Microelectronics for next generation of HEP instrumentation

Farah Fahim

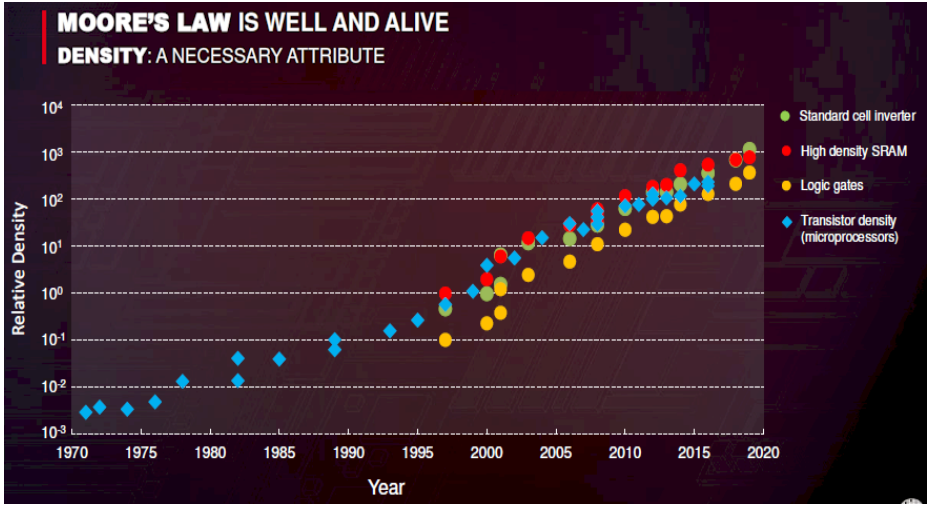
Engineering retreat

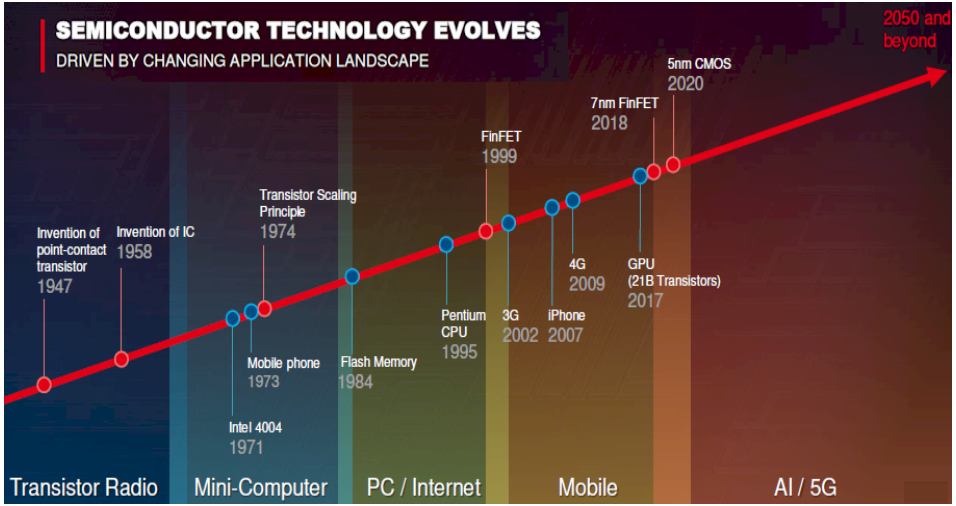
24 Feb 2021

Microelectronics Growth

Traditionally based on Moore's law

Technology scales 2× every 18 months – sustained by transistor scaling

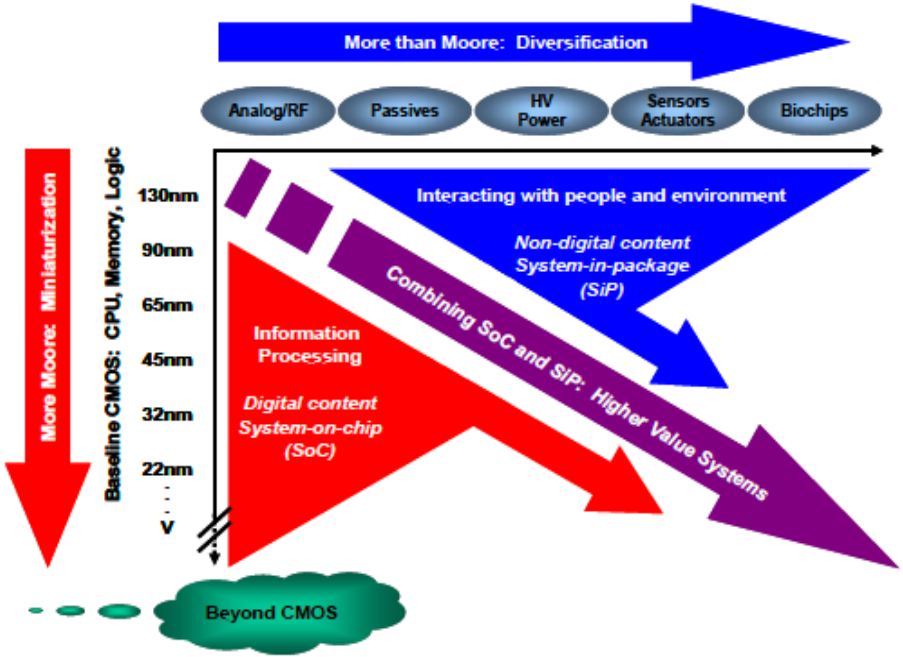


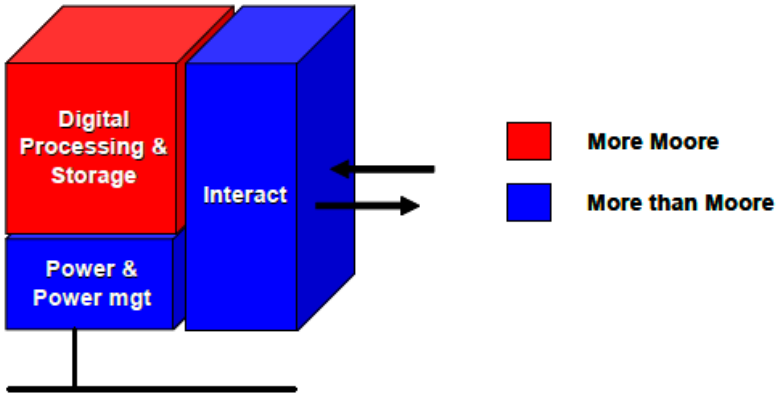


Fermilab

More than Moore?

