The SiPM Physics and Technology - a Review -

G. Collazuol
Department of Physics and Astronomy, University of Padova and INFN

Overview

- Introduction
  - Key physics and technology features
    - I-V characteristics
    - Device response
  - Noises
    - Photo-detection efficiency
    - Timing properties
- Summary and Future
The silicon PM: array of GM-APD

Single GM-APD gives no information on light intensity → MATRIX structure first proposed in the late '80-ies by Golovin and Sadygov

A SiPM is segmented in tiny GM-APD cells and connected in parallel through a decoupling resistor, which is also used for quenching avalanches in the cells.

Each element is independent and gives the same signal when fired by a photon.

In principle output charge is proportional to the number of incident photons.

\[ Q = Q_1 + Q_2 = 2Q_1 \]

Σ digital signals → analog signal !!!
A bit of history

Pioneering work since late 80-ies at Russian institutes

Investigations of various multi-layer silicon structures with local micro-plasma suppression effect to develop low-cost GM-APD arrays

Early devices ageing quickly, unstable, noisy

Dolgoshein - MePhi/Pulsar (Moscow)
Poly-silicon resistor

- Low fill-factor
- Simple fabrication technology

e.g., Dolgoshein, NIMA 563 (2006)

Sadygov – JINR/Micron (Dubna)
Avalanche Micro-channel/pixel Photo Diodes (AMPD)

- high PDE
- very high density of micro-cells

e.g. Sadygov, NIMA 567 (2006)

Golovin - Obninsk/CPTA (Moscow)
Metal-Resistive-Semiconductor (MRS)

- High fill factor
- Good pixel to pixel uniformity

e.g., Golovin NIMA 539 (2005)
Today

Many institutes/companies are involved in SiPM development/production:

- CPTA, Moscow, Russia
- MePhi/Pulsar Enterprise, Moscow, Russia
- Zecotek, Vancouver, Canada
- Hamamatsu HPK, Hamamatsu, Japan
- FBK-AdvanSiD, Trento, Italy
- ST Microelectronics, Catania, Italy
- Amplification Technologies Orlando, USA
- SensL, Cork, Ireland
- MPI-HLL, Munich, Germany
- RMD, Boston, USA
- Philips, Aachen, Germany
- Excelitas tech. (formerly Perkin-Elmer)
- KETEK, Munich, Germany
- National Nano Fab Center, Korea
- Novel Device Laboratory (NDL), Beijing, China
- E2V
- CSEM
- Amplification Technologies (DAPD)
Physics & Technology

Key features

- Closeup of a cell – Custom vs CMOS
- Guard Ring and Optical isolation
- Operation principles of GM-APD and quenching modes
**Shallow-Junction APD**

Example of implementation

*C. Piemonte NIM A 568 (2006) 224*

- **Optical window**
  - note: light absorption in Si, SiO₂

**Substrate**
- low resistivity contact (500 µm)

- **(fully) depleted region** (4 µm)

- **Shallow n⁺ layer** (0.1 µm)

- **n⁺ on p abrupt junction structure**
- **Anti-reflective coating (ARC)**
- **Very thin (100nm) n⁺ layer: “low” doping**
  - minimize Auger and SHR recombination
- **Thin high-field region: “high” doping p layer**
  - limited by tunneling breakdown
  - fixes $V_{BD}$ junction well below $V_{BD}$ at edge

- **$R_Q$ by poly-silicon**
- **Trenches for optical insulation (cross-talk)**
- **Fill factor: 20% - 80%**

**Optimization for blue light (420nm)**

**Critical region:**
- Leakage current
- Surface charges
- **Guard Ring** for
  - preventing early edge-breakdown
  - isolating cells
  - tuning E field shape
  - impact on Fill Factor

**Active volume**
- no micro-plasma's
- high quality epitaxial doping / E field profile engineering
CMOS vs Custom processes

“Standard” CMOS processes
- shallow implant depths
- high doping concentrations
- shallow trench isolation (STI)
- deep well implants (flash extension)
- no extra gettering and high T annealing
- non optimized optical stacks
- design rule restrictions

- high E field (low $V_{bd}$)
- tunneling
- lattice stress (defects/traps)
- high DCR
- limited PDE (often p-on-n)
- limited timing performances (long diffusion tails)

Recent progresses in CMOS APDs due to:
1) high voltage (flash) extension often available in standard processes
   - deep wells (needed for the high voltages used in flash memories)

2) Additional processes (custom) available:
   - buried implants
   - deep trench isolation
   - optical stack optimization

Key elements for CMOS SiPMs
- APD cell isolation from CMOS circuitry
- guard ring (again)
Close up of a CMOS cell

**APD integration into CMOS**

Example of implementation  
*T. Frach in US patent 2010/0127314*

Note
- extended CMOS processes exploited
- careful design of cell isolation and guard ring

**shallow isolation**  
(STI/LOCOS)

- optical window
- anode (p+)

**contact with buried layer**

- epitaxial n (active region)
- buried n (isolation layer)
- epitaxial p

**deep isolation trench**  
(oxide/polysilicon filling)

**buried isolation layer**  
(also protection from substrate radiation induced carriers)

**APD cell isolated by multiple wells**
from CMOS circuitry

- Example of NMOS FET of the RO electronics

- Example of implementation

- substrate (gettering sites)

- Note
  - extended CMOS processes exploited
  - careful design of cell isolation and guard ring

G. Collazuoli - PhotoDet 2012
The Guard Ring structure

- high E field structure, not uniformly distributed
- neutral region (in timing tails)
- limited fill factor
- alternative to Diffused GR
- difficult to implement
- developed by S. Cova and coll. (fully custom)
- state of the art for SPAD timing and PDE (red enhanced)

"enhanced mode structure"

Diffused GR

Virtual GR

Merged Implant GR

Gate bias / Floating GR

"double epitaxy structure"

Timing optimized GR

Shallow Trench Isol. STI GR

- well tuned high E field structure
- no additional neutral regions
- fill factor less limited
- less commonly exploited
- careful modeling required
- physically blocks and confines the high E field in active region
- might cause high DCR due to tunneling and etching induced defects/traps

Guard Ring structures in SiPM

Sul et al, IEEE EDL 31 2010 “G.R. Structures for SiPM”

Virtual guard ring most often used

Maresca et al. Proc. of SPIE Vol. 8072 “Floating field ring ... to enhance fill factor of SiPM”
Operation principle of a GM-APD

Avalanche processes in semiconductors are studied in detail since the '60 for modeling micro-plasma instabilities

*McIntyre JAP 32 (1961), Haitz JAP 35 (1964) and Ruegg IEEE TED 14 (1967)*

**OFF condition:** avalanche quenched, switch open, capacitance charged until no current flowing from $V_{bd}$ to $V_{BIAS}$ with time constant $R_q \times C_d = \tau_{quenching}$ (→ recovery time)

**ON condition:** avalanche triggered, switch closed $C_d$ discharges to $V_{bd}$ with a time constant $R_d \times C_d = \tau_{discharge}$, at the same time the external current asymptotic grows to $(V_{bias} - V_{bd})/(R_q + R_d)$

$P_{01} =$ turn-on probability
probability that a carrier traversing the high-field region triggers the avalanche

$P_{10} =$ turn-off probability
probability that the number of carriers traversing the high-field region fluctuates to 0

Fig. 3. Shape of current pulse for $\theta_d \ll \tau_1(I_0)$. 
Passive Quenching

If $R_Q$ is high enough the internal current is so low that statistical fluctuations may quench the avalanche:

$$\tau_d = R_s C_d << R_q C_d = \tau_q$$

2. turn-off mean time is very short (if $R_q$ is sufficiently high, $I_{latch} \sim 20 \mu A$)

The charge collected per event is the area under the exponential which is determined by circuital elements and bias.

