Center for Ubiquitous Connectivity

JUMP 2.0 Theme 2: Communications and Connectivity

Director: Keren Bergman, Columbia University
Co-Director: Ali Niknejad, UC Berkeley

Columbia University | UC Berkeley | Massachusetts Institute of Technology
University of Illinois Urbana-Champaign | UC Santa Barbara | Princeton University
Duke University | Oregon State University | University of Michigan
University of Southern California | UC San Diego | Cornell University | Stanford University

SRC e-Workshop: CUbiC Center Plan and Vision

JUMP 2.0: Seven Research Themes

Cognition
- Systems & Architectures for next-gen AI
- Distributed: Heterogeneous, Cloud, Edge

Communications & Connectivity
- Orders of magnitude improvements in capacity, energy-efficiency, security, and resiliency

Intelligent Sensing to Action
- Analog & Mixed Signal Systems
- Sensor Fusion, Processing, and Interpretation

Systems & Architectures for Distributed Compute
- Distributed = Heterogeneous, Edge, Cloud
- Energy Efficient Compute
- Accelerators & Accelerator Fabric

Intelligent Memory & Storage
- Full-stack optimization of Intelligent Memory systems
- Emerging memory devices & arrays, including enhanced functionalities

Advanced Monolithic and Heterogenous Integration
- Interconnect Fabrics including Photonics
- Architectures and Applications for Advanced Packaging

High-performance energy efficient devices for Digital & Analog Applications
- Novel Materials, Devices, & Interconnect technologies

System Themes

Novel Materials, Devices, and Interconnect embedded

Technology Themes

Common To All
- Energy Efficiency
- Security & Resiliency: Trusted, Private, Robust, Assured, Safe
- Platform Demonstrator(s)
- Prototyping, & Scenarios & Analysis
- CoDesign Integration Capability
- Modeling & Simulation
- Benchmarking
- Metrology & Characterization

SRC

CUbiC

Agenda

- Challenges & CUbiC's Vision
- Research Plan
  - Theme 1: Connectivity Networks and Systems
  - Theme 2: Wireline and Lightwave Interconnects
  - Theme 3: wireless Circuits and Technology
- Integration & System Testbed Platforms
- Summary

Edge to Cloud Connectivity Challenges

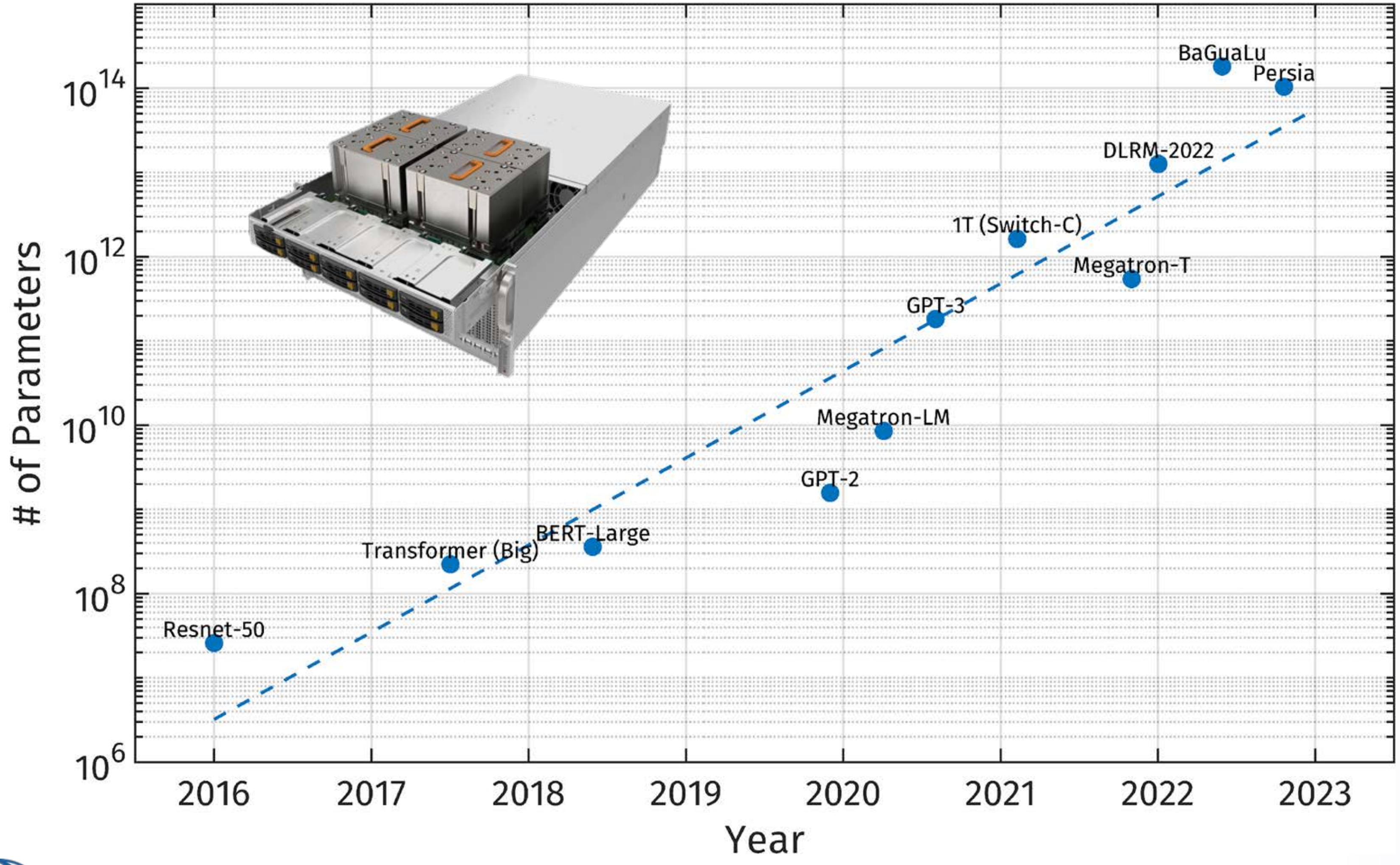
Explosive Growth in Data Communication Demands

Cloud Connectivity Challenges:
- Orders of magnitude gap between on-chip/off-chip BW
- Strong distance-dependent communication energy
- Scalability limited by energy and bandwidth tapering
- Massive heterogeneity – compute/memory/accelerator

Edge Connectivity Challenges:
- Driving mm-Wave capacity to meet data demand with robustness, reliability, mobility, and low cost
- Massive densification, power, loss, thermal cooling
- Long-range links - back-haul, long range front-haul, airborne links - limited by output power

System Connectivity Challenges:
- Seamless connectivity between edge and cloud for optimized cross-layer performance
- Reconfigurable, adaptable connectivity to accelerate heterogeneous applications
- Secure and resilient connectivity across edge and cloud

