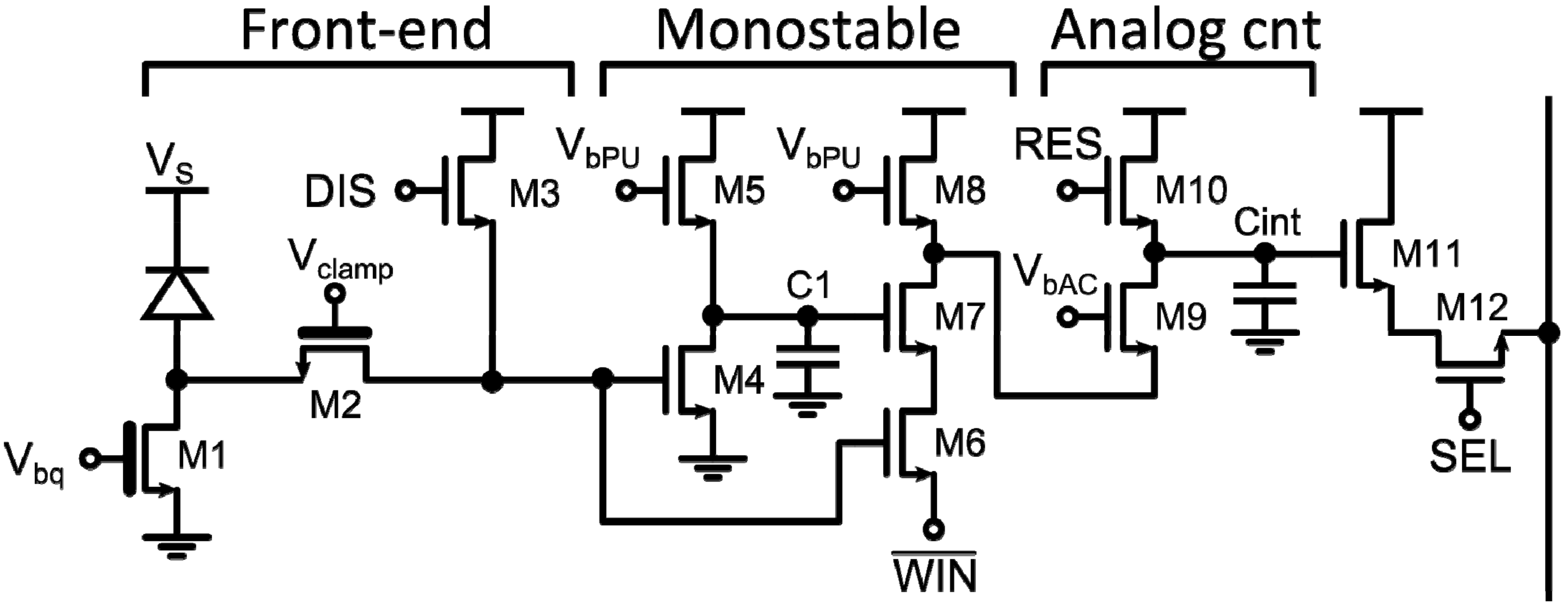
Time-Gating for FLIM **[Pancheri2009]**

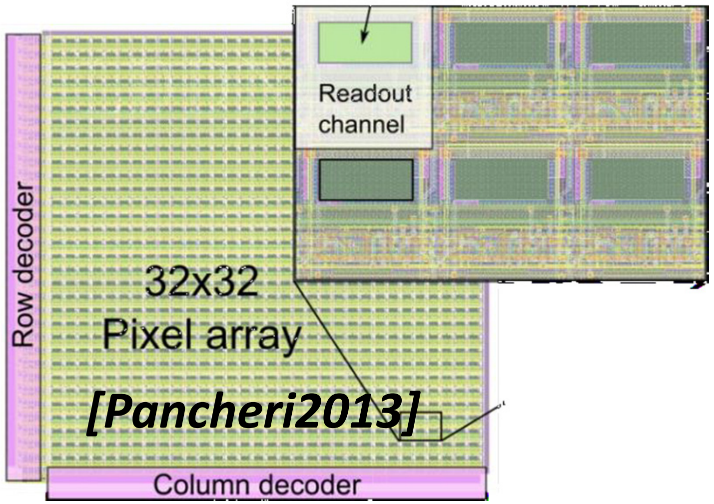




- 64x4 Time-gated linear array
  - 4-SPAD compression (1/4 deadtime)
  - 4 gating windows down to 1ns width
  - 36% fill-factor (linear array), 0.35µm tech
  - High-speed readout: 320Mbps

