JUMP Needs Document

Point of Contact: JUMP-Solicitation@src.org

6 JUMP THEMES TO CREATE 6 JUMP RESEARCH CENTERS

"VERTICAL," APPLICATION-FOCUSED CENTERS

1. RF to THz Sensor and Communication Systems (V)
2. Distributed Computing and Networking (V)
3. Cognitive Computing (V)
4. Intelligent Memory and Storage (V)

"HORIZONTAL," DISCIPLINARY-FOCUSED CENTERS

5. Advanced Architectures and Algorithms (H)
6. Advanced Devices, Packaging, and Materials (H)

"VERTICAL," APPLICATION-FOCUSED CENTERS

1 - RF to THz Sensor and Communication Systems (V)

This theme covers two general, synergistic application areas - RF Sensors and RF Communications Systems - operating at microwave, millimeter wave or THz frequencies in support of consumer, military, industrial, scientific and medical applications. System examples may include radar, communication, reconnaissance and/or mmwave/THz imaging.

As an example, it is envisioned that future RF sensor systems will require novel, energy-efficient devices, circuits, algorithms, and architectures for adaptively sensing the environment, extracting/manipulating/processing information, and autonomously reacting/responding to the information. Another example is cognitive communication systems - systems which will operate in complicated radio environments with interference, jamming and rapidly changing network topology, will obtain (sense) information about their environment (aware of their environment and the available resources) and will dynamically adjust their operation (e.g., efficient spectrum use, interference mitigation, spectrum prioritization) to provide required services to end users.

Elements of these future systems may include but are not limited to:

- Agility: reconfigurable, adaptive, multi-function, multi-mode, self-calibrating sensors with increased degrees of freedom for efficient use of EM spectrum (including: spectrum agility, instantaneous bandwidth/ waveform agility, (very) wide bandwidth, high dynamic range).
- Compact, inexpensive sensors that can sense / measure multiple variables
- Coordinated use of multiple sensor modalities
- Collaboration of sensors and compute through flexible and resilient architectures and interface protocols
- Resilience, trust, privacy
- System level management of spectrum utilization to improve spectrum occupancy and allow "smart" frequency allocation (including non-contiguous band aggregation and use of new waveforms) and/or sharing to overcome interference (including self-induced) and jamming

problems; (e.g., scan and detect transmission across very wide range of frequencies to avoid interference as well as to make efficient use of spectrum resources).

- Scalable, adaptive, behavioral learning and cognitive processing (information extraction) architectures for real-time and reliable sensing, operation and communication (e.g., feature-driven "intelligent" sensing that achieves the ultimate energy efficiency by only detecting salient features/patterns/thresholds).
- Communication and reconnaissance approaches to mitigate threats and enable rapid transmit/receive critical data and constant streaming
- Autonomous operation and decision making (e.g., embedded real-time learning, ability to recognize threat scenarios, ability to do local-processing before transmitting the data/information). Autonomous operation also may require integrated power generation/harvesting, distribution, control/regulation and modulation.
- Super-linear communication links (enabling high modulation formats) and integrated communications components for IoT and distributed sensor systems that enable ultra-low power, high data rate, long-range sensor communications with high linearity in up/down conversion
- In-sensor memory and compute
- Large format (e.g., 3m x 3m) flexible smart sensors/sensor arrays

To address these applications, this vertically integrated theme requires breakthrough research in materials, devices, components, circuits, integration and packaging, connectivity, architectures (e.g., subsystems/arrays), and algorithms principally aimed at efficiently generating, modulating, manipulating, processing (mainly in or very closely coupled to RF/mm wave /THz domain), communicating (transmitting) and sensing/detecting radiated signals.

<u>In addition to the information outlined above, possible research tasks of interest include but are not limited to:</u>

- Novel materials (semiconductors, packages, substrates, magnetics, metamaterials, etc.), devices (diodes, transistors, switches, varactors, etc.), components (High Q inductors, filters, low loss transmission lines, couplers, duplexers, antennas/radiating elements, etc.) and associated fabrication processes and manufacturing methodologies for realization of RF to THz circuits and systems.
- Highly linear, high efficiency (low power dissipation), (ultra) wide-bandwidth, ultra low noise analog/RF devices and circuits (e.g., PAs, LNAs, switches, mixers, IF amplifiers) with practical operating characteristics (e.g., reliable operation at room temperature or higher).
- Breakthroughs in mixed signal and analog circuits (e.g., high dynamic range, low power dissipation ADCs, DACs, etc.) including new approaches (circuits and architectures) for analog signal/information processing.
- New technologies (materials (metals, dielectrics), components (sources, detectors, propagation media, etc.,) and fabrication processes) and architectures to realize (ultra) low loss, wide bandwidth interconnects/connectivity and communication links (e.g., THz wired and wireless interconnects and LOS/NLOS transmission, optical interconnects, RF photonics, free space, backhaul transmission, Innovations in air-interface for zero-overhead and scalable transmission etc.) that dramatically reduce energy consumption in communication or distribution of data/information.