AI Applications Driving Ever Larger Models in Cloud



Model size increases <u>one order of magnitude per year</u>

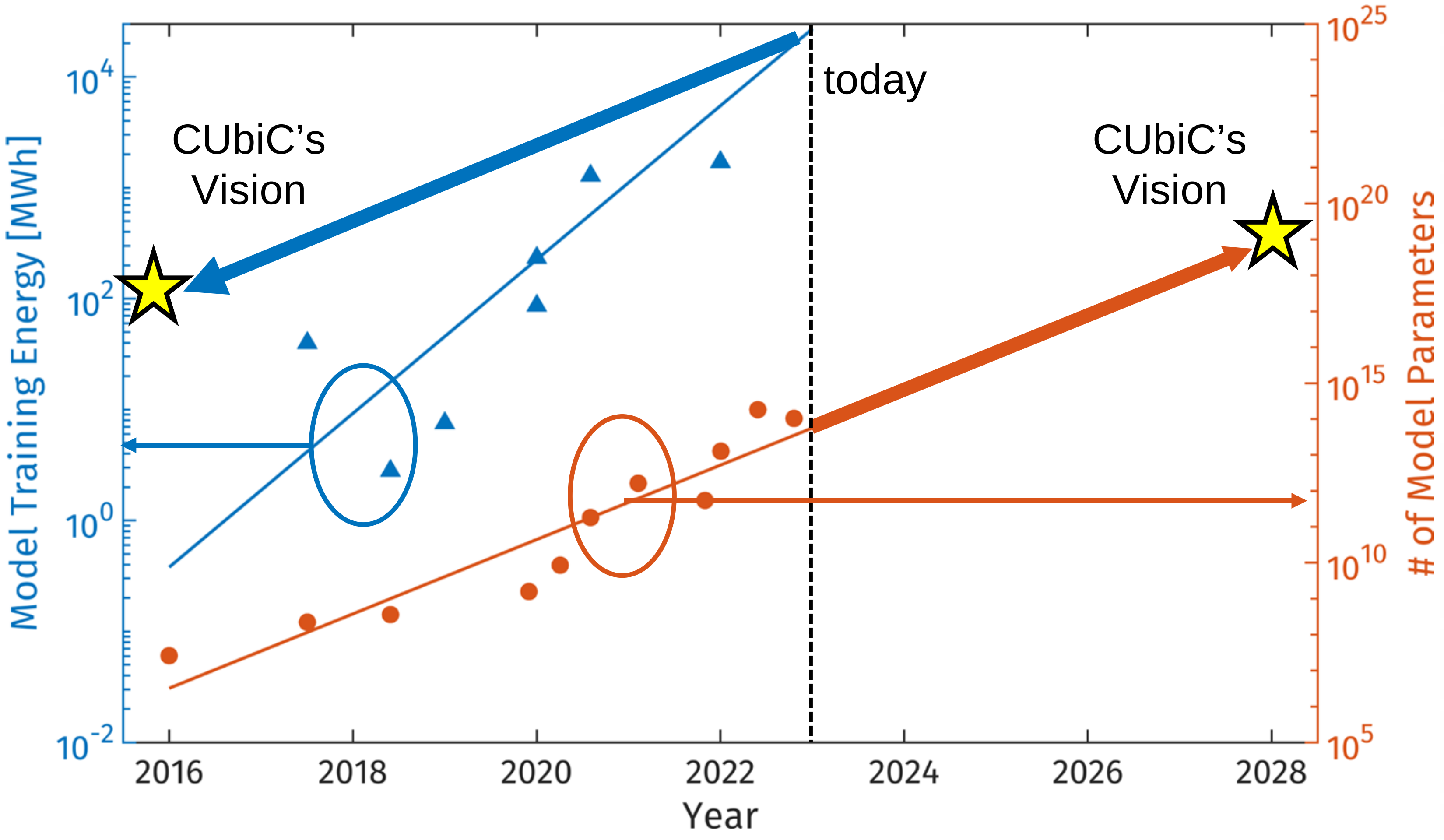
> 100 Trillion parameters

Exceeding memory capacity of any single compute socket

SRC

CUbiC

Cloud Performance Scaling with Reduced Energy



CUbiC enables:

- Scale Performance while
- reducing energy by > 100 X
- Flattened BW/energy across the system by bringing photonics into the socket
- System wide flexible photonic connectivity for accelerating AI/ML/HPC applications

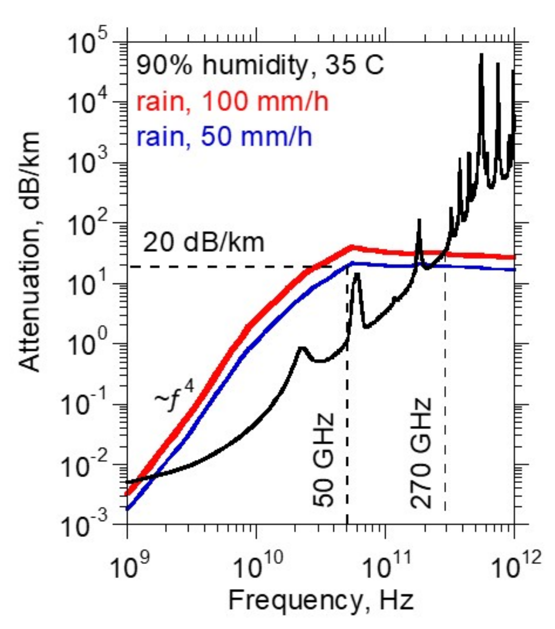
SRC

CUbiC

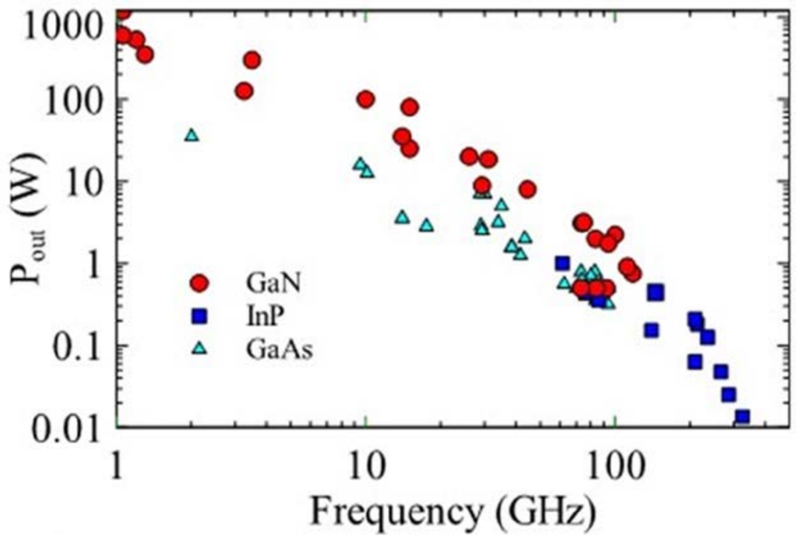
Data source: D. Patterson, et al. "Carbon Emissions and Large Neural Network Training", arXiv:2104.10350

Wireless Edge: Challenges Moving to Higher Frequencies

- **High Frequency Millimeter-Wave Wireless:**
  - More available spectrum; Higher data rates per beam
  - Massive spatial multiplexing; Compact arrays



- **Key Deployment Challenges – even at 5G**



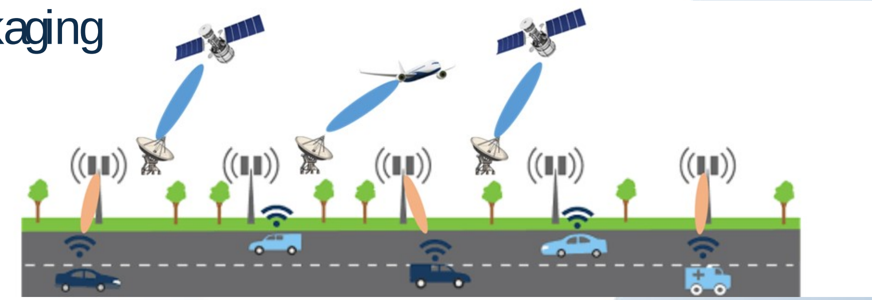
- High atmospheric and λ^2/R^2 losses; limited range
- Massive densification required to provide coverage
- Closely-spaced base stations; high deployment cost
- Extending range requires arrays with many (10^2-10^3) elements
- Dense $\lambda/2$ array pitch; higher power per element; high cost
- Thermal cooling limits practical antenna count and footprint
- Long range applications limited by output power (CMOS/SiGe cannot offer watt level power efficiently) need GaN / III-V

SRC

CUbiC

Road to the Next-G Wireless Connectivity

- **Accelerate Adoption:**
  - Increase 5G systems capacity; all-digital massive MIMO both 28 GHz and 100+ GHz
  - Advanced efficient DSP algorithms for multi-user MIMO
  - Cost-effective densification via O-RAN (low-cost remote radio heads and backhaul)
- **Longer Range, High Capacity, with Low DC Power, and Low Cost:**
  - Massive 2-D arrays with advanced semiconductors and high-density packaging
  - Large-scale, inexpensive CMOS mm-wave arrays
  - Smaller active arrays plus massive steerable passive retroreflectors



- **Highly Flexible, High Spectral Efficiency, Robust Systems:**
  - Favorable to O-RAN distributed architectures; routing massive data efficiently
  - Systems inherently robust to interference in the RF/analog domain; not relying on digital baseband
- **High Frequency GaN Devices:** high output power, thermal cooling, favorable to *wireless* applications

SRC

CUbiC

CUbiC Vision

Flatten the computation-communication gap at both the Edge and the Cloud to deliver seamless Edge-to-Cloud connectivity with transformational reductions in the global system energy consumption.

