InP-based SPADs for Automotive Lidar

Mark Itzler

mitzler@argo.ai

9 June 2020
ISSW2020

Presentation Outline

- Introduction

- Review of legacy cooled 2D SPAD camera focal plane arrays (FPAs)

- Migration from cooled SPAD cameras to automotive lidar SPAD FPAs
  - Performance distributions and scaling to high temperature operation

- Reliability of InP-based SPADs

- Implementation comparison of Geiger-mode lidar and linear-mode lidar

- Summary

Argo AI's Mission

Argo AI is on a mission to build self-driving technology you can trust.

We exist to make getting around cities safe, easy, and enjoyable for all.

This mission is one of the most challenging applications of computer science, robotics, and artificial intelligence.

Argo's Lidar Goal

Deliver high-performance, scalable automotive lidar designed specifically for autonomous vehicles

Argo's SPAD-based lidar

Fundamental detectors in Argo's lidar architecture:

- SPAD-based receivers

- Operating in short-wave infra-red (SWIR)

Technology legacy from SPAD cameras for airborne (long-range) lidar

Why detect short-wave infra-red (SWIR) photons?

- **Minimal loss in optical fiber** (e.g., at 1.5 μm)

- **Covertness** (to human vision and I^2 night vision goggles)

- **Reflective imaging** (+ spectral information)

- <u>**Greater eye safety**</u> for active imaging (beyond 1.4 μm)

- **Environmental factors**
  - Less solar background than visible/NIR
  - Better atmospheric transmission

- **Technological factors**
  - Maturity of pulsed laser sources (e.g., 1.06 μm, 1.5 μm)
  - Maturity of SWIR optics (e.g., from telecom)
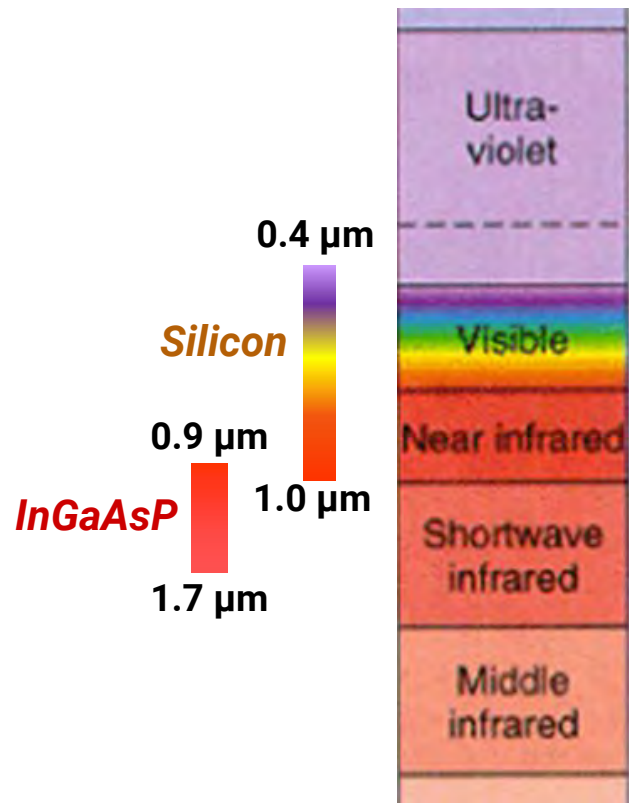
Silicon

0.4 μm

0.9 μm

InGaAsP

1.0 μm

**1.7 μm**

Presentation Outline

- Introduction

- **Review of legacy cooled 2D SPAD camera focal plane arrays (FPAs)**

- Migration from cooled SPAD cameras to automotive lidar SPAD FPAs
  - Performance distributions and scaling to high temperature operation

- Reliability of InP-based SPADs

- Implementation comparison of Geiger-mode lidar and linear-mode lidar

- Summary

Detector structure for InP SPAD

Separate Absorption and Multiplication (SAM) InGaAs(P)/InP avalanche diode structure:

- **Absorption region**: Absorb photon to create electrical carrier

- **Multiplication region**: Create additional carriers by avalanche gain
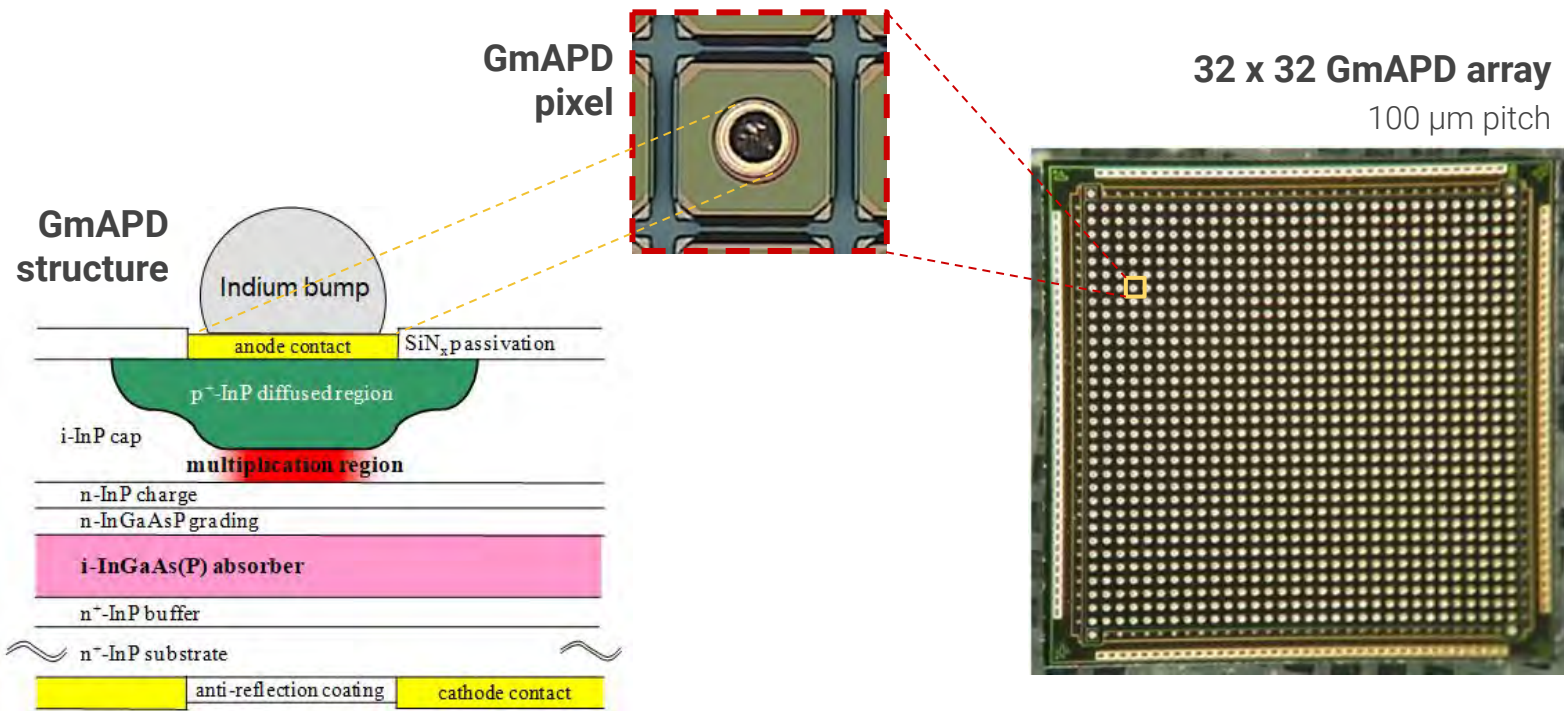
Schematic of an InP/InGaAs avalanche photodiode (APD)
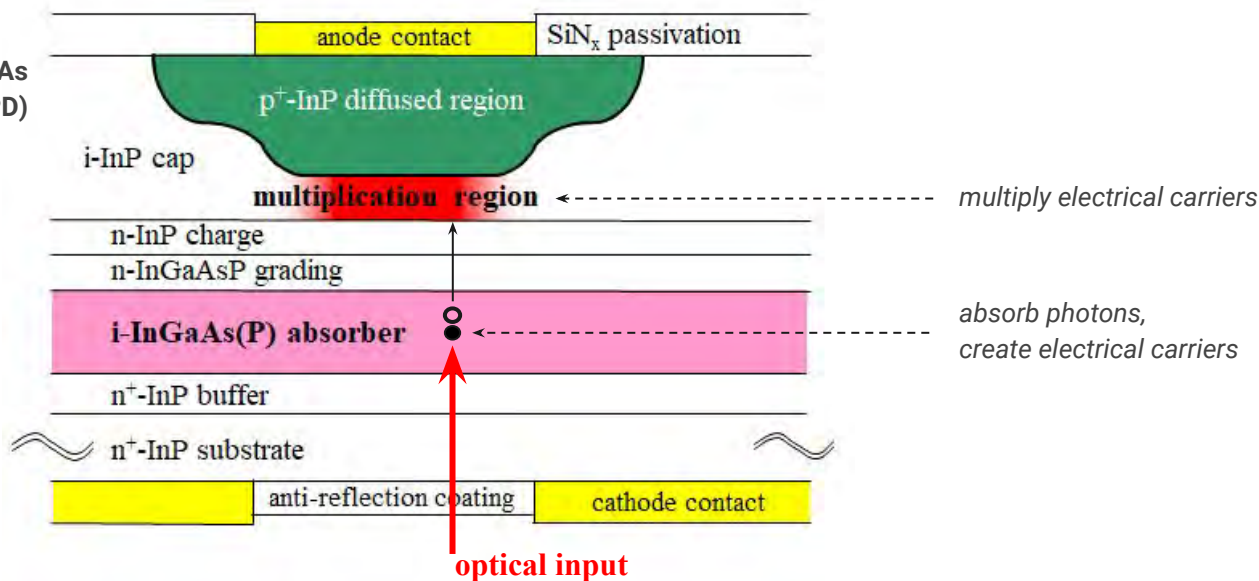
Scaling to 2D SPAD arrays