- Agile (adaptive / dynamically reconfigurable) RF devices, circuits and architectures including self-optimizing and resilient networks and dynamically reconfigurable interconnect fabrics
- Materials/devices/components/packaging for extreme operating conditions (e.g., high temperature) and harsh environments.
- Materials, devices, circuits, subsystems and approaches for package level or integrated (e.g., on-chip) energy generation (harvesting), storage, distribution and management/regulation.
- Material, devices, circuits, packaging, assembly, and systems for spectroscopy and mWave and (sub)THz imaging for DoD and commercial applications (e.g., for vital signs, security, non-invasive bio-signal and characteristic, etc.
- Novel packaging and integration approaches including: 3D and heterogeneous integration of diverse materials (e.g., Si, III-Vs, II-VIs, 2D material, magnetics, ferromagnetic, ferro-electric, multiferroics, etc.), devices and functions to improve performance, reduce cost, create new capabilities and/or increase functional density (e.g., fully integrated logic, RF and mixed signal functions)
- Novel circuits/architecture/algorithms for precision timing for sensor/communication system operation in PNT (Positioning, Navigation, and Timing) denied environment
- Novel, energy efficient transceiver modules, subsystems, arrays and architectures including: high power (high power density) mw to THz transmitters, receivers and communication links (e.g., massive yet scalable MIMO and beam formers).
- Development of rapid, accurate, multi-physics, multi-dimensional, multi-domain (co-)design (RF/analog/EM, thermal, mechanical) and simulation tools and methodologies for complex (e.g., 3D, heterogeneous) multi-function RF systems

Innovations must demonstrate orders of magnitude enhancement in at least one aspect of performance or appropriate proposer defined figure of merit without compromising the other aspects. **In addition, proposers are expected to define a grand challenge in the RF to THz Sensor and Communication Systems space that their Center will address.**

<u>This theme overlaps with / will leverage research from the other 5 themes:</u>

Distributed Computing and Networking

- All aspects of this theme as they apply to Distributed Sensor and Networked Communication Systems

Cognitive Computing

- All aspects of this theme (including: Analog computing (information extraction), Autonomous operation and decision making; real-time behavioral learning and cognitive processing (information extraction) and architectures) as they apply to sensor and communication systems

Intelligent Memory and Storage

- New memory architectures for In-sensor memory and compute (for energy efficient sensors systems)

Advanced Architecture and Algorithms

- Compute in/near data (sensor)
- Energy Efficient Circuits/architectures, real time energy optimization
- Advanced Power management

Advanced Devices, Packaging, and Materials

- Materials, devices, components, packaging for operation at high temperatures and in harsh environments
- Embedded Memory
- Advanced materials and devices for cognitive computing systems
- Monolithic 3D and heterogeneous integration
- Energy harvesting and energy storage devices
- Thermal management

While Centers are not limited to students from these disciplines, students obtaining degrees from the RF to THz Sensor and Communication Systems Center in the following areas of study are of particular interest to our sponsors in the years ahead:

- MS / PhD
- Applied Physics, Chemistry or Chem Eng. (growth, synthesis, fabrication processes), Electrical / Electronics Engineering or Electrical and Computer Engineering (RF devices, circuits, systems, algorithms), Material Science or Engineering (synthesis, growth), Mech. Eng. (thermal, packaging), Physics (modeling)

2 - Distributed Computing and Networking (V)

The use of computing to support enterprises and communities in social interaction, commerce, defense, and governance requires large-scale distributed computing systems. These systems support large numbers of participants and applications, as well as many kinds of networked resources – compute, memory, storage, sensors, actuators, etc. Importantly, new application requirements coupled with physics-based implementation constraints on latency and energy call for novel architectural solutions to computing-at-scale, requiring innovations in interconnect and networking at all levels, from on-chip to between datacenters. There are significant imbalances in the cost, and constraints on performance growth and energy consumption, of compute, storage, and networking that will require novel distributed-system advances, well beyond today's hardware and distributed architectures. The purpose of this theme is to explore the challenges of extremely large-scale distributed architectures. Novel, multi-tier, wired and wirelessly-connected heterogeneous systems are expected; tiers may be sensor/actuator, aggregation, cloud/datacenter, or combinations thereof. All tiers are expected to be highly scalable, and heterogeneity is expected both within and across the tiers. This theme will primarily focus on digital computing. The workload- and system-management aspects of these systems is highly relevant, including privacy-protocol management. Proposals should address the key new challenges in resource and applications management that must be effectively overcome in order for these novel, new systems to support a wide range of existing and emerging workloads.