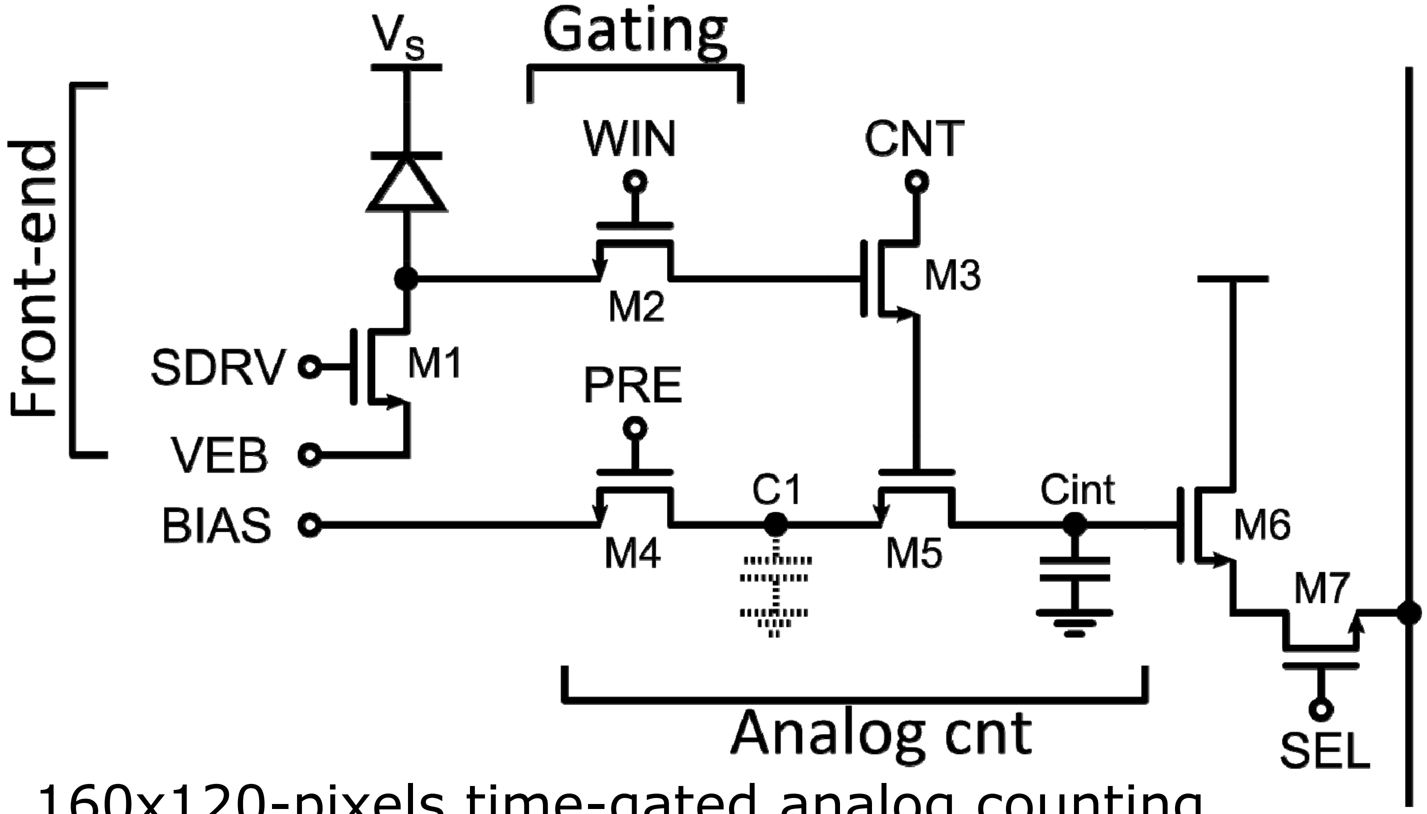
Time-Gating for FLIM **[Pancheri2013]**

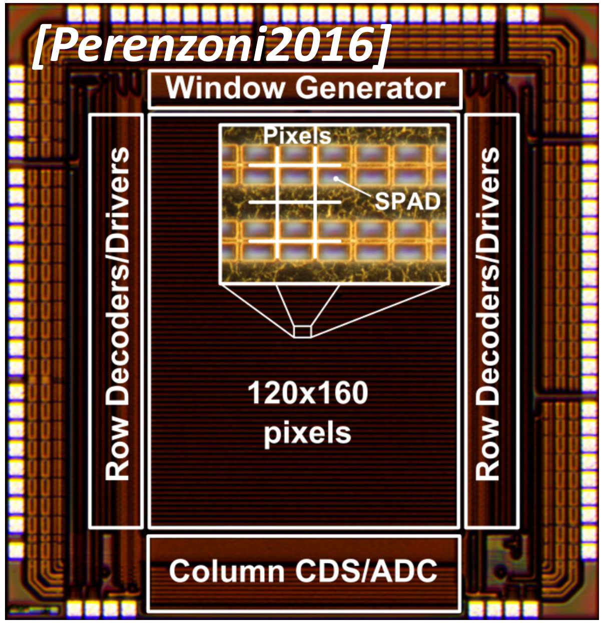




- 32x32-pixels time-gated analog counting
  - 25µm pixel pitch, 20.8% fill-factor, 0.35µm tech
  - Time-gating down to 1.1ns
  - Repetition rate up to 80MHz

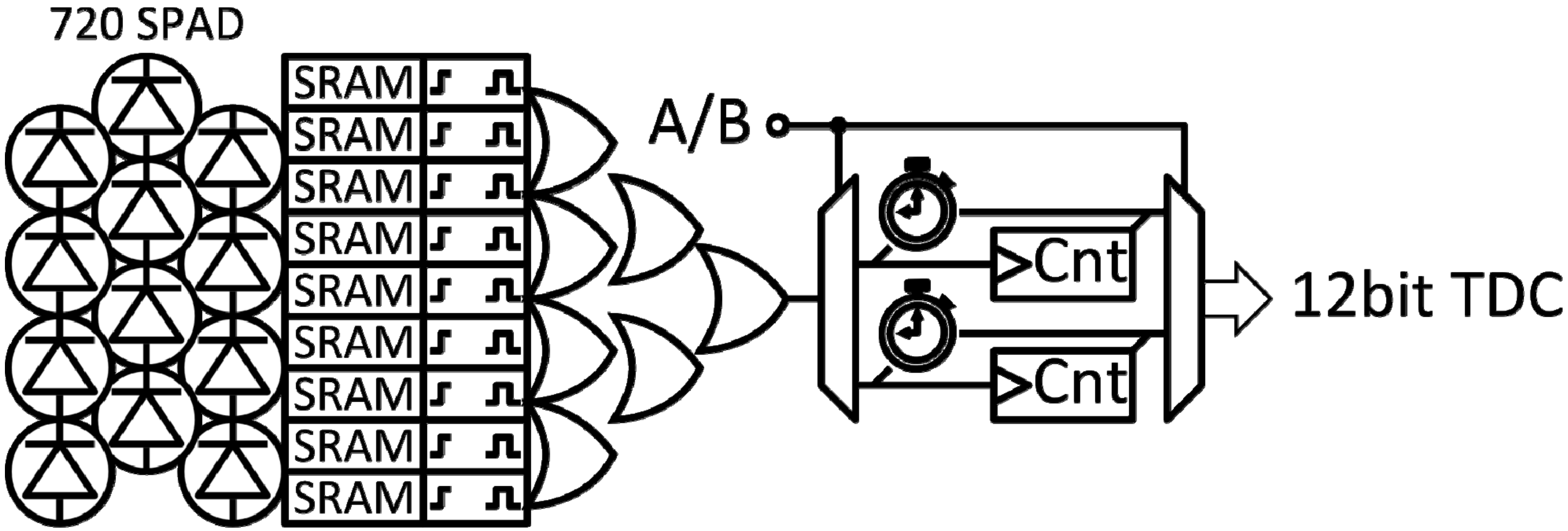
Time-Gating for FLIM **[Perenzoni2016B]**