SPADs scalable to large-format arrays → wafer-level scaling

Optical isolation for crosstalk mitigation

GmAPD pixel

GmAPD structure

Indium bump

anode contact   SiN$_x$ passivation

p$^+$-InP diffused region

i-InP cap

multiplication region

n-InP charge

n-InGaAsP grading

i-InGaAs(P) absorber

n$^+$-InP buffer

n$^+$-InP substrate

anti-reflection coating   cathode contact

32 x 32 GmAPD array
100 μm pitch

SPAD Focal Plane Array Integration: 32 x 32
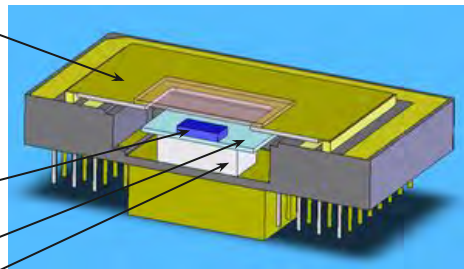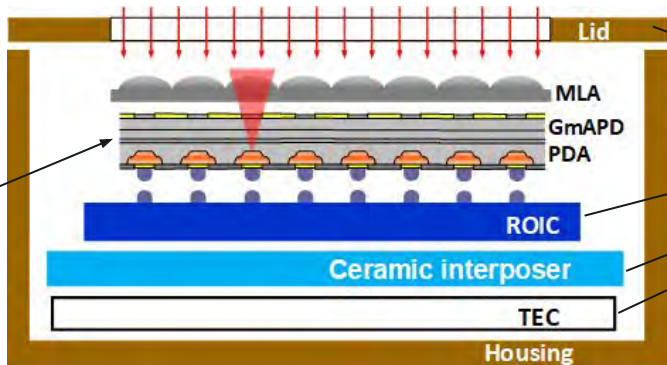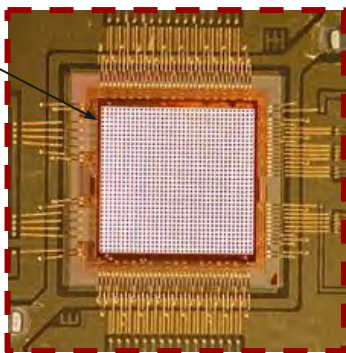
Focal plane array (FPA) integration:

- GaP microlens array (MLA)
- InP GmAPD photodiode array (PDA)
- CMOS readout integrated circuit (ROIC)

32 x 32 PDA

Lid

MLA
GmAPD
PDA

ROIC

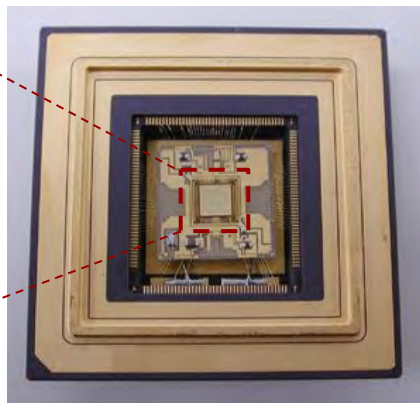Ceramic interposer

TEC

Housing

FPA solid body model

FPA chip stack on interposer
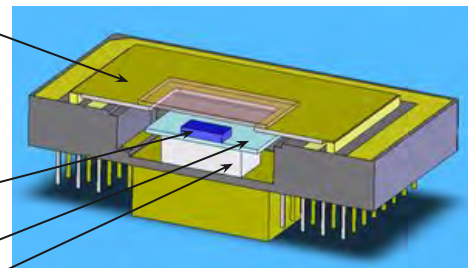(MLA on top)

Full FPA assembly
(no lid)

100 µm pitch

SPAD Focal Plane Array Integration: 128 x 32

Same FPA platform for 128 x 32

- Smaller 50 μm pixel pitch

Lid
MLA
GmAPD
PDA
ROIC
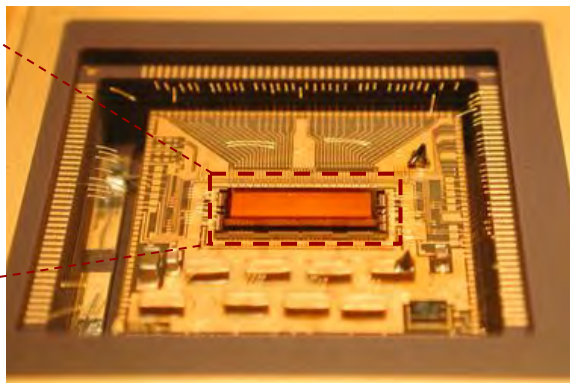Ceramic interposer
TEC
Housing
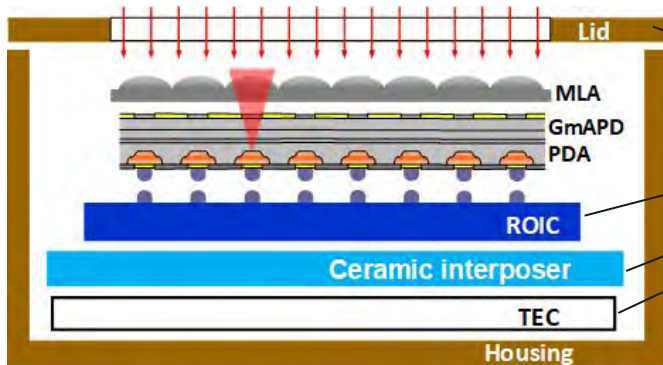
FPA solid body model

128 x 32 FPA chip stack on interposer (MLA on top)
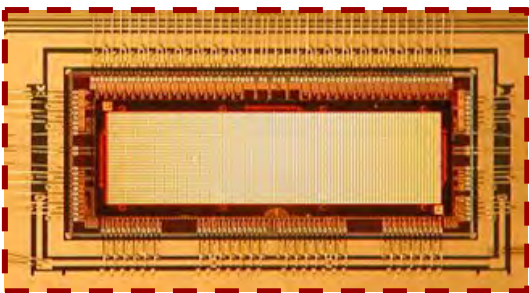
50 μm pitch

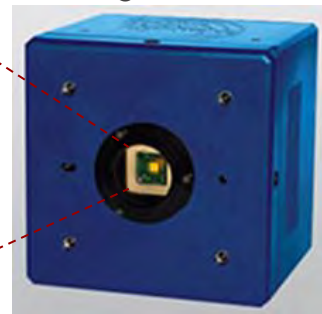**128 x 32 FPA**

Integrated camera

Geiger-mode lidar feasibility proof for automotive

Demonstrate real-time 3D imagery from Geiger-mode lidar using PLI camera

- InP SPAD Camera
- Laser Optics
- Detector Optics
- 1550 nm Fiber Laser
- Scanner & Laser Electronics
- Scanner (Laser)
- Scanner (Detector)
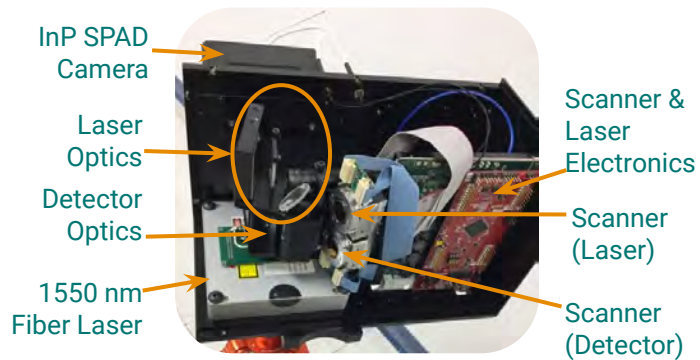
Emulated auto lidar laser/detector resources to demonstrate:

- 50° (H) x 7.5° (V) FOV
- 0.1° x 0.1° H/V angular resolution
- 512 x 64 image format
- ~150 samples/image pixel
- ~5 x 10^6 raw data pts/image
- Real-time coincidence processing
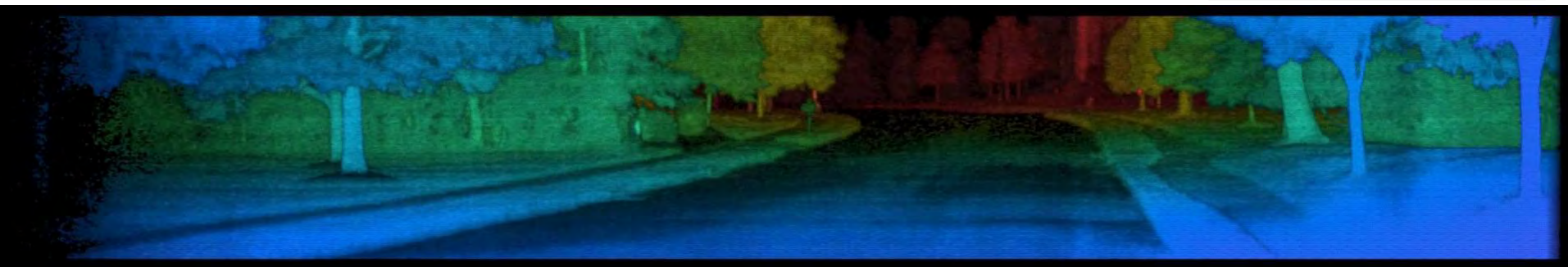
Presentation Outline

- Introduction

- Review of legacy cooled 2D SPAD camera focal plane arrays (FPAs)

- **Migration from cooled SPAD cameras to automotive lidar SPAD FPAs**
  - **Performance distributions and scaling to high temperature operation**

- Reliability of InP-based SPADs

- Implementation comparison of Geiger-mode lidar and linear-mode lidar

- Summary

Migration from SPAD Cameras to Auto Lidar

InP-based SPAD Cameras	InP-based SPAD Automotive Lidar
2D Format (32x32, 128x32)	1D Format
Staring, 1D Scan, 2D Scan	1D Scan
Back-illuminated SPADs	Front-illuminated SPADs
Moderately Cooled FPA (e.g., -25 °C)	Temperature-stabilized FPA (e.g., +40 °C)
Military-style assembly (indium-bump flip-chip bonding)	Low-cost assembly (e.g., wire-bonding)
Large Pitch: 100 μm, 50 μm	Reduced Pitch: 25 μm
Mil-spec Reliability	Automotive Reliability (AEC-Q10X / JEDEC)

Migration from SPAD Cameras to Auto Lidar

InP-based SPAD Cameras	InP-based SPAD Automotive Lidar
Princeton Lightwave	ARGO AI
2D Format (32x32, 128x32)	1D Format
Staring, 1D Scan, 2D Scan	1D Scan
Back-illuminated SPADs	Front-illuminated SPADs
Moderately Cooled FPA (e.g., -25 °C)	**Temperature-stabilized FPA (e.g., +40 °C)**
Military-style assembly (indium-bump flip-chip bonding)	Low-cost assembly (e.g., wire-bonding)
Large Pitch: 100 µm, 50 µm	Reduced Pitch: 25 µm
Mil-spec Reliability	**Automotive Reliability (AEC-Q10X / JEDEC)**

Confirm high-temp scaling of dark count rate

- **First commercial use of InP-based SPAD operating above room temperature**
  - Need to confirm dark count rate (DCR) and photon detection efficiency (PDE) are consistent with operational requirements

- **First InP SPAD arrays with pixel pitch reduced to 25 μm**
  - Drive down cost with smaller array dimensions

- **Compare best-in-class InP SPAD camera performance with new Auto Lidar SPAD FPAs**
  - Confirm expected DCR scaling for different operating temperature and active area size

Migrate to front-illuminated SPAD design

- Legacy back-illuminated structure **modified for front-illuminated operation**
  - Facilitates much more scalable FPA assembly, e.g., using wire-bonding
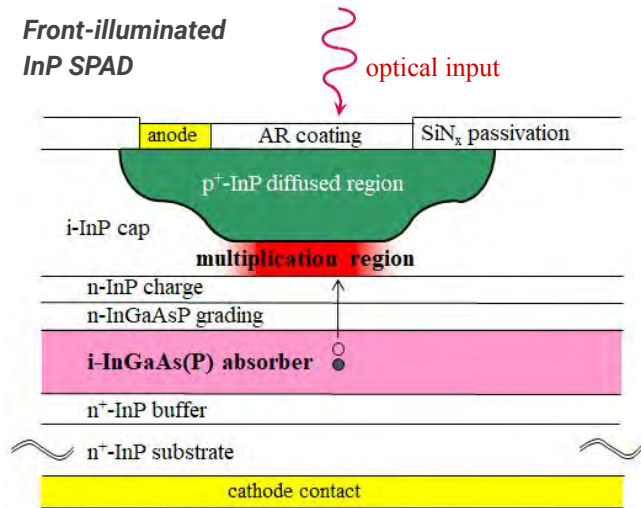
- Reduce device structure dimensions to **reduce pixel pitch from 50 µm to 25 µm pitch**
  - Key focus on crosstalk mitigation

Back-illuminated InP SPAD
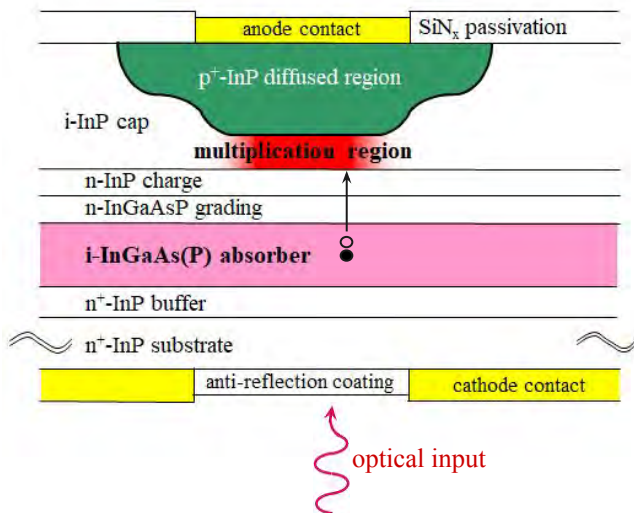
Front-illuminated InP SPAD

Dark count rate mechanisms

- Two primary mechanisms contributing to dark count rate (DCR) in InP/InGaAs SPADs
  - **Trap-assisted tunneling (TAT)** and **thermal generation** can both be important

- Only temperature dependence of TAT is modest decrease in bandgap E_g(InP) as T increases
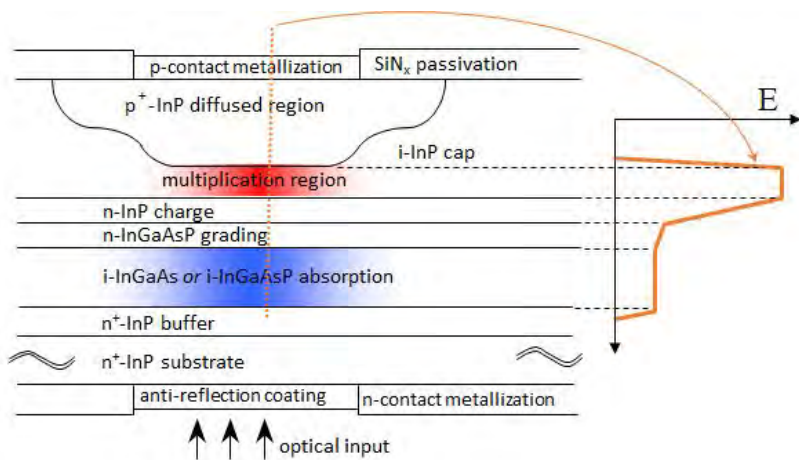
Trap-assisted tunneling in large-bandgap (~1.35 eV) InP multiplier with high field

