RDC4 – Chips research directions at Fermilab

Farah Fahim on behalf of Microelectronics Division

Chips Research Directions

- 3D MPW for the research community

- Extreme Environment PDK development

- Sensor co-design initiatives: Tower, Global Foundries, SkyWater

- AI-on-chip

- Cryogenic readout ICs/ Chiplets

- Superconducting circuits

- Photonics

‡ Fermilab

Design Goals for the next 20 years: Enable Smart Sensors

- Novel sensors; CMOS sensors

- Low power, low noise, ultra-sensitive analog for sensor signals

- Advanced digital architectures and verification methods

- AI-on-chip, Quantum cryoelectronics, 6G and beyond

- Integrated high speed communication between modules and off-detector
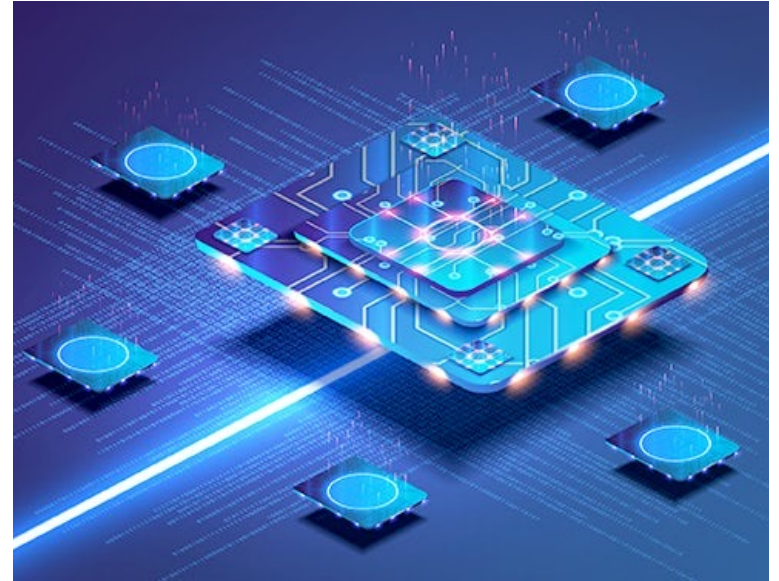
Extremely important to work with industry to leverage production-scale processing

🟦 **Fermilab**

Growth of Integrated Circuit design over ~ 4 decades at Fermilab

- Highly specialized expertise in developing robust custom microelectronics with long-term reliability over decades

- Investigate novel solutions and ensure technology development to enable mid-volume prototyping

- Increased complexity: **2016**: 0.5B transistors in **~cm^2** -> **2021**: 2B transistors in **~mm^2** (shifting the burden from **design to verification**)

Since 1980's

Since 2010's

Ionizing radiation
>1 Grad (1000x higher than outer space)
Extreme flux for single event effects
- Collider Experiments (FCC, HL LHC)

Cryogenic electronics (77K – 100K)
- Neutrino experiments (DUNE), Dark matter experiments (Skipper CCDs)
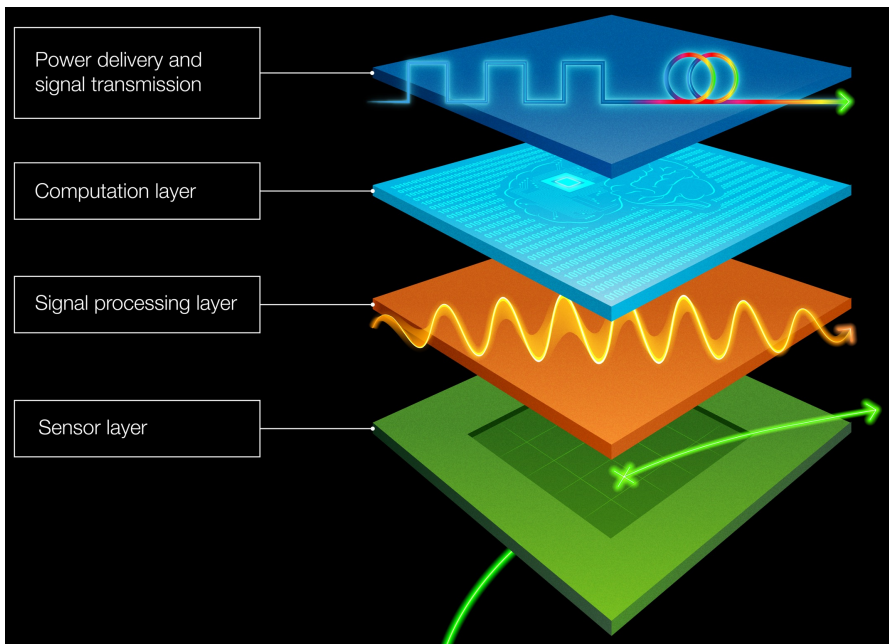
Deep Cryogenic electronics (~ 4K)
- Dark matter experiments (Cryogenic detectors e.g. SNSPDs, TES etc.), Quantum Information Science

Superconducting electronics (~100 mK)
- Quantum Information Science for HEP (TWPAs, JPAs for ADMX)

Since 2019

New 2022

🔷 Fermilab

A PHYSICIST'S DREAM: LOW SWAP, LARGE AREA DETECTORS

Power delivery and signal transmission

Computation layer

Signal processing layer

Sensor layer

- CMOS Sensor Layer: Integrating the sensor in a CMOS process

- Precision Analog Layer: Integrated position ($< 1\mu$m) and timing (~1-5ps) resolution

- At-source energy efficient, high accuracy, low-latency data processing layer

- Photonics layer for power delivery and data transmission

🔬 Fermilab

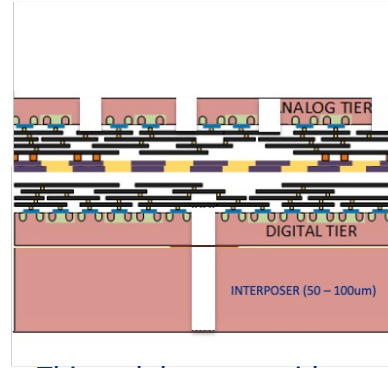# Community Enablement

Fermilab

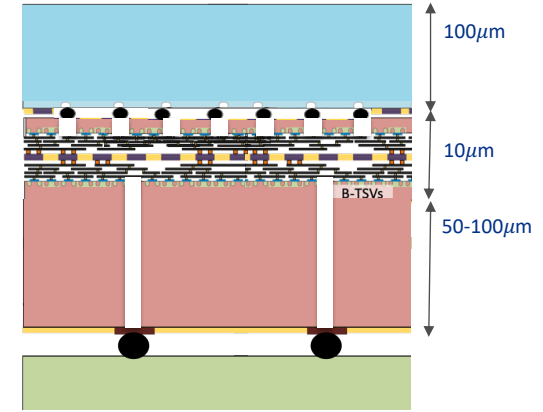# Establishing Chicago 3D Chips Codesign Community C 3D C³
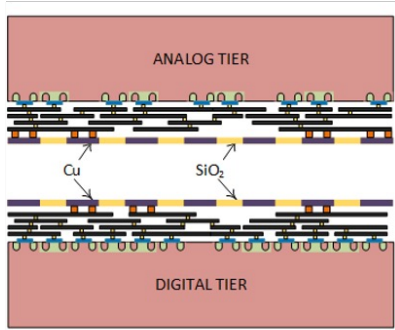
a. Face-2-Face wafer bonding
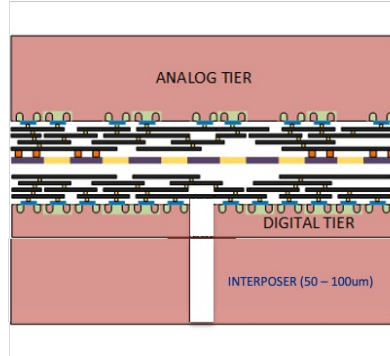
b. B-TSV insertion and bond to TSV interposer wafer

c. Thinned down top side and B-TSV insertion

3D sensor system

- CREATING A COMMUNITY and RELEVANT PARTNERSHIPS
- Create open source ADK (Assembly Design Kits) and distribute to consortium members. Membership agreement is almost ready
- Fermilab assembles MPW work with (IMEC, MUSE and others)
- Currently have a multi party NDA with more than 80 institutions for HEP chips (IMEC-TSMC-CERN-Fermilab-others)
- 1st run: Low cost – same as the cost of silicon; 65 nm; several national labs and university groups have expressed interest and actively working on designs
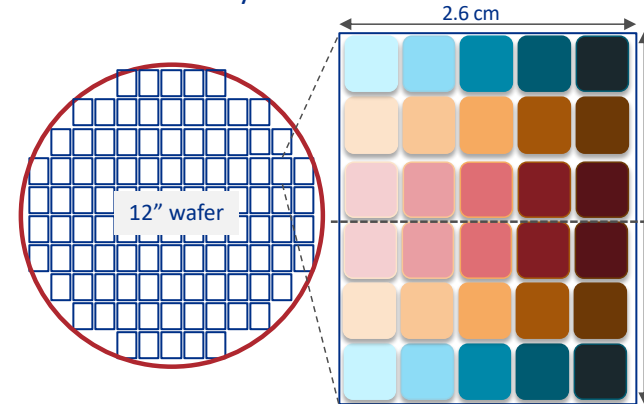
🔷 Fermilab