- World is inherently analog (mandates an analog interface)
- System requires various functions (beyond just digital)
- Multi technology platforms

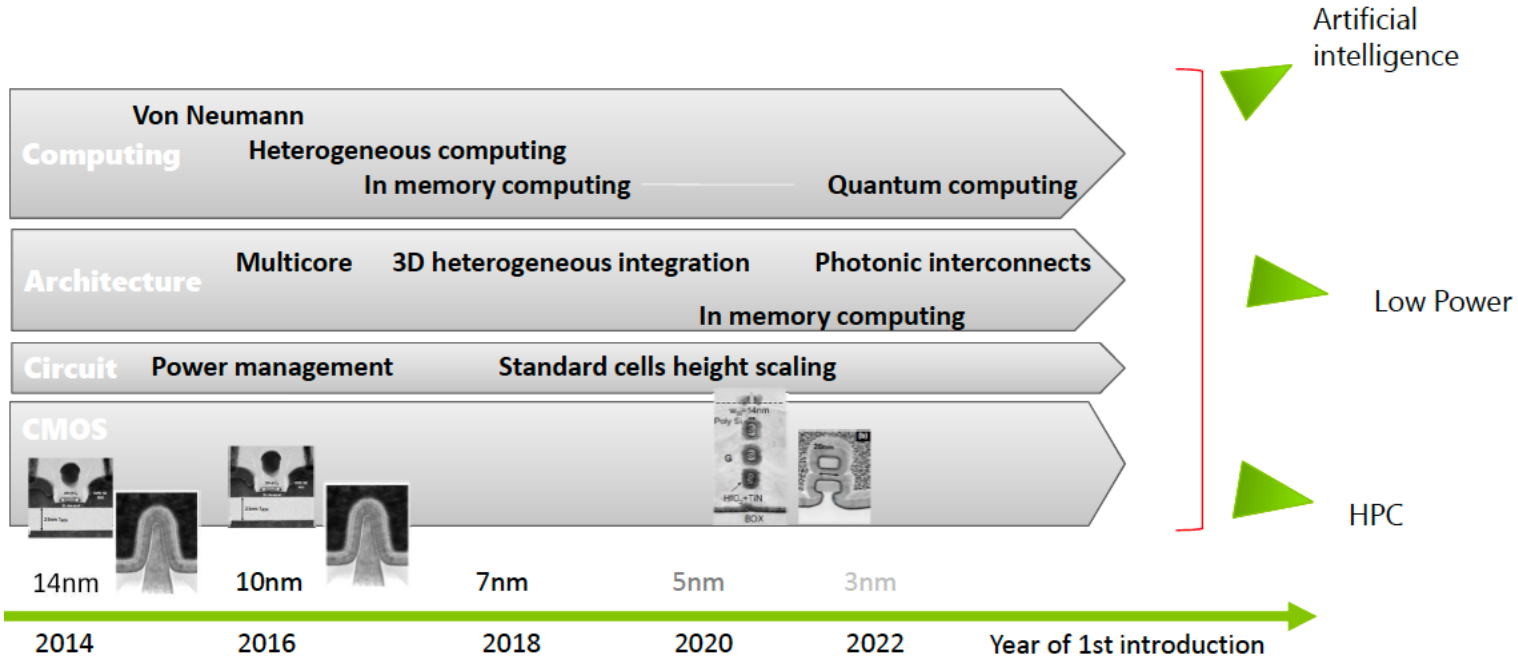




Source: ITRS

Fermilab

Why do we need more than Moore

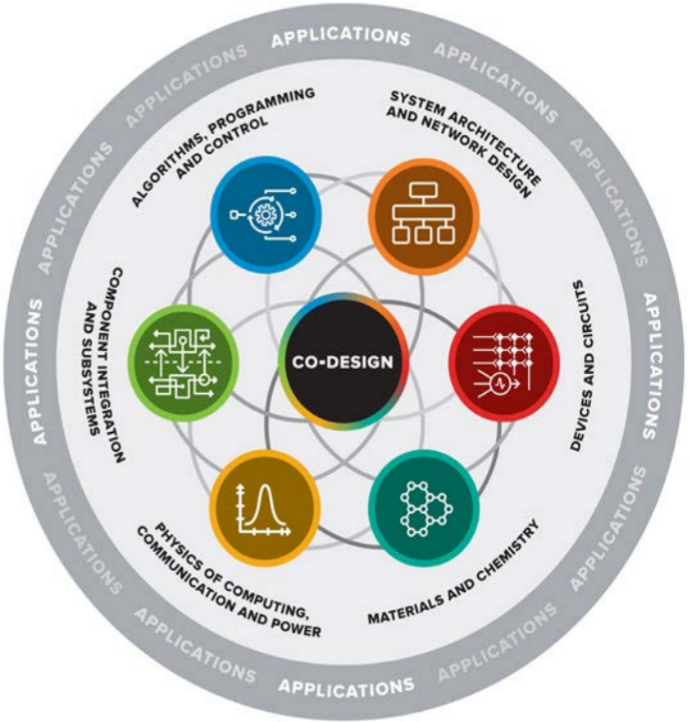


Source: CEA- LETI

Fermilab

Microelectronics enabling next generation HEP instrumentation

- **Novel devices**
  - Skipper CCD-in-CMOS
- **Deep Cryogenic electronics**
  - Quantum Communication & Computing
- **Hybrid integration**
  - Electronic – Photonic Integration
  - 3D integration
- **Hardware – Software codesign to enable edge compute**
  - On-chip machine learning



Fermilab

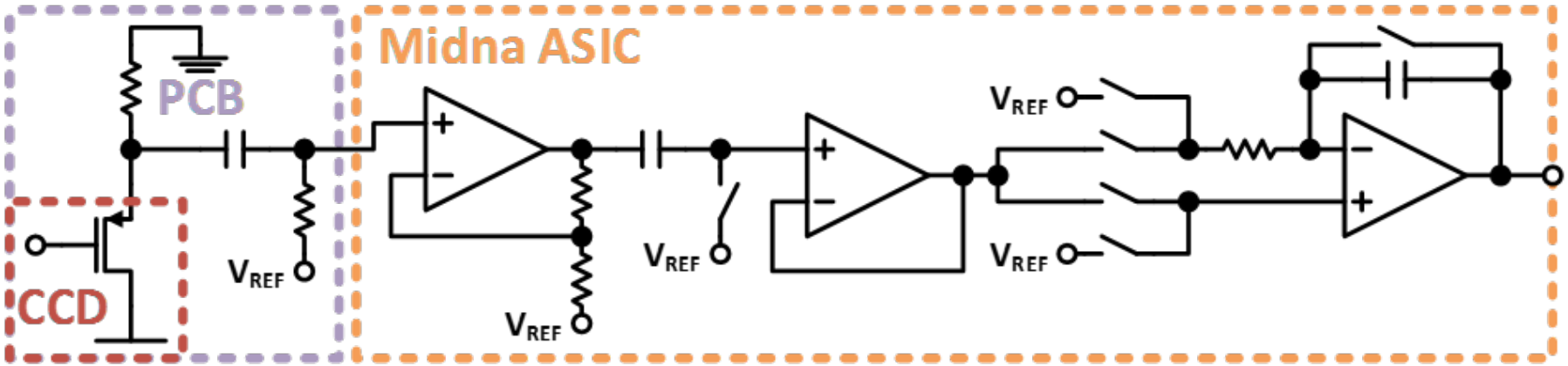
Novel Devices

Future CCD technologies

- **Fermilab has been pioneering Skipper CCD technologies – Averaging multiple samples for ultra-low noise performance (~ 1000 averages for << 1e- noise) – Juan Estrada**

- 1st Step – Enable parallel readout with low-cost multiple channels. Translate PCB design to Readout Integrated Circuit (ROIC) with lower noise performance (1/3rd CCD noise) and ~8mW power per channel.
- 4” × 4” board to ~ 4 × 4 mm^2
- Cost reduction of 100 ×

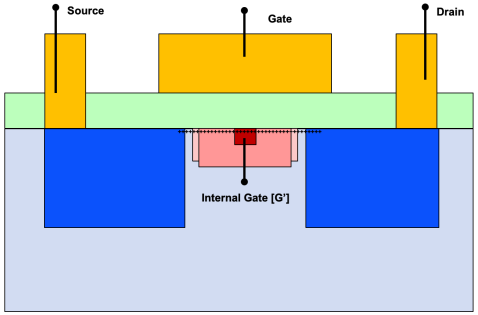
T. England, F. Alcalde Bessia, H. Sun, L. Stefana (SCD)

