Advanced Graphical Intelligence Logical Computing Environment (AGILE)

Dr. William Harrod | Program Manager | December 22, 2020

Proposers' Day

Office of the Director of National Intelligence

IARPA

BE THE FUTURE

Welcome to the AGILE Proposers' Day!

Thank you for your interest

Please send us questions and comments at any time during any presentation through the WebEx "Q&A" tool

Questions may relate to this presentation or the draft BAA Technical Volume posted on beta.sam.gov.

To assure a clear broadcast stream, audio and video are disabled for meeting participants

A recording of the entire Proposers' Day will be posted on the AGILE website

Disclaimers

This presentation is provided solely for information and planning purposes

The Proposers' Day does NOT constitute a formal solicitation for proposals or proposal abstracts

Nothing said at Proposers' Day changes the requirements set forth in a BAA

The BAA language supersedes anything presented or said by IARPA at the Proposers' Day

Goals

Familiarize participants with IARPA's interest in the AGILE program

Foster discussion of complementary capabilities among potential program participants, i.e., TEAMING

- Someone might have a missing piece of your puzzle
- Teaming Information Forms on beta.sam.gov for sharing contact info, areas of expertise, and expertise sought
- Teaming Forms will be posted on iarpa.gov

Please ask questions and provide feedback.
This is your chance to alter the course of events.

Q & A

There will be a 15-minute break and a 30-minute break during the agenda for preparing questions.

Responses to selected questions will be broadcast at **2:30 pm EDT**, so please don't log out or close your WebEx connection

Feedback (but not Qs) about the draft BAA may be submitted to the program email at dni-iarpa-baa-21-01@iarpa.gov

After this Proposers' Day, IARPA will consider all the feedback received in preparing the final BAA, which will be posted on beta.SAM.gov

Agenda

Time	Topic	Speaker
12:00 PM – 12:10 PM	Welcome, Logistics, Proposer's Day Goals	Dr. William Harrod Program Manager, IARPA
12:10 PM – 12:20 PM	IARPA Overview	Dr. Pedro Espina, Director, Office of Collections
12:20 PM – 1:20 PM	AGILE Program Overview	Dr. William Harrod
1:20 PM – 1:35 PM	Break to formulate Questions	
1:35 PM – 2:00 PM	Doing Business with IARPA	Chris Fox, IARPA Contracting Officer
2:00 PM – 2:30 PM	Break Questions after 2:15 PM will not be addressed	
2:30 PM – 3:30 PM	AGILE Questions & Answer	Dr. William Harrod

Note: All times EST (Washington, DC Time)

IARPA Overview

Dr. Pedro Espina, Director, Office of Collections
Intelligence Advanced Research Projects Activity

Office of the Director of National Intelligence

IARPA

BE THE FUTURE

Office of the Director of National Intelligence

Central Intelligence Agency

Defense Intelligence Agency

Department of State

National Security Agency

Department of Energy

National Geospatial-Intelligence Agency

Department of the Treasury

National Reconnaissance Office

Drug Enforcement Administration

Department of the Army

Federal Bureau of Investigation

Department of the Navy

Department of Homeland Security

Air Force

Coast Guard

Marine Corps

IARPA Mission

IARPA envisions and leads *high-risk, high-payoff research* that delivers innovative technology *for future overwhelming intelligence advantage*

- Our problems are **complex** and **multidisciplinary**
- We emphasize **technical excellence** and **technical truth**

IARPA Method

Bring the best minds to bear on our problems

- Full and open competition to the greatest possible extent
- World-class, rotational Program Managers

Define and execute research programs that:

- Have goals that are clear, measureable, ambitious and credible
- Employ independent and rigorous Test & Evaluation
- Involve IC partners from start to finish
- Run from three to five years
- Publish peer-reviewed results and data, to the greatest possible extent
- Transition new capabilities to intelligence community partners

IARPA Snapshot

IARPA's research portfolio is diverse, including math, physics, chemistry, biology, neuroscience, linguistics, political science, cognitive psychology, and more.

- **70% of completed research transitions** to U.S. Government partners
- **3,000+ journal articles** published
- IARPA funded researchers have been awarded the **Nobel Prize in Physics** for quantum computing research, a **MacArthur Fellowship**, a **Bell prize**
- IARPA is a member of the National Science and Technology Council (NSTC) and actively engages with the White House BRAIN Initiative, National Strategic Computing Initiative, and the NSTC Select Committee on Artificial Intelligence, the NSTC Subcommittee on Quantum Information Science (SCQIS), and NSTC Subcommittee on Economic and Security Implications of Quantum Science (ESIX)

How to Engage with IARPA

iarpa.gov | 301-243-1995

info@iarpa.gov

Reach out to our Program Managers.

Schedule a visit if you are in the DC area or invite us to visit you

Opportunities to Engage:

RFIs AND WORKSHOPS

Opportunities to learn what is coming, and to influence programs.

"SEEDLINGS"

Typically a 9-12 month study; you can submit your research proposal at any time. We strongly encourage informal discussion with a PM before proposal submission.

PRIZE CHALLENGES

No proposals required. Submit solutions to our problems – if your solutions are the best, you receive a cash prize and bragging rights.

RESEARCH PROGRAMS

Multi-year research funding opportunities on specific topics.