It is possible to define a GAIN (discharge of a capacitor)

$$G = \frac{I_{max} \cdot \tau_q}{q_e} = \frac{(V_{bias} - V_{bd}) \cdot \tau_q}{(R_q + R_s) \cdot q_e} = \frac{(V_{bias} - V_{bd}) \cdot C_d}{q_e}$$

Gain fluctuations in GM-APD are smaller than in APD essentially because electrons and holes give the same signal.
Passive Quenching Regime

Proper value of quenching resistance $R_q$ is crucial to let the internal current decrease to a level such that statistical fluctuations may quench the avalanche $\rightarrow$ sub-ns quenching time $\rightarrow$ crucial to have well defined gain

Given $R_q$ the proper quenching regime is for $\Delta V$ in the range:

$$0 < \Delta V < R_q I_{latch}$$

where as a rule of thumb $I_{latch} \sim 20\mu A \rightarrow \Delta V_{max} \sim$ a few Volts (typically)
Operative $\Delta V$ Range – $I_{\text{dark}}/\text{DCR}$

Operative $\Delta V$ limited by:
1) $I_{\text{latch}} \sim 20\mu A \rightarrow \Delta V < I_{\text{latch}} R_q$ (non-quenching regime)
2) Dark Count Rate (DCR) acceptable level $\leftarrow$ PDE vs $\Delta V$ $\leftarrow$ E field shape
3) $V_{\text{bd}}$ edge breakdown (usually some 10V above $V_{\text{bd}}$)

A practical method for estimating the operative range (limited by effectc
1) is to measure the ratio $R_I$ of the measured dark current $I_D$ to the dark
current $I'_D$ calculated from the measured dark rate and pixel count
spectra:

$$R_I = \frac{I_D}{I'_D} = \frac{1}{\text{DCR} \cdot \bar{N} \cdot G \cdot q_e}$$

where $\bar{N}$ is the average $N$ of fired cells

Non-quenching regime for values of $\Delta V$
when $R_I$ deviates significantly from 1

Jendrysik et al suggest
$R_I = 2$ as reasonable threshold

after Jendrysik et al NIM A 2011
doi:10.1016/j.nima.2011.10.007
Passive Quenching (Resistive)

1) common solution: poly-silicon

2) alternative: metal thin film
   → higher fill factor
   → milder T dependence

3) alternative principle: bulk integrate resistor
   → flat optical window → simpler ARC
   → fully active entrance window
      → high fill factor (constraints only from guard ring and X-talk)
   → diffusion barrier against minorities
      → less X-talk
   → positive T coeff. (R~ T\(^{+2.4}\))
   → production process simplified → cost

Ninkovic et al NIM A610 (2009) 142
and NIM A628 (2011) 407
Richter et al US patent № 2011/0095388

pro
contra

Rq matching constraints
cells' pitch/wafer thickness

vertical R is JFET
→ non-linear I-V
→ long recovery

Nagano IEEE NSS-MIC 2011

MPI HLL

NDL SiPM

Zhang et al NIM A621 (2010) 116
Passive Quenching (Capacitive)

Quenching feedback due to charge accumulated by means by semiconductor barriers

Amplification Technologies
Shushakov et al US Patents
№ 2004/6885827 and № 2011/7899339

Zecotek
Sadygov et al arXiv 1001.3050
Sadygov RU Patents № 1996/2102820 and № 2006/2316848

Note: induced signal is fast (ns) but recovery quite slow (ms) (non exponential)

a) avalanche at internal high field regions
b) charges accumulated in isolated potential wells
   → E field reduced (locally) → avalanche quenched
   → Fast signal induced (capacitive) outside

c) potential wells discharge slowly by tunneling
   (discharge must be delayed for good quenching)
   → high E field recovered
Active Quenching

Basic circuit elements:
1) quench circuit to detect and stop the avalanche and restore bias conditions
2) buffer (low capacitive load) for isolating the APD from the external electronics capacitance

Configuration with anode to ground potential is best: only $C_{\text{det}}$ is involved → minimum RC load
→ minimum quenching dead-time
→ minimum charge flow in APD (less after-pulses)

(in addition n-well regions (cathode) can be shared among many cells)

Note: use of PMOS to minimize the area wrt NMOS for the same target quenching resistance

buffer → simple inverter as input signal is already digital

Cell electronics area: 120$\mu$m$^2$
25 transistors including 6T SRAM
~6% of total cell area
Modified 0.18$\mu$m 5M CMOS
Foundry: NXP Nijmegen

Cell area ~ 30x50$\mu$m$^2$
Fill Factor ~ 50%

T.Frach at LIGHT 2011
I-V characteristics

- Information from Forward current → - Rq
  - junction Temperature
  ...

- Information from Reverse current → - breakdown $V_{bd}$
  - T coefficient
  ...

G.Collazuol - PhotoDet 2012
I-V characterization: forward bias

1. **Forward current**
   \[ I_{\text{forward}} \sim C(\eta) A(T) \left[ \exp \left( \frac{q V_d}{\eta k T} \right) - 1 \right] \]

   - \( \eta \sim 1 \)

   - **Ohmic behavior at high current**
   - Linear fit \( R_{\text{series}} \sim \frac{R_q}{N_{\text{cells}}} \)

2. **Voltage drop** \( (V_d) \) decreases linearly with \( T \) decreasing (e.g. at 1\( \mu \)A)

3. **FBK devices**

---

**Shockley et al. Proc. IRE 45 (1957)**

**Ideality factor**
- Diffusion current dominating: \( \eta \rightarrow 1 \)
- Recombination current dominating: \( \eta \rightarrow 2 \)
Forward I-V → Junction Temperature probe

Voltage drop at fixed forward current → precise measurement of junction T...

... otherwise not trivially measured!

for $T \to 0$ ideally $V_d \to E_g$

(freeze-out effects apart)

V _drop (mV)

constant current injection $I_{\text{forward}} = 1\mu$A

$V_d = \frac{E_g}{q} - \frac{\eta k T}{q} \ln \frac{C(\eta) A(T)}{I_{\text{forward}}}$

• (almost) linear dependence with slope $dV_{\text{drop}}/dT|_{1\mu A} \sim -3\text{mV/K}^T$ (K)

(we don't see freeze-out effects down to 50K)

• direct and precise calibration/probe of junction(s) Temperature
Two ways for measuring series resistance ($R_s$)

1) Fit at high $V$ of forward characteristic
2) Exponential recovery time (afterpulses envelope)

Measurements (1) and (2) consistent
→ dominant effect from quenching resistor $R_q$
(→ series $R$ bulk gives smaller contribution)

Empirical fit:
$$R_q(T) \sim 0.13 \left(1 + \frac{300}{T} e^{\frac{300}{T}}\right) M \Omega$$

After-pulse envelope

Note: SiPM for low $T$ applications must have appropriate quenching $R$ (not quenching at room $T$ !)
Reverse I-V

Breakdown voltage decreases at low T due to larger carriers mobility → larger ionization rate (electric E field fixed)

Fit: linear + quadratic (V > V_{breakdown})

At high T ~80 mV/K (fit above 240K)

V_{bd} dependence on T

Breakdown voltage decreases at low T due to larger carriers mobility → larger ionization rate (electric E field fixed)
$V_{bd}$ vs $T \rightarrow T$ coefficient ($\Delta V$ stability)

**Breakdown Voltage**

- $V_{br} (V)$ measured by fitting single p.e. charge vs bias voltage (pulsed mode)

**Temperature coefficient**

- $\Delta V_{br} / V_{br} / \Delta T \sim 0.20 \% / K$
- $\Delta V_{br} / V_{br} / \Delta T \sim 0.25 \% / K$

FBK device

Improved stability at low $T$

HPK device (400 pixels)

**Fig. 6.** Breakdown voltage as a function of temperature of the MPPC with 400 pixels.
Depletion layer → $V_{bd}$ dependence on $T$

*Serra et. al. (FBK) IEEE TNS 58 (2011) 1233*

“Experimental and TCAD Study of Breakdown Voltage Temperature Behavior in n+/p SiPMs”

Note: precise agreement simulation/data is not trivial at all. Definition of ionization coefficients is device dependent...