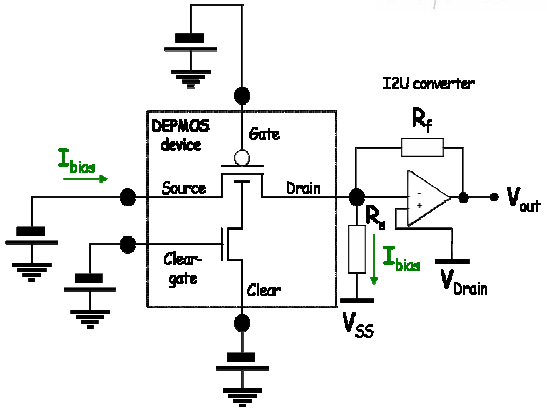


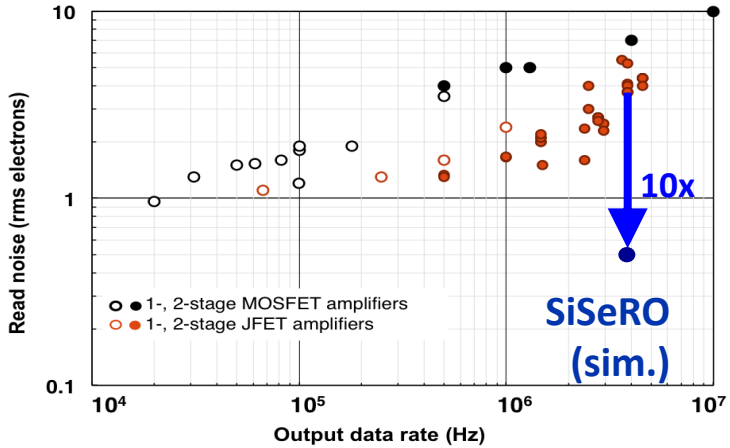
Fermilab

Future CCD technologies

- 2nd Method – Increase readout speed without increasing noise. SiSeRO approach (Collaboration with MIT Lincoln Lab)
- 10 × speed improvement
- Additional advantages – readout is DC coupled at low operating voltages (removes AC coupling capacitors and further increases system integration and reduces footprint)





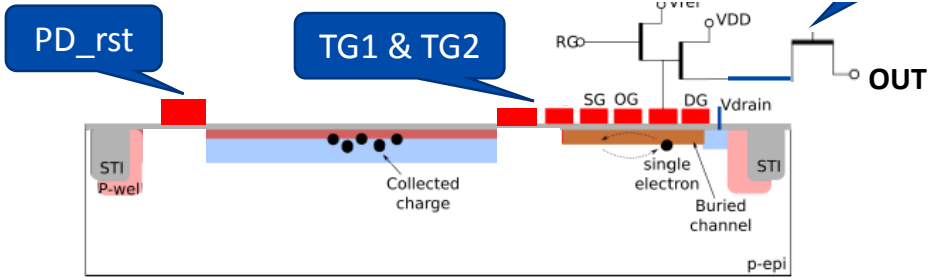


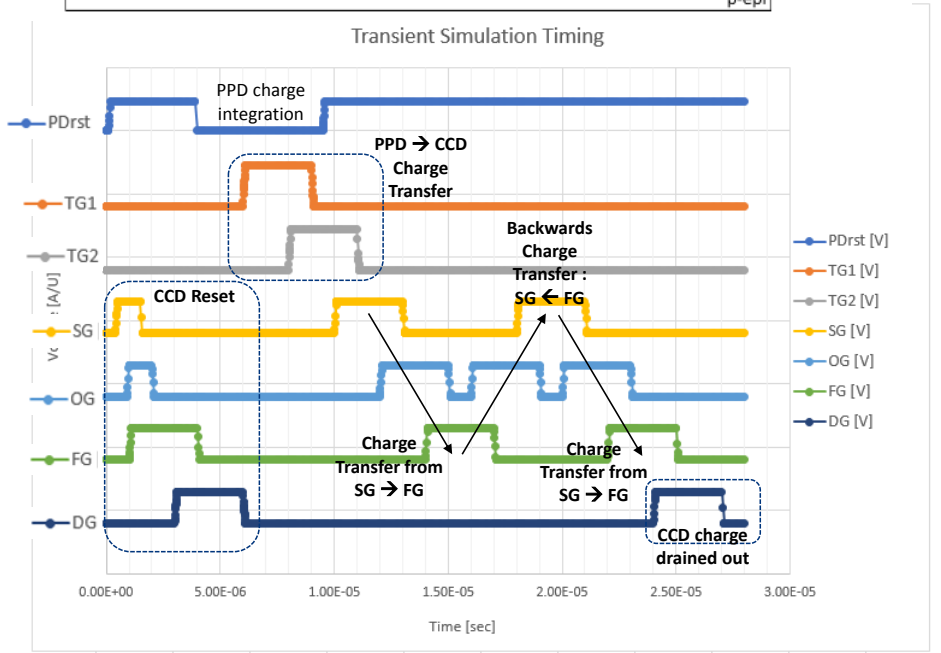
Fermilab

Future CCD technologies

- Ultimate approach – skipper-in-CMOS
- Utilize a commercial CMOS Image Sensor process for lower noise performance (Collaboration with SLAC and Tower Semiconductor)
- Noise of a pinned photodiode (0.7e-)
- With 10 averages we hope to achieve ~0.2e- noise
- Allows hybrid pixel sensor with fully-parallel (per pixel) readout achieving the ultimate goal of 1kfps readout over large areas (6 cm^2) ~ 2.5 Mpixels per chip

M. SofoHaro, A. Birman (TJ)





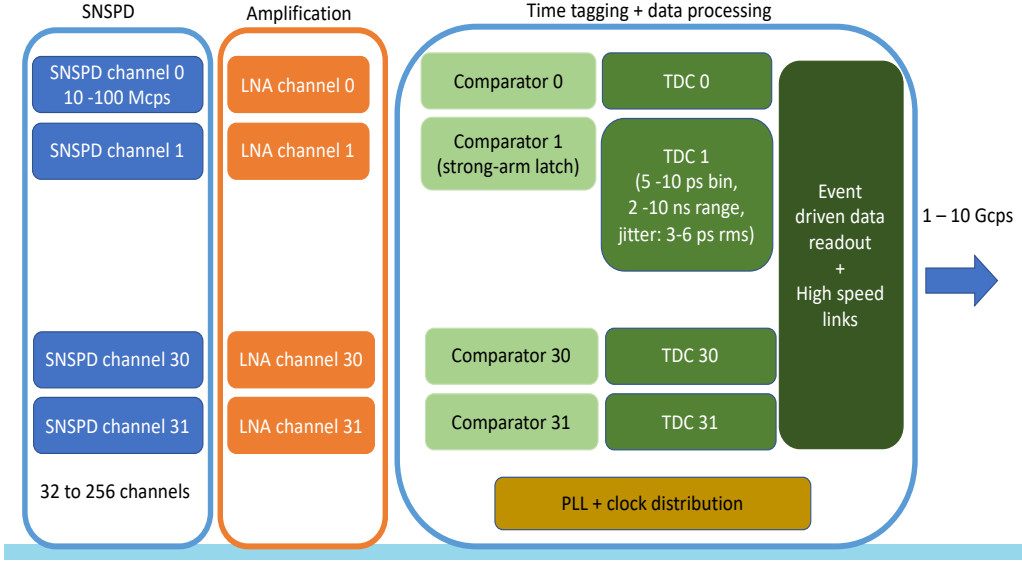
Fermilab

Deep Cryoelectronics for Quantum Computing

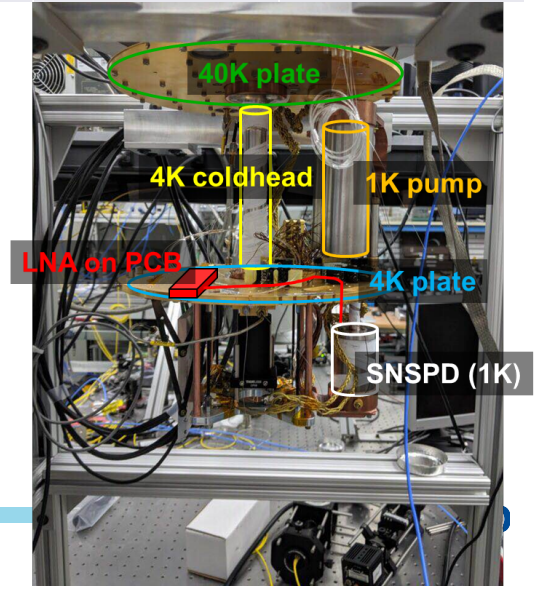
Quantum communication: Superconducting Nanowire Single Photon detectors