Testing, Characterization and Modeling
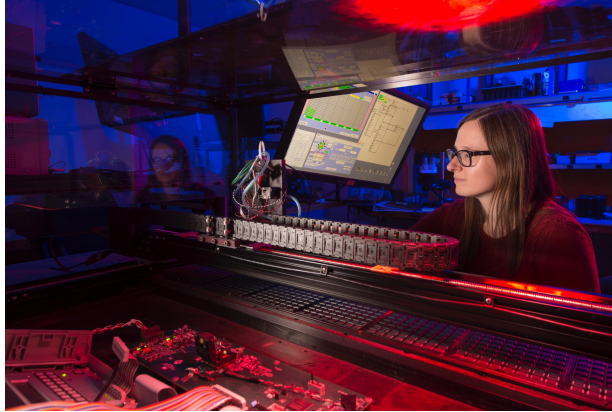
- Testing infrastructure: 12" wafer probing; cryogenic probing

- Robotic testing for mid-volume chip characterization

- Test beam facility at Fermilab

- Development of extreme environment technology models for cryogenic and rad-hard chips

🎇 Fermilab
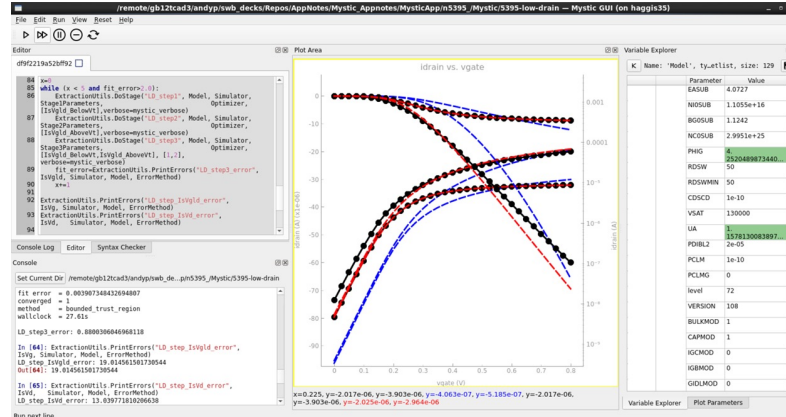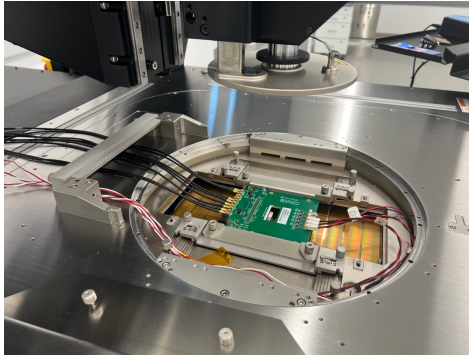
Cryo PDK development with Foundry support

22FDX: (Wafer provided by Global Foundries used for their own room temperature modelling)

In-house:
- Measurement and modeling of high voltage devices at 4K (BOXFET, LDMOS)

With EPFL:
- Measurements of transistors at 4K
- Development of simplified EKV model for analog design
- Low noise test structure measurements

With Synopsys:
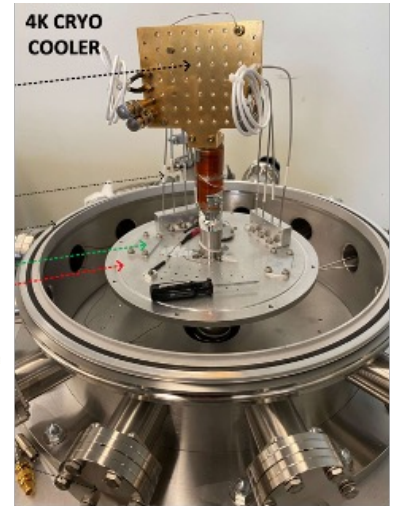- **PDK-compatible BSIM-IMG for 4K**

SiGe 9HP / 28 nm HV, 28 nm HPC+:
With Northwestern University and UTA
- Just starting testing of CMOS test structure for GF 9HP
- DRD 7 collaboration for 28 HPC+

4K CRYO COOLER

GlobalFoundries™

EPFL

SYNOPSYS®
Silicon to Software™

Northwestern University

UTA
THE UNIVERSITY OF TEXAS AT ARLINGTON

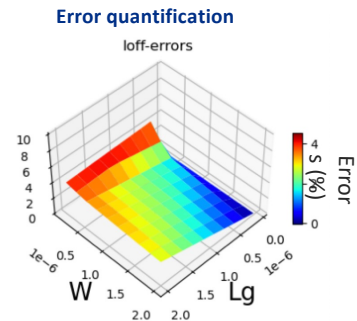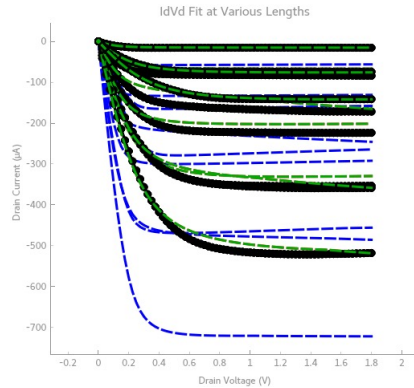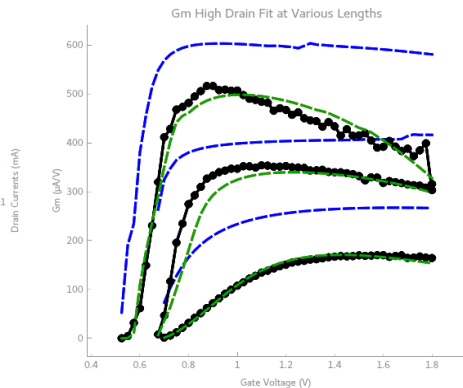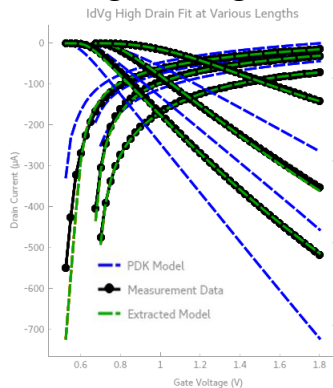🐝 Fermilab

Cryo PDK development with Foundry support

- Nearing completion of BSIM-IMG models for thick-oxide devices
- Measurement data for thin-oxide devices should become available soon
- *Can share with relevant NDAs*

Next steps:
- Targeting completion and **publication** by the end of the year
- 4K timing library for standard cell library next
- Starting to investigate **AI/ML modeling** of devices
- Interested in evaluating layout proximity effects

Awaiting results from:
- **Passive devices**, which would be easier to model
- **Ring oscillators**, to verify timing models
- **High-voltage devices,** to cross check our own test structures

Error quantification

Ioff-errors

Vt-errors

Ion-errors

Custom models for dgnfets

🟦 Fermilab

Sensor Co-design

Fermilab
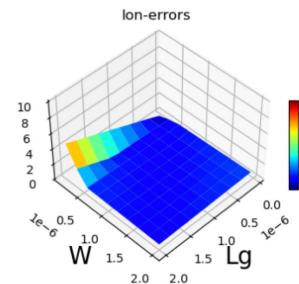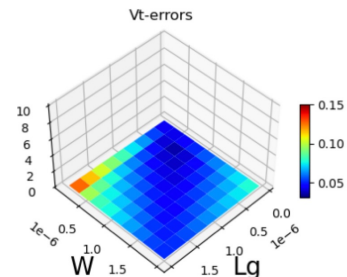
Sensor Co-design with Tower Semiconductor (3D LGADs)

- Goal: Create **LGADs** in 12" CMOS wafers which can be 3D integrated (hybrid bonding) with readout circuits in 28 nm

  Various readout architectures in 28 nm (primary development) and 65 nm (previous IP) process

  Leading edge discrimination (ETROC), Constant Fraction discrimination (FCFD), Waveform sampling architectures (with U. Chicago)

 Low power, picosecond resolution TDCs (ETROC, DILVERT)

  Timing distribution techniques, low power architectures including multi-pixel resource sharing

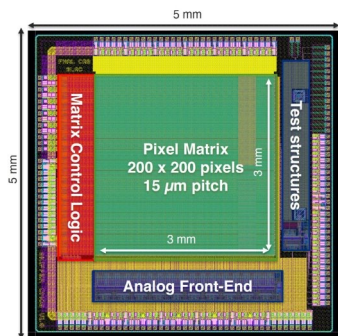- Partners: Fermilab, SLAC, LLNL.