**Narrow depletion layer** (high background doping*) or thin epitaxial layer
→ minimize $V_{bd}$ dependence on $T$
→ gain stability $\frac{\delta V_{bd}/V_{bd}}{\delta T} = \frac{\delta G/G}{\delta T}$

(*) resulting in epitaxial layer not fully depleted at $V_{bd}$

**Trade off:**
→ PDE (thickness)
→ minimum gain (capacity) against after-pulses and cross-talk
Pulse shape, Gain and Response

- Detailed electrical model
- Pulse shape
- Gain and Gain fluctuation
- Response non-linearity

(mostly for passive mode)
Basic electrical model

Fast Capacitor (cell) discharge and slow recharge (roughly speaking)

\[ i = \text{exp}(-t/\tau_d) \]

\[ 1 - \text{exp}(-t/\tau_d) \]

99% recovery time \( \sim 5 \tau_q \)

Recovery time: \( T \) dependence due to \( R_q \)
\( C_d \) is independent of \( T \)

Rise time: \( T \) dependent (to lesser extent) due to \( R_d \)

Gain \( \sim C \Delta V \rightarrow \) independent of \( T \)
at fixed Over-Voltage \( (\Delta V = V_{\text{bias}} - V_{bd}) \)
SiPM equivalent circuit (detailed model)

Single cell model \( \rightarrow (R_d \parallel C_d) + (R_q \parallel C_q) \)

SiPM + load \( \rightarrow (||Z_{cell})||C_{grid} + Z_{load} \)

Signal = slow pulse \( (\tau_d \text{ (rise)}, \tau_{slow} \text{ (fall)}) + \)
+ fast pulse \( (\tau_d \text{ (rise)}, \tau_{fast} \text{ (fall)}) \)

- \( \tau_{d \text{ (rise)}} \sim R_d (C_q + C_d) \)
- \( \tau_{\text{fast} \text{ (fall)}} = R_{\text{load}} C_{\text{tot}} \) (fast; parasitic spike)
- \( \tau_{\text{slow} \text{ (fall)}} = R_q (C_q + C_d) \) (slow; cell recovery)

S.Seifert et al. IEEE TNS 56 (2009) 3726

\( C_q \rightarrow \text{fast current supply path in the beginning of avalanche} \)

Pulse shape

- Rise: Exponential
- Fall: Sum of 2 exponentials

\[ V(t) \approx \frac{Q}{C_q + C_d} \left( \frac{C_q}{C_{\text{tot}}} e^{-t/\tau_{\text{FAST}}} + \frac{R_{\text{load}}}{R_q} \frac{C_d}{C_q + C_d} e^{-t/\tau_{\text{SLOW}}} \right) \]

for \( R_{\text{load}} \ll R_q \)

where \( Q = \Delta V (C_q + C_d) \) is the total charge released by the cell

\( \rightarrow \text{'prompt' charge on } C_{\text{tot}} \text{ is } Q_{\text{fast}} = Q \frac{C_q}{C_q + C_d} \)
Pulse shape

\[ V(t) \approx \frac{Q}{C_q + C_d} \left( C_q e^{-t/\tau_{\text{fast}}} + \frac{R_{\text{load}}}{R_q} C_d e^{-t/\tau_{\text{slow}}} \right) = \frac{Q R_{\text{load}}}{C_q + C_d} \left( C_q e^{-t/\tau_{\text{fast}}} + \frac{C_d}{\tau_{\text{slow}}} \right) \]

→ gain \quad G = \int dt \frac{V(t)}{q_e R_{\text{load}}} = \frac{Q}{q_e} = \frac{\Delta V (C_d + C_q)}{q_e} \quad \text{independent of } R_q

→ charge ratio \quad \frac{Q_{\text{slow}}}{Q_{\text{fast}}} \sim \frac{C_d}{C_q}

\[ V_{\text{max}} \quad \text{→ peak voltage on } R_{\text{load}} \quad V_{\text{max}} \sim R_{\text{load}} \left( \frac{Q_{\text{fast}}}{\tau_{\text{fast}}} + \frac{Q_{\text{slow}}}{\tau_{\text{slow}}} \right) \]

dependent on \( R_q \) (increasing with \( 1/R_q \))

\[ V_{\text{max}}^{\text{slow}} \sim \frac{C_d C_{\text{tot}} R_{\text{load}}}{C_q^2 R_q} \quad \text{increasing with } C_d \text{ and } 1/R_q

C_d = 10fF
C_q = C_d
C_g = 10pF
R_q = 400k\Omega
R_q = 50\Omega

Note: valid for low impedance load \( R_{\text{load}} \ll R_q \)

- \( \tau_{\text{fast}} = R_{\text{load}} C_{\text{tot}} \)
- \( \tau_{\text{slow}} = R_q (C_q + C_d) \)
Pulse shape: dependence on Temperature

The two current components behave differently with temperature:
→ fast component is independent of T because $C_{\text{tot}}$ couples to external $R_{\text{load}}$
→ slow component is dependent on T because $C_{\text{d,q}}$ couples to $R_q(T)$

H. Otono, et al. PD07

Akiba et al. Optics Express 17 (2009) 16885

HPK MPPC

high pass filter / shaping → recover fast signals
Pulse shape vs T

HPK MPPC: 25µm, 50µm, 100µm

Measurements by Adam Para at Light 2011
Gain and its Fluctuations

\[ G = \Delta V \left( C_q + C_d \right) / q_e \]

→ Gain is linear if \( \Delta V \) in quenching regime

but

there are many sources for non-linearity of response (non proportionality)

SiPM gain fluctuations (intrinsic) differ in nature compared to APD where the statistical process of internal amplification shows a characteristic fluctuations

\[ \frac{\delta G}{G} = \frac{\delta V_{bd}}{V_{bd}} \oplus \frac{\delta C_{dq}}{C_{dq}} \]

\( \delta V_{bd} \) → cell to cell uniformity (active area and volume) control at % level

• doping densities (Poisson):
  \( \delta V_{bd} \geq 0.3V \)

*Shockley, Sol. State Ele. 2 (1961) 35*

• doping, epitaxial, oxide (processing):
  \( \delta V_{bd} \sim O(0.1V) \)

In addition \( \delta G \) might be due to fluctuations in quenching time

... and of course after-pulses contribute too (not intrinsic → might be corrected)
Response Non-Linearity

Non-proportionality of charge output w.r.t. number of photons (i.e. response) at level of several % might show up even in quenching regime (negligible quenching time), depending on $\Delta V$ and on the intensity and duration of the light pulse.

Main sources are:
- finite number of pixels
- finite recovery time w.r.t. pulse duration
- after-pulses, cross-talk
- drop of $\Delta V$ during the light pulse due to relevant signal current on (large) series resistances (e.g. ballast)

\[ T.\text{van Dam IEEE TNS 57 (2010) 2254} \]

Detailed model to estimate non-lin. corrections

Finite number of cells is main contribution in case number of photons $\sim O$(number of cells) (dynamic range not adequate to application)

$\rightarrow$ saturation \[ n_{\text{fired}} = n_{\text{all}} \left(1 - e^{-\frac{n_{\text{phot}} PDE}{n_{\text{all}}}}\right) \]

$\rightarrow$ loss of energy resolution

see Stoykov et al JINST 2 P06500 and Vinogradov et al IEEE NSS 2009 N28-3
New high dynamic range SiPMs

Different types available or in preparation:

- **tiny cells**
  - HPK, FBK, NDL, MPI-LL

- **micro cells**
  - Zecotek, AmplificationTech

SiPMs NDL (Beijing)

*Zhang et al NIM A621 (2010) 116 Han at NDIP 2011*

- type: n-on-p, Bulk Rq
- high cell density (10000/mm²)
- fast recovery (5ns)
- low gain
  - dynamic range
  - radiation hardness

**Latest MPPC tiny cell by Hamamatsu**

![Image of tiny cells with 20 µm and 15 µm pitch](image)

**Measurements by Y. Musienko**

![Graph showing equivalent number of fired pixels vs. N_PDE](image)
Noise sources:

Dark counts

After-pulsing

Cross-Talk

"optical"

pulses triggered by non-photo-generated carriers (thermal / tunneling generation in the bulk or in the surface depleted region around the junction)

carriers can be trapped during an avalanche and then released triggering another avalanche

photo-generation during the avalanche discharge. Some of the photons can be absorbed in the adjacent cell possibly triggering new discharges
Dark Count Rate