SNSPD best performance – (operating at 1 - 4K)
Time-correlated single photon counting from the deep UV to the mid-infrared
Extremely low dark counts and very high precision
QUANTUM INTERNET - High bandwidth communication

1st Step – Low Noise amplification in collaboration with Georgia Tech and JPL



Parameter	Goal by 2025	SOA 2019
Efficiency	>80% @ 10 μm	98% @ 1550 nm
Dark Counts	< 1e-6 cps / mm^2	< 1e-4 cps / mm^2
Energy Threshold	12.5 meV (100 μm)	0.125 eV (10 μm)
Timing Jitter	< 1 ps	2.7 ps
Active Area	100 cm^2	0.92 mm^2
Max Count Rate	100 Gcps	1.2 Gcps
Pixel Count	1.6e7 (4096x4096)	1024 (32x32)



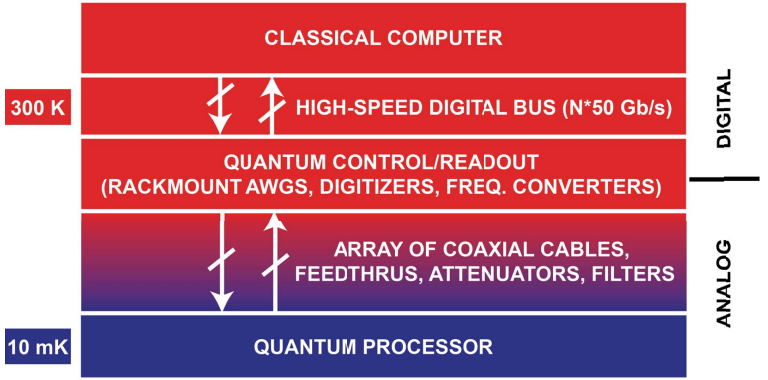
D. Braga

Beyond NISC era QC utilizing cryo-electronics

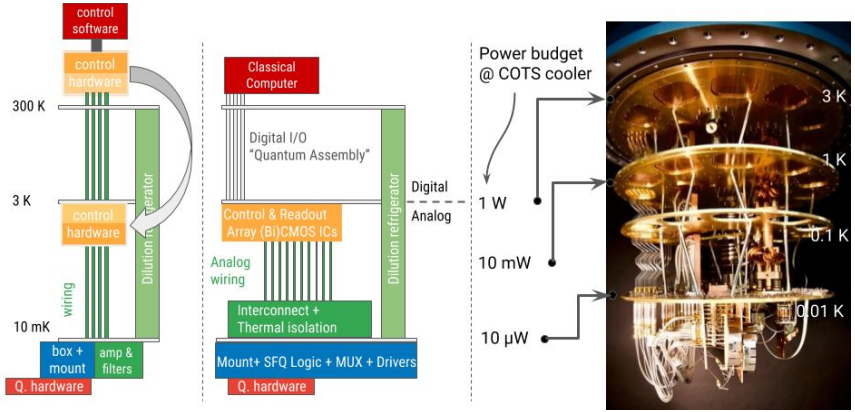
- Collaboration with industry (Microsoft – High speed ADC)
- Achieving high-speed and high-resolution are often conflicting goals
- Key transistor behavior – such as low noise performance improves at cryogenic temperature
- Why National Lab? [Cryogenic electronics for DUNE – full cycle from modelling to testing; very similar approach for radiation environment]
- Modelling is key (Collaboration with EPFL/ TUDELFT)

T. England, D. Braga

Architecture Suitable for 72 Qubit Computer



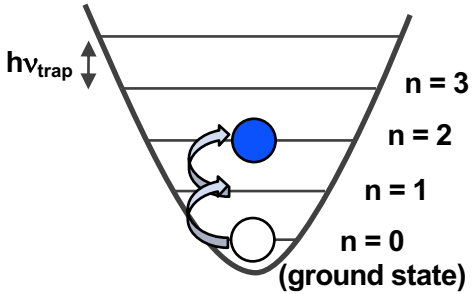
Scale by Integrating Control Electronics

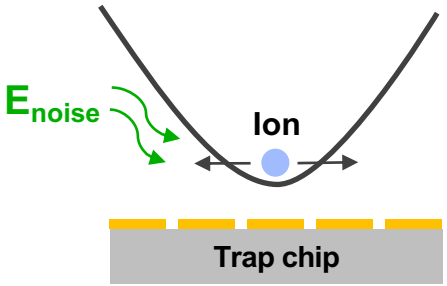


Google AI Quantum

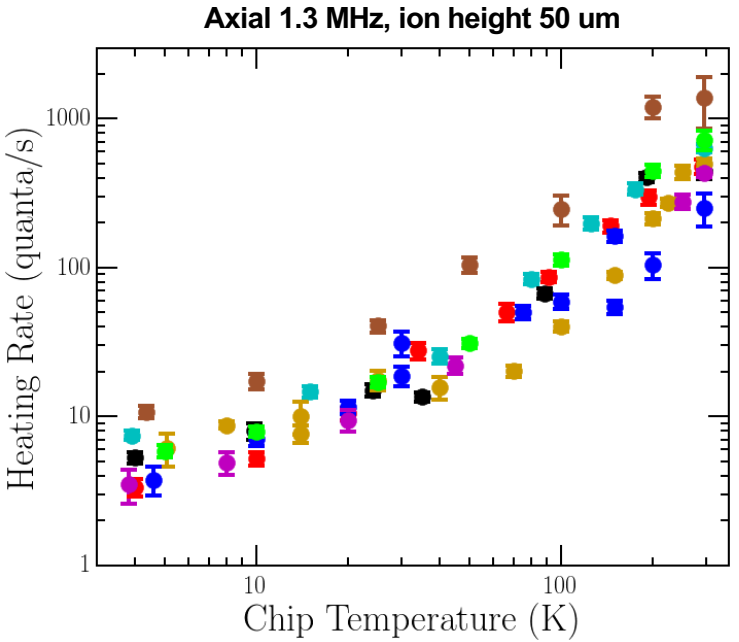
Benefits of Cryogenics for Trapped-Ion QIP

J.Chiaverini,
MIT LL





- **Greatly reduced electric-field noise**
  - **This noise is a limiting factor in error in trapped-ion 2QGs in small traps**
  - **Measured to be much larger than JN ("anomalous")**
  - **Source unknown**
  - **Empirically, 2 orders of magnitude lower at ~5K when compared to room temp.**
  - **This is true when technical noise is under control**