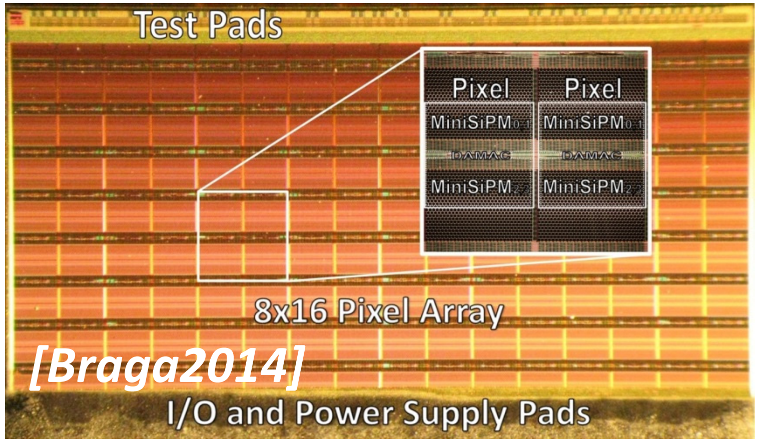




- 160x120-pixels time-gated analog counting
  - 15µm pixel pitch, 21% fill-factor, 0.35µm tech
  - Time-gating down to 750ps
  - Self-referenced A/D conversion with non-uniformity cancellation

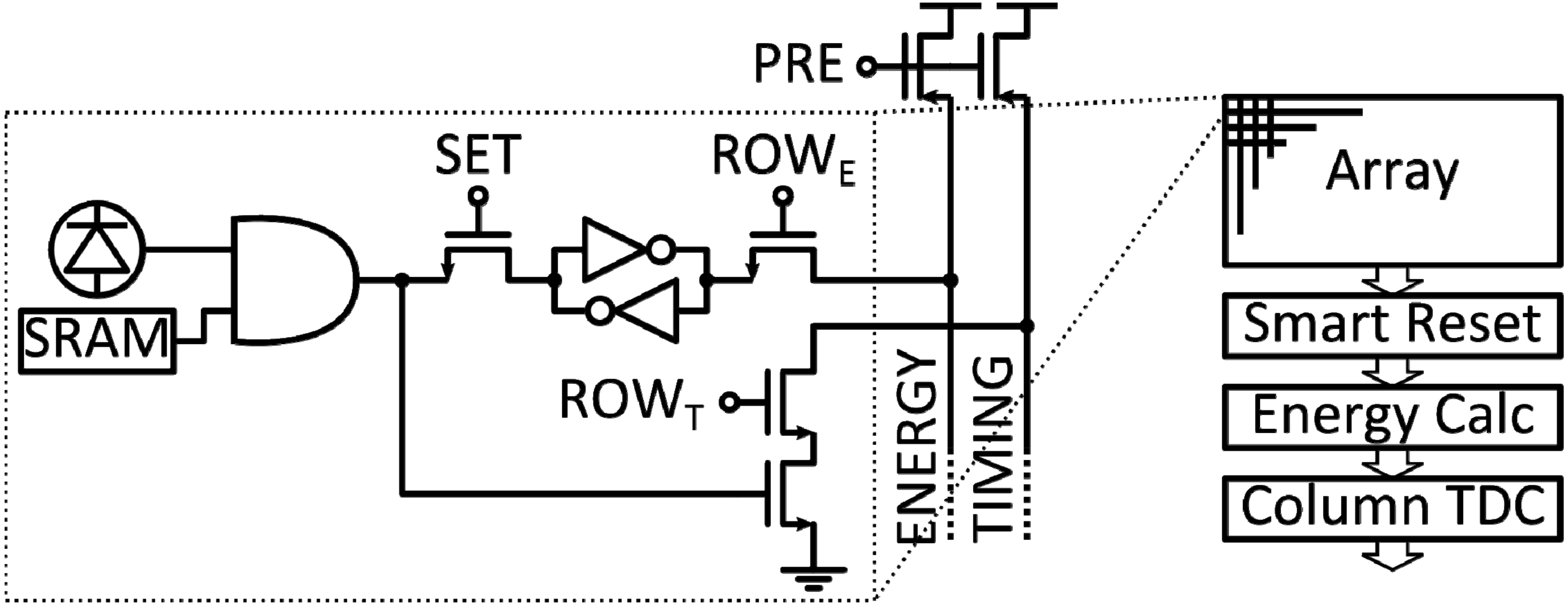
Digital SiPM for PET **[Braga2014]**

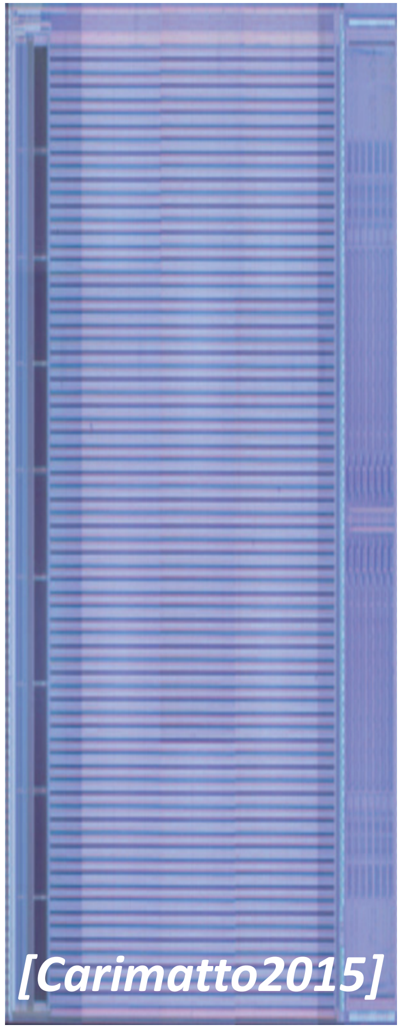




- 16x8 digital-SiPM with energy and timing information
  - 720-SPAD dSiPM per-pixel, 43% fill-factor (whole array)
  - Global 100MHz energy (counts) output
  - Per-pixel timestamping (64ps), energy
  - On-chip validation algorithm for events with energy above threshold