- DCR $\rightarrow$ linear dependence due to $P_{01} \propto \Delta V$ ($\rightarrow$ same as PDE vs $\Delta V$)
  $\rightarrow$ non-linear at high $\Delta V$ due to cross-talk and after-pulsing $\rightarrow \propto \Delta V^2$
- DCR scales with active surface (not with volume: high field region dominating)

$N.Dinu$ $et$ $al.$ $NIM$ $A$ $($2008$)$
Electro-optical characterization of SiPM: a comparative study
Dark Count Rate

**KETEK** PM 3350 (p⁺-on-n, shallow junction)
3x3mm² active area pixel size 50x50 µm²

Critical issues:
- quality of epitaxial layer
- gettering techniques
- Efield engineering (low T)

**Exelitas** 1st generation SiPM 2011
(p⁺-on-n) 1x1mm²

\[ V_{bd} \sim 25V \]

\[ V_{bd} \sim 140V \]

Latest Hamamatsu devices reached ~80kHz/mm²

HPK claiming for additional improvements coming
(*HPK at LIGHT 2011*)

\[ 100 \text{ um, GE} = 74\% \]
\[ 100 \text{ um, GE} = 52\% \]
\[ 50 \text{ um, GE} = 51\% \]
\[ 50 \text{ um, GE} = 39\% \]
\[ 25 \text{ um, GE} = 29\% \]
Control over individual SPADs enables detailed device characterization

- Over 90% good diodes (dark count rate close to average)
- Typical dark count rate at 20°C and 3.3V excess voltage: ~150cps / diode
- Low dark counts (~1-2cps) per diode at -40°C

T. Frach at NDIP 2011
Dark current vs T sources of DCR

Noise mainly comes from the high E Field region (no whole depletion region)

1) Generation/Recombination SRH noise (enhanced by trap assisted tunneling)

\[ I_{\text{reverse}} \sim T^{1.5} \exp\left(-\frac{E_{\text{act}}}{K_B T}\right) \]

Conventional SRH trap assisted tunneling contribution to DCR from diffusion of minority carriers negligible below 350K

Noise mainly comes from the high E Field region (no whole depletion region)

Tunneling noise dominating for T<200K (FBK devices have E field quite peaked)

2) Band-to-band Tunneling noise (strong dependence on the Electric field profile)

Efield engineering is crucial for min. DCR (esp. at low T)
Dark Count Rate vs T (constant $\Delta V$)

Measurement of counting rate of $\geq 1$ p.e. at fixed $\Delta V = 1.5$V ($\rightarrow$ constant gain)

$$DCR \sim T^{1.5} \exp \left( \frac{-E_{\text{act}}}{2K_B T} \right)$$

Activation energy $E_{\text{act}} \sim 0.72$eV

Note: $E_{\text{act}}$ should be $\sim E_g$ but tunneling makes effective gap smaller

Additional structure carriers freeze-out (?)

(carrier collection losses at very low T due to ionized impurities acting as shallow traps $\rightarrow$ drop in PDE)

G.C. et al NIM A628 (2011) 389
Dark Count Rate vs T

- Hamamatsu (100µm pixels)
- Comprehensive MPPC characterization at low T

J. Csathy et al. NIM A 654 (2011) 225

Akiba et al. Optics Express 17 (2009) 16885
After-Pulsing Carrier trapping and delayed release

\[ P_{\text{afterpulsing}}(t) = P_c \cdot \frac{\exp(-t/\tau)}{\tau} \cdot P_{01} \propto \Delta V^2 \]

\( \tau \) : trap lifetime depends on trap level position

\( P_c \) : trap capture probability

\( \propto \) carrier flux (current) during avalanche \( \propto \Delta V \)

\( \propto N \) traps

\(~\text{Few \% level at 300K}\)

\( \Delta V(t) \) : avalanche triggering probability

\( \propto \Delta V \)

\( \Delta V \) : quadratic dependence on \( \Delta V \)

Fast components

Slow components

Fig. 10. Spectrum of the delay time from the primary pulse to the after-pulse.

Only partially sensitive to after-pulsing during recovery

ie recovery hides After-pulses (does not cancel them)


not trivial dependence on \( T \)
After-Pulses vs T (constant $\Delta V$)

$\Delta V = 1.5V$

FBK devices

Measurement by waveform analysis:
- trigger on single carrier pulses (with no preceding pulses within $\Delta t=5\mu s$), **count subsequent pulses** within $\Delta t=5\mu s$
  (find the after-pulsing rate $r_{AP}$)
- **Subtract dark count** contribution
- extract **after-pulsing probability** $P_{AP}$
corrected for after-pulsing cascade

$P_{AP} = \frac{r_{AP}}{1 + r_{AP}}$

- Few % at room T
- $\sim$constant down to $\sim$120K

$T$ decreasing: increase of characteristic time constants of traps ($\tau_{traps}$) compensated by increasing cell recovery time ($R_{q}$)

- several % below 100K

$T<$100K: additional trapping centers activated possibly (?) related to onset of carriers freeze-out

$\rightarrow$ Analysis of life-time evolution vs T of the various traps (at least 3 types at $T_{room}$)

G.C. et al NIM A628 (2011) 389
Optical cross-talk  Avalanche luminescence (NIR)

Carriers' luminescence (spontaneous direct relaxation in the conduction band) during the avalanche: probability $3 \times 10^{-5}$ per carrier to emit photons with $E > 1.14$ eV

A.Lacaita et al. IEEE TED (1993)

Photons can induce avalanches in neighboring cells. Depends on distance between high-field regions

$\Delta V^2$ dependence on over-voltage:
- carrier flux (current) during avalanche $\propto \Delta V$
- gain $\propto \Delta V$

Counteract:
- optical isolation between cells by trenches filled with opaque material
- low over-voltage operation helps

It can be reduced to a level below % in a wide $\Delta V$ range
Optical cross-talk: reflections from the bottom

Measured Emission spectrum

(1) Cross-talk due to narrow $\lambda$ range (<100nm)

(2) Main component due to total reflection internal from the bottom (substrate)

(3) Isolation implants are sufficient to stop direct component

$\rightarrow$ Crosstalk can’t be eliminated simply by means of trenches
$\rightarrow$ Main contribution to crosstalk comes from bottom reflections (using trenches)

A. Ingargiola – NDIP08
Rech et al Proc. of SPIE Vol. 6771 677111-1
DCR, AP, Gain, X-talk vs $\Delta V$ (various T)

- Dark Noise Rate dumped at low T
- After-Pulsing swift increase below 100K
- P$_{AP}$ $\sim$ independent of T above 100K
- (slight reduction expected due to lower P$_{01}$ for large $\lambda$ at low T)

Slopes changing with T:
- different mechanisms
  - SRH$\sim\Delta V^2$ / Tunneling $\sim\Delta V^3$
- $P_{01}$ changing with T

Gain and Cross-Talk are independent of T

FBK devices
G.C. et al NIM A628 (2011) 389
Photo-Detection Efficiency (PDE)
\[ \text{PDE} = \text{QE} \cdot \text{P}_{01} \cdot \text{FF} \]

**QE:** carrier Photo-generation
- probability for a photon to generate a carrier that reaches the high field region
- \(\rightarrow \lambda\) and \(T\) dependent
- \(\rightarrow \Delta V\) independent if full depletion at \(V_{bd}\)

**\(P_{01}\):** avalanche triggering probability
- probability for a carrier traversing the high field to generate the avalanche
- \(\rightarrow \lambda\), \(T\) and \(\Delta V\) dependent

**FF:** geometrical Fill Factor
- fraction of dead area due to structures between the cells, e.g., guard rings, trenches
- \(\rightarrow\) mild \(\Delta V\) dependence (cell edges)
optical $T, A, (R)$ of the entrance window (dielectric on top of silicon surface) → angular and polarization dependence
carrier recombination loss: collection efficiency front, depl. region, back

front region critical for $60\text{nm} < \lambda < 400\text{nm}$
→ C eff. depends on surface recombination velocity $S_f$
→ freeze-out at low $T$

internal quantum efficiency: probability to photo-generate an e-h pair $\sim$ photon $E$ (above threshold)

eg of QE optimization (blue)
• Anti-reflective coating (ARC)
• Shallow junctions for short $\lambda$
• Thick epi layers for long $\lambda$
**QE single cell**