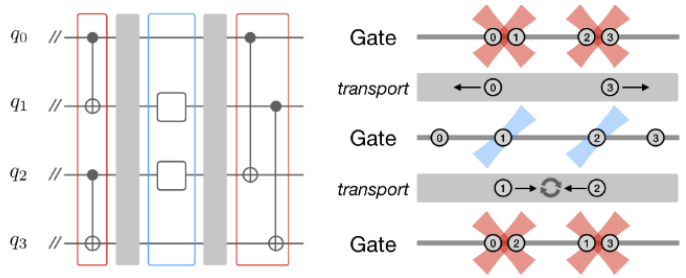


MIT-LL measurements

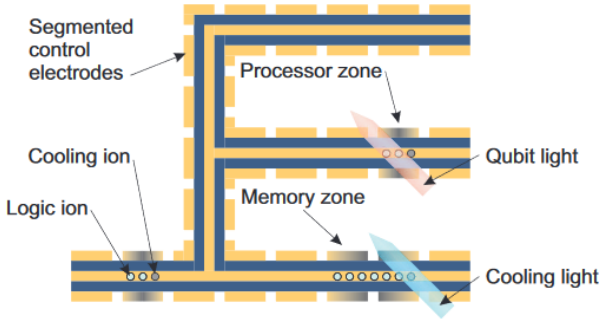
QIP: Quantum information processing
2QG: Two-qubit gate
JN: Johnson noise

Motivation for Integrated Electronics

J.Chiaverini,
MIT LL



Pino et al., arXiv:2003.01293 (2020) [HQS]



JC et al., Quant Inf. Comp. 5, 419 (2005) [NIST]

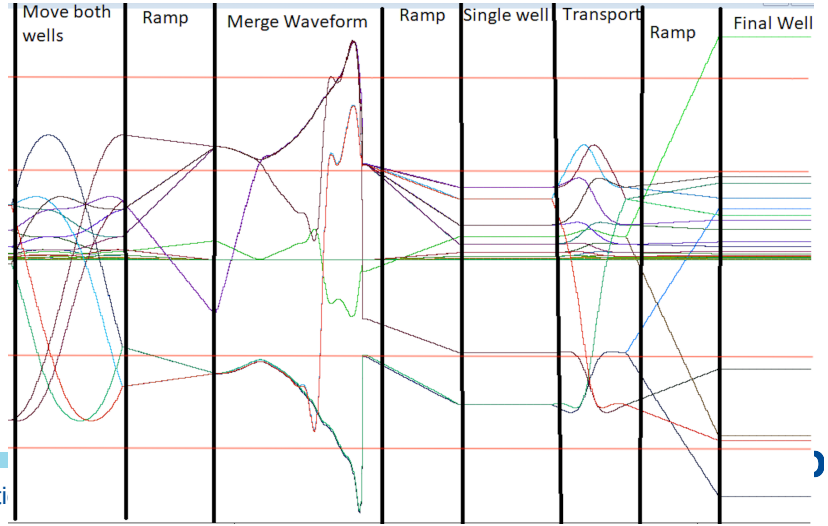
- Controlled ion motion through variation of electrode potentials
  - Each electrode segment requires a dedicated voltage for ion array control
  - For large arrays, electrical interconnects will become a limit
- Multiplexing can reduce wiring overhead, but at a cost of speed
- On chip analog voltage production can directly address this issue
  - Further level of on-chip control: microprocessor to implement time-dependent voltage updates
    - Standard motion subroutines, calibration, etc.

QCCD: Kielpinski, Monroe, and Wineland, Nature 417, 709 (2002) [NIST]

Cryo-electronics control for Ion-Traps (QSC - ORNL)

S. Li

- **Design challenges:**
  - Low output noise: < 100nV/sqrt(Hz) around a wide frequency range (0.5 - 5 MHz) and at low frequency.
  - Low power: < 5 mW/DAC (limited by the cooling power of the cryostat) while driving a wide rage of load capacitance (70 – 1800 pF) of +/- 10 V full scale at 10 MHz waveform updating rate.
  - High resolution: 14-16 bit for precise control and not disturbing RF electrodes.
  - Memory: 100 electrodes * 14 bit * 5000 points ~ 2.5 MB



Hybrid Integration

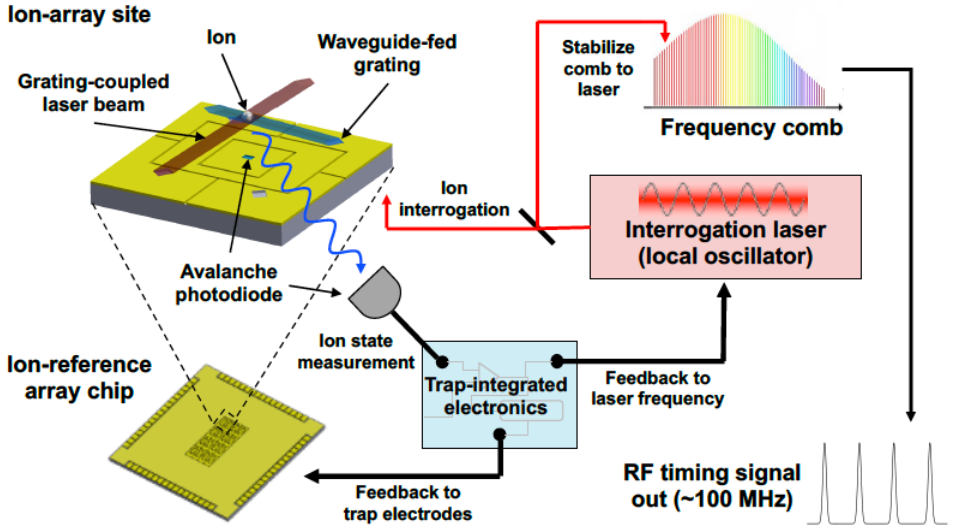
Atomic Clock: Joint DOD – DOE project

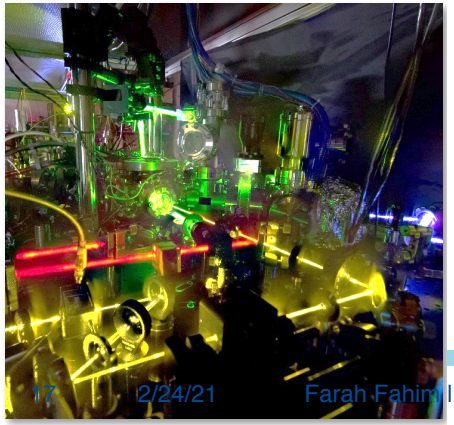
H. Sun, S. Li

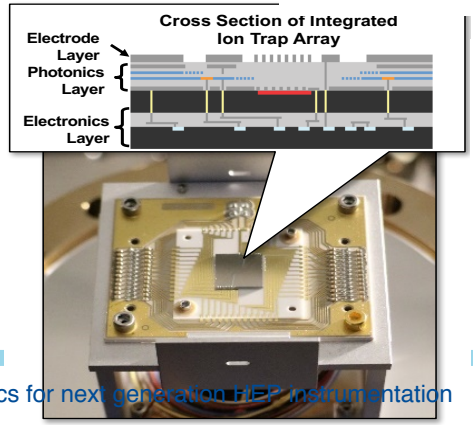
Portable optical atomic clock with frequency instability of 10^{-16} over 10,000 sec

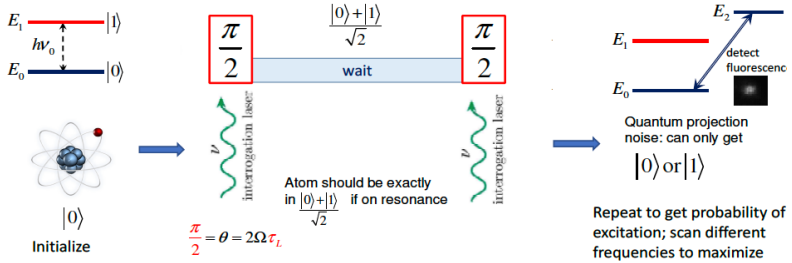
DOD – Atomic Photonic Integration
DOE – Electronic Photonic Integration