Dramatic advances over today's systems (cloud, mobile, etc.) and capabilities are required. **Proposers are expected to define and tackle a grand challenge in the Distributed Computing and Networking space; the grand challenge should focus attention on research issues that would benefit a broad range of civilian and defense applications (e.g. society-scale digital currencies; battlefield command-and-control in denied environments; smart grid optimization; disaster management in digital cities).**

<u>In addition to the broad theme described above, possible research tasks of interest includes but are not limited to:</u>

- Demonstrate breakthrough levels of scalability and efficiency for large-scale systems of multiple, heterogeneous tiers (e.g., sensor, aggregation, datacenter/cloud)
- Achieve major improvements in system performance, energy efficiency, and robustness across compute, network, and storage
- Resilient distributed computing fundamentals
- Architectures, protocols, algorithms and systems to support >10X improvement of power efficiency and latency for wired and wireless communication
- Explore datacenter-level interconnect and networking innovations that dramatically improve scalability, and reduce latency and energy consumption.
- Self-optimizing and resilient networks, reconfigurable interconnect fabrics, high-speed, secure data links
- Deliver and demonstrate innovations for management of privacy and authentication protocols across the system
- Deliver and demonstrate innovations for configuration, workload, and data management for large-scale, multitenant systems
- System instrumentation and analytics for prognosis, diagnosis, reconfiguration, optimization, and repair
- Programming paradigms and languages for distributed and networked systems
- Software-defined infrastructure and resource virtualization
- Distributed decision and optimization support
- Novel computing architectures to reduce the energy and time used to process and transport data, locally and remotely for hyperspectral sensing, data fusion, decision making, and safe effector actuation in a distributed computing environment
- Provide cooperative and coordinated distributed-system concepts that are scalable and function in communications-challenged environments (where both wired and wireless environments are not guaranteed to be available, reliable, or safe); address approaches to allow for proper operation in isolation environments, and that can intelligently synchronize when communications are restored, including only partial restoration
- Development of new distributed computing systems for new applications besides IoT and big data
- Tasks should include both non-traditional and traditional architectures to push the boundaries for software research and application.

<u>This theme will benefit from research of all remaining themes</u>

While Centers are not limited to students from these disciplines, students obtaining degrees from the Distributed Computing and Networking Center in the following areas of study are of particular interest to our sponsors in the years ahead:

- MS / PhD

- Applied Math, Comp Sci. or Comp Eng., Electrical / Electronics Engineering or Electrical and Computer Engineering, Systems Engineering

3 - Cognitive Computing (V)

The Cognitive Computing theme aims to create cognitive computing systems that can learn at scale, perform reasoning and decision making with purpose, and interact with humans naturally and in real-

time. Realizing these novel systems may heavily leverage non-traditional computing methods, such as analog computing, stochastic computing, Shannon inspired computing, approximate computing, and bio/brain-inspired models including neuromorphic computing for a broad application space.

This theme seeks to explore multiple approaches for building machine intelligent systems with both cognitive and autonomous characteristics. Such systems can be solely non-traditional, solely von-Neumann or a combination of both elements. A key goal is creating systems that, without explicit objectives, operate in the natural world on their own by forming and extending models of the world they perceive around them, and by interacting with local human decision makers and with global distributed intelligent networks in performing actions to achieve useful yet complex goals. When exploiting non-traditional, non-von Neumann paradigms for cognitive computing, fundamental improvements must be demonstrated in performance, capabilities, and energy efficiency through breakthroughs in information processing, programming paradigms, algorithms, architectures, human machine interfaces, circuits, and device technologies. Benchmarking should be done to prove that the potential of the novel system is more than 100x better than traditional computing for the same task for at least one of the key metrics (energy, speed, area of the chip etc.). This research must span the gamut of theory, validation and demonstration of a complete scalable compute system, which leads to a cross-disciplinary effort embracing the programming structure, algorithms, architectures, circuits, devices, and materials.