- Progress: Secured funding through DOE Accelerate

  - First simulations in Tower Semiconductor 65 nm epi process without any modifications look extremely promising

 - Several discussions with Foundry
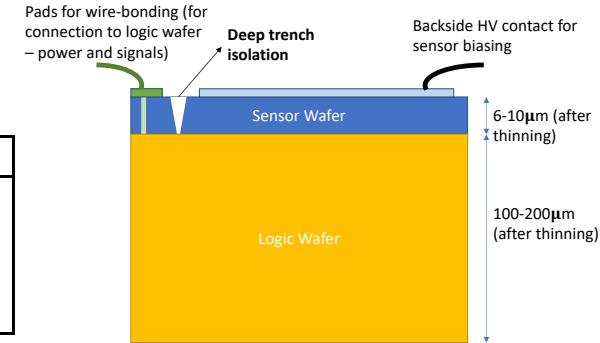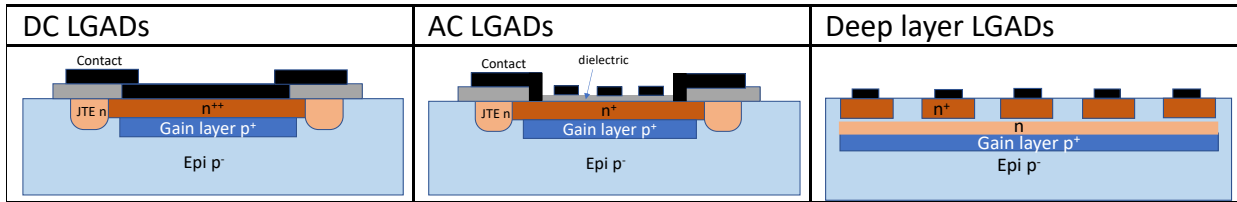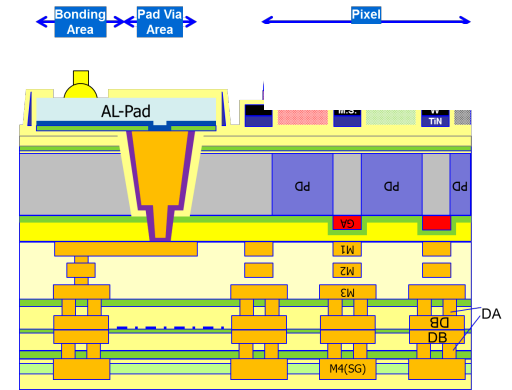
🔷 **Fermilab**

Sensor Co-design with Tower Semiconductor (Skipper-in-CMOS)

- Goal: Skipper-in-CMOS: co-developed novel ultra-low noise sensors with Tower Semiconductor 180 nm CIS process

- Partners: Fermilab, SLAC, University of Bariloche, CONICET Argentina, Tower Semiconductor

- Progress: Testing at Fermilab of more than 400 devices will help determine structures with highest Quantum efficiency

Demonstrated Skipping in CMOS with sub-electron noise performance. Funded by DOE ME Co-design team

In preparation for engineering run
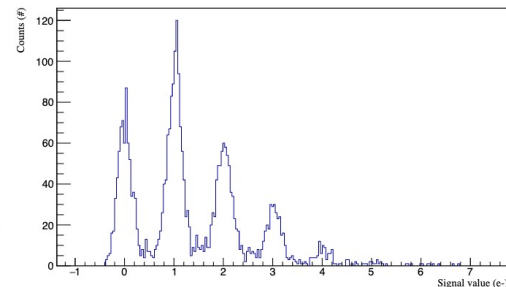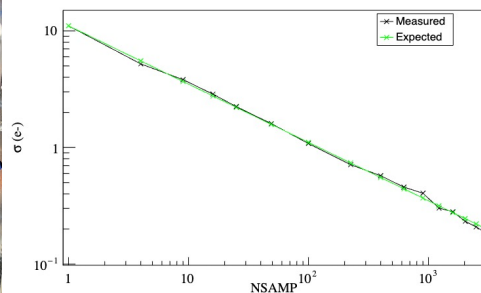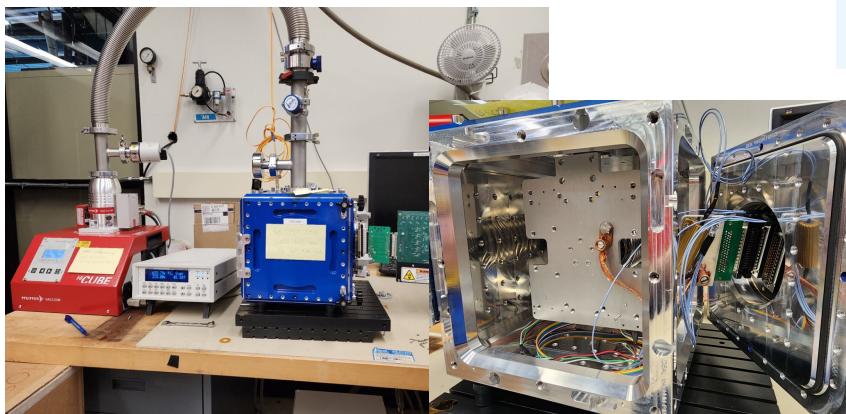
Sensor Co-design with MIT LL - SiSeRO

- Goal:

  - Modify output stage of Skipper-CCD to improve readout speed for Astrophysics and Dark Matter experiments.

  - Demonstrated 6x improvement in speed

- Partners: Fermilab, MIT LL, LBL, Centro Atomico Bariloche

- Progress:

  - Successful demonstration of fabricated structures at MIT LL , Measurement shows 6x improvement in speed without trading off noise

  - DURIN: Readout IC in development, fully simulated chip frontend

Fermilab
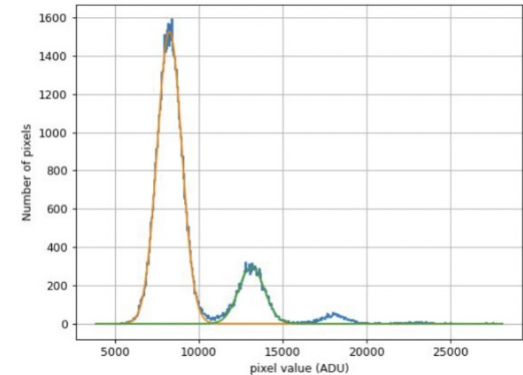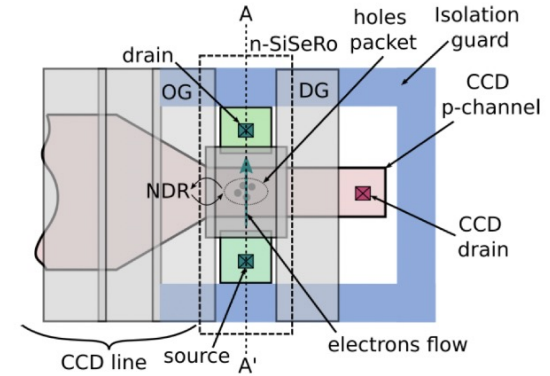
Skipper CCD readout chips

Skipper CCD readout: MIDNA

- State-of-the-art noise performance (~3e- noise performance)

- Cryogenic operation (100K)

- 100x lower power, extremely small footprint, significantly reduced cost

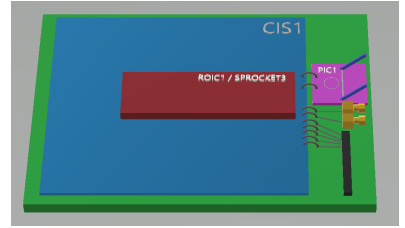- Excellent test performance

- 2nd chip with on-chip pile-up up to 7000 without saturation

- Version 2.1 with lower frontend gain to increase dynamic range

- 3rd chip 4 to 16 channels

- Collaboration with Atomico Centro Bariloche
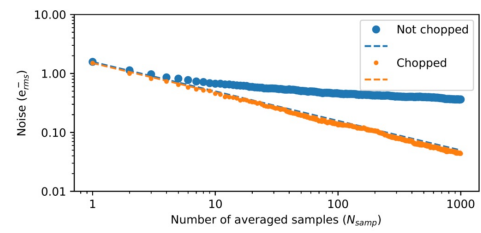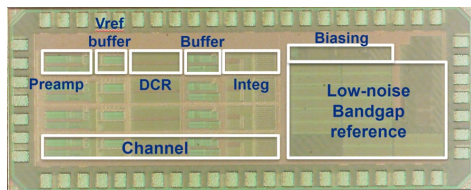
RDC4 Meeting

Highly-parallel readout ASIC for Skipper-on-CMOS: SPROCKET

- Developed low-power in-pixel ADC for highly parallel readout ($\rightarrow$ high frame rates)

- 1st and 2nd prototype fully functional

- 3rd large area chip (~ 1.3 cm^2): integrated photonic readout (with University of Washington)

- Full-reticle ASIC in 2024 (~ 6 cm^2)
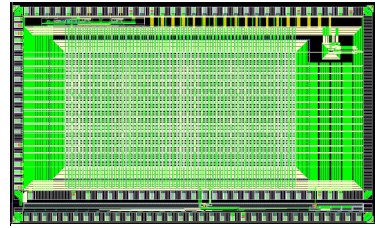
Initial (Warm) Test Results