Create a closed loop compact system









Fermilab

Hardware-Software codesign: AI

Why do we need data processing on the edge

POWER: CV^2f x (data volume) problem

- Total power consumption to move data from pixel to periphery: 1 pJ/bit (~ 5mm distance)
- Total power consumption to move data off-chip: > 0.1 nJ/bit

Minimize C,V

- 3D Integration (high density, low capacitance interconnect)
- Low voltage signaling

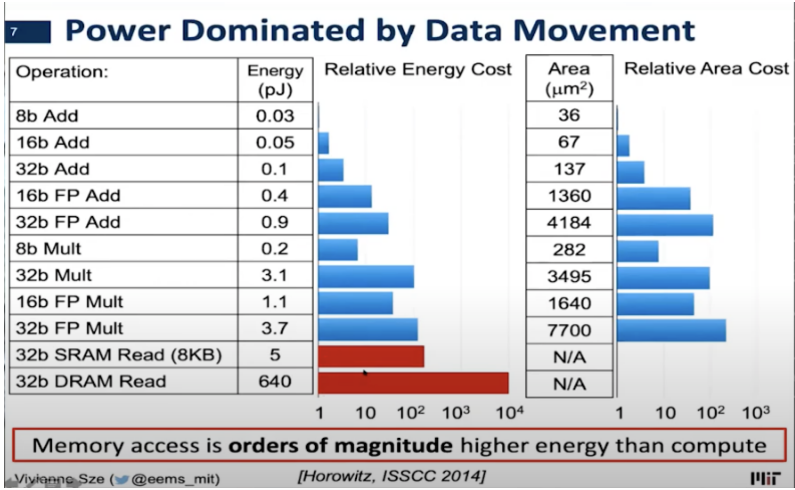
Reduce data

- Typically just zero-suppression for on-detector sparce data

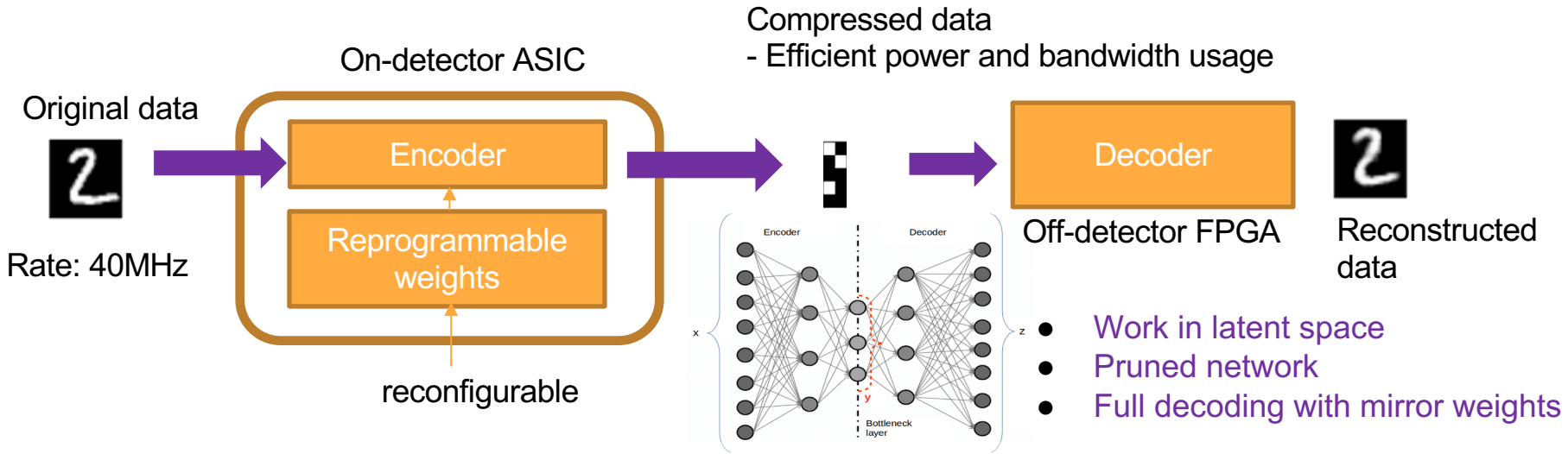
HL LHC:

Higher granularity, higher occupancy, higher precision

=> needs NEW APPROACH



Deep Neural Network: Autoencoder for data-compression



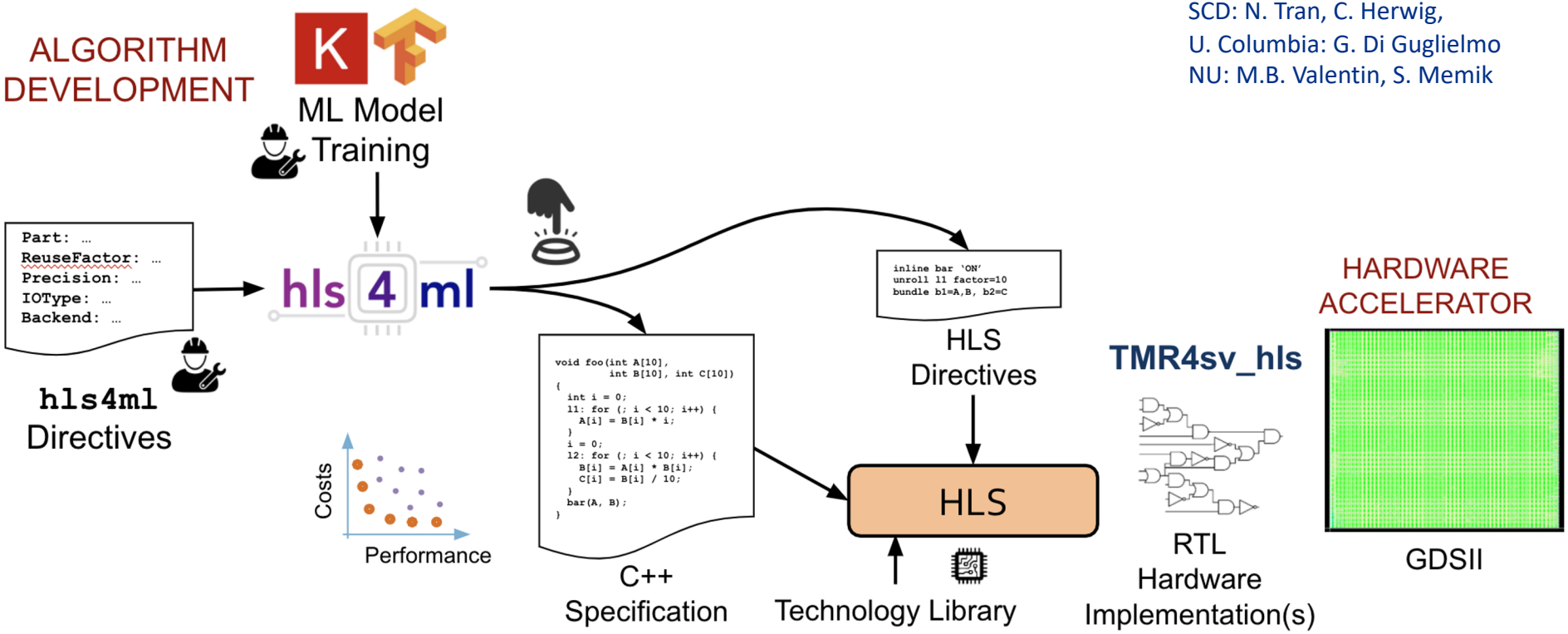
- Enable edge compute : Data compression for efficient usage of power and bandwidth
- Programmable and Reconfigurable: ability to reprogram weights to adjust for detector conditions and eventually lead to self-learning intelligent detectors
- Hardware – Software codesign : Algorithm driven architectural approach
- Optimized : Low power and Low latency
- Operating in extreme radiation environment: 200 M rad