Grand Challenge: Creation of a stretch objective or grand challenge can be a useful tool in articulating and pushing the boundaries in application of non-traditional computing and cognitive & autonomous systems. **Proposers are expected to define a grand challenge in the Cognitive Computing space that their Center will address.**

<u>In addition to the broad theme described above, possible research tasks of interest includes but are not limited to:</u>

1) <u>Advancements in Cognitive Capabilities</u>

- **Knowledge generation, reasoning, prediction, planning, decision making:**
  - Create insights, actionable intelligence, knowledge, wisdom and machine assisted decision from raw sensor data by next generation cognitive nanoelectronics computing systems and architectures
  - Anomaly detection for prediction, decision making, planning
  - Architectures to accelerate tactical decision-making. Delegating decisions to machines in the situations require faster-than-human reaction times and nontraditional chain-of-command responses.
    - Computation on encrypted data to support safety of scalable decision making systems
- **Learning and perceiving:**
  - Pursue breakthrough advances in sensing technologies for cognitive systems
  - Develop universal algorithms using e.g. context-aware learning applicable to a wide range of cognitive workloads. Develop hardware (architectures/circuits/technologies) and software to realize the said universal learning system.
    - Develop perception algorithms (which enable understanding of the environment), authentication enablers, and learning enablers

- Explore the fundamental limits and capabilities of existing machine learning algorithms, for instance, deep neural network, for cognitive workloads
  - Develop new architectures and algorithms that allow hitting the required test accuracy with a significantly reduced training set.
    - Big data and distributed analytics for supporting cognitive and autonomous computing optimizations across local and global machine intelligence systems
    - Develop algorithmic and hardware speedup for sparse coding
    - Develop inference systems orders of magnitude more efficient than today's inference systems, including real-time computing for classification and clustering.
    - Develop energy-efficient, low-cost techniques for cognitive workloads

- **Human-Machine Interface:**
  - Develop seamless Human-Machine Interface for autonomous systems. Innovative cooperative operations and collaborations between manned and unmanned platforms.
  - Develop seamless Human-Machine Interface/collaboration for sensing and feedback (haptic, audio, vision, etc.) with high accuracy both in HW and SW levels. The scope could be with bio sensors or stand-alone / away-from-human sensors.
- **Network of cognitive sub-systems:**
  - Develop architectures for resilient self-optimizing and self-healing networks, memories and compute elements for connecting billions of devices in intelligent systems
  - Develop society-scale applications and intelligent data collections systems, which can interact with local cognitive and autonomous machine intelligence systems to optimize decision support.
- **Other Important Aspects:**
  - Develop programming paradigms and languages that enable the cognitive capabilities
  - Investigate the nature of malicious and/or destructive AI systems

2) <u>Advancements in approximate/stochastic and Shannon inspired computation</u>

Modern computing has operated on a nearly error free device layer since the advent of Von Neumann computing. Computational architectures relying on approximate computing (AC) or/and on stochastic computing (SC) can allow a significant scalability in error rates. Application of approximate computing (AC) or/and on stochastic computing (SC), Shannon inspired information framework can provide significant benefits in energy, delay, error rate or enable scalable architectures on erroneous hardware layers. Subtasks can include:

- Theoretical considerations:
  - Computational Complexity (time and memory) of stochastic and random computing
  - Complexity of classifiers, complexity of associate processors
  - Scalability of Stochastic computation
- Computing at variable accuracy approximate computing, HW and SW
- Computing with device and hardware dynamic variations

- Statistical error correction
- Algorithmic noise tolerance
- Shannon derived computation
- Computation in projected spaces (increased and reduced dimensionality)
- Application of Restricted isometry and dimensionality reduction for computing and sensing
- Compressive sensing paradigms for computation and classification
- Computing for classification and clustering

3) <u>Advancements in neuromorphic computation</u>

Advancements in neuromorphic hardware and software computing, which could include mimicking the actions of a neuron through new memory elements and circuits, new topologies like spiking neural-nets and inference augmented neural-nets which significantly advance the efficiency, accuracy, scalability, and latency. Subtasks can include:

- Develop computational theory for various neuromorphic computing methods
    - Computational Complexity (time and memory)
    - Self-Learning
    - Scalability
- Develop energy-efficient low-cost techniques, including analog, for neural networks and other brain-inspired computing applications, in addition to cognitive workloads.
- Develop new architectures and algorithms for neuromorphic computing systems that would enable improved test accuracy using significantly reduced training sets for targeted applications (cognitive, non-cognitive, logic based etc.).
- Develop reconfigurable networks suitable for a class of NN applications (ANN, RNN, CNN, etc.)
- Develop materials and devices designed and optimized for neural network applications, such as artificial neurons and synaptic memory devices
- Develop compressive paradigms for classifiers and computing for classification and clustering

4) <u>Integration of non-traditional methods</u>

Demonstrate practical mechanisms for integration and verification of these new methods with Von Neumann systems to support realistic, complex workflows. Thus the work proposed under this theme are strongly encouraged to include:

- The theoretical basis and support for non-traditional methods
- The breadth of applicability of the proposed non-traditional methods
- Their integration into relevant workloads and systems.
- The resiliency and reliability of the proposed non-traditional methods

<u>This theme overlaps with / will leverage research from:</u>

- RF to THz Sensor and Communication Systems
- Distributed Computing and Networking
- Intelligent Memory and Storage
- Advanced Architecture and Algorithms

While Centers are not limited to students from these disciplines, students obtaining degrees from the Cognitive Computing Center in the following areas of study are of particular interest to our sponsors in the years ahead:

- MS / PhD
- Applied Math, Applied Physics, Bioengineering, Computational Neuroscience, Comp Sci. & Comp. Eng., Electrical / Electronics Engineering or Electrical and Computer Engineering, Material Science or Engineering, Mathematics, Neuroscience, Physics, Systems Engineering

4 - Intelligent Memory and Storage (V)

Advances in information technology have pushed data generation rates and quantities to a point where memory and storage are the focal point of optimization of computer systems. Transfer energy, latency and bandwidth are critical to performance and energy efficiency of these systems. The solutions to many modern computing problems involve many-many relationships that can benefit from high cross-sectional bandwidth of the distributed computing platform. As an example, large scale graph analytics involve high cross-data-set evaluation of numerous neighbor relationships ultimately demanding high the highest possible cross-sectional bandwidth of the system.

This research vector seeks a holistic, vertically-integrated, approach to high-performance Intelligent Storage systems encompassing the operating system, programming models, memory management technologies, and a prototype system architecture. Significant advancements may be made by utilizing distributed compute elements and accelerators in close physical proximity to the location of data storage. An optimal hardware platform may contain a large fabric of interconnected memory and storage devices with large cross-sectional bandwidth, integrated logic and accelerators in conjunction with islands of conventional high performance computing elements. Gains of orders of magnitude in volumetric density, information processing density, power or performance should be pursued.

A primary focus area for this center will be in establishing an operating system framework allowing run-time optimization of the system based on system configuration preferences, programmer preferences, and the current state of the system. There will be many run-time optimization challenges in such a heterogeneous platform containing Near-Memory Von Neumann and non-Von Neumann elements as there are inherently numerous means to achieve the same end (e.g. compute slowly locally, transfer data and compute quickly remotely, compute very efficiently but approximately in an accelerator) and proper optimization may differ depending on the current state of the system. As an envisioned system example, execution time and data movement are automatically balanced in a near-optimal fashion on simple hardware through OS and hardware hooks without the requirement of, but with the allowance of, programmer intervention; where built-in controls allow a system administrator to tradeoff performance, power efficiency, response latency, etc. to operate within the constraints of the specific system installation. Scheduling decisions may involve proper use of local and global metrics of bandwidth, power, latency, and temperature, etc.

In order to reasonably establish run-time optimization algorithms, suitable metrics for performance of such a system must be established to measure information processing density. E.g. ("decision rate" or "correct decision rate") per ("Kg of silicon", "cm^3", or "Watt"). For a simple example, imagine a 1U enclosure containing > 100K memory dice interconnected in a high dimensional fabric, each capable of accessing and operating on data locally and each consuming 1 watt of power in its most active state. Algorithms for power throttling based on temperature and power consumption would be required for this system and these algorithms likely involve local and global aspects of the control system. System theoretic bounds on information processing density should be established to demonstrate the research areas of highest potential. It is foreseen that replication will be a required aspect of this type of system

for performance and for resilience and that this system should be designed to co-optimize this replication with retention and endurance management and authentication strategies (which users are allowed to run which programs on which data).

This holistic approach must keep in focus the relevant, emerging memory technologies and their respective, novel system architectures and hierarchies (including subsystems and their caches), as well as the advanced materials and processes required to manufacture them. An eye towards backward compatibility where possible is also of interest to facilitate migration of applications to this new framework.

Proposers are expected to define a grand challenge in the Intelligent Memory and Storage space that their Center will address.