Thermal generation in small-bandgap (~0.77 ev) InGaAs absorber with low field
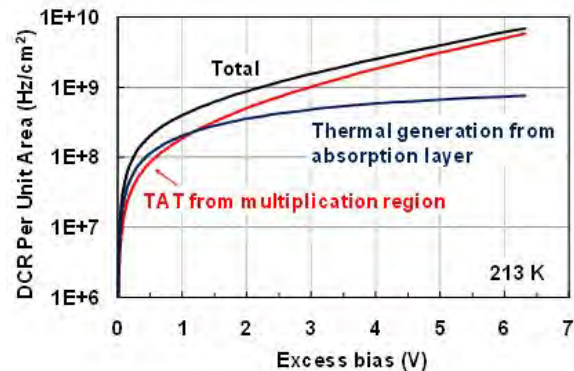
X. Jiang, et al., *Proc. of the SPIE* **6771**, 677127 (2007)

Dominance of thermal generation beyond 250K

- **Strong temperature dependence of Shockley-Read-Hall generation** across InGaAs bandgap via mid-gap states

- Modeling shows transition of activation energy **from TAT to thermal generation from ~210K to ~250K**
  - Experimentally verified by DCR(T) data

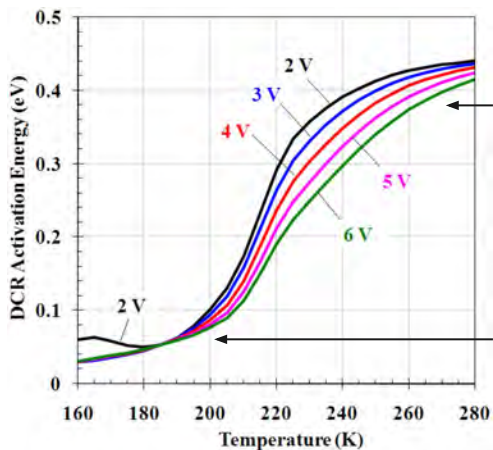- Operation **beyond ~250K dominated by thermal generation**

For mid-gap trap states at energy E_t,
$E_t - E_v \cong E_c - E_t \cong E_g/2$

$$DCR_{thermal} \propto \exp\left(\frac{-E_g/2}{kT}\right)$$

Dominated by thermal generation

For InGaAs/InP SPADs,
E_g(InGaAs, 280K) ~ 0.75 eV
so $E_g/2$ ~ 0.38 eV

Dominated by TAT

M. A. Itzler, et al., "Advances in InGaAsP-based avalanche diode single photon detectors," *J. Modern Optics* **58**, Nos. 3–4, 174 – 200 (2011).

DCR and PDE distributions from cooled FPAs

- Examples of **DCR and PDE distributions for best-in-class InGaAs/InP 128 x 32 SPAD FPAs** with 50 μm pitch
  - FPA-level data (includes microlens array)

FPA 1223
-25 °C

<DCR> = 4.1 kHz
<PDE> = 20.5%
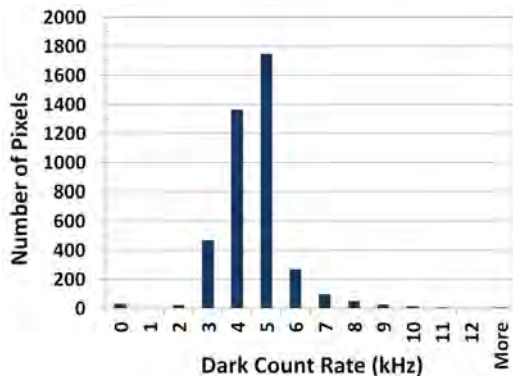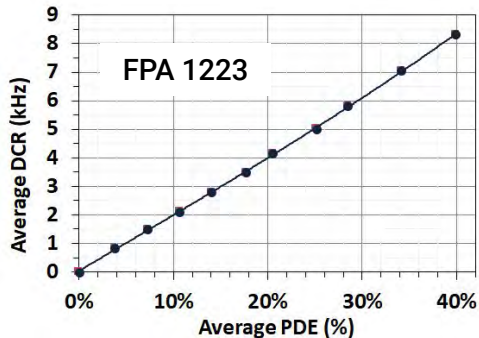
FPA 1226
-20 °C

<DCR> = 3.6 kHz
<PDE> = 21.5%
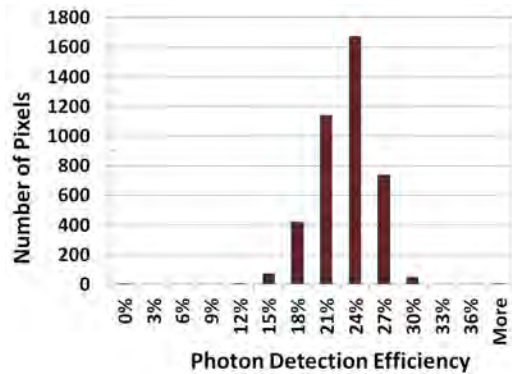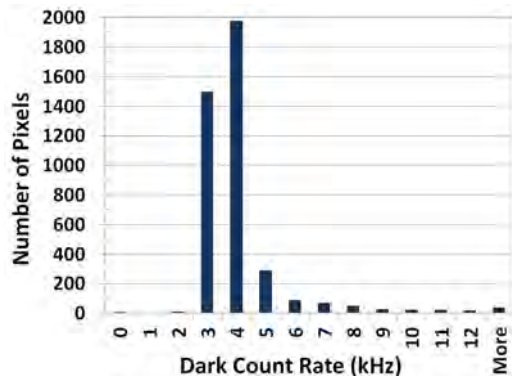
Calculation of expected DCR scaling with T

- **Estimate expected DCR increase** at temperatures above typical camera FPA operation (e.g., -25°C)

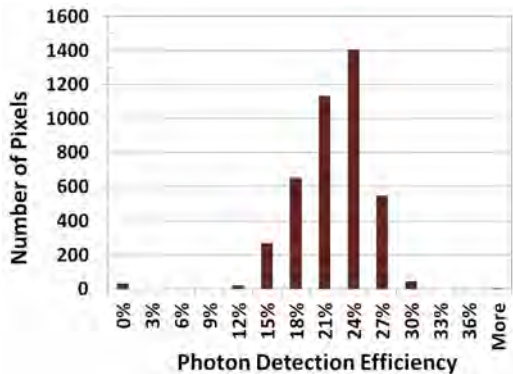 - Beyond exponential scaling with kT, must also consider temperature dependence of E_g(InGaAs)

$$\exp\left(\frac{-E_g/2}{kT}\right) \rightarrow \exp\left(\frac{-E_g(T)/2}{kT}\right)$$

- DCR ratio $DCR(T_2)/DCR(T_1)$ has **factor of 63X increase for T_2 = 40°C operation** at relative to T_1 = -25°C
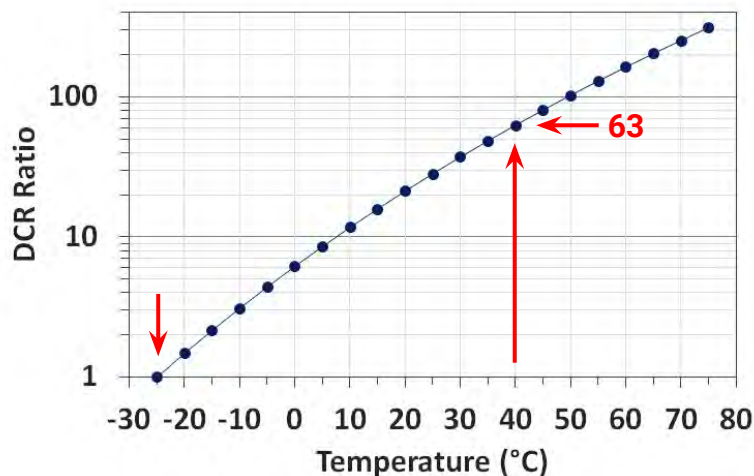
$$\frac{DCR(T_2)}{DCR(T_1)} = \frac{\exp\left(\frac{-E_g(T_2)/2}{kT_2}\right)}{\exp\left(\frac{-E_g(T_1)/2}{kT_1}\right)}$$

$E_g(T)$ for $In_{0.53}Ga_{0.47}As$: S. Paul, et al., J. Appl. Phys. **69**, p. 827 - 829 (1991).

Near 0°C, the ~0.75 eV bandgap decreases by ~0.5% for every +10°C.