ADC Transfer Function

Midpoint DNL vs CapTrim Code

ADC Input Noise Histogram @ Vtest = 3 mV
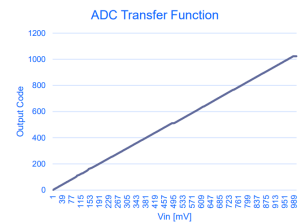
Gain	~ 1.037 codes/mV
Dynamic Range	~ 985 mV
RMS Noise	< 0.25 LSB (~ 250 μV)

🔷 Fermilab
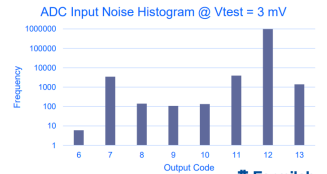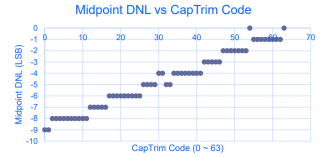
Sensor Co-design with SkyWater

GOALS

- Enable US manufactured sensor capability for HEP experiments
- Optimize the process to enable various types of sensors ubiquitously used in HEP (MAPS, MAPS with timing, Digital SPADs, LGADs, CMOS LGADs)
- Co-design sensor and readout electronics

- Enable the broad adoption of the development across HEP community

PROGRESS

PARTNER with SKYWATER TECHNOLOGIES. Multi party NDA with LBL, ANL has confirmed participation

Strong academic support from UC, UIC, Purdue U., NU, UIUC, for device simulation and testing

Engineering run with various designs on a high resistivity wafer. Secured partial funding from JTFI award

High-throughput testing of sensors at Fermilab and other institutions

28 μm
collection electrode

2 x 2 pixel volume

0.3 pJ / bit

$C_{in} \approx 5$ fF

Artistic view of a SEM picture of ALPIDE cross section
Not to scale

Q_{in} (MIP) ≈ 1300 e $\Rightarrow$ V ≈ 40mV

🔷 **Fermilab**
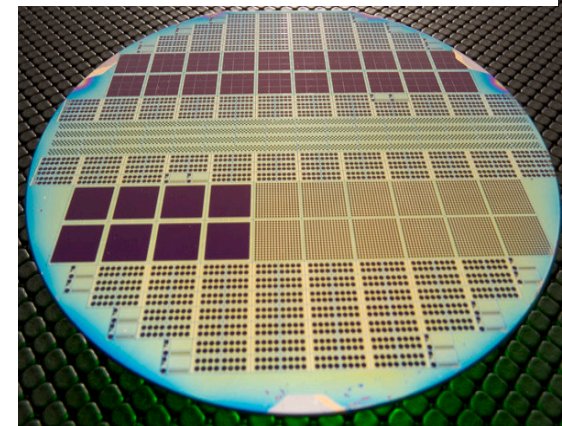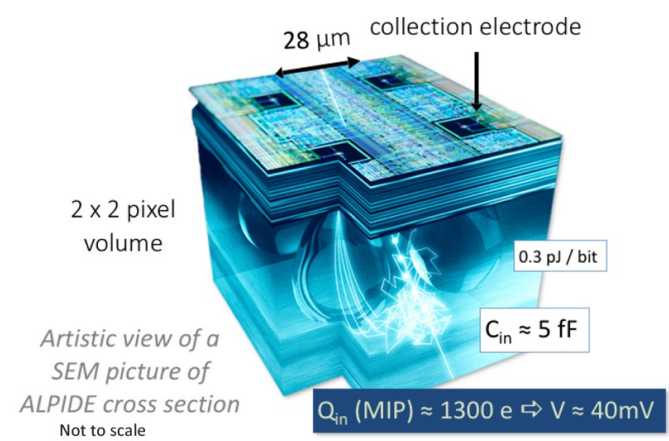
Sensor Co-design with SkyWater

- Process details: Epi on bulk CMOS

  - Epi thickness: > 25um

  - Epi resistivity: ~ 1k-Ohm-cm

- Create a HEP specific MPW

- Divide reticle into 24 dies (5 x 5 mm^2)

  - ½ wafer with only sensors

  - ½ wafer with sensors & readout circuits

  - Test dies with single transistors

- Sensors: would require process splits for optimization

8" wafer

2.6 cm

3.2 cm

reticule

🧬 **Fermilab**
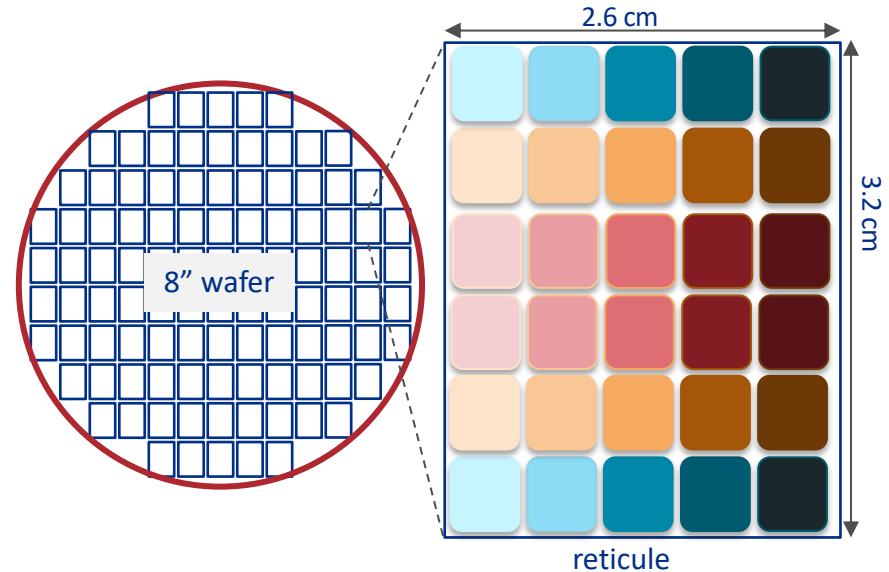
Sensor Co-design with Global Foundries

R&D for DUNE photo detector

GOAL – Cryo Digital SiPM (77K) eventually (3K)

- Overcome limitations of Analog SiPMs which requires either

  Cold ADC or Analog Signal over fiber transmission

- Digital SiPM - Well established concept (SPAD integrated in CMOS process). Active or passive quenching circuit can control the quench time

- Can provide added position resolution and/or timing resolution if needed.

- Simple electronics to count the number of photons.

- Focused on cryo Digital SiPM (currently evaluating first samples with various pixel sizes from GF to establish testing and characterisation procedures)

- Integrated photonics for power delivery and data transfer

- Collaborate with industry (GF) and academic groups (EPFL, University of Washington)

SPAD = Single Photon Avalanche Diode
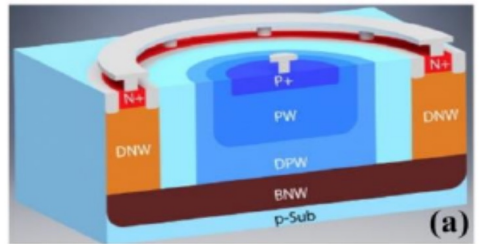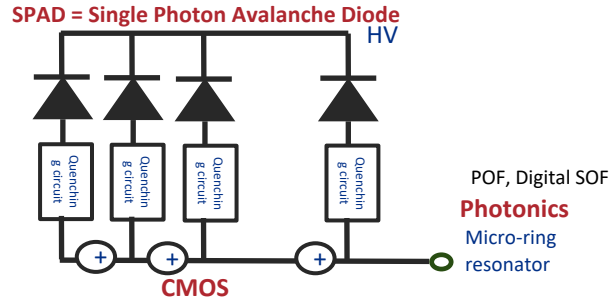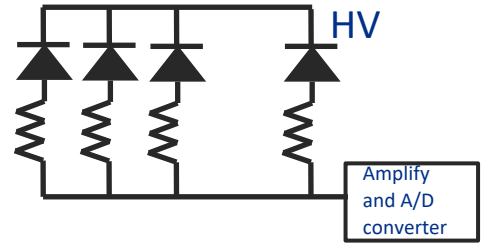
🐝 Fermilab

SOI MAPS

- Goal: Integrated sensing-computing-communication

- Technology nodes: Skywater 90nm, GF 90 nm, GF 45 nm

- Working with SBIR company to improve radiation performance of SOI process

- Need process with "Thru Box Vias".