<u>In addition to the broad theme described above, possible research tasks of interest include, but are not limited to:</u>

Compute in memory systems

- Memory at a higher level
- Compute in Memory – Von Neumann (e.g. PIM)
- Compute in Memory – Non Von Neumann (e.g. Brain-Inspired)
- Heterogeneous Systems
    - Systems on chip with novel memory hierarchy
    - Nonvolatile computing (HW and SW)
    - In-memory computing
    - Reconfigurable computing
- Inference kernels
    - Learning algorithms at a low level
    - Sensitivity analysis on how process variation might affect accelerators

New Architecture and Programming paradigms, Self-optimizing Systems Allowing for Appropriate Programmer Control

- New programming and architecture paradigms to exploit non-volatile and low latency emerging memory systems
- Algorithms and memory architectures that facilitate decision making based on approximate, or error-prone, stored data
- Address the myriad of architecture, software, and applications implications of new memory devices and alternative memory hierarchies over the cache hierarchy, system memory, and storage tiers
- Scalable, reconfigurable, energy efficient information extraction (circuits/ systems/ algorithms/ architecture). Includes Distributed, cognitive, non-traditional computing and analog
- 10X more power efficient computing platform scalable from high performance application processors to less-demanding processors for IoT/sensors/etc. with cost awareness. The technology can span across material, devices, packaging, circuits/systems techniques, computer architecture including but not limited to heterogeneous computing, memory technology (including NVM) and high-speed interface (on-chip and off-chip), etc.

- Heterogeneous Systems - Computing on non-traditional fabrics (e.g. cross points, FPGA as a compute unit. etc.)
- Co-designed HW and algorithms
    - Computational theory for Neuromorphic
    - Computational complexity (time and memory) of non von Neumann
    - Computational Complexity (time and memory) of neuromorphic
    - Computational Complexity (time and memory) of stochastic and random computing
    - Scalability of Neuromorphic and Stochastic computation

Small, Probably Low Cost, Compute+Memory+Sensor Node Capable of making Basic Decisions/observations and Reporting to a Larger System

- Small scale localized memory and storage systems (e.g. IOT)
- In-sensor memory and compute (for energy efficient sensors systems)

Intersection of Big-data Analytics and Big-data for Sensor Data/Lazy Computation of Sensor Data Local to Data Storage

- Architectures to optimize data movement and memory-compute partitioning for energy and performance
- Big data, large-scale, distributed memory and storage systems (e.g. Data Center)
- Heterogeneous big data and distributed analytics
- Intersection of Big-Data Analytics and Big-Data for sensor data / lazy computation of sensor data local to data storage

New Technologies, Materials and Processes

- Identification and development of new memory-device technologies
    - Highly-scalable, non-volatile memory exhibiting high durability with low variation/stochasticity
        - Low latency
        - Energy-efficient
    - Non-volatile memories suitable for embedded implementations including low-power computing and SOC
    - Ultra-low power, moderate-density, integrated memory for sensors
    - Access devices for cross-point and 3D geometries
- Advanced charge-based and steep-slope devices.
    - Phase change materials
    - Metal insulator transition, mott insulators
    - 2D materials (e.g. TFET)
    - New ferro-electrics (highlighting reliability)
- Materials and interfaces for spin-based logic and memory
    - High polarization magnets/interfaces for spin injection
    - High polarization and low coercivity ferroelectrics
    - Magnetoelectrics, multiferroics, and uniferroics with ME
    - Spin-orbit coupled materials
    - Topological materials

    - Physic of critical oxide technologies (e.g. MgO, MgAlO, BiFeO) including fatigue and failure
- Novel processing for advanced devices and scaled geometries (incl. ALD and Atomic Layer Etching)
- Novel materials and means for reducing processing temperatures, especially B/E processes.
- Advanced materials and processes for self-assembly.
- Novel organic electronic materials (semiconductor & conductor)
- Functional 3D integration (monolithic 3D ICs)
- Holistic approach to harnessing advanced memory/storage devices and complimentary CMOS (and beyond) with supporting interconnect and packaging technologies targeted and optimized for non-traditional and cognitive computing systems.
- Demonstration of novel, vertical integration and leveraging of advanced CMOS and beyond, memory/storage, and interconnect device technologies, to achieve dramatically improved energy/performance versus state-of-the-art CMOS. Demonstrate these techniques on realistic computational problems by prototyping functional units/cores/accelerators.

Authentication & Resilience

- Memory management and control technologies including error management and authentication
- Resilient/self-healing compute and memories
- Disruptive memory architectures and technologies to optimize intelligent and flexible storage design, management, processing and system scalability to mitigate data explosion challenges with radiation hardening capability.