Grand Challenge:

Realize robust, scalable Edge to Cloud connectivity at > 10 Tbps with sub-pJ/bit energy efficiencies while enhancing bandwidth densities by >100X over capacity-constrained channels

SRC

CUbiC

CUbiC Team: 23 PIs from 13 Universities

Columbia University in the City of New York
Berkeley University of California
MIT Massachusetts Institute of Technology
University of Illinois Urbana-Champaign
UC Santa Barbara
Princeton University
Duke University
Oregon State University
University of Michigan

USC University of Southern California
UC San Diego
Cornell University

Stanford University

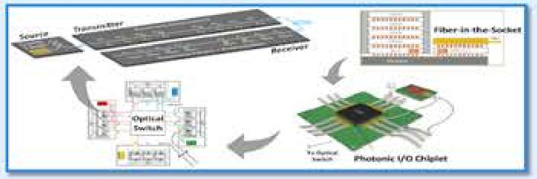
SRC

CUbiC

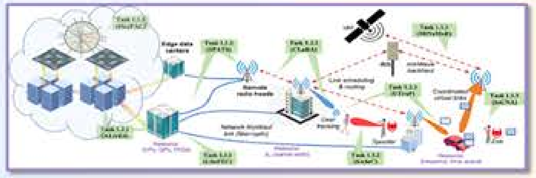
CUbiC Research Plan

Vertically Integrated Research Organization

Theme 2: Wireline and Lightwave Interconnects



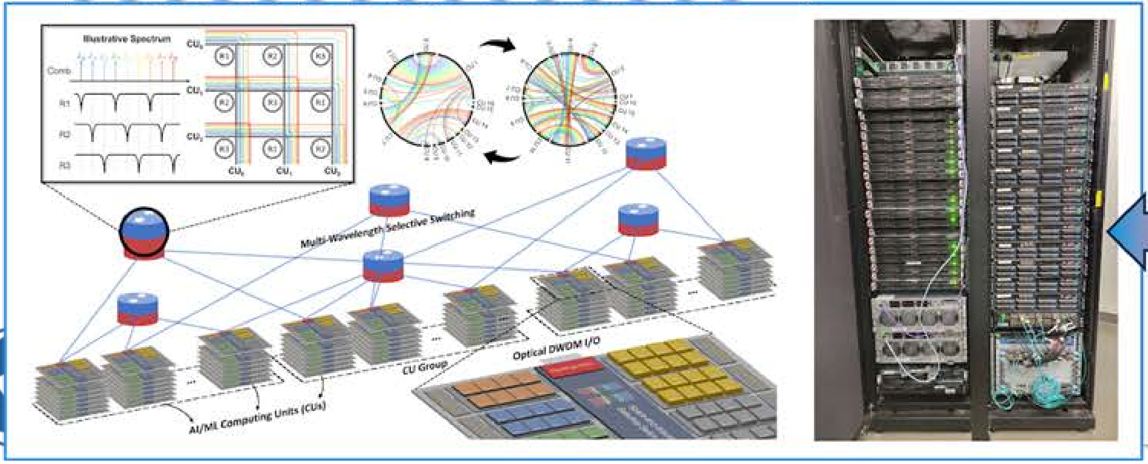
Theme 1: Connectivity Networks and Systems



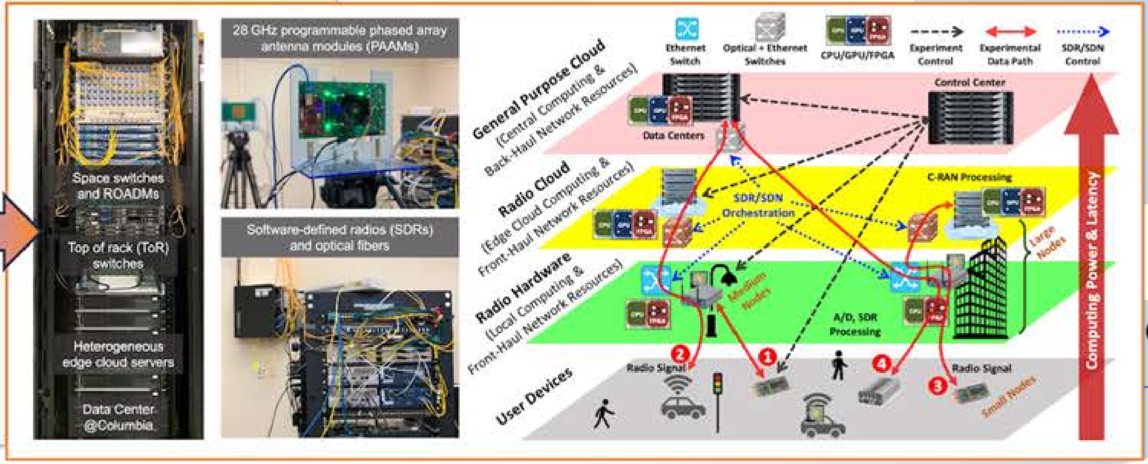
Theme 3: Wireless Circuits and Technology

- 1.1 Terabit/s PHY Systems
- 2.1 Systems & Algorithms for Connectivity
- 3.1 Large-Scale Millimeter-wave Arrays
- 2.2 Circuits & Architectures for Links
- 3.2 Wideband HDR Analog Signal Processing Interfaces
- 2.3 Circuits & Architectures for Switches
- 1.2 Cross-layer Design of Terabit/s Networks
- 3.3 High-Dimensional DSP for Emerging Wireless
- 1.3 Security and Resiliency
- 2.4 Photonic Devices for Connectivity
- 3.4 Heterogeneous Technologies for Next Generation Wireless
- 1.4 System Connectivity Platforms

SoSFab

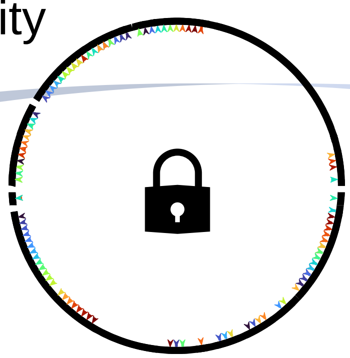


ReACT



Theme 1: Addressing System Connectivity Challenges

Enabling Cross-Layer Seamless, Adaptable, Resilient, and Secure Connectivity



System applications driven connectivity challenges:

- Inform and drive technologies in Themes 2 and 3
- Uncover fundamental system bottlenecks
- Enable co-design of hardware, algorithms, system, network architectures, and applications.

Bandwidth Tapering vs. Data Demand Informing Theme 2

Edge Ubiquity vs. High-Frequency Challenges Informing Theme 3

SRC

CUbiC

Theme 1: Task Organization

1.1 Terabits/s PHY Systems

Task 1.1.1: Flexible Photonic Accelerated Computing (FlexPAC)
Bergman, Ghobadi, Lipson

Task 1.1.2: Signal Processing Architectures for Terabit/s Scaling (SPATS)
Madhow, T. Chen, Zhang

Task 1.1.3: Millimeter Wave Networking at Extreme Mobility and Range (MiNxMoR)
Madhow, Ghasempour, Krishnaswami, Rodwell

1.2 Cross-layer Design of Terabit/s Networks

Task 1.2.1: Application-aware Links for Edge and Cloud (ALfrEd)
Ghobadi, Zhang, Stojanovic, Mahdavifar

Task 1.2.2: Cross-Layer Resource Allocation for Terabit vRANs (CLaRA)
T. Chen, Madhow, Ghasempour

Task 1.2.3: User Tracking and Propagation Mapping for Seamless Connectivity (UTraP)
Ghasempour, Madhow, T. Chen

1.3 Security and Resilience

Task 1.3.1: Lightweight Forward Error Correction (LiteFEC)
Mahdavifar, Zhang, Shanbhag

Task 1.3.2: Coding for Authenticated Secure Connectivity (CASeC)
Mahdavifar, Ghasempour

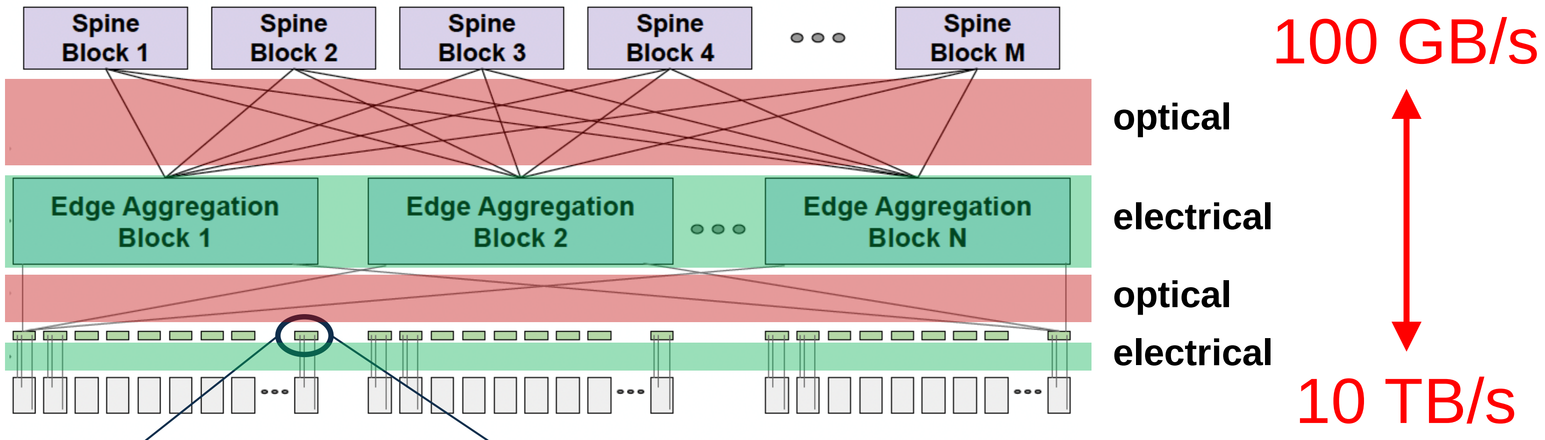
Task 1.3.3: Secure Cross-layer Network Architectures (SeCNA)
Ghasempour, Mahdavifar, Madhow, T. Chen