- Fermilab has previously explored SOI MAPS with OKI/ LAPIS 0.15um process

Labels on figure: Met 2, Met 1, Poli, Device layer n type, SiO2, PMOS Tr, P well, NMOS Tr, pn junction, Support layer n type, E, particle track

Fermilab
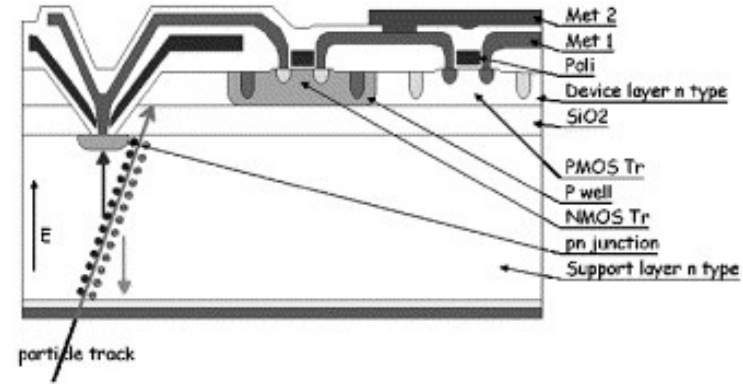
AI Co-design

Design Methodology: Physics driven hardware co-design

- Algorithm development based on Physics data
- **hls4ml** simplifies the design of on-chip ML accelerators

  | hls4ml directives | << | HLS directives |

  C++ library of ML functionalities optimized for HLS

- TMR4sv_hls: Triple Modular Redundancy tool for System Verilog & HLS

ALGORITHM
DEVELOPMENT

ML Model
Training

```
Part: …
ReuseFactor: …
Precision: …
IOType: …
Backend: …
```

hls4ml
Directives

Costs

Performance

```
void foo(int A[10],
         int B[10], int C[10])
{
  int i = 0;
  l1: for (; i < 10; i++) {
    A[i] = B[i] * i;
  }
  i = 0;
  l2: for (; i < 10; i++) {
    B[i] = A[i] * B[i];
    C[i] = B[i] / 10;
  }
  bar(A, B);
}
```

C++
Specification

```
inline bar 'ON'
unroll l1 factor=10
bundle b1=A,B, b2=C
```

HLS
Directives

HLS

Technology Library

TMR4sv_hls

RTL
Hardware
Implementation(s)

HARDWARE
ACCELERATOR

GDSII

Edge computing – algorithm and implementation approach

What is the best approach?

- Data compression, Data filtering, Data featurization, Neuromorphic

Where should the algorithm be implemented?

- **Near sensor: AI/ML mixed-foundry chiplet:**

  Standalone general purpose

  Does not overcome IO bottleneck (1 Tbps off-sensor)

- **Edge of sensor/ Readout integrated circuit:**

  Easier implementation (larger area and power), optimally utilizes chip bandwidth
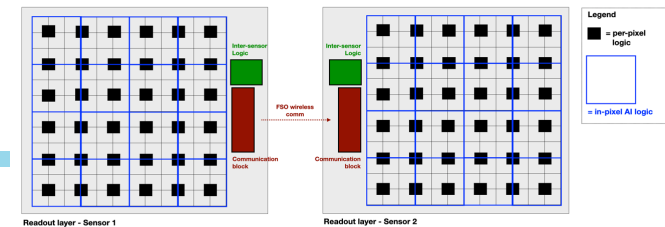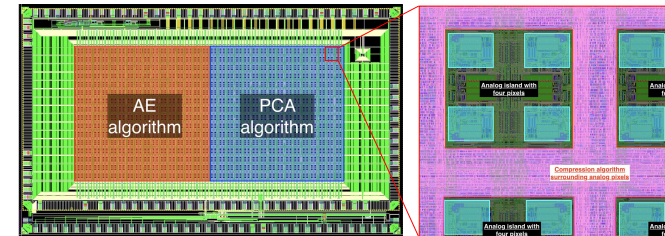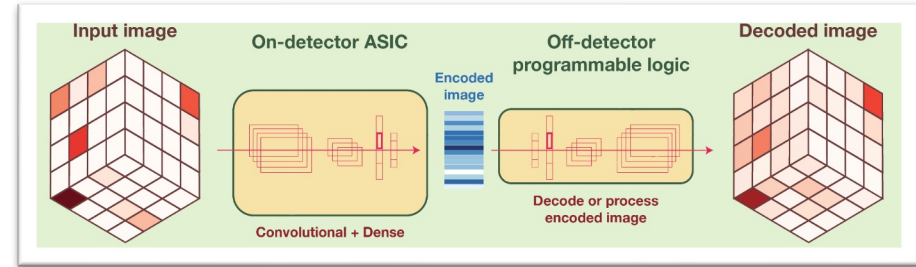
  On-chip transfer

- **In-pixel: At the data source:**
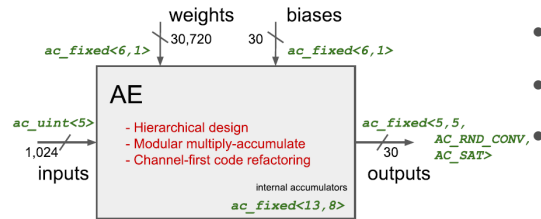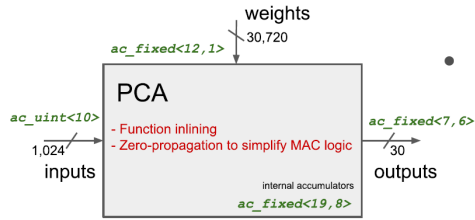
  Minimizes data movement

  Power and area constraint

- **Distributed AI: Across sensors/chips:**

  Share data across chips for higher accuracy/ performant AI algorithm implementation

AI-in-Pixel for at-source data processing

weights
30,720
ac_fixed<12,1>

PCA

ac_uint<10>

- Function inlining
- Zero-propagation to simplify MAC logic

ac_fixed<7,6>

1,024
inputs

30
outputs

internal accumulators
ac_fixed<19,8>

weights   biases
30,720   30
ac_fixed<6,1>   ac_fixed<6,1>

AE

ac_uint<5>

- Hierarchical design
- Modular multiply-accumulate
- Channel-first code refactoring

ac_fixed<5,5,
AC_RND_CONV,
AC_SAT>

1,024
inputs

30
outputs

internal accumulators
ac_fixed<13,8>

- **Lossy data compression at source, 65nm**
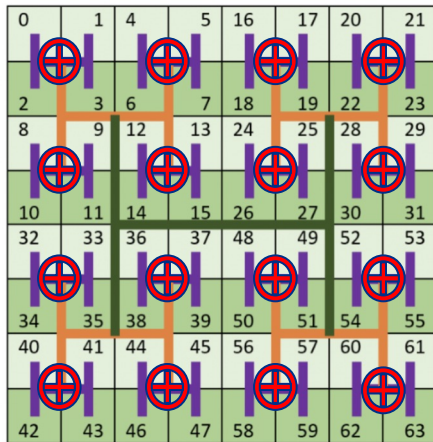 - Comparison of two algorithms synthesized with Catapult HLS (+ hls4ml)
    - AI/ML based (Auto)Encoder, 70x compression, 30 clk lat., +21% area
    - Principal Component Analysis, 50x compression, 1 clk lat., + 21% area
  - Design methodology: Learnt best practice for HLS and PnR
  - Fully functional chips
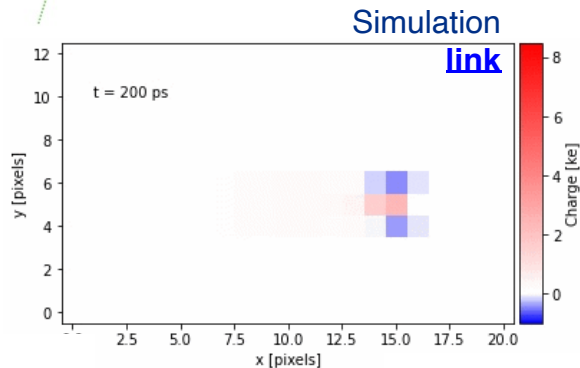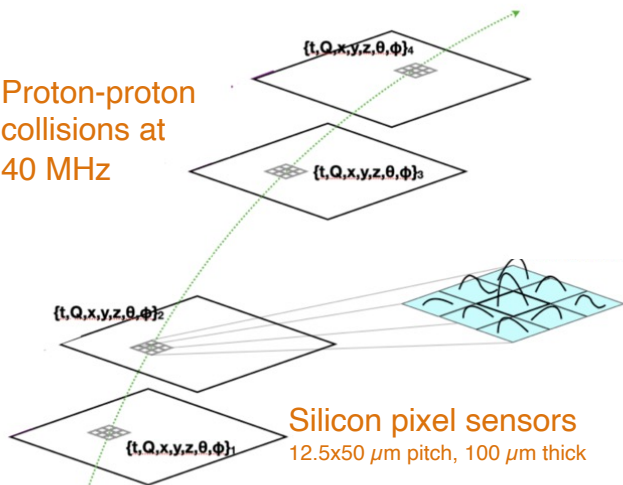 - Partners: Fermilab, Northwestern University, Columbia U.