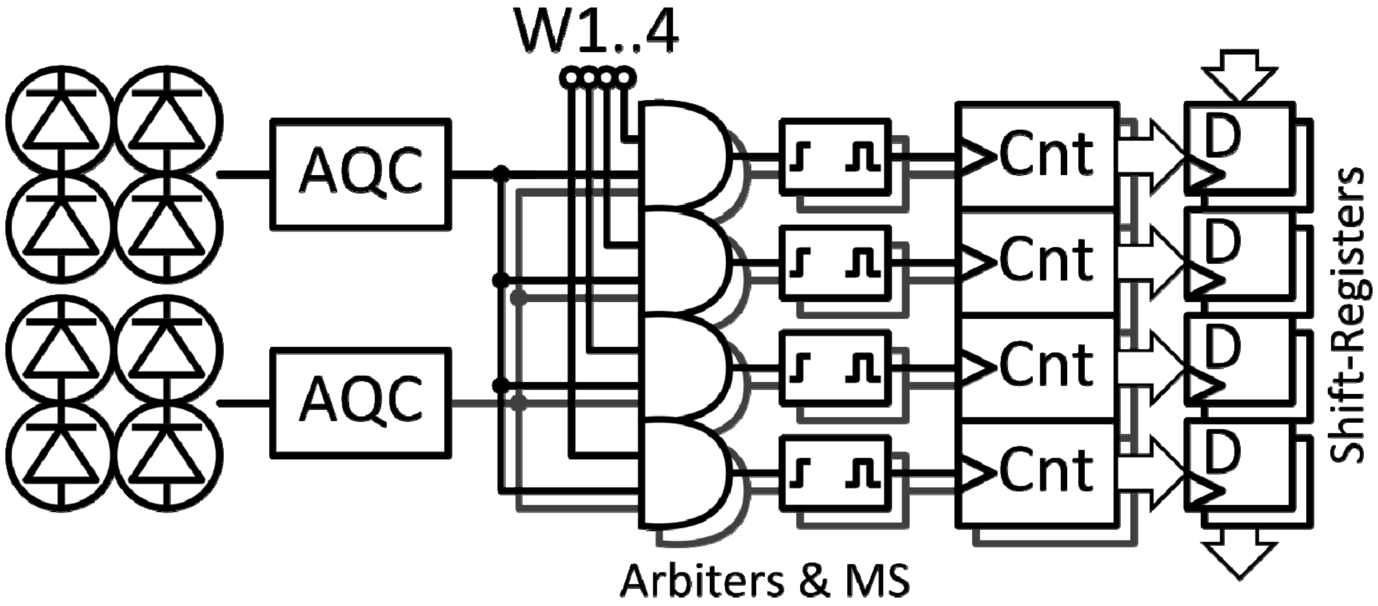
Digital SiPM for PET **[Carimatto2015]**

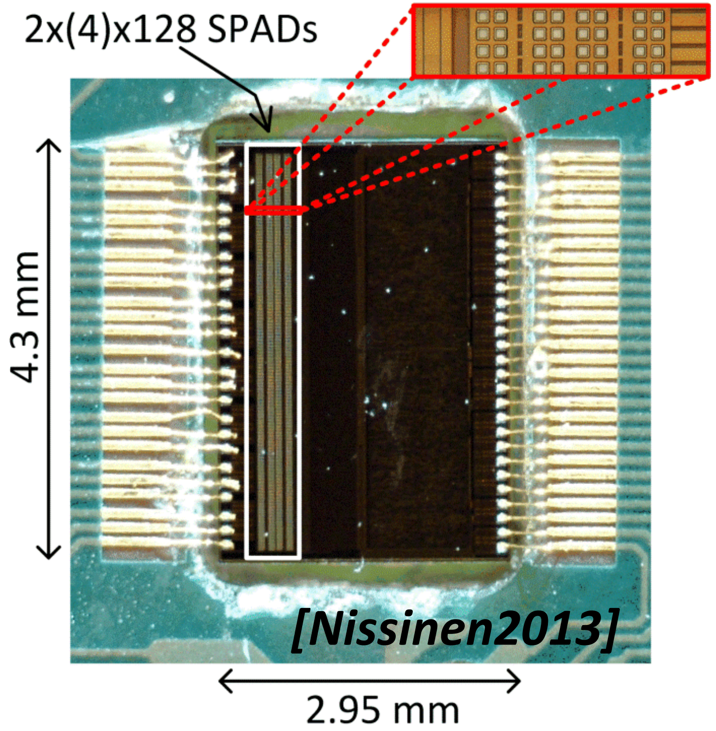




- 9x18 multichannel (16x26) d-SiPM with energy and timestamping
  - Smart-reset technique for reuse of column TDCs
  - Per-column timestamping (48.5ps TDC)
  - 57% fill-factor (single SiPM), 0.35 tech

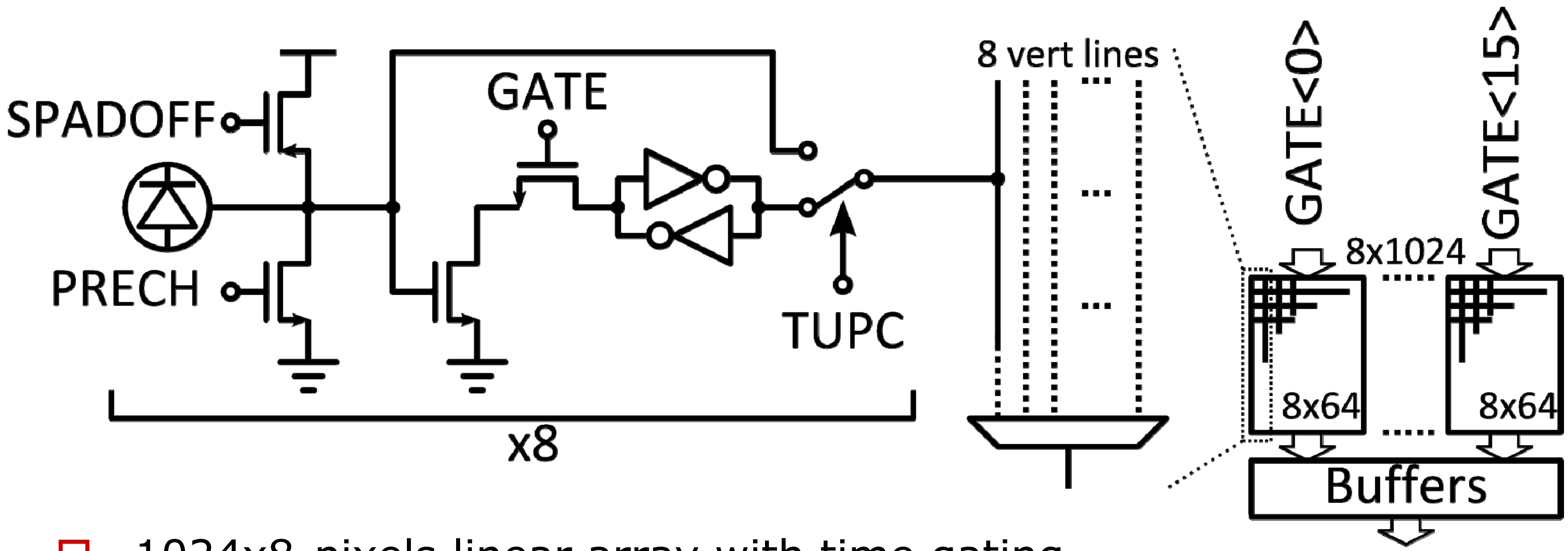
Raman Spectroscopy [Nissinen2013]