IO & Networking

- In-Memory networking structures
- Virtual=physical large memories (no translation Software)

<u>This theme overlaps with / will leverage research from:</u>

- Distributed Computing and Networking
- Cognitive Computing
- Advanced Architectures and Algorithms
- Advanced Devices, Packaging, and Materials

While Centers are not limited to students from these disciplines, students obtaining degrees from the Intelligent Memory and Storage Center in the following areas of study are of particular interest to our sponsors in the years ahead:

- MS / PhD
- Applied Math, Applied Physics, Bioengineering, Chemistry or Chem Eng. (growth, synthesis, fabrication processes), Comp Sci. & Comp. Eng., Electrical / Electronics Engineering or Electrical and Computer Engineering, Material Science or Engineering, Physics, Systems Engineering

"HORIZONTAL," DISCIPLINARY-FOCUSED CENTERS

5 - Advanced Architectures and Algorithms (H)

Today's system architectures, including distributed clusters, symmetric multiprocessors (SMPs), and communications systems, are generally comprised of homogeneous hardware components that are difficult to modify once deployed. Heterogeneous architectures and elements, such as accelerators, will increasingly be needed to enable scaling of performance, energy efficiency, and cost. This theme must lay the foundations for new paradigms in scalable, heterogeneous architectures, co-designed with algorithmic implications and vice versa. A major goal of this theme is to address the design and integration challenges of a broad variety of accelerators, both on-chip and off-chip, along with the algorithmic and system software innovations needed to readily incorporate them into both existing and future systems (e.g, information processing, communications, sensing/imaging, etc.).

As an example, any new acceleration approach must demonstrate significantly higher system value, in terms of efficiency and/or performance, over known approaches such as CPU+GPU. Broadly-applicable architectures for compute in and near data (such as in-sensor or in-memory computing) are in scope. These advanced architectures should also provide mechanisms for workload-driven runtime composition of hardware resources, as well as auto-configuration and auto-tuning of system parameters. Reliability, resiliency, and energy-efficiency of the new architectures must be addressed. Architectures and algorithms that are significantly lower in power and capable of scaling to 10^12 devices/chip, 10^12 nodes/network using CMOS and beyond-CMOS technologies are in the scope of this theme. Employing novel co-design to bridge the gap between architectures and algorithms for optimization, combinatorics, computational geometry, distributed systems, learning theory, online algorithms and cryptography are within scope. Along with implementation, a strong component of theoretical computer science and modeling is expected. The innovations that result from this foundational theme are expected to impact a broad variety of workloads and system applications.

Proposers are expected to define a set of key metrics that their center will use to benchmark and drive efforts in the Advanced Architectures and Algorithms space.

<u>In addition to the broad theme described above, possible research tasks of interest includes but are not limited to:</u>

- Novel Architectures:
    - Explore architectural mechanisms, microarchitectures, and alternative protocols to minimize communication and communication cost at all levels
    - In memory computing
    - Reconfigurable computing
- Hardware-software co-design:
    - Design and integration of accelerators, from circuit level up to system & software level for heterogeneous systems, based on both existing and emergent devices
    - Demonstrate dynamic control of operating points in order to manage the reliability of the computation in addition to co-optimization of power and performance
    - Develop non-conventional technologies for accelerators
- System Integration:
    - Develop software mechanisms to enable transparent usage of accelerators

    - Architectural and software innovations for enhanced system flexibility and composability, including for the integration of novel accelerators, allowing workload-driven composition of systems at runtime
    - Develop mechanisms for auto-configurations and auto-tuning of heterogeneous systems
- Software:
    - Software-defined infrastructure and resource virtualization
    - Programming paradigms and languages for emerging technologies and architectures
    - Novel approaches to software engineering and developer operations, including computer-supported programming and program synthesis.
- Energy efficiency:
    - Energy efficient circuits/architecture, algorithms and software: hardware/software/algorithm co-design methodologies and tools and Includes underlying physics, physics of failure. Required for energy efficient sensor systems, information extraction and autonomous systems
    - Real-time energy optimization. System-level design or circuit-based techniques that provide granular power gating. This may include feed-back or feed-forward with high- or low-resolution depending on the scenario or system.
    - Advanced power management methods
- Scaling:
    - Architectures that explicitly scale to 1 tera devices per processor (not including memory)
    - Fundamental understanding of the new architectures for scaling
    - Computational theory of Beyond CMOS
- Memory related:
    - Systems on chip with novel memory hierarchy
    - Nonvolatile computing (HW and SW)
    - Computing on non-traditional fabrics e.g. cross points
- Privacy and encryption:
    - Accelerated homomorphic encryption
    - Authentication protocols and reliability supported by both HW and SW systems.
    - Computational theory for encryption and privacy
- Modeling of the new architectures:
    - Verification and Validation
    - Modeling – simulating the performance of the new heterogeneous architectures, benchmarking them against von Neumann machines
    - Modeling of performance: needs demonstration for a broad enough class of workloads and applications.
    - Modeling: validate modeling results against hardware performance.
        - Prove the relevance of models to real-world applications (The deliverable of the theme is system hardware, not just a simulator or modeling results.)