milab
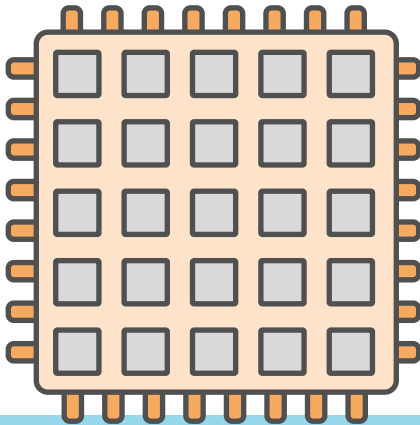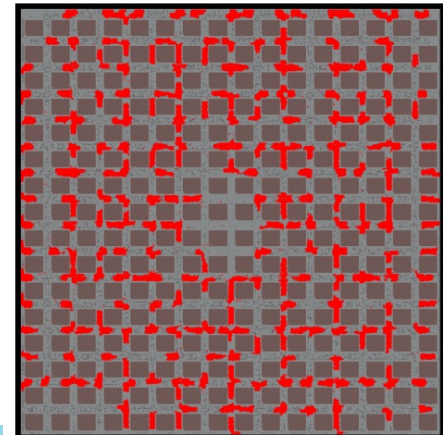
AI-in-pixel: Real-time tracking

Proton-proton collisions at 40 MHz

$\{t,Q,x,y,z,\theta,\phi\}_4$

$\{t,Q,x,y,z,\theta,\phi\}_3$

$\{t,Q,x,y,z,\theta,\phi\}_2$

$\{t,Q,x,y,z,\theta,\phi\}_1$

Silicon pixel sensors
12.5x50 μm pitch, 100 μm thick

Simulation
link

Cluster shapes and Pulse information for filtering out low p_T particles

- NN classifier identifies and saves clusters from tracks with $p_T > 2$ GeV

- $\geq$ 95% data reduction by saving only high p_T

- Low power implementation

**Compact algorithms for data reduction through featurization**

- **Predict** physics information (x, y, θ, ϕ) and **meaningful error** (UQ) on particle position, angle

- Potential for reduction of track seeds $\rightarrow$ saves time & computing resources down the line

**Technology development to enable on-sensor computing**

- Ultra low power in-memory compute chips

- 3D integration for optimized data processing

- Leverage emerging technologies such as novel CMOS compatible memory

- Integrated Photonics for data transfer between modules (performant AI)

Farah Fahim | AI-Enhanced Co-Design for Next Generation Microelectronics

🐿 **Fermilab**

AI-in-pixel

- Smart Pixels: CMS pixel detector replacement R&D: $25\mu m$ pixel pitch in TSMC 28 nm with on-chip neural networks for data filtering and data compression for readout at 40 MHz

- On-chip binary classifier for rejecting tracks with momentum >0.3GeV to achieve > 50% data reduction in innermost layers – 2nd chip submission

- Now focusing on compact inference – regression model for prediction x,y, angle and error

- Partners: ORNL, KU (Neuromorphic algorithm), Georgia Tech (Analog Floating gate) Sandia National Lab for ReRAM and ECRAM implementation of the algorithm (Sandia Grand Challenge), UIC, UC (testing and algorithm development), JHU (detector simulation, data generation)
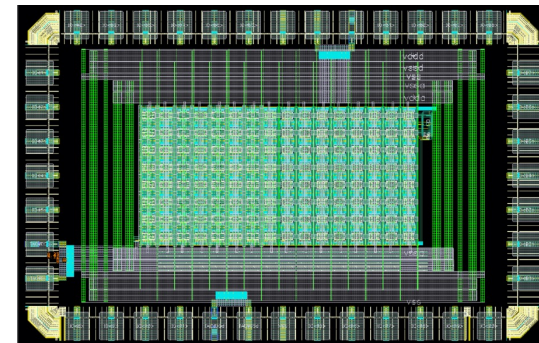
28 nm chip with 32 x 16 pixels
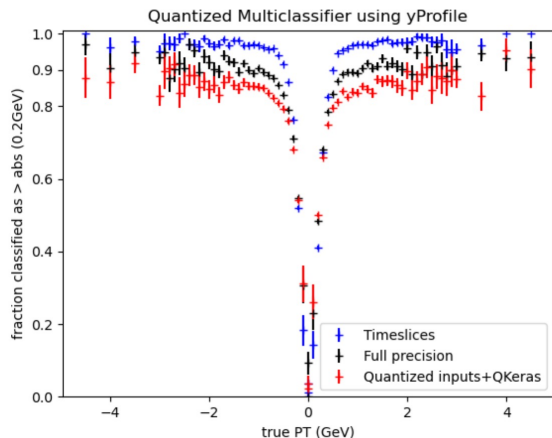
	Fraction correctly predicted	
	> 1 GeV	> 2 GeV
Timeslices	97.30%	97.60%
Full Precision	91.00%	92.60%
Quantized Inputs + QKeras	85.80%	87.20%

Conservatively reject:

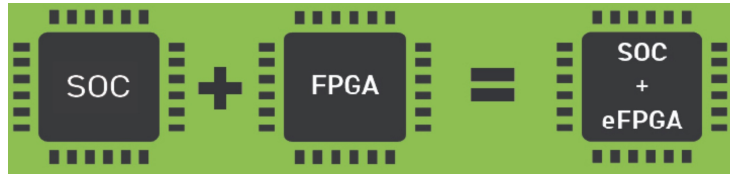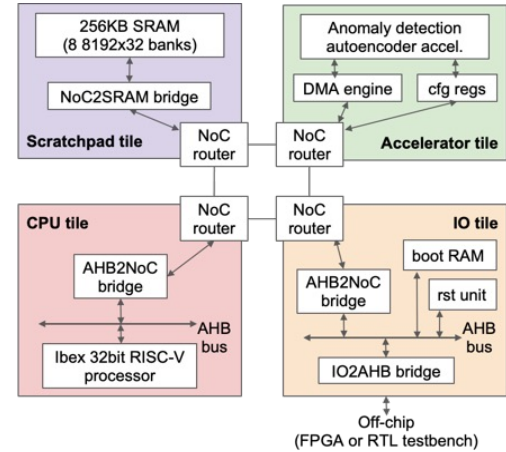| < 0.2 GeV	≥ 6%
< 0.5 GeV	≥ 36%
< 1 GeV	≥ 70%
< 2 GeV	≥ 94%

~20x reduction

🟦 Fermilab

Reconfigurable Edge AI – with eFPGAs

- Collaboration with Columbia U. & Northwestern U.

- Edge AI: Combining two established open-source platforms (ESP and HLS4ML) into a new system-level design flow to build and program a System on chip

  In the modular tile-based architecture, we integrated a low-power 32-bit RISC-V microcontroller (Ibex), 200KB SRAM-based memory, and a neural-network accelerator for anomaly detection utilizing a network-on-chip.

- Embedding FPGAs on detector: Radhard/ cryogenic eFPGA on-chip – with Flex Logix (22nm / 28nm).

- Establishing design flow (with ESP) and investigating extreme environment performance

Quantum & Cryoelectronics

Using Ion-traps as dark matter sensors

- Optical Atomic Clocks (Joint DOE-DOD project): Fermilab + MIT LL + Global Foundries + Stonybrook University + NU

- Custom analog Ion-trap simulator: Fermilab + ORNL + Global Foundries + NU

3D-integrated trapped-ion control and readout system

● ○ Ions

11/9/23
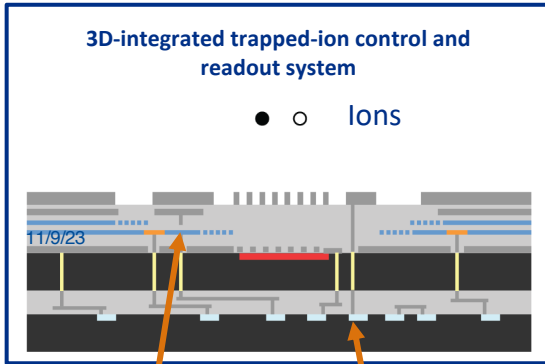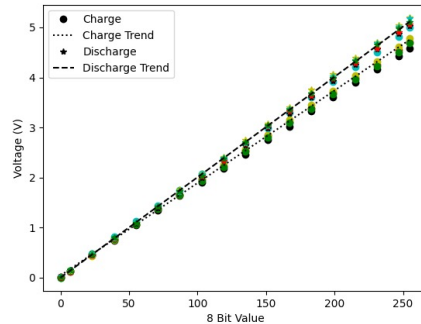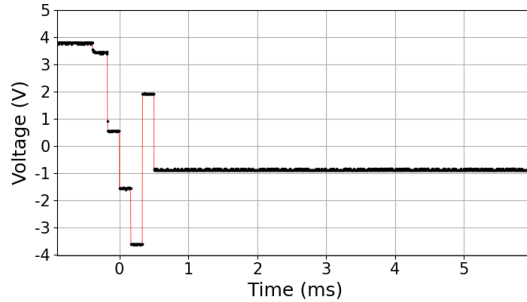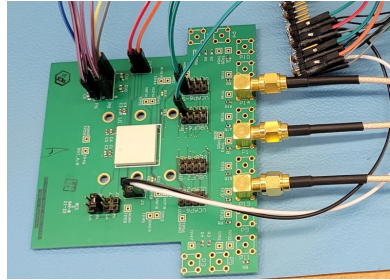
Trap/photonics chip   Electronics chip