1.4 Platforms/Testbeds

Task 1.4.1: Socket-to-Socket Distributed AI/ML/HPC Fabric Platform (SoSFab)
Ghobadi, ***Stojanovic***, Bergman, Wu

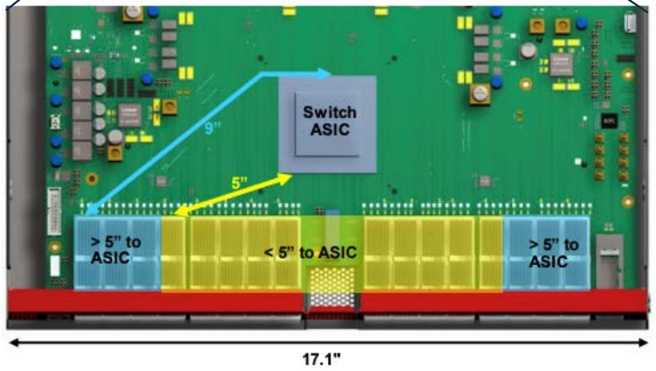
Task 1.4.2: CUbiC Real-time Antenna-to-Compute Testbed (ReACT)
T. Chen, ***Krishnaswamy***, Niknejad, Ghasempour

Theme 2:
Addressing Connectivity Challenges Within the Cloud



Transform wireline technologies: <u>electrical and optical</u> connectivity to address **bandwidth tapering**

Switch Fabric

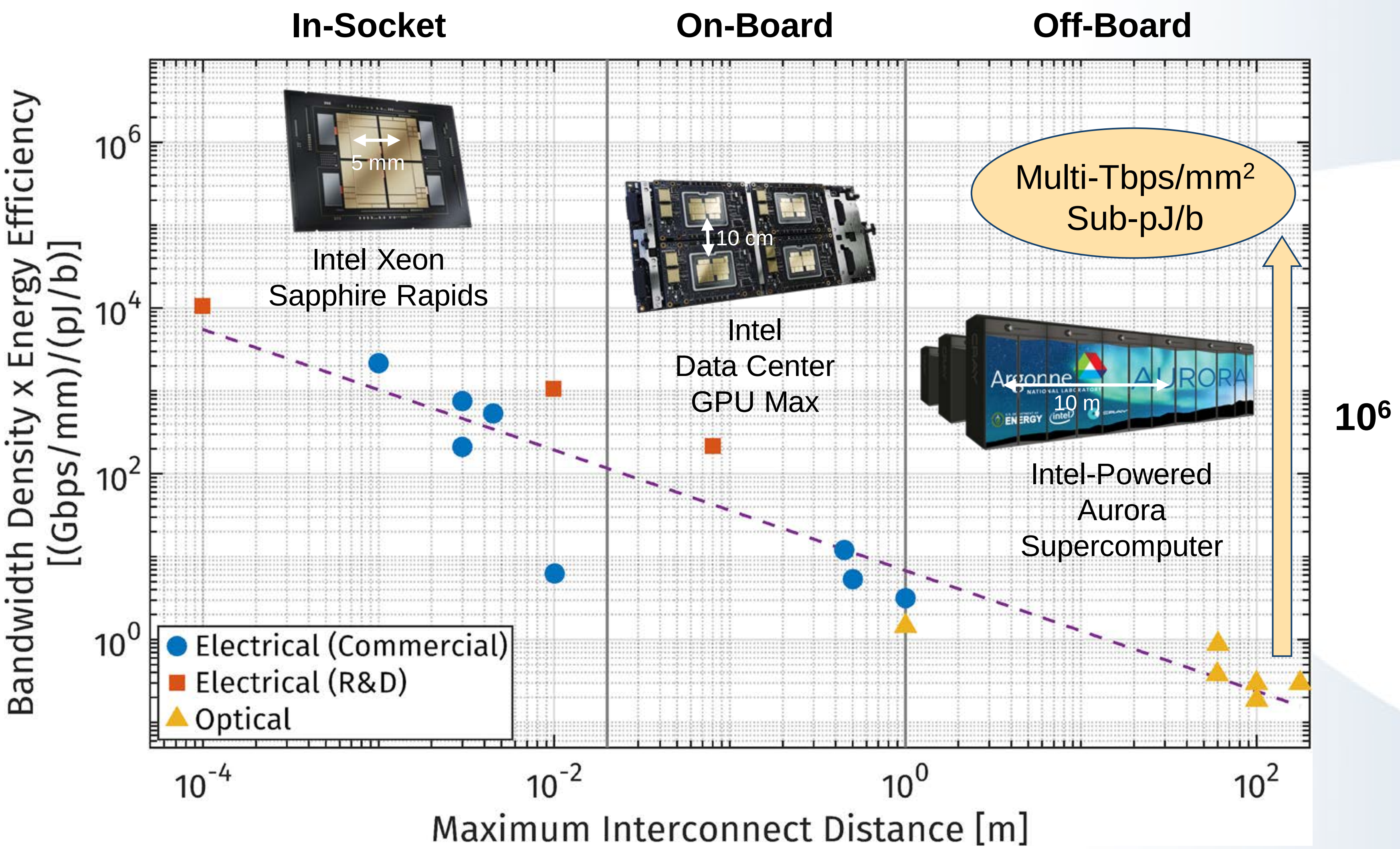


- AI/ML applications & heterogeneous integration drive ever increasing BW requirements;
- Application-level efficiency and performance limited by huge compute vs. comm. gap in BW (10^2X) and energy efficiency (10^3X)

SRC

CUbiC

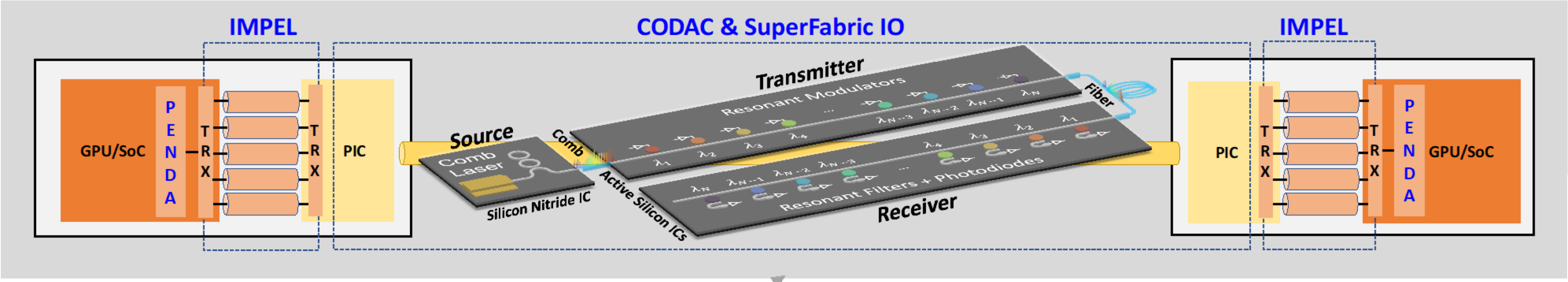
Theme 2: Bringing Photonics to the Socket