- Autoencoder for data compression is the first use case towards a DNN based on-chip learning and inference

Tool-kit development and Operation in rad-hard environment

- Integration of HLS generated and expert RTL
- Design code agnostic approach for implementation of various triplication methods

SCD: N. Tran, C. Herwig,
U. Columbia: G. Di Guglielmo
NU: M.B. Valentin, S. Memik

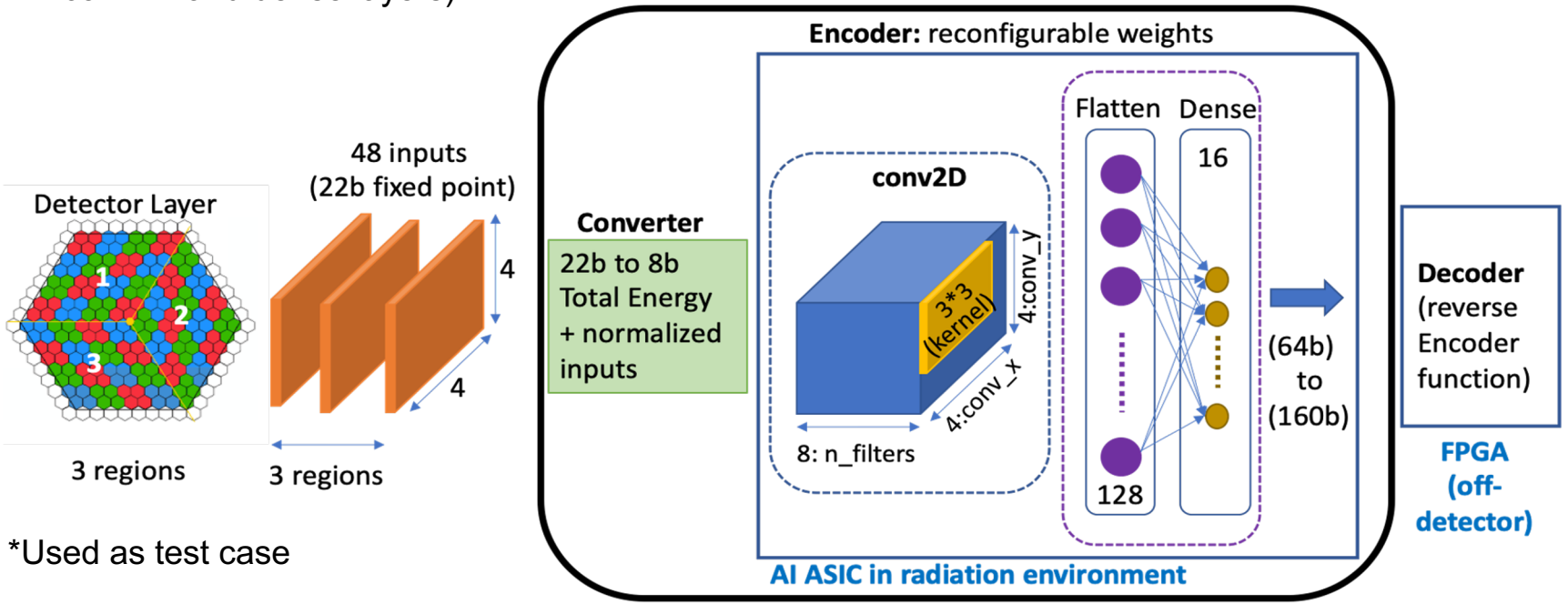


HL LHC High Granularity Calorimeter*: Data flow

J. Hirschauer
SCD: N. Tran, C. Herwig,
U. Columbia: G. Di Guglielmo
NU: M.B. Valentin, S. Memik

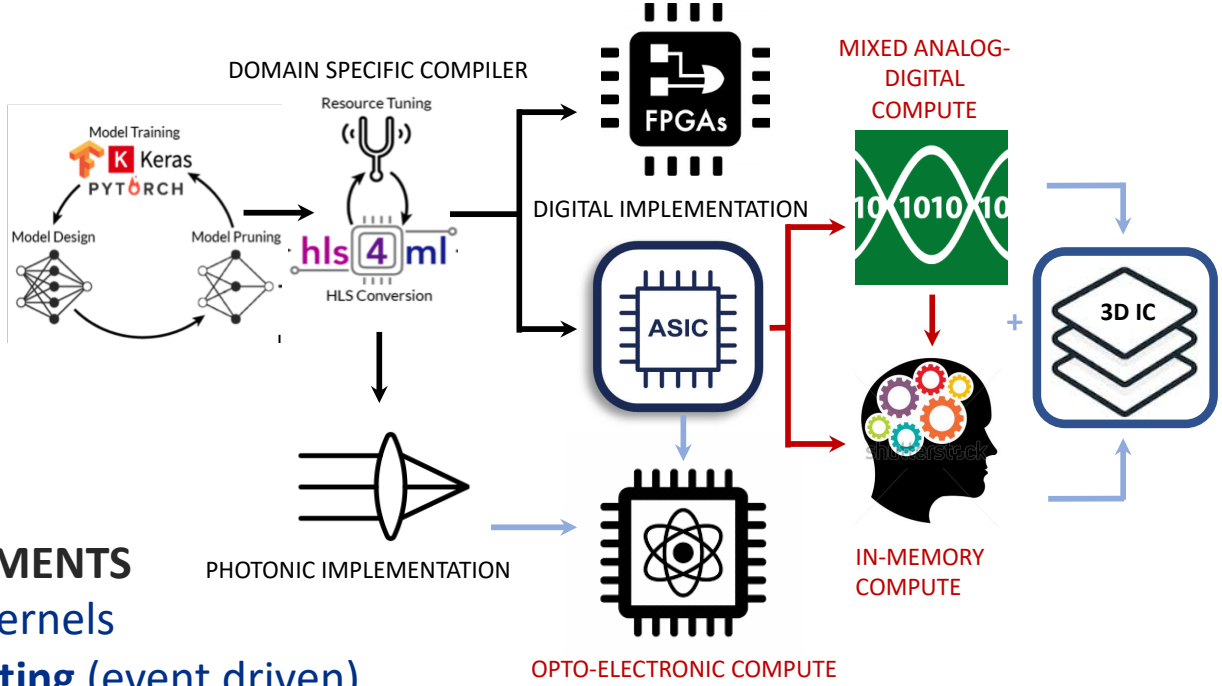
CNN: Encodes information by correlating spatial features

- **conv2D layer** – extract spatially corelated geometric features
- **Flatten layer** – Vectorizes the 2D image from the conv2D layer [8 x 4 x 4 = 128 x 1]
- **Dense layer** – aggregates the various features to provide higher order information
- **ReLU** – an activation function which introduces non-linearity by applying thresholds (part of both the conv2D and dense layers)



*Used as test case

Towards heterogenous system on-chip



OPTIMIZATION REQUIREMENTS

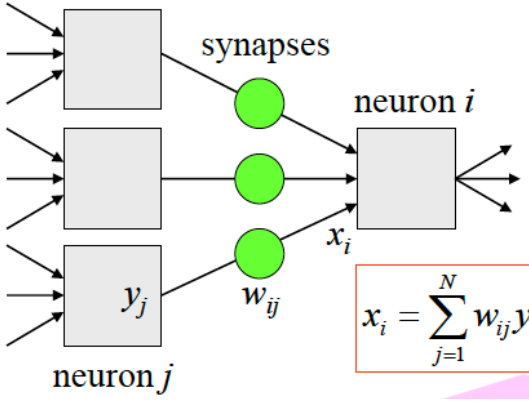
- Analog Mixed-Signal Kernels
- **Neuromorphic computing** (event driven)
- In-memory compute (**non-Von Neumann approaches**) – new materials
- **Electronic-Photonic conversion**
- **Hybrid integration**