DCR and PDE distributions for new FPAs at +40°C

- Examples of **DCR and PDE distributions for recent prototype 1D FPAs** with 25 µm pitch
  - DCR ~ 100 - 200 kHz is **well below solar noise background**

FPA A
+40°C

<DCR> = 204 kHz
<PDE> = 21.3%
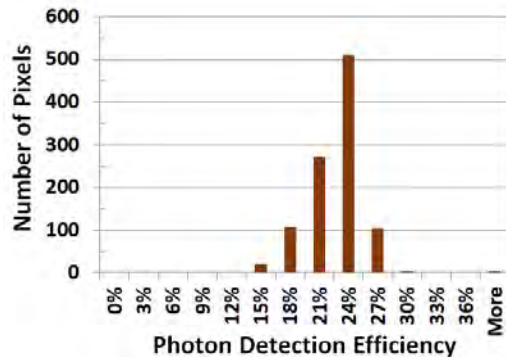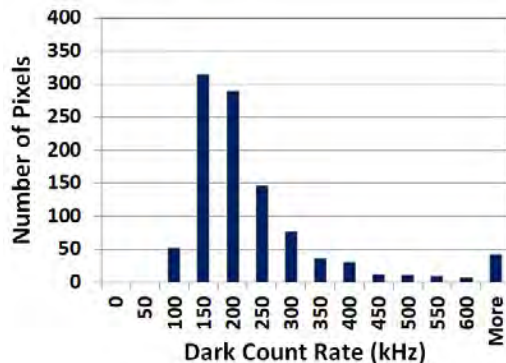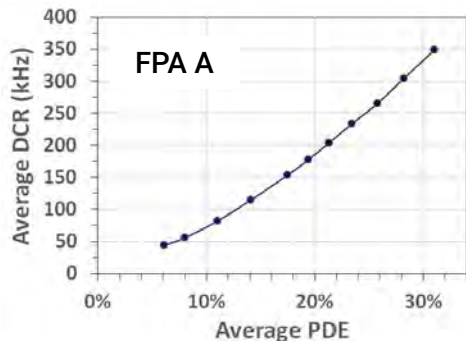
FPA B
+40°C

<DCR> = 118 kHz
<PDE> = 21.8%
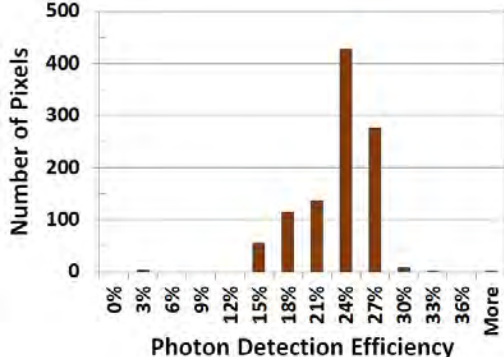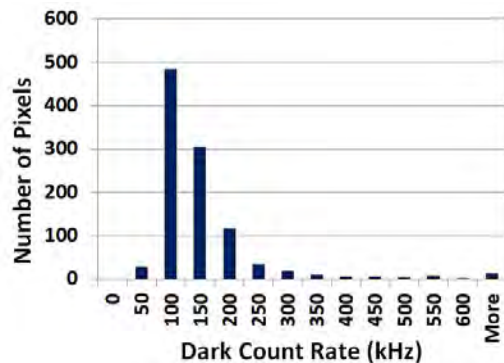
DCR scaling for 2D cooled and 1D new FPAs

Device	Operating Temperature	Avg PDE	Avg DCR	Temp Scale Factor to +40°C	DCR for Same Area	Active Area Scale Factor	**Expected DCR** (Temp and Area Scaling)
FPA 1223	**-25 °C**	20.5%	**4.1 kHz**	63	258 kHz	~0.7	**180 kHz**
FPA 1226	**-20 °C**	21.5%	**3.6 kHz**	43	155 kHz	~0.7	**110 kHz**

Device	Operating Temperature	Avg PDE	**Avg Measured DCR**
FPA A	**+40 °C**	21.3%	**204 kHz**
FPA B	**+40 °C**	21.8%	**118 kHz**

- **DCR scaling** of legacy cooled FPAs for temperature and active area **predicts DCR of 100 - 200 kHz**

- Measured data for **new prototypes operating at +40 °C are consistent with expected scaling**

Consistency of distribution attributes

Beyond agreement of average DCR scaling, obtain **quantitatively similar distributions** for old and new FPAs

- **DCR** histograms contain **most pixels within factor of ~2X - 3X** with exponential tail at higher DCR

- PDE histograms have approximately normal distribution with most pixels within ±3% of mean PDE
    - Asymmetry in FPA B is directly related to spatial variation at right side of FPA
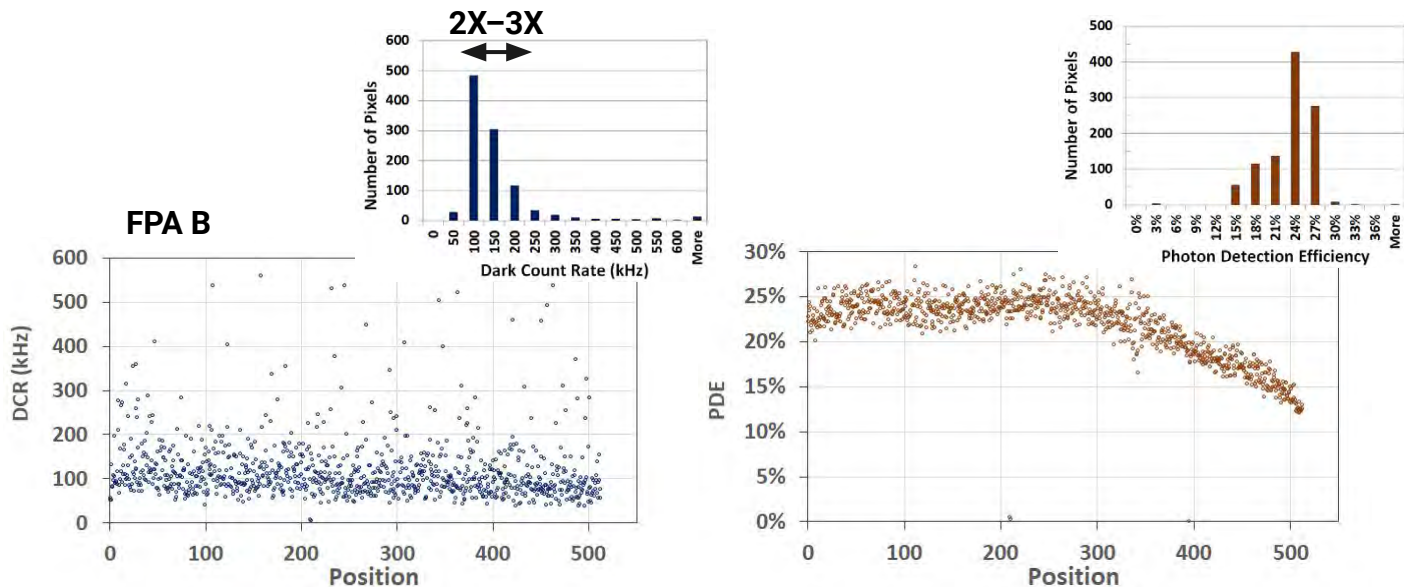
Presentation Outline

- Introduction

- Review of legacy cooled 2D SPAD camera focal plane arrays (FPAs)

- Migration from cooled SPAD cameras to automotive lidar SPAD FPAs
  - Performance distributions and scaling to high temperature operation

- **Reliability of InP-based SPADs**

- Implementation comparison of Geiger-mode lidar and linear-mode lidar

- Summary

Reliability of InP-based SPADs

- **Lack of industry standards for reliability of SPAD operation**
  - Linear-mode APD reliability testing is mature (e.g., for telecom)
  - Question as to whether Geiger-mode operation induces different failure mechanism(s)

- **Only reported reliability work on InP-based SPADs is MIT Lincoln Lab study** from ~10 years ago
  - Failure rate model involving total charge flow, but may be specific to mesa devices

- **Silicon SPAD/SiPM producers have adopted JEDEC standards or similar methodologies**

- **Argo initiating highly accelerated aging tests** at higher temperatures (cf. telecom)
  - Initial studies using discrete test devices to define limits
  - Will adopt JEDEC standards, relevant to automotive qualification (e.g., AEC Q10X)

MIT-LL study of InP-based SPAD reliability

- No degradation with 1000 hours of linear-mode aging (V_b - 2), but **rapid degradation in Geiger-mode** (4 V overbias)
  - Small change in E-field intensity → consider different mechanism

- **Hypothesized failure rate dependent on total charge flow** → accelerate aging by inducing more charge per pulse
  - To accelerate aging, used 300 ns gating pulses at 30 kHz rate (~1% duty cycle)
  - From three lots, found E_a ~ 0.8, 0.8, and 0.45 eV
  - Ultimately estimated 500 FIT reliability for 23 °C operation
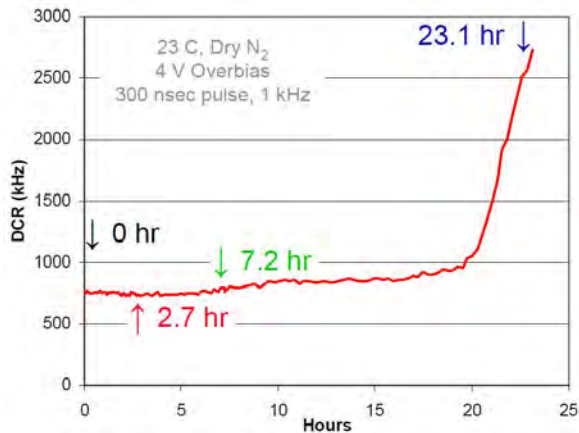
$$R_{fail} \propto I \times \exp\left(\frac{-E_a}{kT}\right)$$

Hypothesized failure mechanism:

- surface charge accumulation at mesa passivation interface
- charge injection into avalanche region
- consequent increase in DCR

Possibly specific to mesa structures?