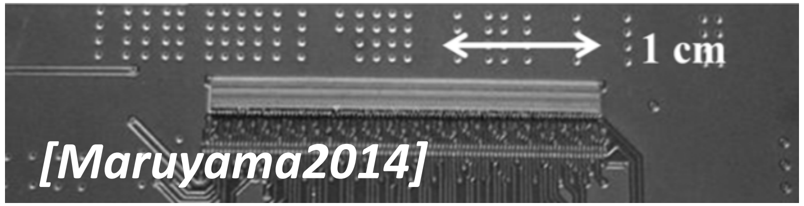


- 2x(4)x128 linear array with time gating and counting
  - Active quenching for short deadtime
  - Very precise time-gating (100ps) with low skew (±35ps)
  - Linear array fill-factor 23%

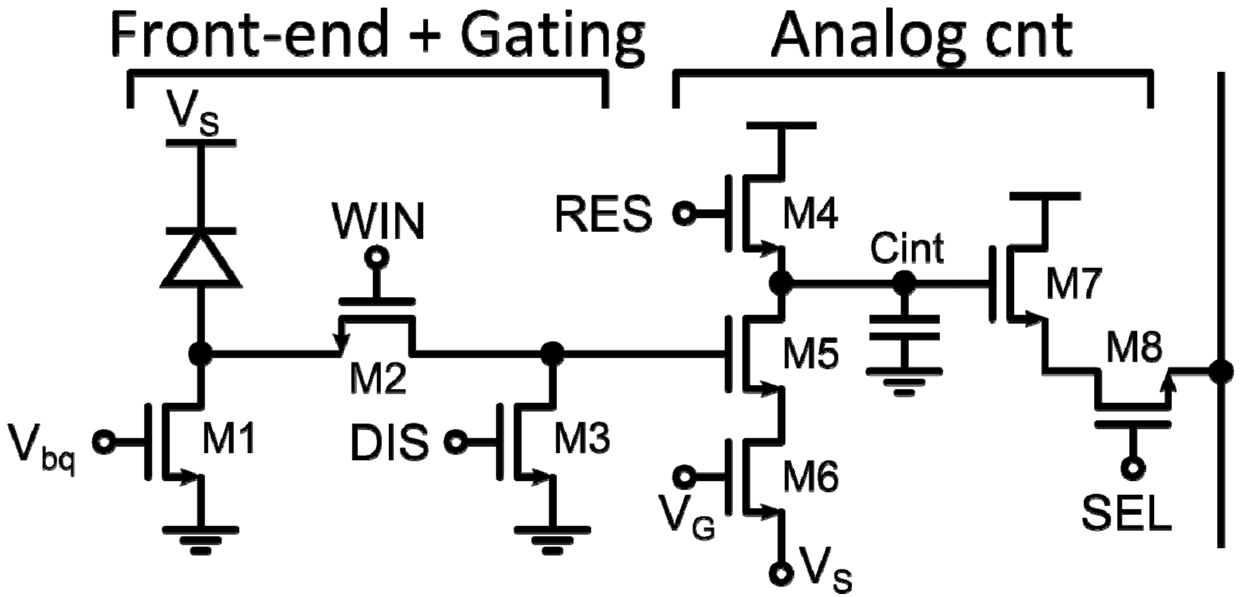
Raman Spectroscopy **[Maruyama2014]**

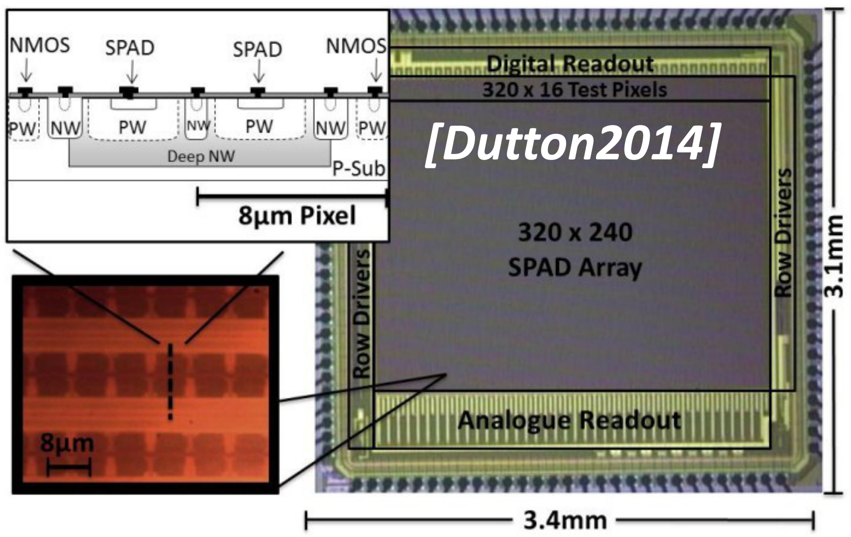


- 1024x8-pixels linear array with time gating
  - Short time gate down to 700ps
  - One-bit counter per pixel, external readout
  - 24μm pixel pitch, 44.3% fill-factor