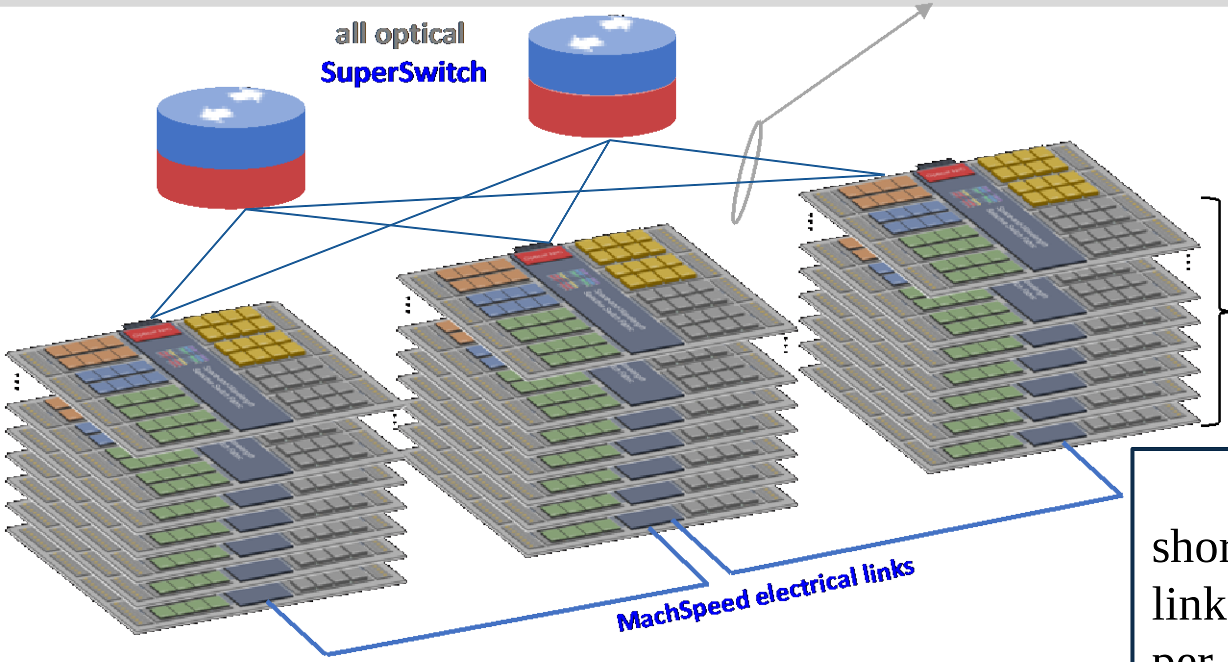


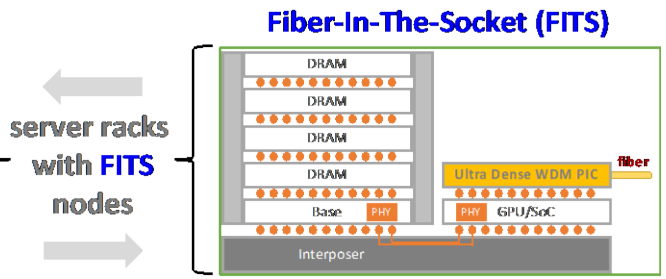
Adapted from Gordon Keeler, DARPA

Theme 2: Wireline and Lightwave Interconnects

pervasive system-wide socket-to-socket photonics-based connectivity







Target Metrics

shoreline BW density: > 5-10 Tb/s/mm

link energy efficiencies: 0.25-1 pJ/b

per-fiber bandwidths :1-10 Tb/s

off-package bandwidth :100+ Tb/s

SRC

CUbiC

Theme 2: Task Organization

2.1 Systems & Algorithms for Connectivity

Task 2.1.1: Algorithms for Energy-efficient Connectivity (ALEC)
Shanbhag, Hanumolu, Madhow

Task 2.1.2: Programmable Energy-efficient DSP Architectures (PENDA)
Shanbhag, Hanumolu, Zhang

Task 2.1.3: Adaptive Low-Cost High-Speed ADC (ALoHA)
M. Chen, Niknejad, Shanbhag, Hanumolu

2.2 Circuits & Architectures for Links

Task 2.2.1: SuperFabric IO
Stojanovic, Wu, Ghobadi, Hanumolu

Task 2.2.2: Coherent Optics in the Data Center (CODAC)
Hanumolu, Shanbhag, Bowers, Stojanovic

Task 2.2.3: Integrated Massively Parallel Electrical Links (IMPEL)
Hanumolu, Shanbhag, Bergman, Bowers

Task 2.2.4: Machine Learning-inspired High-Speed Links (MachSpeed)
Anand, Mahdavifar

2.3 Circuits & Architectures for Switches

Task 2.3.1: SuperSwitch – A high-radix silicon photonic switch
Wu, Stojanovic, Ghobadi

Task 2.3.2: SuperSwitch Controller – A controller for SuperSwitch
Stojanovic, Wu, Ghobadi

Task 2.3.3: Optical Packaging of SuperSwitch and SuperSwitch Controller
Wu, Stojanovic, Bergman, Ghobadi

2.4 Photonic Devices for Connectivity

Task 2.4.1: Fiber-In-The-Socket (FITS)
Bowers, Bergman, Stojanovic

Task 2.4.2: High-bandwidth Mode Coupling (HaMoC) ***Lipson***, Bergman

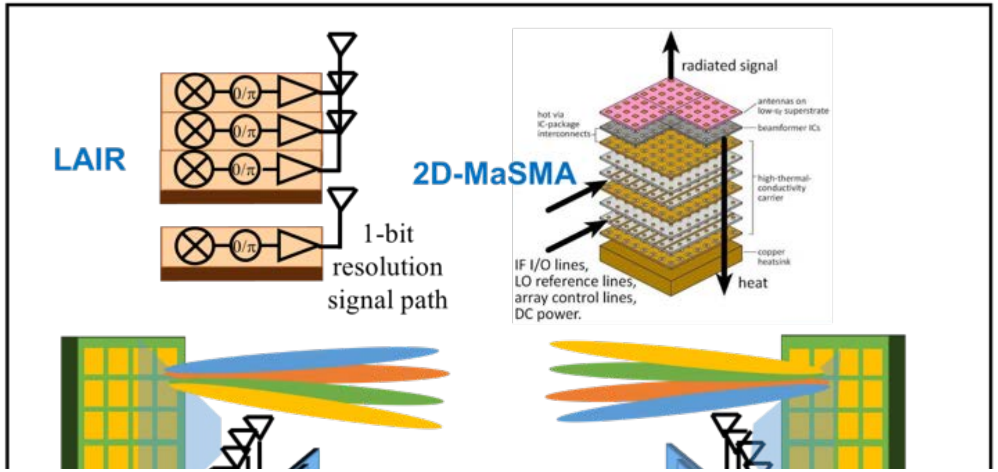
Task 2.4.3: Photonic Resonators for Ultra High-Bandwidth and Efficiency (PRUnE)
Lipson, Krishnaswamy

SRC

CUbiC

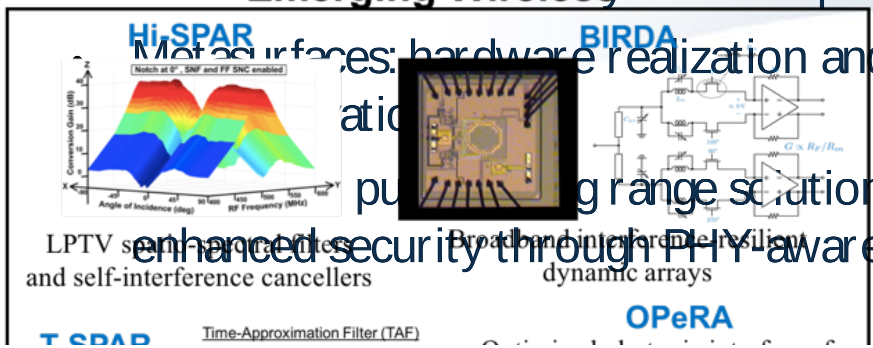
Theme 3: Wireless Circuits and Technology

Large-Scale Millimeter-Wave Arrays



Wideband High-Dynamic Range Analog Signal Processing Interfaces and High-Dimensional Digital Signal Processing for Emerging Wireless

- Massive 1D and 2D arrays: RFIC + packaging
- Metasurfaces: hardware realization and system-...atio...
- ...pu...g range solutions with enhanced security through PHY-aware O-RAN



- Drive innovation in III-V technology for higher gain, output power and efficiency
- Thermal cooling to enable compact integrated solutions at sub-wavelength pitch

- T... dynamic ra... arrays for reduced ... & power
- S... cessing inte... critical sou... ...erence
- C... hemes for r... ...ble performance at low energy