G. Smith, et al., "Reliable InP-based Geiger-mode Avalanche Photodiode Arrays", *Proc. of the SPIE* **73200R** (2009)

Si SPAD approach to reliability testing

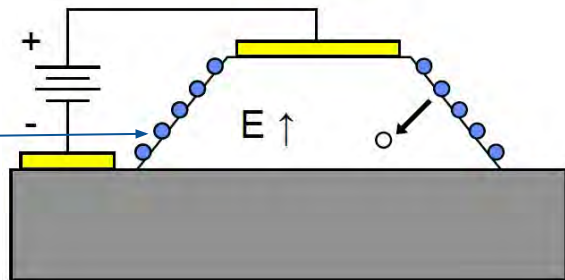- Lack of reliability standards for SPADs → **Si SPAD community leveraging integrated circuit standards**
  - e.g., JEDEC standard conditions for Temperature, Bias, and Operating Life (JESD22-A108F)
  - Accelerated aging for 1000 hours high-temperature operating life (HTOL) and other stress tests

"…standard reliability stress procedures do not exist for SiPM…adopted integrated circuit industry standard tests… carried out as per Joint Electron Device Engineering Council (JEDEC) standard conditions.

Table 3   Reliability stress test program for SensL B-Series sensors.

Test	Objective	Required condition	Lot size	Duration/acceptance	Status
High-temperature operating life	Junction stability	Ambient temperature = 125°C; bias = 30 V	3 lots of 77 units	1000 h/no change in any parameter > 10%	100% Pass
High-temperature operating life	Junction stability over longer stress time	Ambient temperature = 125°C; bias = 27 V	256 units	2000 h/no change in any parameter > 10%	100% Pass
High-temperature operating life	Package stress to examine chemical stability (e.g., discoloration of package)	Ambient temperature = 85°C; bias = 27 V	1 lot of 77 units	1000 h/no change in any parameter > 10%	100% Pass
Unbiased highly accelerated stress	Package stress to examine delamination, transmission loss and wire bond failure	110°C, 85% relative humidity; passive no bias	3 lots of 25 units	264 h/no change in any parameter > 10%; no critical package delamination	100% Pass
Temperature cycling	Package stress to examine delamination, transmission loss and wire bond failure	−40°C to 85°C cycle, 15 s transition, 15 min dwell time; passive no bias	3 lots of 77 units	500 cycles/no change in any parameter > 10%; no critical package delamination	100% Pass
High-temperature storage test	Package stress to examine chemical stability (e.g., discoloration of package)	504 h at 125°C; passive no bias	3 Lots of 25 units	504 h/no change in any parameter > 10%	100% Pass

Carl Jackson, et al., "High-volume silicon photomultiplier production, preformance, and reliability," *Optical Engineering* **53**, 081909 (2014).

1000 hrs at 60 C accelerated aging "considered to be representative of the lifetime of the product in a consumer application"
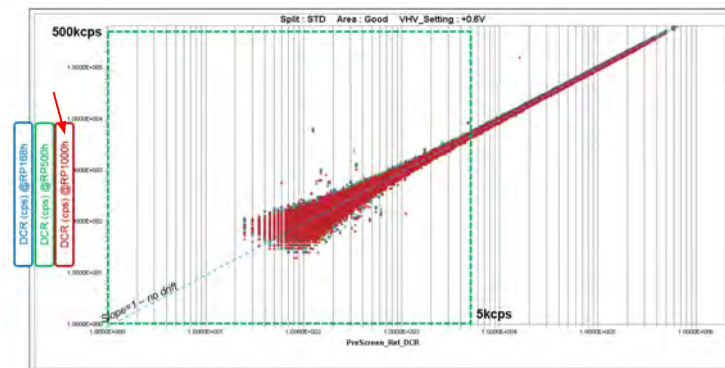

Figure 8. DCR evolution of each SPAD between PreScreen and stress test point. Different colours correspond to different testing point. Most of the SPADs sit along the line with a slope of 1.

Sara Pellegrini and Bruce Rae, "Fully industrialised single photon avalanche diodes," *Proc. of the SPIE* **102120D** (2017).

Aggressive initial high-temp aging study

- **Use 0.1 mA constant current (DC) bias → significant fraction of ns-scale avalanche peak currents of ~1 mA**
  - Difficult to implement full FPA Geiger-mode operation at 175 °C

- **Highly accelerated aging at 175 °C with 0.1 mA DC current** to obtain initial reliability feedback
  - Assume E_a = 0.7 eV as representation of metal ion electromigration
  - 175 °C provides factor of ~10X acceleration relative to 125 °C aging

- Assuming E_a = 0.7 eV, **every 100 hours at 175 °C is ~28 years of 40 °C operation**

E_a	T_1	T_2	Acceleration factor A_T	Years at T_1 per 100 hours at T_2
0.7 eV	**40 °C**	**175 °C**	2491	**28.4**
0.7 eV	125 °C	175 °C	9.8	-

$$R_{\text{fail}} \propto \exp\left(\frac{-E_a}{kT}\right)$$

$$A_T = \frac{\exp\left(\frac{-E_a}{kT_2}\right)}{\exp\left(\frac{-E_a}{kT_1}\right)} = \exp\left(\frac{-E_a}{k}\left(\frac{1}{T_2} - \frac{1}{T_1}\right)\right)$$

A_T: acceleration factor with temperature
E_a: apparent activation energy
R_{fail}: failure rate
k: Boltzmann constant ~ 8.617 x 10^{-5} eV/K

- Following up with lower temperature (125°C) aging for 1000+ hours

- Also need to confirm no new failure mechanisms introduced at 175 °C

Initial high-temp aging results

- **Discrete SPAD test devices with design and processing equivalent to PDA pixels**

  - Group 1 and Group 2 devices are representative of pixels on PDAs described earlier but had modest differences resulting in higher DCR

- **Aging** Conditions: **175 °C, 0.1 mA DC** current

- **DCR Test** Conditions: **25 °C, 2.5 V** overbias

- **120 hours at 175 °C** equivalent to **34 years operation at 40 °C** ($E_a = 0.7$ eV)

  - **Breakdown voltage essentially constant**

  - **Observed DCR increase is within acceptable bounds** of operating requirements (solar background can be multi-MHz)
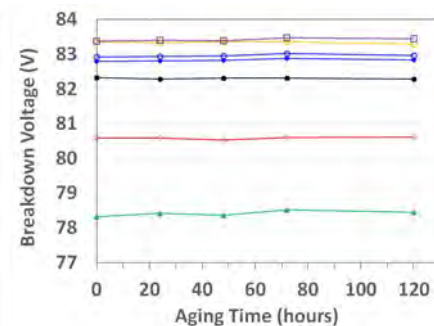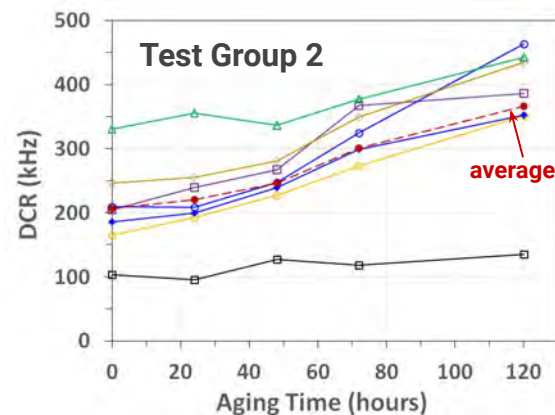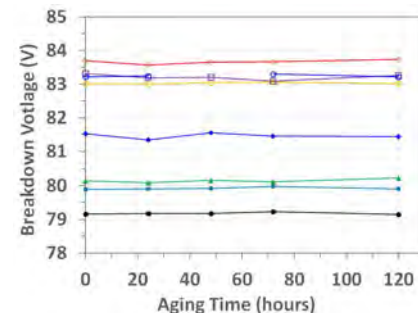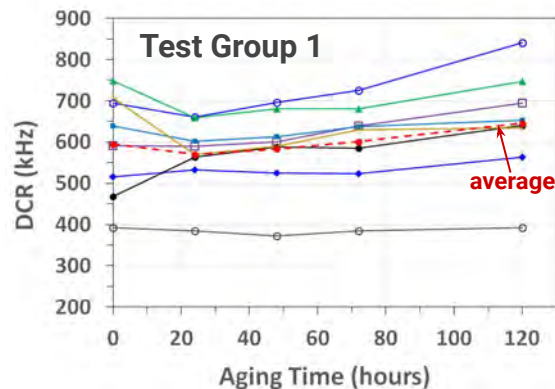
Presentation Outline