FBK single cell

- **photo-voltaic regime** ($V_{bias} \sim 0$ V)

- **Limited by**
  - ARC Transmittance
  - Superficial Recombination

- **Limited by the small π layer thickness**

- Most critical issue for **Deep UV SiPM**

  - note: reduced superficial recombination in n-on-p wrt p-on-n

---

**Graph Details**

- **Wavelength (nm):** 300 to 800
- **QE (%)**
  - 0V (solid blue line)
  - -2V (dashed orange line)
  - Simu ARC (dotted purple line)

---

G.Colluzzo - PhotoDet 2012
Avalanche trigger probability ($P_{01}$)

Probability calculations after W. Oldham et al. IEEE TED (1972)

Example with constant high-field:
(a) only holes trigger the avalanche
(b) both electrons and holes trigger
(c) only electrons trigger

$P_{01} = \frac{P_{DE}}{QE} \times FF$
PDE vs $\Delta V$

Ionization rate in Silicon

$P_{01}$ optimization (n-on-p)

- high over-voltage
- photo-generation in the p-side of the junction

Ionization Rates (lum)

- Electrons
- Holes

SiPMs

DATA

- HPK 500nm
- FBK 500nm

E field profile $\rightarrow$ the slope of PDE vs $\Delta V$

note: $P_{01}$ fixes also the slope of DCR vs $\Delta V$ $\rightarrow$ working range

$\Delta V/V$ (%)
Fig. 5a) The PDE vs. λ of the Photonique, FBK-irst and SensL devices and b) HPK

**n-on-p structures**
- **n+**
- **p- epi**
- **p-substrate**

**p-on-n structure**
- **p+**
- **n**
- **n- epi**
- **n-substrate**

Note: geometrical fill factor included
Improving PDE

→ PDE peak constantly improving for many devices
→ every manufacturer shape PDE for matching target applications
→ UV SiPM eg from MePhi/Excilta (see E.Popova at NDIP 2011)
→ DUV SiPMs in development too

\[ V_{bd} = 25V \quad \Delta V = 3.3V \]

\[ FF \sim 50\% \]

\[ FF \sim 60\% \]

\[ \Delta V \sim 6V \]

\[ dSiPM \text{ (latest sensor 2011)} \]
→ up to now no optical stack optimization
→ no anti-reflecting coating
→ potential improvement up to 60% peak PDE

(Y.Haemish at AIDA 2012)

Barlow – LIGHT 2011

T.Frach 2012 JINST 7 C01112

F.Wiest – AIDA 2012 at DESY

Excilta

PM1150 Standard Technology Type

KETEK

31.0 V
28.5 V
PDE vs T (ΔV constant)

When T decreases:

1) silicon $E_{\text{gap}}$ increasing
   → larger attenuation length
   → lower QE (for larger $\lambda$)

2) mobility increasing
   → larger impact ionization
   → larger trigg. avalanche $P_{01}$

3) carriers freeze-out onset below 120K
   → loss of carriers

G.C. et al NIM A628 (2011) 389

FBK devices

Halogen lamp (CW)

- 400 nm
- 600 nm
- 800 nm

Normalization to PDE (room T)

?? interplay between (1) and (2): modulation
... drop in 250<T<300 not well understood
(common feature with APDs')

RMD APD at 400nm < $\lambda$ < 700nm
Johnson et al, IEEE NSS 2009

Additional effects in APD
(depletion region depends on T, ...)

lines are for eye guide

freeze-out (3)
PDE dependences, changing with T

PDE vs $\lambda$ ($\Delta V$ constant)

- Data
- $T=295\,K$
- $T=238\,K$
- $T=150\,K$
- $T=55\,K$

PDE spectrum at low T peaks at shorter $\lambda$

$\Delta V = 2V$

- Simulation
- $T=300\,K$
- $T=250\,K$
- $T=150\,K$
- $T=50\,K$

PDE $\Delta V$ vs ($\lambda$ constant)

- Simulation
- $\lambda=400\,nm$
- $T=50, 150, \ldots, 300\,K$
- saturation starts earlier at low T

Data

Pulsed laser (405nm)

- Data
- $T=297\,K$
- $T=193\,K$
- $T=123\,K$
- $T=60\,K$
Timing

- SiPM are intrinsically very fast
  - jitter (gaussian) below 100ps, depending on $\Delta V$
  - but also → non-gaussian tails up to $O(\text{ns})$, depending on wavelength

- Timing measurement:
  - use of fast signal shape component
  - use waveform, better than CFD (much than ToT)
**GM-APD avalanche development**

(1) Avalanche “seed”: free-carrier concentration rises exponentially by "longitudinal" multiplication.

(1') Electric field locally lowered (by space charge R effect) towards breakdown level.

Multiplication is self-sustaining. Avalanche current steady until new multiplication triggered in near regions.

(2) Avalanche spreads "transversally" across the junction.

(diffusion speed ~ up to 50µm/ns enhanced by multiplication).

(2') Passive quenching mechanism effective after transverse avalanche size ~ 10µm.

If no quench, avalanche spreads over the whole active depletion volume → avalanche current reaches a final saturation steady state value.

_A. Spinelli Ph.D thesis (1996)_

Simulation w/o quenching: → steady current reached.

---

**Diagram Notes:**
- Longitudinal multiplication:
  - Duration ~ few ps
  - Internal current up to ~ few µA

- Transverse multiplication:
  - Duration ~ few 100ps
  - Internal current up to ~ several 10µA

---

**Graphs:**
- Current vs. time (ps)
- Current vs. time (ns)
Avalanche transverse propagation by a kind of shock wave: the wavefront carries a high density of carriers and high E field gradients (inside: carriers' density lower and E field decreasing toward breakdown level)

\[ \frac{dS}{dt} = \frac{d}{dt} 2\pi r(t) \Delta r = 2\pi v_{\text{diff}} \Delta r = 4\pi \Delta r \sqrt{\frac{D}{\tau}} \]

Rate of current production:

\[ \frac{dI}{dt} = \frac{dI}{dS} \frac{dS}{dt} \sim \frac{\sqrt{D}}{R_{sp} \sqrt{\tau}} \]

\[ \frac{dI}{dS} = J = \frac{V_{\text{bias}}}{R_{sp}(S)} \]

Internal current rising front: 
the faster it grows, the lower the jitter 
\( \frac{dI}{dt} \rightarrow \) understand/engineer timing features of SiPM cells

→ timing resolution improves at high \( V_{\text{bias}} \) 
→ E field profile affects \( \tau \) and \( R_{sp} \) (wider E field profile \( \rightarrow \) smaller \( R \)) 
(should be engineered when aiming at ultra-fast timing)
→ T dependence of timing through \( \tau \) and \( D \) 
→ slower growth at GAPD cell edges \( \rightarrow \) higher jitter at edges 
reduced length of the propagation front
GM-APD timing jitter: fast and slow components

1) Fast component: gaussian with time scale $O(100\text{ps})$

Statistical fluctuations in the avalanche:

- **Longitudinal** build-up (minor contribution)
- **Transversal** propagation (main contribution):
  - via multiplication assisted diffusion (dominating in few $\mu$m thin devices)
    
    \[ A.\text{Lacaita et al. APL and El.Lett. 1990} \]

  - via photon assisted propagation (dominating in thick devices – $O(100\mu m)$)
    
    \[ PP.\text{Webb, R.J. McIntyre RCA Eng. 1982 A.Lacaita et al. APL 1992} \]

**Fluctuations** due to

- a) impact ionization statistics
- b) variance of longitudinal position of photo-generation: finite drift time even at saturated velocity note: saturated $ve \sim 3 \text{vh}$
  (n-on-p are faster in general)

$\rightarrow$ Jitter at minimum $\rightarrow O(10\text{ps})$
(very low threshold $\rightarrow$ not easy)