Heterogeneous Technologies for Next Generation Wireless

SRC

CUbiC

Theme 3: Task Organization

3.1 Large-Scale Millimeter-wave Arrays

Task 3.1.1: Large Arrays of Inexpensive Radios (LAIR)
Rodwell, Niknejad

Task 3.1.2: Reconfigurable Energy-efficient Active Intelligent Metasurfaces (RE-AIM)
Krishnaswamy, Rebeiz

Task 3.1.3: 2D Massively-Scalable Millimeter-wave Arrays (2D-MaSMA)
Rebeiz, Rodwell, Mishra

3.2 Wideband High-Dynamic-Range Analog Signal Processing Interfaces

Task 3.2.1: High-dynamic-range Interference-tolerant Wideband Signal Processing at Analog/RF (HI-SPAR)
Molnar, Krishnaswamy, M. Chen

Task 3.2.2: Broadband Interference Resilient Dynamic Arrays (BIRDA)
Niknejad, Rebeiz, M. Chen

Task 3.2.3: Time-based Signal Processing at Radio Frequencies (T-SPAR)
M. Chen, Molnar, Krishnaswamy

Task 3.2.4: Optimized RF-to-Photonic Interfaces for mmWave Massive MIMO Arrays (OPeRA)
Molnar, Bergman, Niknejad

3.3 High-Dimensional Digital Signal Processing for Emerging Wireless

Task 3.3.1: Adaptive and Distributed Array Processing for Massive MIMO Transceivers (ADAPT)
Zhang, Madhow, Shanbhag

Task 3.3.2: Pay As You Go FEC Decoder for Array Processing (PAYGO)
Zhang, Mahdavifar, Ghobadi, Shanbhag

3.4 Heterogeneous Technologies for Next-Generation Wireless

Task 3.4.1: Devices Incorporating Ensemble Velocity Overshoot (EVO)
Mishra, Ahmadi, Chowdhury, Rodwell

Task 3.4.2: N-polar GaN Growth on Bulk GaN (GoG)
Ahmadi, Mishra, Chowdhury, Niknejad

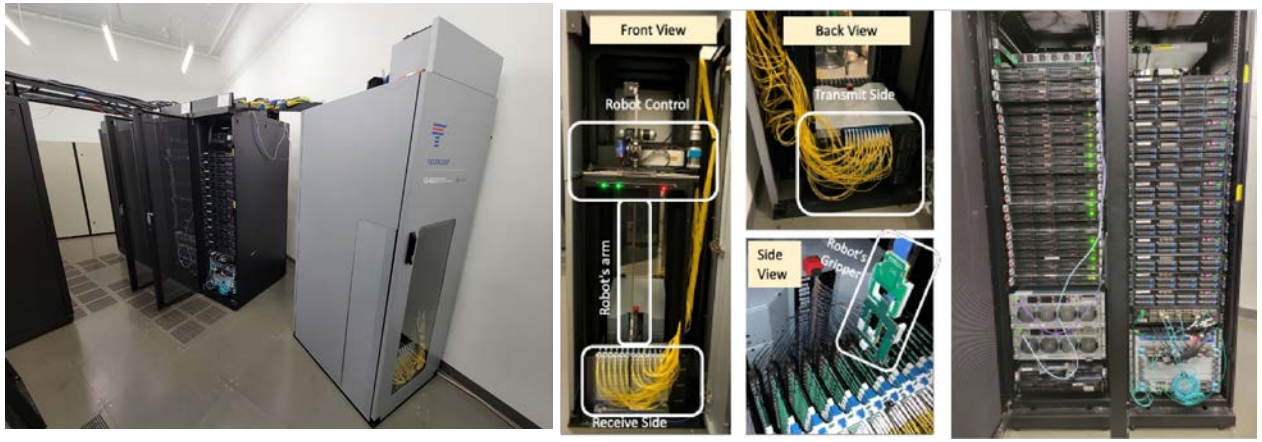
Task 3.4.3: High-K and Ferroelectric Dielectrics (HfZrO and in-situ ScAlN) (HiFi)
Ahmadi, Mishra, Chowdhury, Rodwell

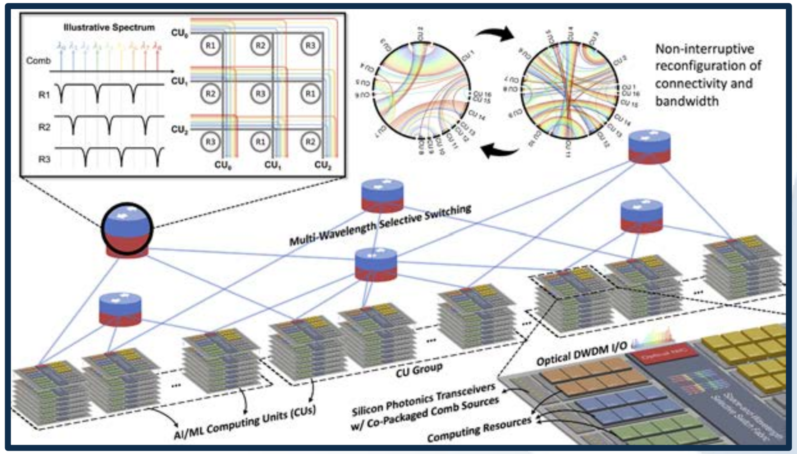
Task 3.4.4: Diamond-based Thermal Management for GaN and InP (D-Therm)
Chowdhury, Mishra, Ahmadi

CUbiC Integration & System Testbed Platforms

CUbiC System Connectivity Platform 1: SoSFab: Distributed AI/ML/HPC Fabric

- **Vision:** A system-wide energy-efficient demonstration of CUbiC data center platform
- **Approach:** Co-optimize and co-design from applications to devices to exploit synergies across the system stack, application requirements, and device capabilities.
- **Outcome:** End-to-end application demonstration CUbiC's ubiquitous connectivity that scales to meet the needs of emerging distributed data-centric applications, such as machine learning training and inference.



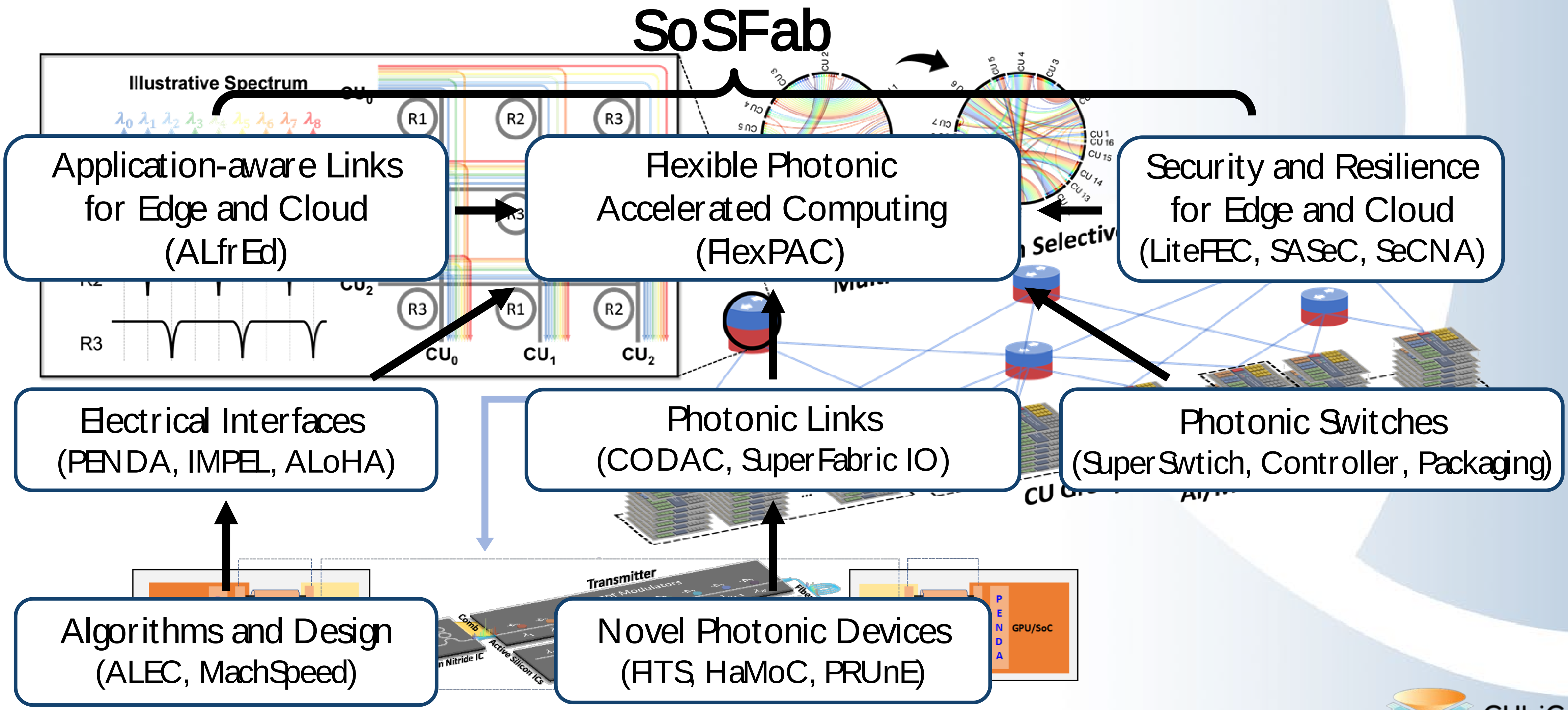


- Only known platform that enables adjusting the network topology, communication collective, parallelization strategy, and workload scheduling for AI/ML/HPC workloads.
- Potential for cooperation with the JUMP2.0 ACE Center (UIUC)