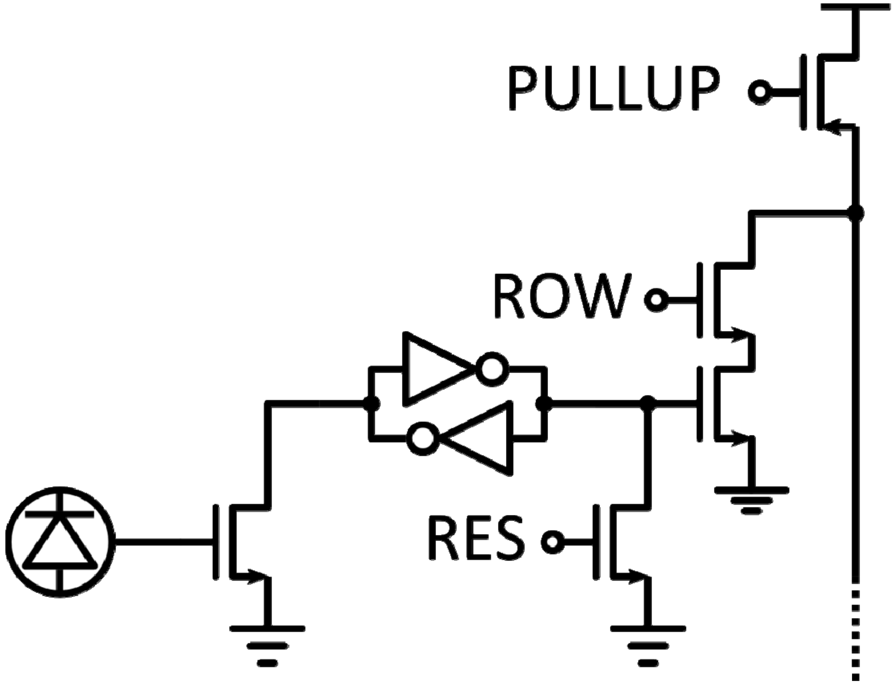
Quanta Image Sensor **[Dutton2014]**

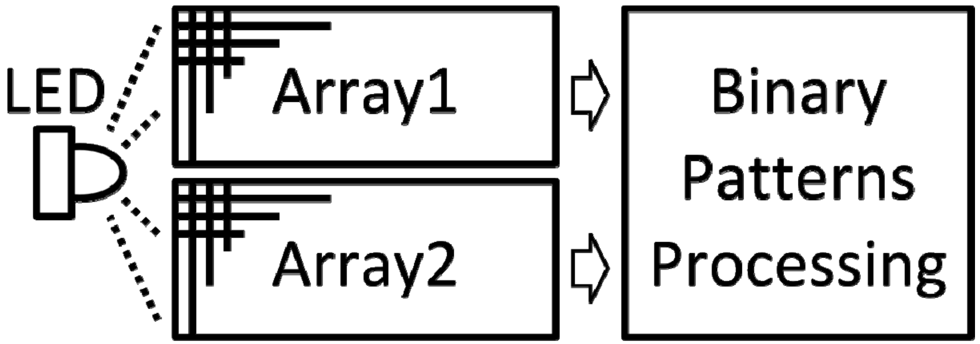


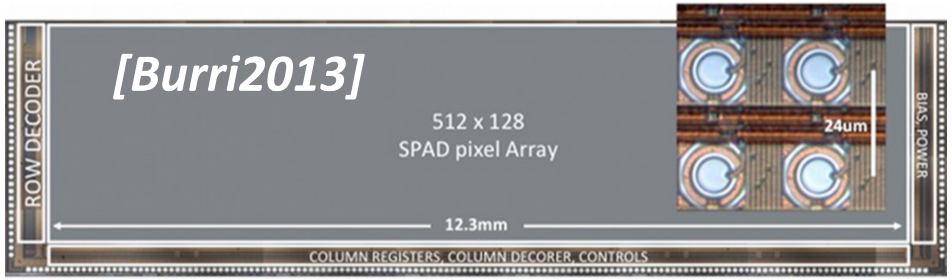


- 320x240-pixels quanta image sensor (QIS)
  - Analog counting or single digital count (“jot”)
  - 8µm pixel pitch, 26.8% fill-factor, 0.13µm tech
  - 5.14kfps readout at 1bit/pixel

Quantum RNG **[Burri2013]**

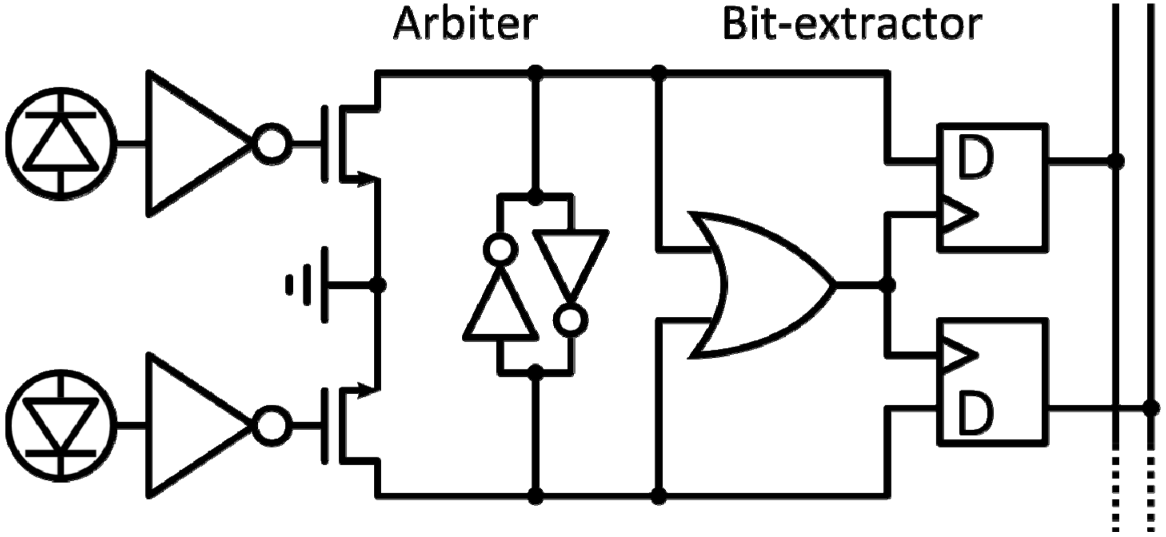


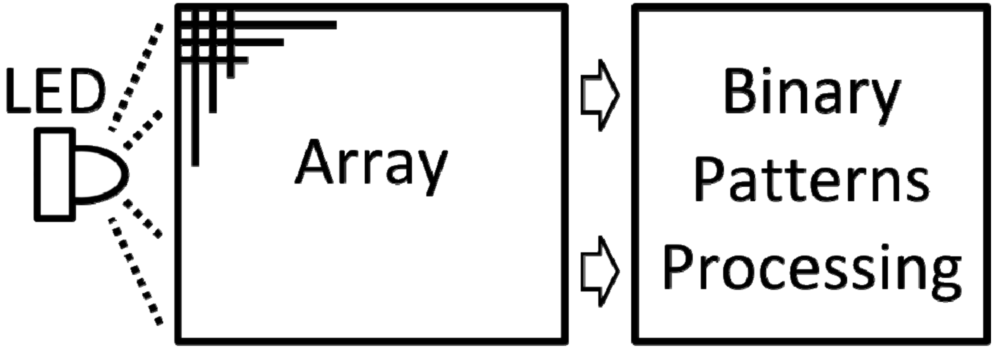




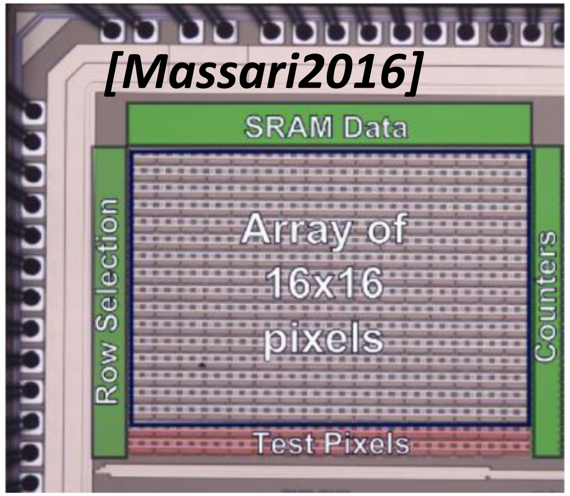
- 512x128 (x2) SPAD array used for random number generation
  - Simple 1-bit pixel
  - Illumination of two arrays (to ensure 50% probability of patterns)
  - Readout of binary patterns, 5Gbit/s QRNG generation