- c) variance of the transverse diffusion speed $v_{diff}$

**Fluctuations** due to

- d) variance of transverse position of photo-generation: slope of current rising front depends on transverse position

$\rightarrow$ Jitter $\rightarrow O(100\text{ps})$
(usually threshold set high)
GM-APD timing jitter: fast and slow components

2) Slow component: non-gaussian tails with time scale $O(\text{ns})$

Carriers photo-generated in the neutral regions above/beneath the junction and reaching the electric field region by diffusion

$G.\text{Ripamonti, S. Cova} \text{ Sol. State Electronics (1985)}$

![Diagram of GM-APD with neutral regions and guard ring]

$\tau \sim \frac{L^2}{\pi^2 D} \sim \text{up to some ns}$

$L = \text{effective neutral layer thickness}$

$D = \text{diffusion coefficient}$

$S. \text{Cova et al. NIST Workshop on SPD (2003)}$

→ **Neutral regions** underneath the junction: timing tails for long wavelengths
→ **Neutral regions** in APD entrance: timing tails for short wavelengths
PDE vs timing optimization


E field

k = ratio of hole (β) to electron (α) ionization coefficient (increasing with E field)

w = high field region width

narrow avalanche region, high E:
- small w
- high k = b/a

better for TIMING

w = high field region width

better for PDE

wide avalanche region, low E:
- wide w
- small k = b/a

Plots are courtesy of C.H. Tan
Waveform analysis: optimum timing filter

Example of intrinsic SPTR measurement from $\Delta t$ of consecutive pulses by laser shots

Different algorithms to reconstruct the time of the pulses:

✗ parabolic fit to find the peak maximum
✗ CFD (digital)
✗ average of time samples weighted by the waveform derivative
✔ digital filter: weighting by the derivative of a reference signal
→ optimum against (white) noise (if signal shape fixed)

Digital filter to minimize N/S for timing measurements:
solve the following equation on $t_0$:

$$\int V_a(t) \frac{\partial V_r(t-t_0)}{\partial t} dt = 0$$

$V_a =$ measured signal (includes noise)
$V_r =$ reference signal
$t_0 =$ reference time

see e.g. Wilmshurst “Signal recovery from noise in electronic instrumentation”
Waveform (single p.e.)

Average waveform (the band is rms)

FBK device $\Delta V = 3V$

Falling signal shape fluctuates considerably (due eg to after-pulses) → signal tail is non useful for timing, if not detrimental

For comparison about waveform method and various digital algorithms see Ronzhin et al NIM A 668 (2012) 94
Waveform analysis: 1 p.e. reference signal

Average waveform
(the band is rms)

FBK device
$\Delta V = 3\text{V}$

G.C. (2011, unpublished)

Rise time (10%-90%)
(dominated by electronics contribution)

Reminder:

$$\frac{dI}{dt} \sim \frac{\sqrt{D}}{R_{sp} \sqrt{\tau}}$$

$$\tau \sim \frac{1}{1 - \left( \frac{E_{max}}{E_{breakdown}} \right)^n}$$

For comparison about rise-time of HKP devices see
Single Photon Time Resolution = \textit{gaussian + tails}

Time resolution of SiPM is not just a gaussian, but gaussian + tails (in particular at long wavelengths)


Data at $\lambda=400\text{nm}$

A simple \textit{gaussian component} fits fairly

Data at $\lambda=800\text{nm}$

fit gives reasonable $\chi^2$ in case of an \textbf{additional exponential term} \newline \hspace{1cm} $\exp(-|\Delta t|/\tau)$ summed with a weight

- $\tau \sim 0.2 \div 0.8\text{ns}$ (depending on device) in rough agreement with diffusion tail lifetime: $\tau \sim L^2/\pi^2 D$ wher $L$ is the diffusion length

- Weight of the \textbf{exp. tail} $\sim 10\% \div 30\%$ (depending on device)

\begin{itemize}
  \item Overvoltage=4V \textbf{FIT: gauss+const} \newline
  \hspace{1cm} $\lambda=400\text{nm}$
  \newline
  \hspace{1cm} $\alpha=0.2$ \div 0.8\text{ns} \newline
  \hspace{1cm} \text{mod}(\Delta t, T_{\text{laser}}) [\text{ns}]$
  \newline
  \hspace{1cm} Gauss \hspace{1cm} + \hspace{1cm} Tails (long $\lambda$) \newline
  \hspace{1cm} $\sim \exp (-t / \Omega (\text{ns}))$
  \newline
  \hspace{1cm} contrib. several % for long wavelengths

  \item Overvoltage=4V \textbf{FIT: gauss+const} +\textbf{exponential} \newline
  \hspace{1cm} $\lambda=800\text{nm}$
  \newline
  \hspace{1cm} $\alpha=0.02$ \div 0.03\text{ns} \newline
  \hspace{1cm} \text{mod}(\Delta t, T_{\text{laser}}) [\text{ns}]$
\end{itemize}

Distributions of the difference in time between successive peaks
**SPTR: FBK devices – shallow junction**

In general due to drift, resolution differences

1) high field junction position
   - shallow junction: $\sigma_{t}^{\text{red}} > \sigma_{t}^{\text{blue}}$
   - buried junction: $\sigma_{t}^{\text{red}} < \sigma_{t}^{\text{blue}}$

2) $n^+\text{-on-}p$ smaller jitter than $p^+\text{-on-}n$ due to electrons drifting faster in depletion region (but $\lambda$ dependence)

3) above differences more relevant in thick devices than thin

**NOTE:** good timing performances kept up to 10MHz/mm² photon rates
SPTR: Hamamatsu

- **SPTR Parameters**
  - \( \lambda = 800 \text{ nm} \)
  - \( \lambda = 400 \text{ nm} \)

**Suggested Operating Range**

- **Electron Injection**
- **Hole Injection**

**Graphs**

- **1600 cells (25x25\(\mu\text{m}^2\))**
- **400 cells (50x50\(\mu\text{m}^2\))**

**Legend**

- \( \bullet \): eye guide
- \( \circ \): hole injection
- \( \bigcirc \): electron injection

**Diagram**

- HPK-3
- HPK-2

**Layers**

- p+
- n-epi
- n-substrate
SPTR: CPTA/Photonique – thick structures

- thick structures
- deep junctions

a) Green-Red sensitive
SSPM 050701GR_TO18

b) Blue sensitive
SSPM 050901B_TO18

\[ \lambda = 800 \text{ nm} \]
\[ \lambda = 400 \text{ nm} \]

- eye guide

a) \( n^+ \)-on-p
→ electrons drift

b) \( p^+ \)-on-n
→ holes drift \((v_e/3)\)
dSiPM timing resolution

ΔV = 3.3V

- Sensor triggered by attenuated laser pulses at first photon level
- Laser pulse width: 36ps FWHM, λ = 410nm
- Contribution to time resolution (FWHM):
  - SPAD: 54ps, trigger network: 110ps, TDC: 20ps
- Trigger network skew currently limits the timing resolution

T. Frach at LIGHT 2011
SPTR: position dependence $\rightarrow$ cell size

Data include the system jitter (common offset, not subtracted)

### Table

<table>
<thead>
<tr>
<th></th>
<th>FWHM (ps)</th>
<th>FWTM (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>199</td>
<td>393</td>
</tr>
<tr>
<td>2</td>
<td>197</td>
<td>389</td>
</tr>
<tr>
<td>3</td>
<td>209</td>
<td>409</td>
</tr>
<tr>
<td>4</td>
<td>201</td>
<td>393</td>
</tr>
<tr>
<td>5</td>
<td>195</td>
<td>383</td>
</tr>
</tbody>
</table>

K. Yamamoto PD07

Larger jitter if photo-conversion at the border of the cell

Due to:
1) slower avalanche front propagation
2) lower E field at edges
   → cfr PDE vs position
SPTR: timing at low T

**Timing**: improves at low T
Lower jitter at low T due to higher mobility:

- a) avalanche process is faster
- b) reduced fluctuations

(Over-voltage fixed)