SRC

CUbiC

SoSFab: Distributed AI/ML/HPC Fabric

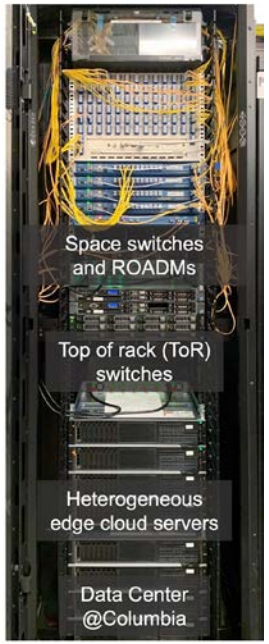


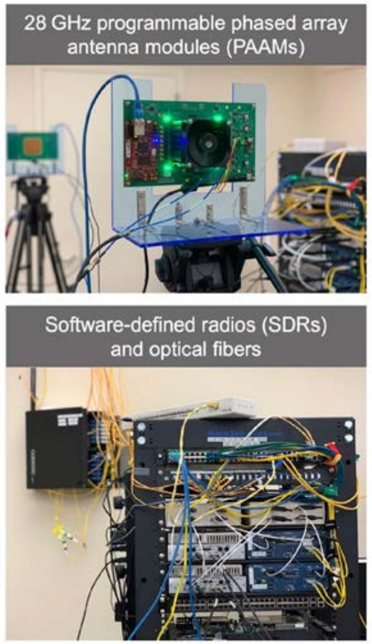
SRC

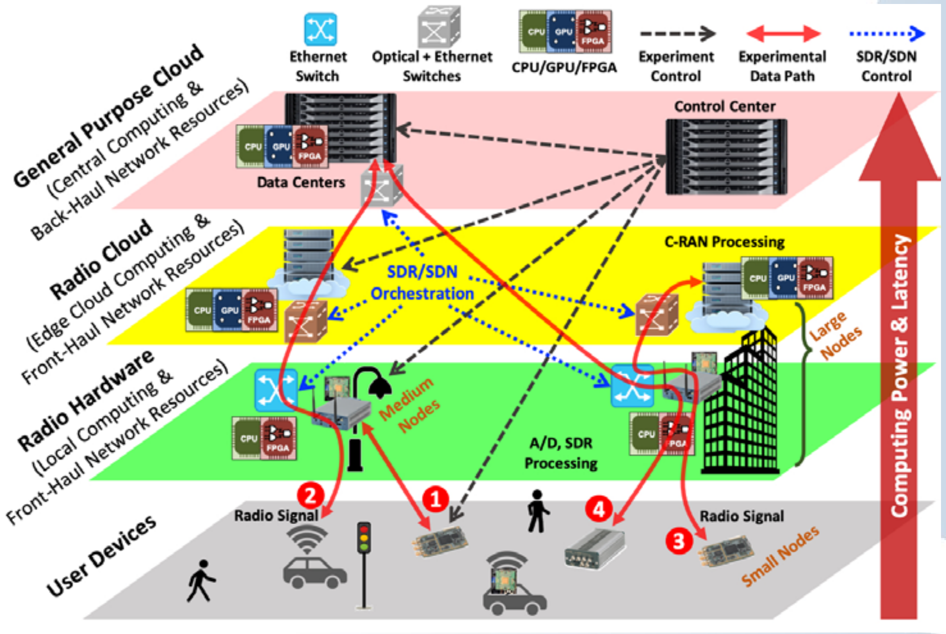
CUbiC

CUbiC System Connectivity Platform 2:
ReACT: Realtime Antenna to Compute

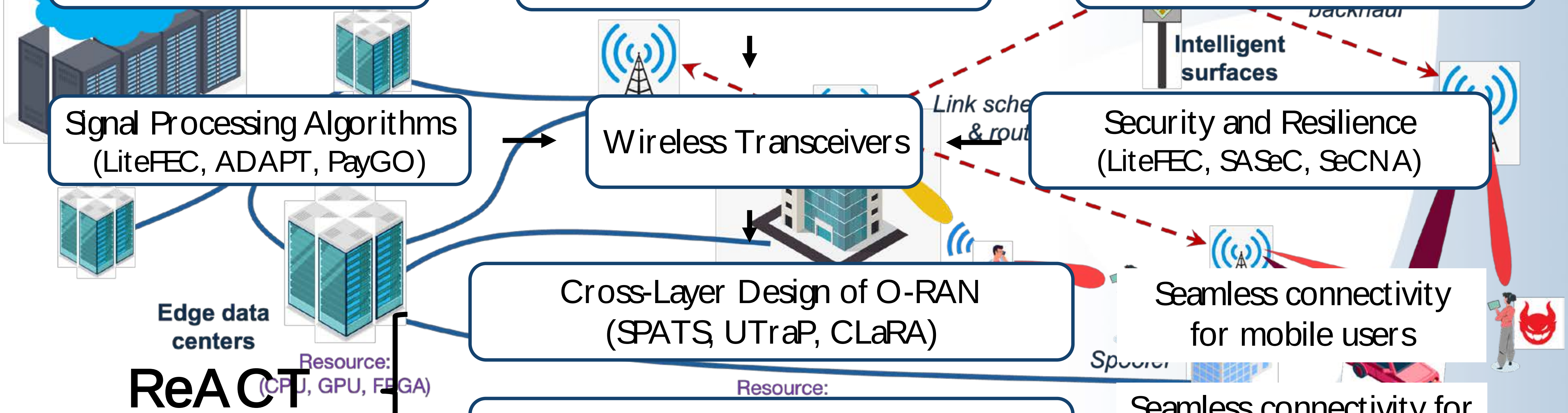
- **Vision:** a center-wide demonstrations of wireless connectivity from antenna to compute
- **Approach:** translation from advanced mm-Wave ICs developed in CUbiC labs to a programmable radio platform, O-RAN for system-level evaluation and network-level experimentation.
- **Outcome:** center-wide demonstrations that will take a holistic system approach to integrate the unique mm-Wave frontend and digital circuits, and provide the evaluation of the advanced algorithms and control plane – robust, secure







Connectivity Platform 2 – ReACT: Realtime Antenna to Compute



Wireless Device Physics
(EVO, GoG, HiFi, D-Therm)

→

Large Scale Antenna Arrays
(LAIR, 2D-MaSMA, RE-AIM)

←

Filter Design/Signal Processing
(HiSPAR, T-SPAR, BIRDA, OPeRA)

Signal Processing Algorithms
(LiteFEC, ADAPT, PayGO)

→

Wireless Transceivers

←

Security and Resilience
(LiteFEC, SASeC, SeCNA)

ReACT

Cross-Layer Design of O-RAN
(SPATS, UTraP, CLaRA)

Networks at Extreme Mobility and Range
(MiNxMoR)

Seamless connectivity for mobile users

Seamless connectivity for rapid moving platforms

Cross-Center Collaboration

Jump 2.0 Centers	Topics of Interactions
Cognition **COCOSYS (GaTech)**	• Data movement and communication challenges in next-Gen AI systems • **Naresh Shanbhag**
Communications and Connectivity **CUbiC (Columbia)**	
Intelligent Sensing to Action **CogniSense (GaTech)**	• Analog and mixed signal; massive array technology • **Al Molnar**
Systems & Architectures for distributed Compute **ACE (UIUC)**	• Photonic interconnection networks in distributed computing architectures; SoSFab testbed • **Manya Ghobadi and Zhengya Zhang**
Intelligent Memory and Storage **PRISM (UCSD)**	• High bandwidth photonic connectivity to memory; deeply disaggregated connectivity architectures
Advanced Monolithic and Heterogeneous Integration **CHIME (Penn State)**	• Advanced heterogeneous assembly and packaging; initiated joint effort on models (Shanbhag/Hanumolu) • **Michal Lipson**
High-Performance Energy-Efficient Devices for Digital and Analog Applications **SUPREME (Cornell)**	• Advanced materials and GaN devices for wireless

SRC

CUbiC

Summary

- CUbiC will strive to flatten the computation-communication gap, delivering seamless Edge-to-Cloud connectivity with transformational reductions in the global system energy consumption.