- Introduction

- Review of legacy cooled 2D SPAD camera focal plane arrays (FPAs)

- Migration from cooled SPAD cameras to automotive lidar SPAD FPAs
  - Performance distributions and scaling to high temperature operation

- Reliability of InP-based SPADs

- **Implementation comparison of Geiger-mode lidar and linear-mode lidar**

- Summary

Argo Geiger-mode lidar vs. linear-mode lidar

- Distinctions of Geiger-mode lidar design approach from most "legacy" lidars

Argo Geiger-mode lidar	"Legacy" linear-mode lidar
Sensitive to **single photons** (Geiger mode)	Sensitive to **10s - 100s photons** (Linear mode)
Statistical sampling of scene	**Single-shot image** of scene
Short-wave infrared (**>1400 nm**)	Near infrared (~**900 nm**)
Arrays of lasers and detectors	**Discrete** lasers and detectors
Space-filling ("gapless") imaging	**Gaps** between imaged points
Design entire lidar down to **custom chips**	Integrate **commodity components**

Single-photon sensitivity w/ Geiger-mode lidar

- **Single-photon sensitivity** provided by **Geiger-mode** avalanche photodiodes (GmAPDs)
  - Can measure 1 photon

Geiger-mode lidar can see further with a given laser energy
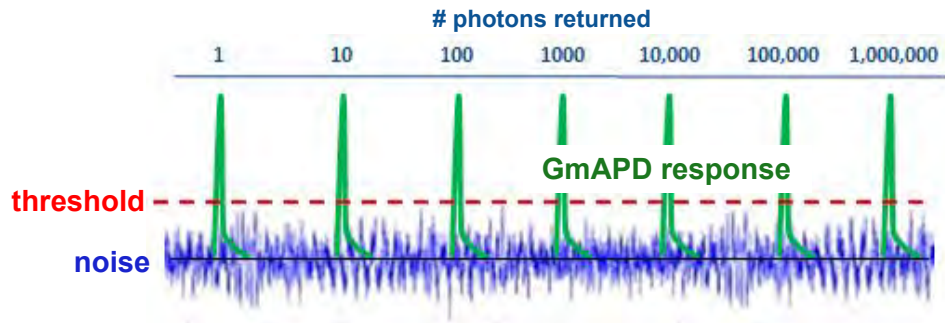
- Most **commercial lidars** employ **linear-mode** avalanche photodiodes (LmAPDs)
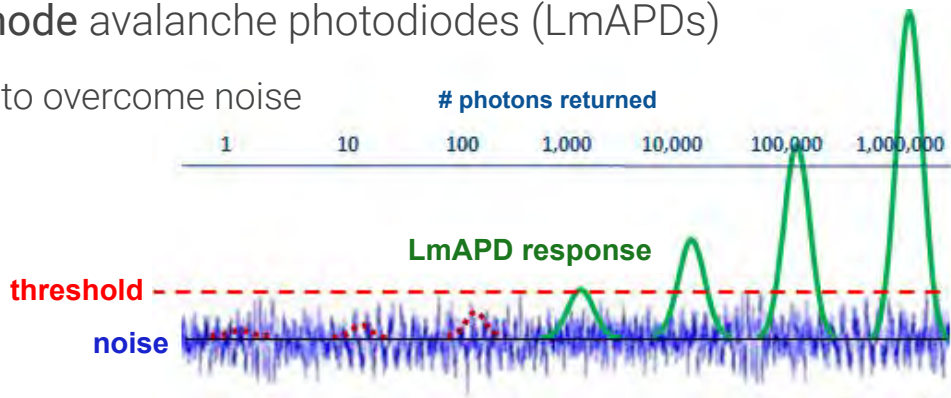 - **Needs 10s - 100s of photons per return** to overcome noise

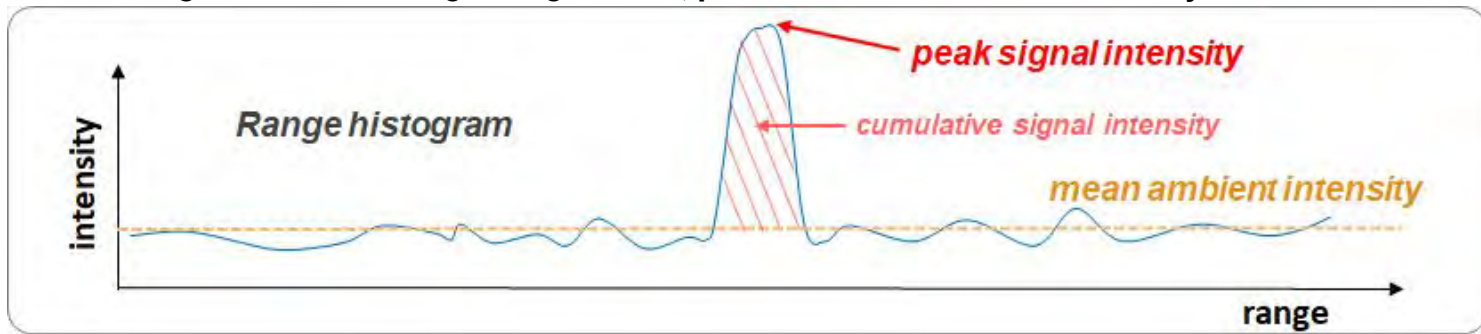| Argo Geiger-mode lidar	"Legacy" linear-mode lidar
Sensitive to **single photons** (Geiger mode)	Sensitive to **10s - 100s photons** (Linear mode)
Statistical sampling of scene	**Single-shot image** of scene
Short-wave infrared (**>1400 nm**)	Near infrared (~**900 nm**)
Arrays of lasers and detectors	**Discrete** lasers and detectors
Space-filling ("gapless") imaging	**Gaps** between imaged points
Design entire lidar down to **custom chips**	Integrate **commodity components**

Statistical sampling with Geiger-mode lidar

- **Statistical analysis** of time-of-flight returns with high repetition-rate pulses

  - Each point cloud point can be created from **~100 - 200 samples** (**temporal and spatial**)

  - **Correlated counts** provide high detection probability for signal returns

Consecutive image frames

Photons incident on LiDAR aperture

aperture   ambient light   laser pulse

Build histogram of time-of-flight range values, peaks indicate reflections from objects

peak signal intensity

Range histogram

cumulative signal intensity

intensity

mean ambient intensity

range

temporal averaging

spatial averaging

Argo Geiger-mode lidar	"Legacy" linear-mode lidar
Sensitive to **single photons** (Geiger mode)	Sensitive to **10s - 100s photons** (Linear mode)
Statistical sampling of scene	**Single-shot image** of scene
Short-wave infrared (>**1400 nm**)	Near infrared (~**900 nm**)
Arrays of lasers and detectors	**Discrete** lasers and detectors
Space-filling ("gapless") imaging	**Gaps** between imaged points
Design entire lidar down to **custom chips**	Integrate **commodity components**

Short-wave infrared Geiger-mode LiDAR

- **More common 905 nm LiDAR constrained by eye-safety limits**
  - Allows use of GaAs laser diodes and Silicon detectors, but...
  - **...inadequate long-range detection (<200 m)** for low-reflectance objects

- **Greater eye-safety with longer wavelength (>1400 nm) detectors**
  - Detection using InP-based compound semiconductors
  - Allows use of more laser energy to **achieve detection range >>200 m**