Note:
\[
\frac{dI}{dt} \sim \frac{\sqrt{D}}{R_{sp} \sqrt{\tau}}
\]

![Graph showing the relationship between temperature (T) and sigma (probability distribution)]

FBK devices

G.C. (2011, unpublished)
Optimizing signal shape for timing

Timing by (single) threshold:
- time spread proportional to 1/rise-time and noise

\[
\sigma_{\text{time}} = \frac{\sigma_{\text{amplitude}}}{\frac{df(t)}{dt}}
\]

Timing with optimum filtering:
- best resolution with \( f'(t) \) weighting function

\[
\sigma^2_{\text{time}} = \frac{\sigma^2_{\text{amplitude}}}{\int dt \left[ \frac{df(t)}{dt} \right]^2}
\]

Pulse sampling and Waveform analysis:
- Sample, digitize, fit the (known) waveform
- get time and amplitude

\[
\sigma^2_{\text{time}} = \frac{\sigma^2_{\text{amplitude}}}{N_{\text{samples}} \int dt \left[ \frac{df(t)}{dt} \right]^2}
\]

Fig. 7. Optimum filter for timing in presence of white noise (method of derivation).
(a) signal waveform
(b) optimum filter for amplitude measurements.
(c) optimum filter for timing - derivative of (b).
(d) output waveform.

V. Radeka IEEE TNS 21 (1974)
Optimizing signal shape for timing

Single cell model \( (R_d||C_d) + (R_q||C_q) \)

SiPM + load \( (||Z_{cell})||C_{grid} + Z_{load} \)

Signal = slow pulse \( (\tau_d \text{ (rise)}, \tau_q \text{-slow (fall)}) \) + fast pulse \( (\tau_d \text{ (rise)}, \tau_q \text{-fast (fall)}) \)

- \( \tau_d \text{ (rise)} \sim R_d (C_q + C_d) \)
- \( \tau_q \text{-fast (fall)} = R_{load} C_{tot} \) (fast; parasitic spike)
- \( \tau_q \text{-slow (fall)} = R_q (C_q + C_d) \) (slow; cell recovery)

Pulse shape

\[
V(t) \approx \frac{Q}{C_q + C_d} \left( C_q e^{-t/\tau_{FAST}} + \frac{R_{load}}{R_q} C_d e^{-t/\tau_{SLOW}} \right)
\]

\[
\frac{Q_{fast}}{Q_{slow}} \sim \frac{C_q}{C_d}
\]

\[
\frac{V_{\text{max fast}}}{V_{\text{max slow}}} \sim \frac{C_q^2 R_q}{C_d C_{tot} R_{load}}
\]

Increasing \( C_q/C_d \) or/and \( R_q/R_{load} \)

\( \rightarrow \) spike enhancement
\( \rightarrow \) better timing

- \( C_d = 10\text{fF} \)
- \( C_q = C_d \)
- \( C_q = 10\text{pF} \)
- \( R_q = 400\text{k}\Omega \)
- \( R_q = 50\Omega \)
Optimizing signal shape for timing (SPTR)

\[ \frac{V_{\text{max, fast}}}{V_{\text{max, slow}}} \sim \frac{C^2_q R_q}{C_d C_{\text{tot}} R_{\text{load}}} \]

Enhancing \( C_q \) does improve timing performances

Yamamura et al. at PD09

Analogous method for timing optimization proposed in C. Lee et al. NIM A 650 (2010) 125 “Effect on MIM structured parallel quenching capacitor of SiPMs”

Note:
The steep falling front of the fast peak could be exploited too for optimum timing

\[ \sigma_{\text{time}}^2 = \frac{\sigma_{\text{amplitude}}^2}{N_{\text{samples}} \int dt \left[ f'(t) \right]^2} \]
Summary

- **Operative ΔV over-bias range:** from 2V (eg HPK) to 10V (eg FKB) depending on E field profile and Rq

- **T coefficient:** low, below 0.3%/°C for many devices → might be lower, but tradeoff against PDE and noise

- **Pulse Shape** and **Gain:** tuned for matching application requirements (tradeoffs) → photon counting and timing vs energy measurement (signal spike, E\textsubscript{field} profile)

- **Dynamic range:** Large, up a few x10000 pixels (eg NDL, Zecotek) → improved **radiation hardness** (not covered in this review) is relevant bonus → trade-off with Fill Factor

- **PDE:** up to 60% for blue-green light (eg. KETEK) → easily tuned to match applications (but only in visible optical range)

- **DCR** at T room can be < 100kHz/mm\(^2\) (eg. Hamamatsu)

- **Cross-Talk:** can be as low as 1% in operative range (eg. FBK, MePhi/Pulsar)

- **After-Pulsing:** still at some % level for many devices → exploiting higher Rq “just” to hide A-P is not a good practice... → Digital SiPM is prone too, though less affected (active quenching)

- **Timing:** intrinsically fast, SPTR < 50ps in operative range → but mind the diffusion non-gaussian tails in temporal response (long λ)

- **Calibration:** precise, thanks to existing detailed operative models

**Significant development** of SiPMs over the last few years and new players
Still missing and Future threads

- Avalanche **detailed physical models are still missing**. In particular for → **ultra-fast timing** applications there is room for device improvement → techniques for reducing **long timing tails** might be exploited

- Physical models might be of help also in further reducing **DCR** and **A-P** → eg: E field engineering for reducing tunneling

- **PDE**: expected soon are → improvements the **UV, VUV, EUV region** → devices with **through vias** → coupling with scintillators, fast imaging!

- GM-APD arrays for **NIR, IR sensitivity**: **different semiconductors** → InGaAs GM-APD arrays from AmplificationTechnologies do exist but... small area, noise and cost (!)

- **DCR**: → expected in 2012 a factor **x3 improvement** → larger area devices will follow → in the mean time devices tuned for working at **cryogenic T** easy to devise

- **Low T**: SiPM perform ~ideally in the range $100K < T < 200K$ → $R_q$ should be tuned shorter recovery (ad hoc devices) → lower gain (small cells) might be desired to mitigate after-pulses
Thanks for your attention
Additional material
The building block of a SiPM: GM-APD

**APD: Linear-Proportional Mode**
- Bias BELOW $V_{BD}$ ($V_{APD} < V < V_{BD}$)
- It's an **AMPLIFIER**
- Multiplication: in practice limited to $10^4$ by fluctuations
- No single photo-electron resolution
  ...except at low T with slow electronics,
  *Dorokhov et.al. J.Mod.Opt. 51 (2004)*

**GM-APD: Geiger Mode**
- Bias ABOVE $V_{BD}$ ($V - V_{BD} \sim$ a few volts)
- It's a **TRIGGER** (BINARY) device
- Multiplication: $\infty$... in practice limited by macroscopic parameters (R,C)
- Limited by dark count rate
- Single photo-electron resolution
- Need Reset (Feedback - Quenching)
Readout Mode

High Z node

Voltage Mode

R_s = 50 Ω

Avalanche trigger time

-V_{A}

Low Z node

Current Mode

R_L

Avalanche trigger time

R_s = 50 Ω

+ V_{A}

hv
Key elements in SiPM cell

Doping and Field profiles

Optical window

Light absorption in Silicon

Guard Ring:
- for avoiding early edge breakdown
- for isolating cells
- for tuning E field shape
- has important impact on fill factor (more than Rq and metal grid)

Doping and Field profiles

Optical window

Doping concentration (10^11 cm^-3)

Field (V/cm)

Depth (um)

Absorption length, μm

Wavelength (nm)

77 K
173 K
300 K

Surface effects dominate

Transparency, interference are issues

Light absorption in Silicon
PDE vs $\lambda$ ($\Delta V$ fixed, various T)

FBK devices

PDE spectrum at low T peaks at shorter wavelengths
RPL model: fast simulation

“Statistics of Avalanche Current Buildup Time in Single-Photon Avalanche diodes”
C.H. Tan, J.S. Ng, G.J. Rees, J.P.R. David (Sheffield U.)