Quantum RNG **[Massari2016]**

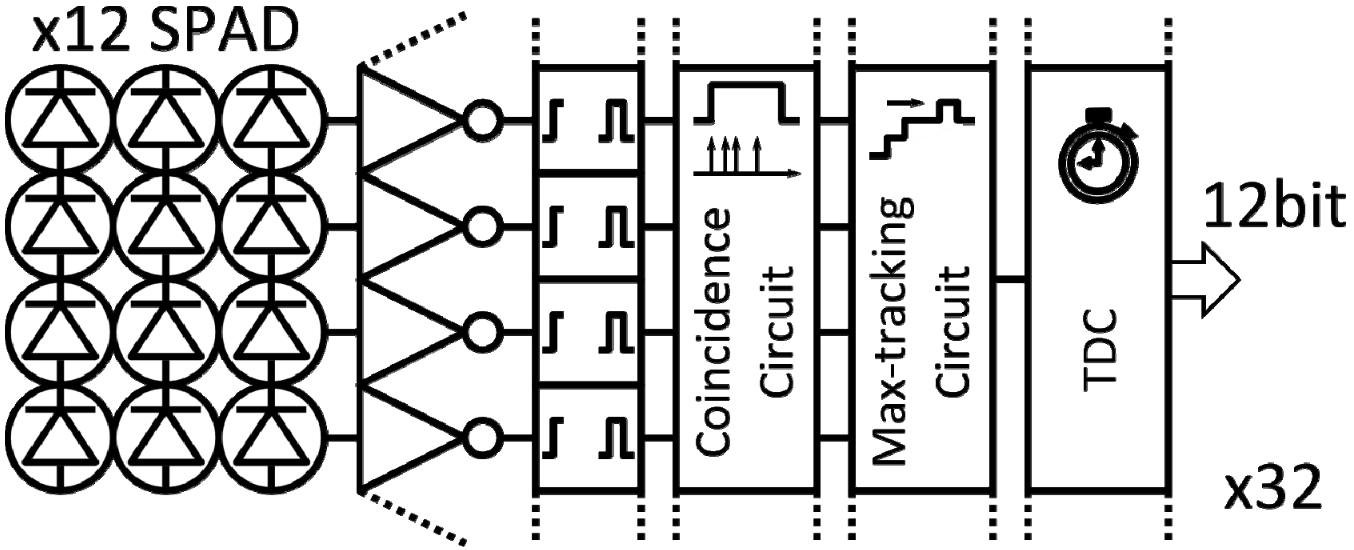


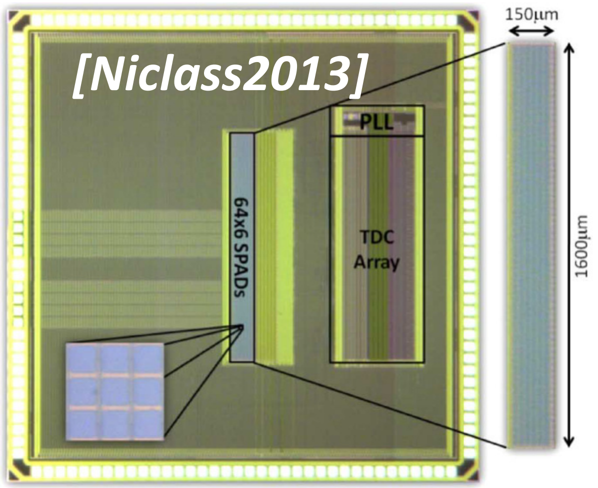


- 16x16-pixels with arbiter for QRNG
  - Dual-SPAD pixel with time arbiter (first photon wins)
  - Low sensitivity to temperature: -6.7ppm/deg
  - Wide illumination range up to 74dB
  - 128Mbps QRNG generation rate



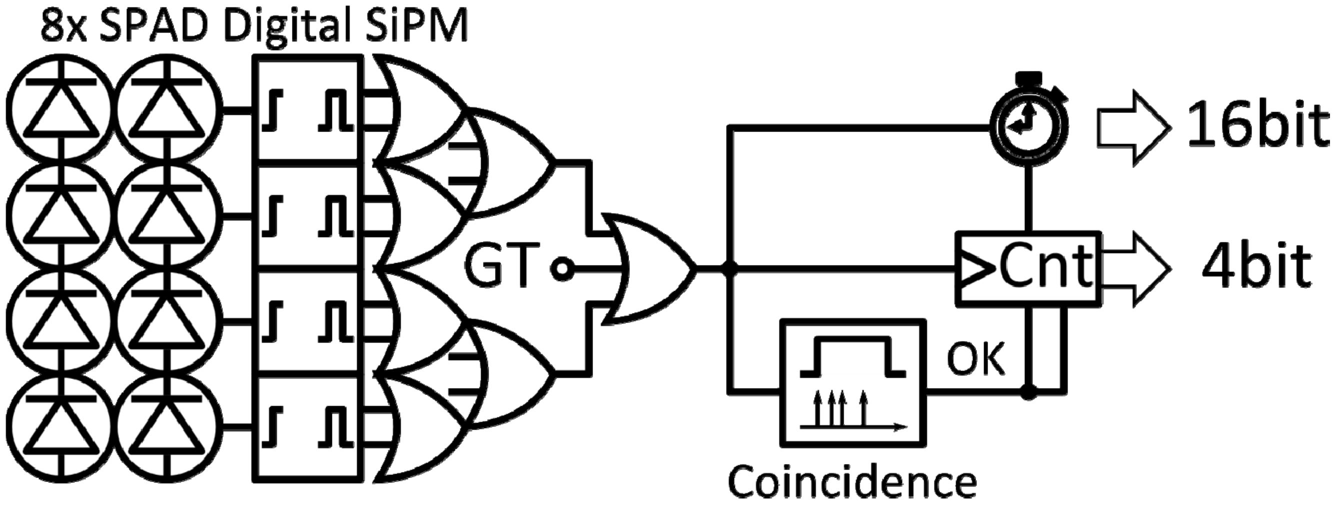
Time-of-Flight Imaging **[Niclass2013]**