Fermilab

Vector Matrix Multiplication for Neural Networks

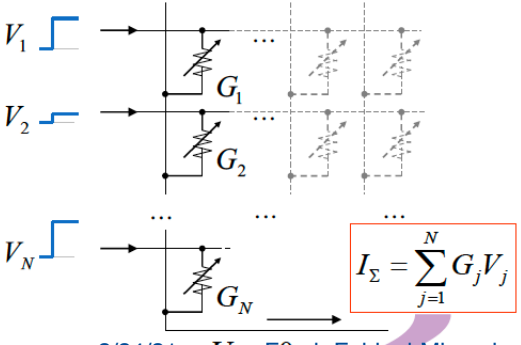
D. Strukov, UCSB

Vector-by-Matrix Multiplication …



$$x_i = \sum_{j=1}^{N} w_{ij} y_j$$

… by Analog Circuit

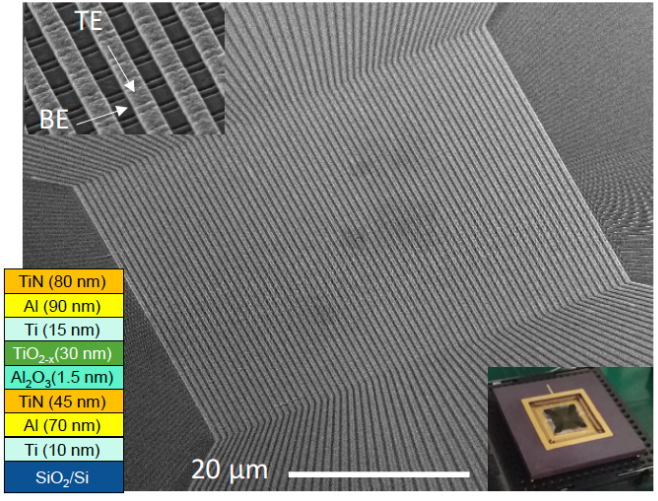


$$I_\Sigma = \sum_{j=1}^{N} G_j V_j$$

$U_\Sigma \approx 0$

UC Santa Barbara's Metal-Oxide Memristors

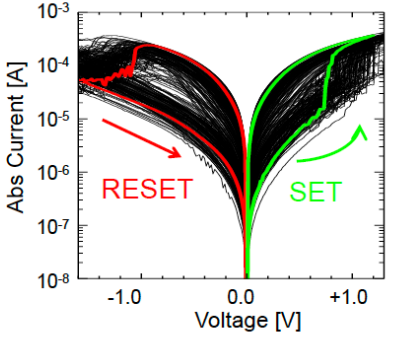
- 64 × 64 passive crossbar circuit



H. Kim et al. arXiv 2019

Background work: *M. Prezioso et al., Nature 521, 61 2015, M. Prezioso et al. IEDM'15 p. 17.4.1, 2015, F. Merrikh Bayat et al. Nature Comm., 2018*

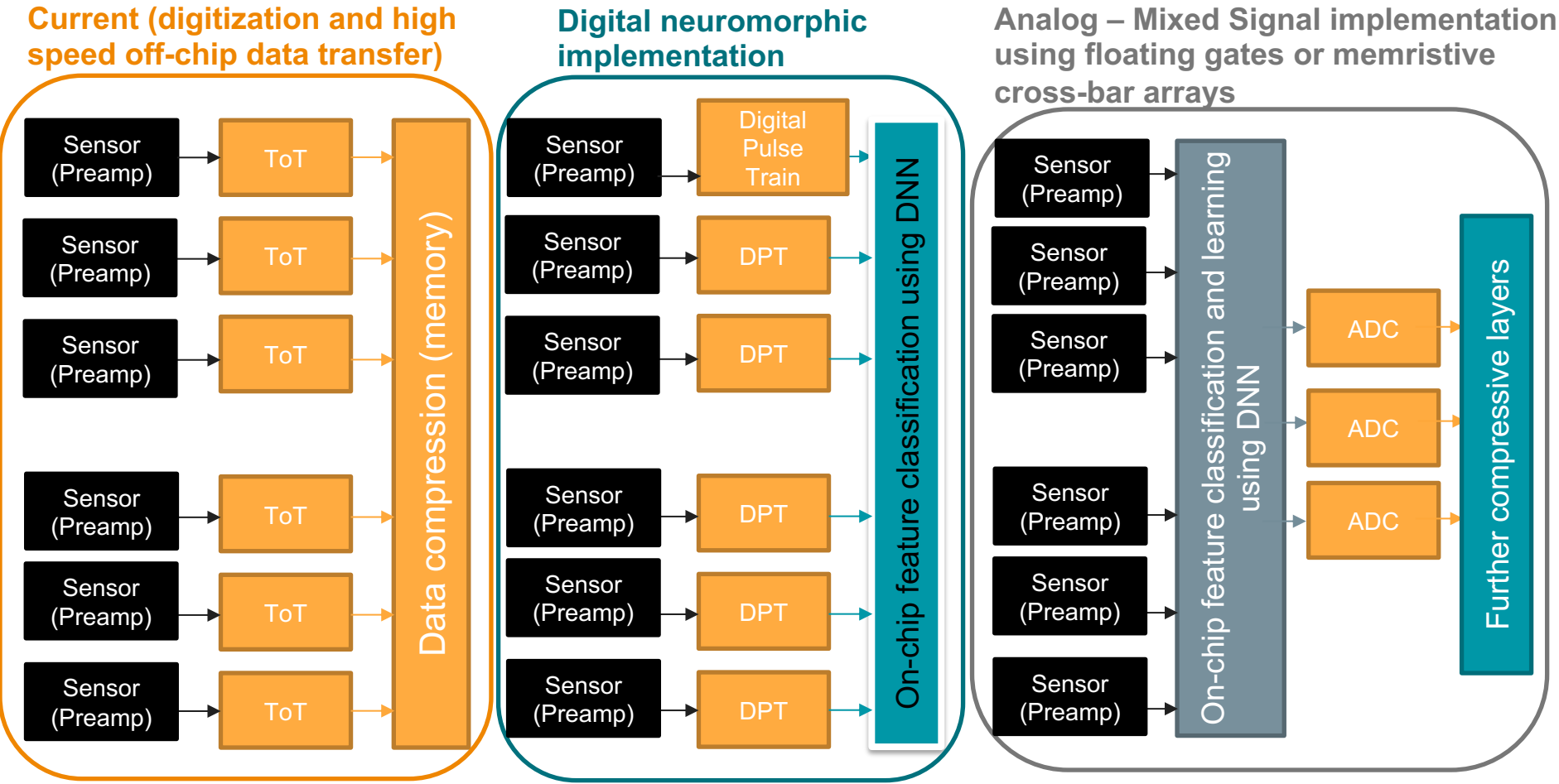
- Typical I-V characteristics



Details:

- Al_2O_3/TiO_{2-x} active bilayer by reactive sputtering
- CMOS-compatible CMP/dry etching process and TiN/Al electrodes for higher conductance
- ~250 nm wide lines
- The largest functional analog-grade passive memristor crossbar circuit supported by proper statistics

In-Pixel AI


Current (digitization and high speed off-chip data transfer)
Sensor (Preamp)
ToT
Data compression (memory)
Digital neuromorphic implementation
Digital Pulse Train
DPT
On-chip feature classification using DNN
Analog – Mixed Signal implementation using floating gates or memristive cross-bar arrays
ADC
On-chip feature classification and learning using DNN
Further compressive layers

Thankyou