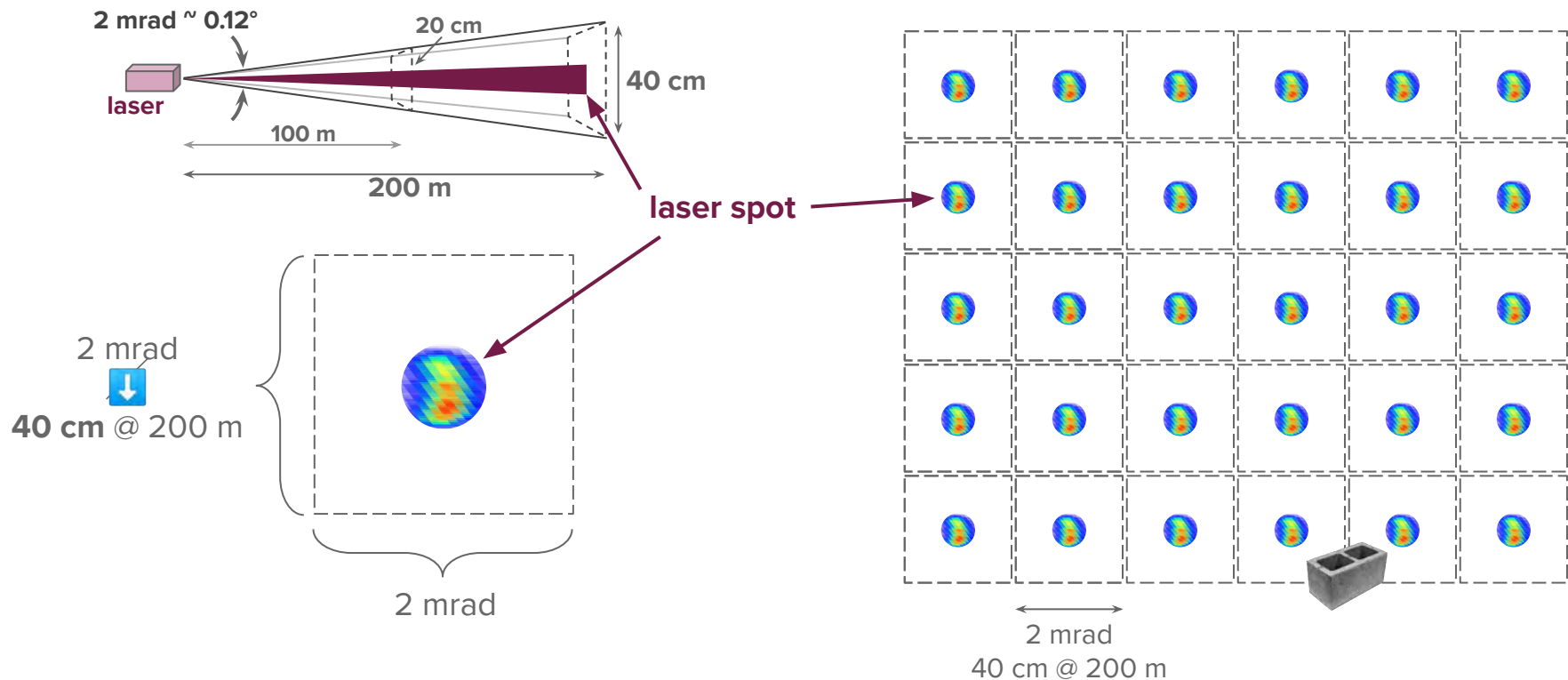> **Greater eye-safety for >1400 nm allows longer range operation**

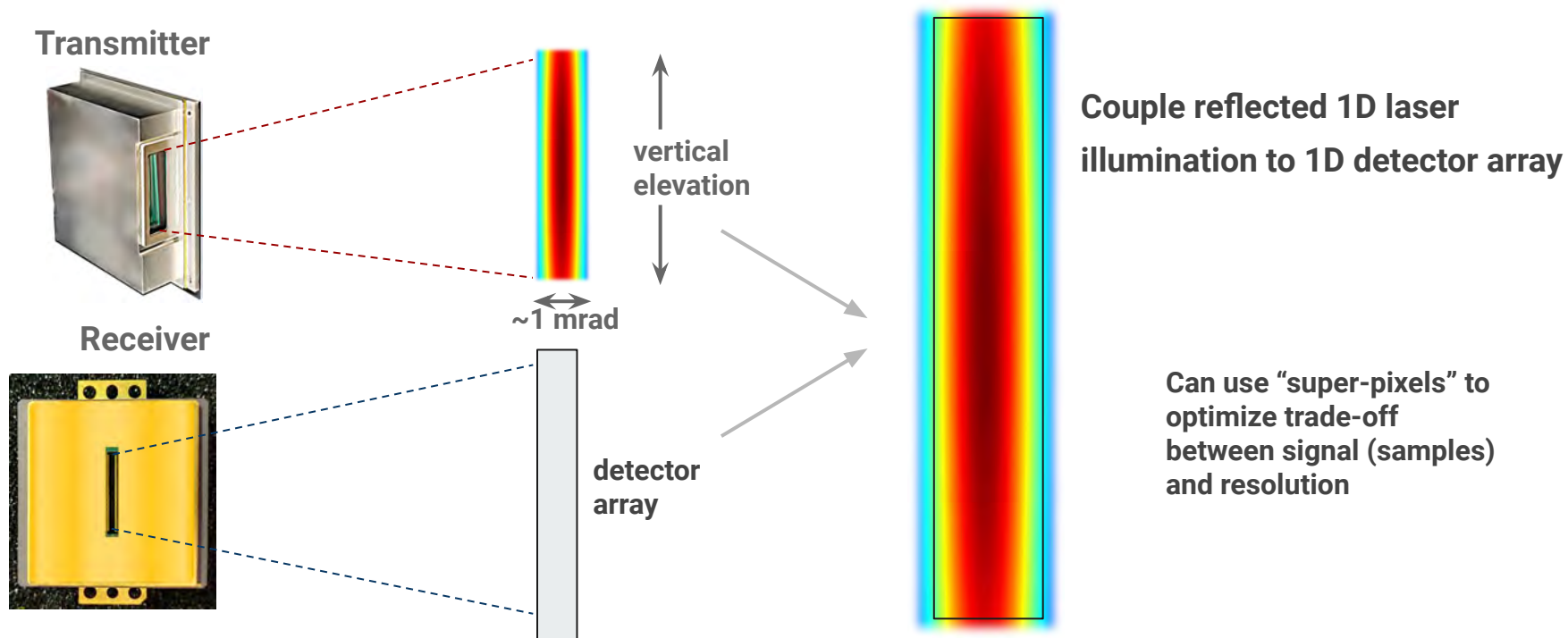| Argo Geiger-mode lidar	"Legacy" linear-mode lidar
Sensitive to **single photons** (Geiger mode)	Sensitive to **10s - 100s photons** (Linear mode)
Statistical sampling of scene	**Single-shot image** of scene
Short-wave infrared (**>1400 nm**)	Near infrared (~**900 nm**)
Arrays of lasers and detectors	**Discrete** lasers and detectors
Space-filling ("gapless") imaging	**Gaps** between imaged points
Design entire lidar down to **custom chips**	Integrate **commodity components**

Sparse imaging

Sparse laser spots leave large non-illuminated (and non-imaged) "gaps" in scene

Space-filling ("gapless") imaging

Vertical space-filling using laser array illumination and detector array sensing

Transmitter

vertical elevation

~1 mrad

Receiver

detector array

Couple reflected 1D laser illumination to 1D detector array

Can use "super-pixels" to optimize trade-off between signal (samples) and resolution

| Argo Geiger-mode lidar	"Legacy" linear-mode lidar
Sensitive to **single photons** (Geiger mode)	Sensitive to **10s - 100s photons** (Linear mode)
Statistical sampling of scene	**Single-shot image** of scene
Short-wave infrared (**>1400 nm**)	Near infrared (~**900 nm**)
Arrays of lasers and detectors	**Discrete** lasers and detectors
Space-filling ("gapless") imaging	**Gaps** between imaged points
Design entire lidar down to **custom chips**	Integrate **commodity components**

Design of custom components

- **SPAD arrays at >1400 nm** wavelength are unique to Argo

- **Argo laser diode array** is an innovative solution not available commercially
  - Low current, high power solution suitable for automotive deployment

- Provide significant **flexibility for system-level design**
  - Format of detector array
  - Distribution of power from laser diode array

- If in-house design has minimal value-add, procure third-party components

Argo is the only auto lidar developer with SPAD lidar technology at >1400 nm

To sum up…
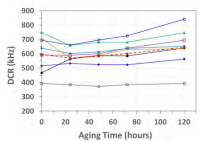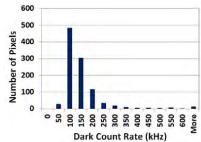
Demonstration of **SPAD performance capability for automotive environment**

   Initial on-vehicle point cloud data is excellent

Initial reliability results consistent with **long life at target operating conditions**

SPAD-based lidar has **significant benefits for automotive platform deployment**

- **Single photon sensitivity** → Potential for sensing objects at **greater range**

- **>1400 nm wavelength** → **greater eye-safety** provides system design flexibility

- **Monolithic arrays** of lasers/detectors → **single alignment** and **gapless imaging**

- **Custom chip design** → greater **flexibility in system functionality and performance**

Acknowledgment

This work is the output of an incredible team of scientists and engineers that I have worked with for many years.

I'm grateful to all of my current colleagues at Argo AI and my former colleagues at Princeton Lightwave.

Thank you!