<u>This theme will benefit from research of all remaining themes</u>

While Centers are not limited to students from these disciplines, students obtaining degrees from the Advanced Architectures and Algorithms Center in the following areas of study are of particular interest to our sponsors in the years ahead:

- MS / PhD
- Applied Math, Comp Sci. & Comp. Eng., Electrical / Electronics Engineering or Electrical and Computer Engineering, Mathematics, Systems Engineering

6 - Advanced Devices, Packaging, and Materials (H)

Advanced active and passive devices, interconnect, and packaging concepts, based on physics of new materials and unconventional syntheses are needed to enable the next breakthrough paradigms in computation (including analog) and information sensing, processing, and storage that will provide for further scaling and energy efficiencies. These devices will need showcase materials with new functionalities and properties that can augment and/or surpass conventional semiconductor technologies, and potentially with 3D options. Material development, device demonstration with viable process integration are within the scope. Efforts can provide guidance to future manufacturing. Along with experimental demonstrations, ab-initio material and process modeling is expected.

Proposers are expected to define a set of key metrics that their Center will use to benchmark and drive efforts in the Advanced Devices, Packaging, and Materials space.

In addition to the broad theme described above, possible research tasks of interest includes but are not limited to:

- Memory and embedded memory: new memory physics, scalable embedded memory with high density and low power consumption
- Cognitive computing: advanced materials and devices for cognitive computing systems (like synaptic memory devices and artificial neurons), that can allow the control of symmetric, multi-level, gradual resistance variations with programming pulses.
- Non-volatile logic: novel devices that enable non-volatile logic, hysteretic devices that improve SRAM noise margins
- Memory access devices: back-end compatible access devices with very low leakage and moderate to high drive
- Monolithic 3D: integration of logic, memory, and analog devices in back-end, block partitioning for 3D
- Heterogeneous integration: low cost and low defect integration of materials with silicon
- Integration of cognitive computational devices with CMOS: on-chip integration schemes, advanced packaging and 3D integration schemes for neuromorphic computing devices with silicon devices in cognitive computing systems.
- Energy harvesting and energy storage devices: novel materials for high efficiency energy harvesting, supercapacitors, integrated batteries, power delivery
- Thermal management: novel materials and processing for heat removal, thermoelectric cooling, novel heat sinks for packaging
- Packaging for non-traditional devices: advanced packaging with system modularity, small form factor packaging for IoT devices

- Steep devices: super steep devices for low V_{dd}, novel ferroelectrics enabling NCFET, novel MIT materials
- Transduction mechanisms: novel materials for transduction between electric, magnetic, or strain states, novel magnetoelectric materials
- Spin devices: novel spin injection materials, highly spin polarized magnets, novel spin filters
- Van der Waals devices: novel van der Waals materials for logic, memory and electro-optical devices, with large-scale, controlled growth while minimizing defectivity
- Parasitics reduction: low resistance lines for memories, novel materials for lowering parasitic resistance and capacitance for CMOS devices
- Device reliability: reliability of novel ferroelectrics, spin transfer torque devices
- Passive devices: materials and scaled passives (inductors, capacitors, transformers, antennae) with improved performance
- Harsh environment devices: power devices for 3D integration, materials and devices to enable high temperature, high power devices, electronics for extreme operating conditions
- Material, Devices for analog/RF functions and applications
- Frequency and energy sources: RF oscillators and sources, coupled oscillators for image recognition
- Flexible devices: materials and processes for flexible displays and wearable electronics (logic, connectivity, energy harvesting)
- Electro-optical devices: materials and devices for low cost, low power, high resolution displays, novel silicon and 2D material photonic devices
- Devices for cyberattack prevention, theft, and counterfeiting
- Develop advanced materials and devices for cognitive computing systems, which may include (but are not restricted to) synaptic memory devices and artificial neurons. As an example, in synaptic devices, materials and device mechanisms that allow the control of symmetric, multi-level, gradual resistance variations with programming pulses are sought after.
- Develop on-chip integration schemes for the said neuromorphic computing devices with silicon devices in cognitive computing systems.
- When on-chip integration is impractical, develop advanced packaging and 3D integration schemes for the new devices in cognitive computing systems.

<u>This theme will benefit from research of all remaining themes</u>

While Centers are not limited to students from these disciplines, students obtaining degrees from the Advanced Devices, Packaging, and Materials Center in the following areas of study are of particular interest to our sponsors in the years ahead:

- MS / PhD

- Applied Physics, Chemistry or Chem Eng. (growth, synthesis, fabrication processes), Electrical / Electronics Engineering or Electrical and Computer Engineering (devices, circuits, systems), Material Science or Engineering (synthesis, growth), Mech. Eng. (thermal, packaging), Physics (modeling)