Test chip

9 channels   9 channels

Charging circuits   Bias

Discharging circuits   Level shifters
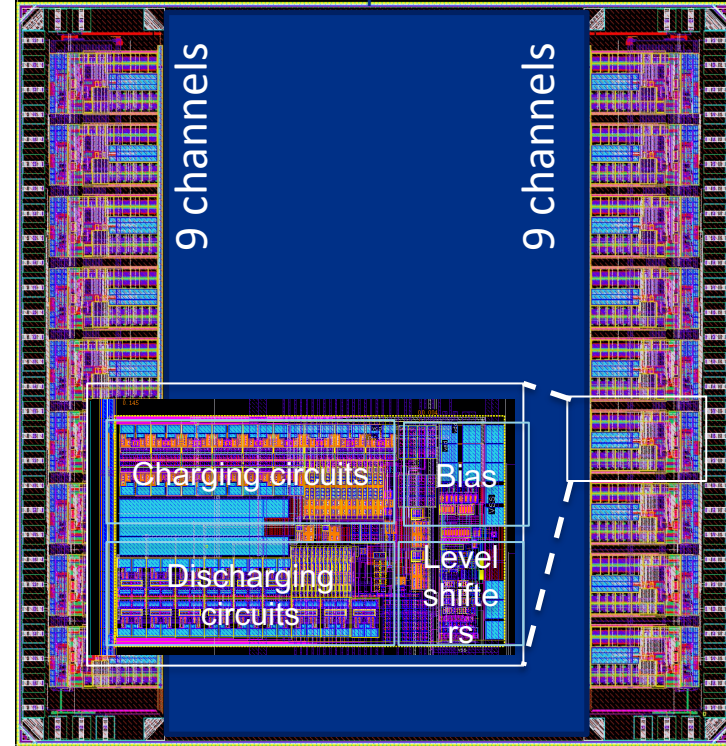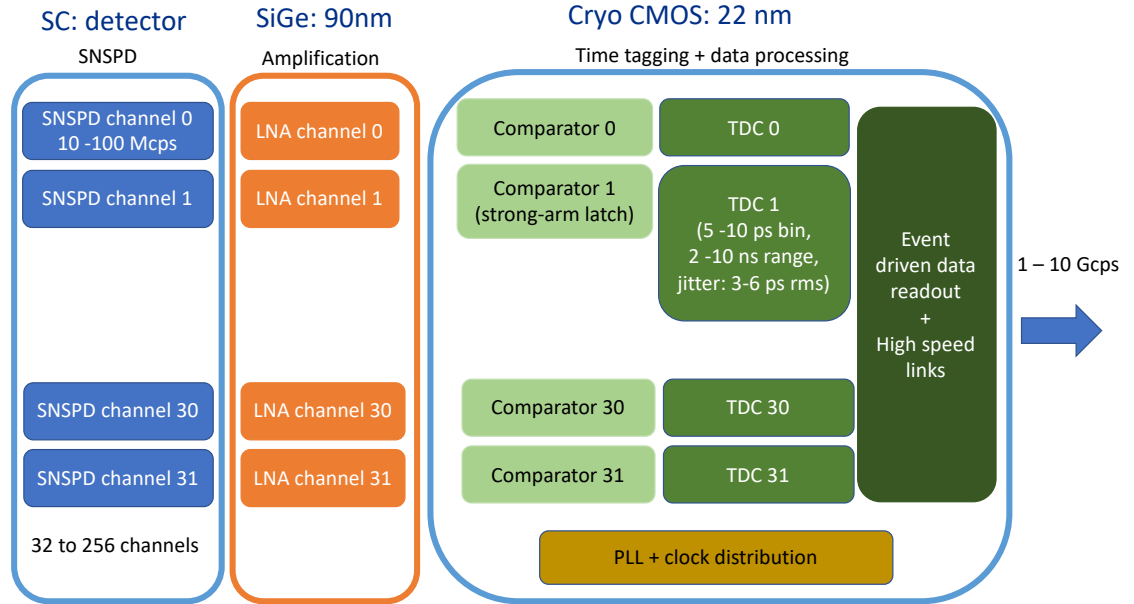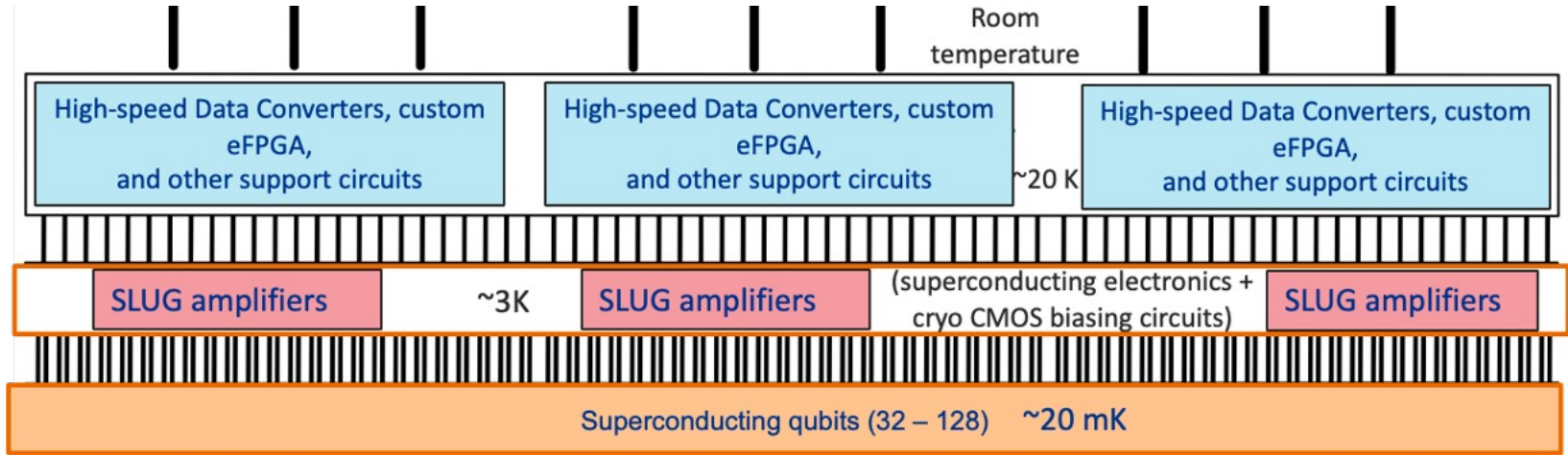
3 mm

🎗 Fermilab

Chiplets for (Cryogenic detectors) Sensing : SNSPD

- SNSPD for dark matter detection and other space science and quantum applications
- Enable scaling (kpixel to Gpixel)
- Energy Efficiency (limited cryogenic power budgets)
- Chiplets are easier to adapt as the detector performance improves across generations
- Partners: Fermilab + ANL + JPL + MIT

SC: detector
SNSPD

| SNSPD channel 0 10 -100 Mcps |
| SNSPD channel 1 |

| SNSPD channel 30 |
| SNSPD channel 31 |

32 to 256 channels

SiGe: 90nm
Amplification

| LNA channel 0 |
| LNA channel 1 |

| LNA channel 30 |
| LNA channel 31 |

Cryo CMOS: 22 nm
Time tagging + data processing

Comparator 0	TDC 0
Comparator 1 (strong-arm latch)	TDC 1 (5 -10 ps bin, 2 -10 ns range, jitter: 3-6 ps rms)
Comparator 30	TDC 30
Comparator 31	TDC 31

Event driven data readout + High speed links

PLL + clock distribution

1 – 10 Gcps

| **SNSPD** (improved gain) | **Nano-cryotrons** (Muxing and data reduction) | **Time tagging** (comparator TDC) | **Photonic chiplet** (electronic to photonic conversion) |

🌀 **Fermilab**

# Multi-tier Cryogenic Chiplets for Quantum Sensing/Computing

- Distribute readout and control across temperature domains 20mK to Room Temp

- Overcome Wiring challenge (high density flex: poor thermal conductor, good electrical conductor)

- Chiplets: Superconducting electronics (SLUGs: Qolab) + cryoCMOS biasing circuits (Fermilab) + High speed ADCs (Fermilab + Microsoft) + eFPGAs (FlexLogix) + GF (cryo)

Superconducting electronics

- Focus on design and investigate mature fab processes such as MIT LL and SkyWater

- TWPAs and JPAs for ADMX-BREAD using a super conducting fab at MIT LL

- Collaborating with Washington University at St. Louis

- Established design flow for Josephson Parametric Amplifiers (JPA)

- Using JPA design flow as a foundation to study Traveling-Wave Parametric Amplifiers (TWPA)