Numerical model (MC): Random distribution of impact ionization Path Length (RPL)

Analysis of breakdown probability, breakdown time and timing jitter as functions of avalanche region width (w), ionization coefficient ratio \( k = \beta_{\text{hole}} / \alpha_{\text{electron}} \) and dead space parameter (d) (uniform E field, constant carrier velocity)

1) increasing \( k \):
   - improves timing performances
   - but breakdown probability \( P_{br} \) increases slowly with overvoltage

1a) hole injection results in better timing than electron injection (in Si devices)

2) dead space effects worsen timing performances (the more at small \( k \)) Important for devices with small \( w \)
Many photons (simultaneous)

Dependence of SiPM timing on the number of simultaneous photons

Poisson statistics: $\sigma_t \propto 1/\sqrt{N_{pe}}$

![Graph showing the dependence of SiPM timing on the number of simultaneous photons. The graph includes a curve fit to $c/\sqrt{N_{pe}}$.](image)
Signal shape for timing - many photons

Single p.e. signal slow falling-time component $\tau_{\text{fall}} = R_q (C_d+C_d)$ strongly affects multi-photon signal risetime

PMT - 1 p.e.  
various gaussian signal shapes

SiPM - 1 p.e.  
changing risetime

SiPM - 1 p.e.  
changing falltime

PMT - 511keV in LYSO  
convolution 1pe $\otimes$ scint.exp.

SiPM - 511keV in LYSO  
convolution 1pe $\otimes$ scint.exp.

SiPM - 511keV in LYSO  
convolution 1pe $\otimes$ scint.exp.
Optimizing shape for timing - many photons

Enhancing $C_q$ and $R_q$ does improve timing performances

$\frac{V_{\text{max fast}}}{V_{\text{max slow}}} \sim \frac{C_q}{C_d C_{\text{tot}} R_q R_{\text{load}}}$

FBK devices type:
- Active area: 4x4mm$^2$;
- Cell size: 67x67μm$^2$;
- Fill factor: 60%;
- $C_Q+C_D$: about 180fF;
- $R_Q$: 1.1M;
- Dark noise rate: $\sim$100MHz at DV> 4V

C.Piemonte et al IEEE TNS (2011)

Fig. 2. Test set-up consists of two similar gamma ray detectors (LYSO crystal + SiPM) in coincidence. A $^{22}$Na source (disc in the middle) was used to generate two opposite 511keV photons in coincidence.

- Signal rise-time $< 5$ns
- CRT $\sim 320$ps (*) FWHM triggering at 5% height

Both are much better than for different structures with high $C_{\text{tot}}$ and/or lower $C_q$, $R_q$ (risetime up to several x 10ns, CRT $> 400$ps)

??? peak shape is not scaling with $\Delta V$ (non linearity in the F.Corsi et al electrical model)

Can be corrected $\rightarrow$ energy resol. $\sim 11$

(*) $\sim 40\%$ from light propagation in crystals
Radiation damage

Note:
- small cells smaller charge flow (small gain, high dynamic range)
- small epi-layed width
Radiation damage: two types

- **Bulk damage** due to Non Ionizing Energy Loss (NIEL) $\leftrightarrow$ neutrons, protons
- **Surface damage** due to Ionizing Energy Loss (IEL) $\leftrightarrow \gamma$ rays

(accumulation of charge in the oxide (SiO2) and the Si/SiO2 interface)

**Assumption:** damage scales linearly with the amount of Non Ionizing Energy Loss (NIEL hypothesis)

**Expectations:**
- protons / $\gamma$-ray $\sim 100$
- protons / neutrons $\sim 2\sim 10$

**Examples of radiation tolerances for HEP and space physics**
- ATLAS inner detector $\ldots 3 \times 10^{14}$ hadrons/cm$^2$/10 year $\sim 10^4$ hadrons/mm$^2$/s
- General satellites $\ldots \sim 10$ Gy/year

**Graph:**
- Reactor neutrons (T.Matsumura-PD07)
- Protons 53.3 MeV (Matsumura)
- Protons 200 MeV (Danilov-VCI07)
- Protons 400 MeV (Musienko - NDIP08)
- $\gamma$-ray $^{60}$Co (Matsumura)
- Electron $e^+ 28$ GeV (Musienko)

G.Lindstrom et al. NIM A426(1999)1-15
Radiation damage: effects on SiPM

1) Increase of dark count rate due to introduction of generation centers

Increase ($\Delta R_{DC}$) of the dark rate:

$$\Delta R_{DC} \sim P_{01} \alpha \Phi_{eq} V_{eff} / q_e$$

where $\alpha \sim 3 \times 10^{-17} \text{ A/cm}$ is a typical value of the radiation damage parameter for low E hadrons and $V_{eff} \sim \text{Area}_{SiPM} \times \varepsilon_{geom} \times W_{epi}$

NOTE:
The effect is the same as in normal junctions:
- independent of the substrate type
- dependent on particle type and energy (NIEL)
- proportional to fluence

2) Increase of after-pulse rate due to introduction of trapping centers

$\rightarrow$ loss of single cell resolution $\rightarrow$ no photon counting capability

Indications from measurements:

1) no dependence on the device
   similar effects found for SiPM from MePHY (Danilov) and HPK (Matsumura)
   (normaliz. to active volume)

2) no dependence on dose-rate
   HPK (Matsumura)

3) n similar damage than p

4) p $\times 10^{1-10^2}$ more damage than $\gamma$

Sample #20 ($130 \text{ Gy/h}$)
Sample #21 ($16 \text{ Gy/h}$)
Damage comparison

HPK devices
T. Matsumura – PD07

Damage effect ...
almost the same for protons and neutrons

Damage effect ...
1~2 orders larger with protons than γ-ray irradiation

2.3×10⁵ p/mm²/s (130 Gy/h)
I_{\text{leak}} @ (V_{\text{op}}, 1.4×10⁸ p/mm²) = 6.7 µA

4.2×10⁵ n/mm²/s
I_{\text{leak}} @ (V_{\text{op}}, 1.0×10⁸ n/mm²) = 8.5 pA

100 pixel MPPC
V_0 = 69.0V

Bias Voltage (V)
Current (µA)

Current after 1 hour (µA)

I_{\text{leak}} @ (V_{\text{op}}, 1.4×10⁸ p/mm²) = 6.7 µA

Bias Voltage (V)
Current (µA)

2.8×10⁸ p/mm²

2.3×10⁵ p/mm²/s (130 Gy/h)

Bias Voltage (V)
Current (µA)

Current after 1 hour (µA)

Bias Voltage (V)
Current (µA)

2.8×10⁸ p/mm²

Bias Voltage (V)
Current (µA)

Current after 1 hour (µA)

Bias Voltage (V)
Current (µA)

1.4×10⁸ p/mm²

Bias Voltage (V)
Current (µA)

Current after 1 hour (µA)

Bias Voltage (V)
Current (µA)

1×10⁸ n/mm²

Bias voltage (V)
Current (µA)

Current after 1 hour (µA)

Bias Voltage (V)
Current (µA)

4.2×10⁵ n/mm²/s

Bias Voltage (V)
Current (µA)

Current after 1 hour (µA)

Bias Voltage (V)
Current (µA)

1×10⁸ n/mm²

Bias Voltage (V)
Current (µA)

Before irradiation

I_{\text{leak}} @ (V_{\text{op}}, 1.4×10⁸ p/mm²) = 6.7 µA

I_{\text{leak}} @ (V_{\text{op}}, 1.0×10⁸ n/mm²) = 8.5 pA

Before irradiation
Radiation damage: neutrons (0.1 - 1 MeV)

8.3×10^4 n/mm^2  3.3×10^5 n/mm^2  1.0×10^8 n/mm^2

I-V drastically change. No signal
Signal pulse is still there, but continuous pulse height. (No photon-counting capability)

Nakamura at NDIP08
Radiation damage: neutrons 1 MeV $E_{eq}$

- No change of $V_{bd}$ (within 50mV accuracy)
- No change of $R_q$ (within 5% accuracy)
- $I_{dark}$ and DCR significantly increase

SiPMs with high cell density and fast recovery time can operate up to $3 \times 10^{12}$ n/cm$^2$ ($\delta G < 25\%$)

Y. Musienko at SiPM workshop CERN 2011