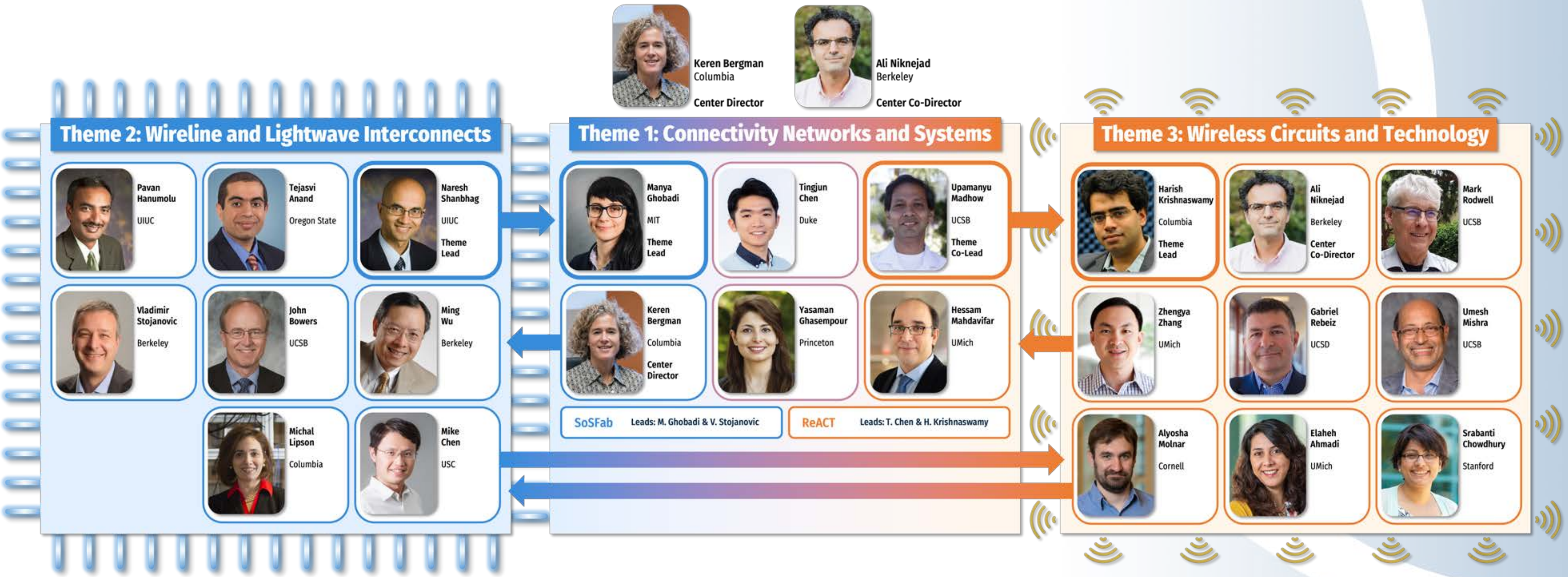
- Vertically integrated research agenda cross-cutting 3 technical themes
- Outstanding team of 23 PIs from 13 Universities
- 37 Research Tasks
- Expected: >85 graduate students per year

- June 27-28 CUbiC Annual Review

SRC

CUbiC

CUbiC's Integrated Team



SRC

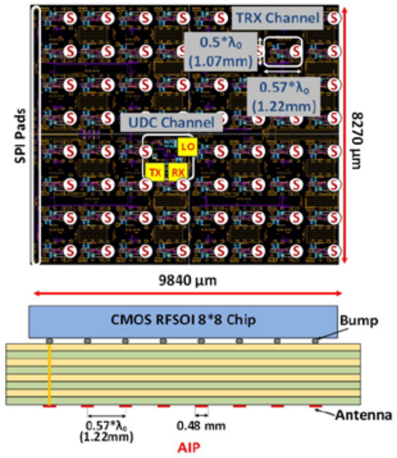
CUbiC

arm
IBM
ANALOG DEVICES
AHEAD OF WHAT'S POSSIBLE™
SK hynix
Micron
BOEING
EMD Electronics
HRL LABORATORIES
tsmc
qorvo
NSF
intel
DARPA
GlobalFoundries™
Raytheon Technologies
SAMSUNG
MEDIATEK

Building on ComSenTer Groundbreaking Outcomes

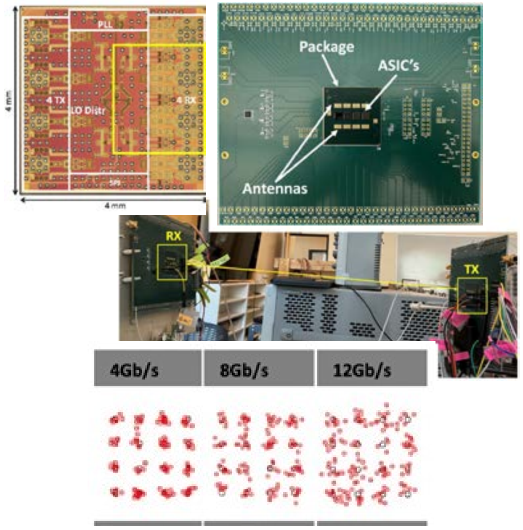
- Validation of core mmWave technology
  - 100+ GHz doable in CMOS, CMOS + III/V for increased range
  - RF and Fully digital beamforming for massive antenna arrays
  - 140 GHz Hub demo; Beamspace ICs

140 GHz 8x8 Wafer-Scale TRX Array



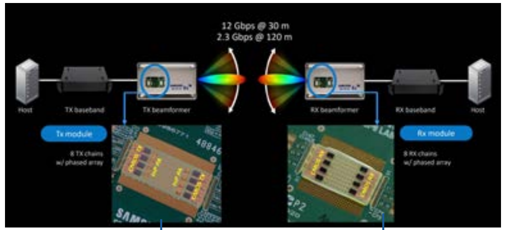
Rebeiz, UCSD

All-CMOS 140GHz MIMO hub modules & demo



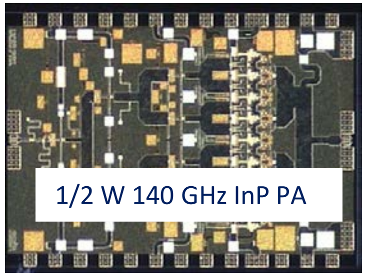
Niknejad, UCB

ComSenTer 140 GHz demo



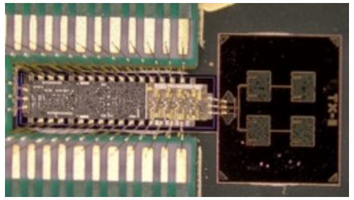
140 GHz InP/CMOS hub arrays

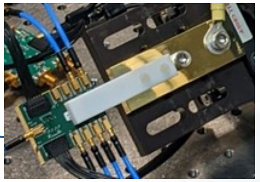
Rodwell, UCSB

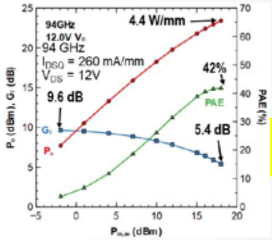


1/2 W 140 GHz InP PA

200 GHz MIMO backhaul modules



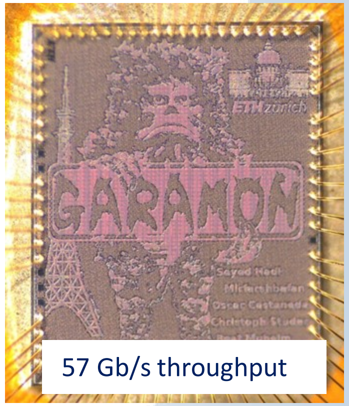




Record power, efficiency GaN

Mishra, UCSB

MIMO Hub Digital Beamformer ICs



57 Gb/s throughput

Molnar, Cornell
Madhow, UCSB
Studer, Cornell (ETHz)