- Expand beyond TWPAs to other superconducting circuits

- Investigate integration with cryoCMOS for optimized hybrid platform
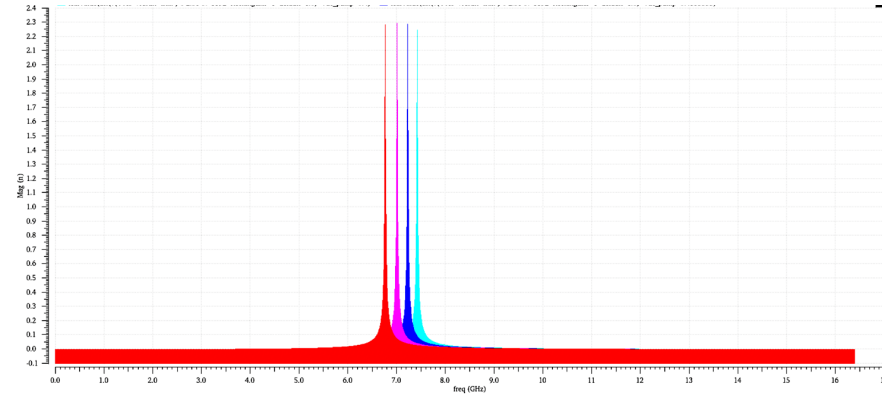

JPA Simulation
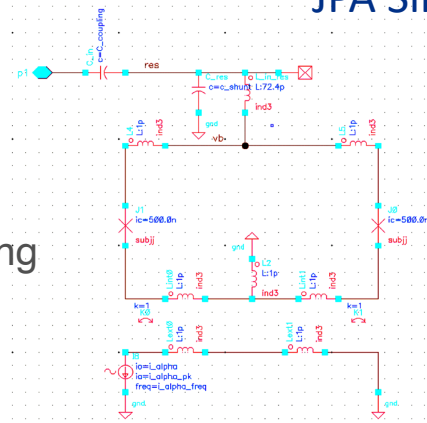
JPA Schematic

JPA Layout

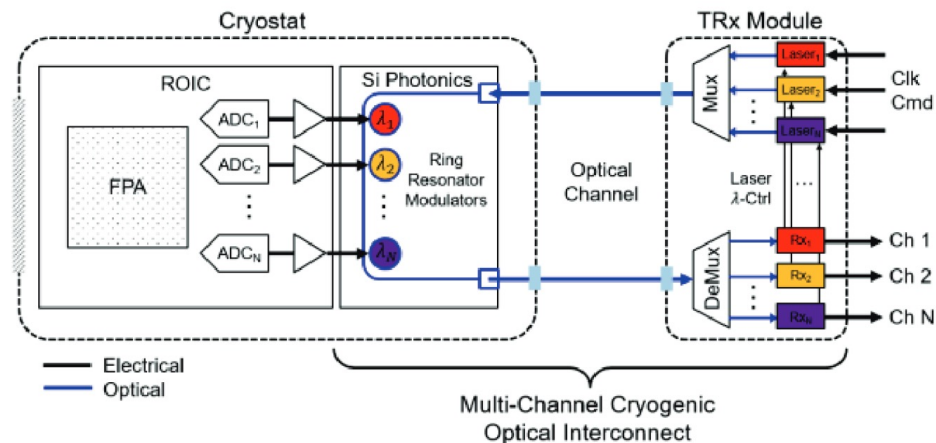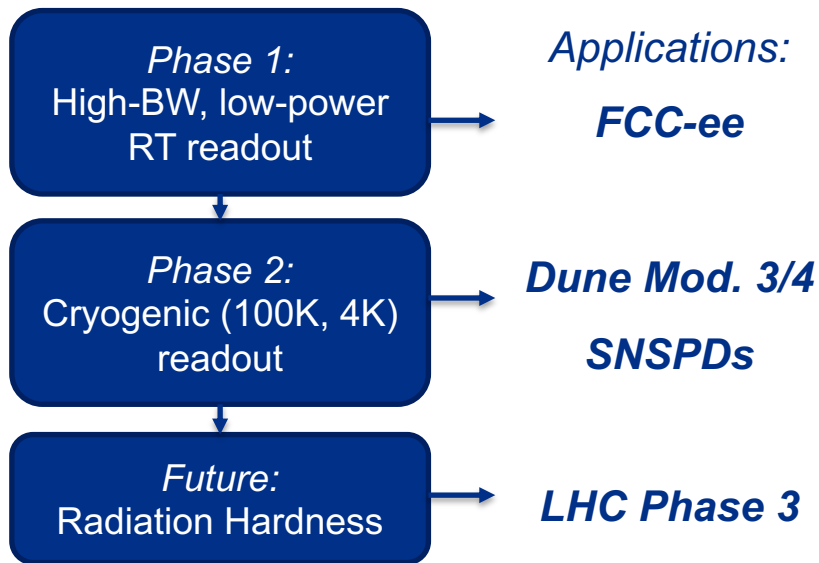Fermilab

New capabilities

Fermilab
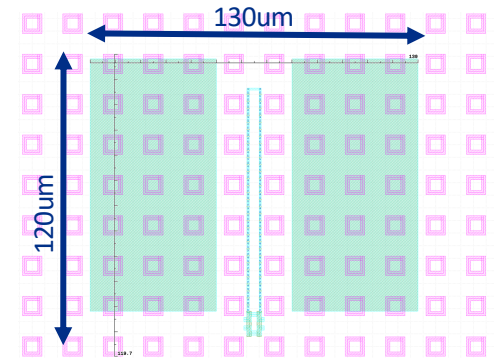
Microelectronics Hardware Discovery platforms

- Wafer scale platform for in-memory compute devices (advanced node wafers in 28 nm):

  - Accelerate development of materials/devices for in-memory compute

  - Compact Algorithms: Reduce the number of operations to exploit new materials/devices

  - Integration and co-design with CMOS circuits

  - Integration with sensing architectures

- Develop criteria to ensure that the wafer-scale platform can be seamlessly used for integrated memory development.

  - surface preparation,

  - types and sizes of contact

- Establish benchmarks to assess maturity

 - statistical insights into growth and performance of the device

🔷 **Fermilab**
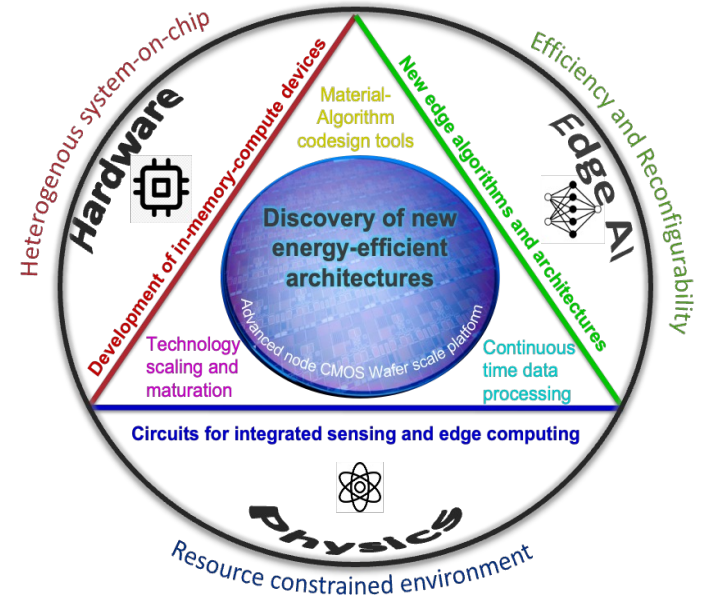
Silicon Photonics: Developing Expertise in a Critical Technology for Future Detectors

100x bandwidth, **100~1000x** lower heat load, **10~100x** channel density, EMI/cross-talk immunity.

Phase 1: High-BW, low-power RT readout	→	Applications: **FCC-ee**
Phase 2: Cryogenic (100K, 4K) readout	→	**Dune Mod. 3/4** **SNSPDs**
Future: Radiation Hardness	→	**LHC Phase 3**

Conceptual schematic of a scalable cryogenic readout architecture based on silicon photonics [1]

Phase 1 Prototype Block Diagram

Fermilab
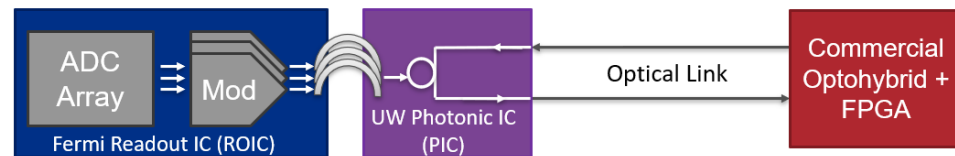
Workforce development

Microelectronics Ecosystem

- *CAD-EDA tool initiative for growth of microelectronics teams across the DOE complex (led by Helmut Marsiske)*

- Co-designing with other applications more cross agency collaborations (DARPA: new extreme environment initiative, DOD ME Commons)

- Collaborative cross disciplinary teams with academia, national labs, international partners and industry

- Microelectronics workforce:

  - Career pipeline for research engineers

  - University Internship program

- Focus on technology transfer and enable spin-offs

Increasing Leadership Capacity

Resources + Potential + Support = Increased Leadership Capacity

🐾 **Fermilab**