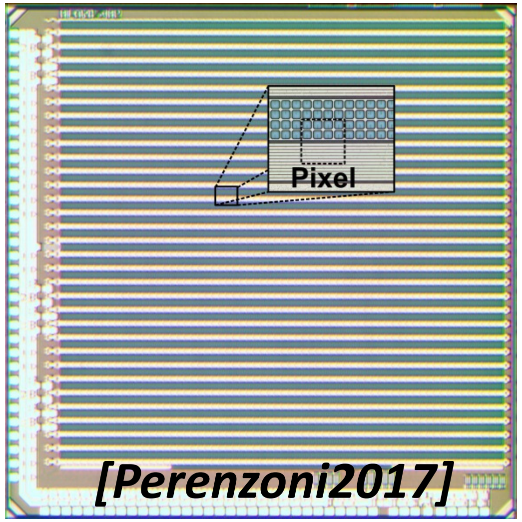




- 64x6-pixel for scanning lidar with photon correlation
  - Up to 100m with 0.1% relative precision
  - 70% fill-factor (linear array), 0.18μm tech
  - 10fps at 340x96 pixel image resolution
  - Rejection up to 80klux background light

Time-of-Flight Imaging [Perenzoni2017]





- 64x64-pixel for flash lidar with photon correlation
  - Up to 300m (imaging) and 6km (single point) with 0.14% relative precision
  - 60μm pixel pitch, 26.5% fill-factor, 0.15μm tech
  - 17.9kHz single timestamp frame rate
  - Rejection up to 100Mph/s/pix background light

Summary

- Single-photon avalanche diode in CMOS allow
  - Photon counting
  - Photon timestamping

- The "SPAD Toolbox" includes
  - Various front-end circuits
  - Processing blocks
  - Specific architectures

- Time resolution and sensitivity + CMOS processing
  - New applications and excellent performance!

SPAD-related Papers at ISSCC 2018

- **Session 5: Image Sensors (Monday),** afternoon:
- 5.7 A 20ch TDC/ADC Hybrid SoC for 240×96-Pixel 10%-Reflection <0.125%-Precision 200m-Range Imaging LiDAR with Smart Accumulation Technique DS1
- 5.9 A 256×256 45/65nm 3D-Stacked SPAD-Based Direct TOF Image Sensor for LiDAR Applications with Optical Polar Modulation for up to 18.6dB Interference Suppression
- 5.10 A 32×32-Pixel Time-Resolved Single-Photon Image Sensor with 44.64μm Pitch and 19.48% Fill-Factor with On-Chip Row/Frame Skipping Features Reaching 800kHz Observation Rate for Quantum Physics Applications DS1

References (I)

- [McIntyre1972] R.J. McIntyre, "The distribution of gains in uniformly multiplying avalanche photodiodes: Theory", IEEE TED, 1972.
- [Xu2017] H. Xu, et al, "Design and characterization of a p+/n-well SPAD array in 150nm CMOS process", Optics Express, 2017.
- [AlAbbas2017] T. Al Abbas, et al, "8.25µm Pitch 66% Fill Factor Global Shared Well SPAD Image Sensor in 40nm CMOS FSI Technology", IISW, 2017.
- [Cova1996] S. Cova, "Avalanche photodiodes and quenching circuits for single-photon detection", Applied Optics, 1996
- [Piemonte2016] C. Piemonte, et al, "Performance of NUV-HD Silicon Photomultiplier Technology", IEEE TED, 2016.
- [Braga2014] L.H.C. Braga, et al, "A fully digital 8×16 SiPM array for PET applications with per-pixel TDCs and real-time energy output", IEEE JSSC, 2014.
- [Perenzoni2016A] M. Perenzoni, et al, "Compact SPAD-based pixel architectures for time-resolved image sensors", MDPI Sensors, 2016.
- [Perenzoni2016B] M. Perenzoni, et al, "A 160 x 120 Pixel Analog-Counting Single-Photon Imager With Time-Gating and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging", IEEE JSSC, 2016.
- [Pancheri2013] L. Pancheri, et al, "SPAD image sensor with analog counting pixel for time-resolved fluorescence detection", IEEE TED, 2013.
- [Stoppa2009] D. Stoppa, et al, "A 32x32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain", ESSCIRC, 2009.
- [Veerappan2011] C. Veerappan, et al, "A 160×128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," ISSCC, 2011.

References (II)

- [Villa2014] F. Villa, et al, "CMOS Imager with 1024 SPADs and TDCs for Single-Photon Timing and 3-D Time-of-Flight", IEEE JSTQE, 2014.
- [Pancheri2009] L. Pancheri, et al, "A SPAD-based pixel linear array for high-speed time-gated fluorescence lifetime imaging", ESSCIRC, 2009.
- [Carimatto2015] A. Carimatto, et al, "A 67,392-SPAD PVTB-compensated multi-channel digital SiPM with 432 column-parallel 48ps 17b TDCs for endoscopic time-of-flight PET", ISSCC, 2015
- [Nissinen2013] I. Nissinen, et al, "2×(4×)128 time-gated CMOS single photon avalanche diode line detector with 100 ps resolution for Raman spectroscopy", ESSCIRC, 2013.
- [Maruyama2014] Y. Maruyama, et al, "A 1024x8, 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS", IEEE JSSC, 2014
- [Dutton2014] N.A.W. Dutton, et al, "320x240 oversampled digital single photon counting image sensor", IEEE VLSI Symp, 2014
- [Burri2013] S. Burri, et al, "Jailbreak Imagers: Transforming a Single-Photon Image Sensor into a True Random Number Generator", IISW, 2013.
- [Massari2016] N. Massari, et al, "A 16x16 pixels SPAD-based 128-Mb/s quantum random number generator with -74dB light rejection ratio and -6.7ppm/C bias sensitivity on temperature", ISSCC, 2016.
- [Niclass2013] C. Niclass, et al, "A 100-m Range 10-Frame/s 340x96-Pixel Time-of-Flight Depth Sensor in 0.18-um CMOS", IEEE JSSC, 2013.
- [Perenzoni2017] M. Perenzoni, et al, "A 64x64-Pixels Digital Silicon Photomultiplier Direct TOF Sensor With 100-MPhotons/s/pixel Background Rejection and Imaging/Altimeter Mode With 0.14% Precision Up To 6 km for Spacecraft Navigation and Landing", IEEE JSSC, 2017.