Agenda

Time	Topic	Speaker
12:00 PM – 12:10 PM	Welcome, Logistics, Proposer's Day Goals	Dr. William Harrod Program Manager, IARPA
12:10 PM – 12:20 PM	IARPA Overview	Dr. Pedro Espina, Director, Office of Collections
12:20 PM – 1:20 PM	AGILE Program Overview	Dr. William Harrod
1:20 PM – 1:35 PM	Break to formulate Questions	
1:35 PM – 2:00 PM	Doing Business with IARPA	Chris Fox, IARPA Contracting Officer
2:00 PM – 2:30 PM	Break Questions after 2:15 PM will not be addressed	
2:30 PM – 3:30 PM	AGILE Questions & Answer	Dr. William Harrod

Note: All times EST (Washington, DC Time)

Advanced Graphical Intelligence Logical Computing Environment (AGILE)

Dr. William Harrod | Program Manager | December 22, 2020

Proposers Day

Office of the Director of National Intelligence

IARPA

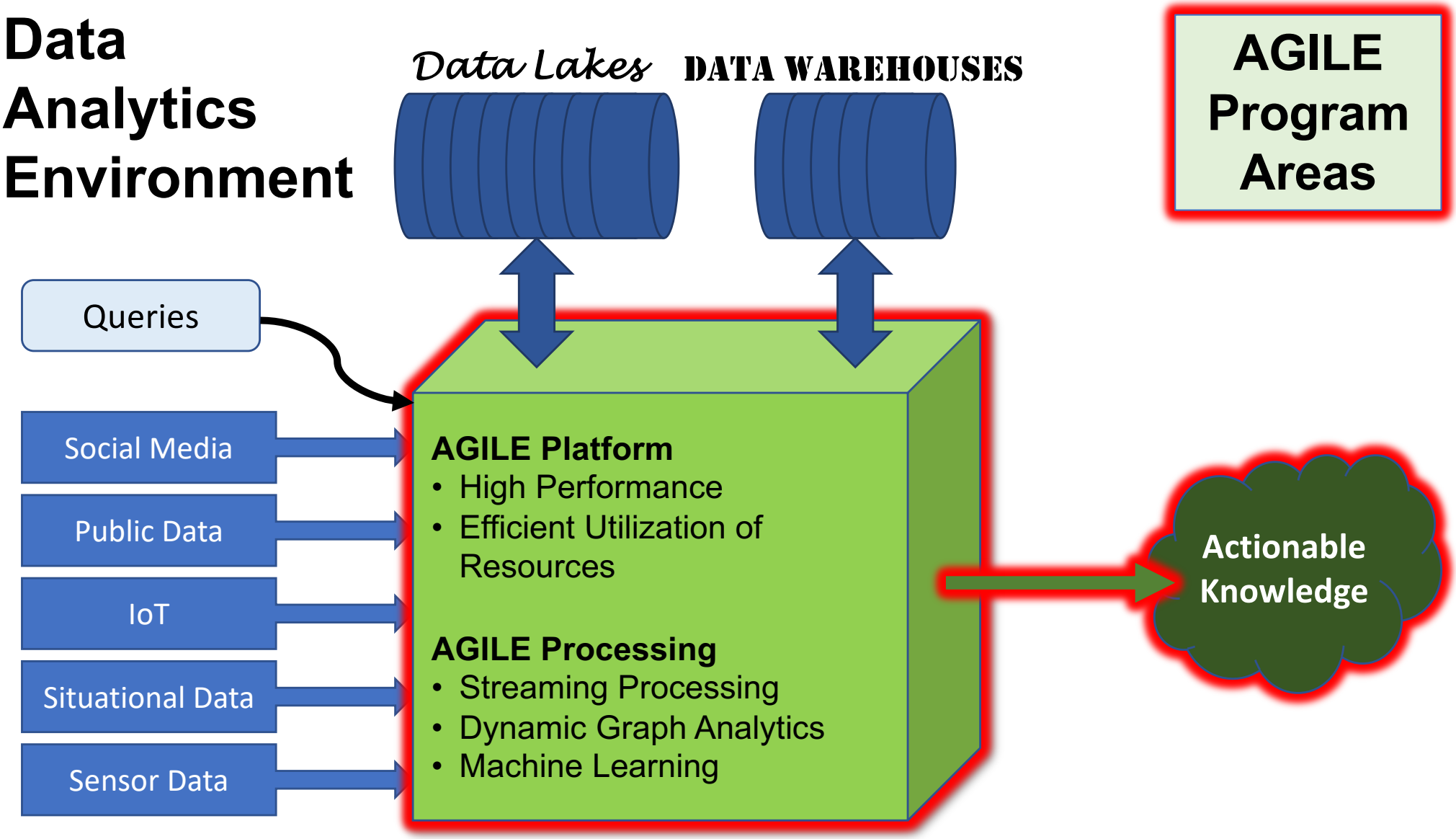
BE THE FUTURE

My agenda for today

Discuss the following topics:

1. What is AGILE and why are we doing this program?
2. What are the technical challenges?
3. What are possible research strategies?
4. What is the design process and expectations?
5. How is success measured?
6. What is the schedule?

Data Analytics Environment



What is AGILE?

- **A**dvanced **G**raphical **I**ntelligence **L**ogical Computing **E**nvironment (AGILE) program.
- IARPA is seeking revolutionary new computer architectures for Intelligence Community (IC) analysts to solve critical data-intensive problems.
- AGILE is envisioned to be a 3-year program, beginning approximately June 2021 and extending through June 2024.
- Performers will develop Register Transfer Level (RTL) designs for their proposed AGILE architecture, including runtime system
- Designs will be evaluated using an AGILE enhanced modeling and simulation toolset

Why AGILE?

- **AGILE problem** – emerging data problems that can't be transformed into actionable knowledge in a reasonable timeframe
- **AGILE data problems** – data is often represented by a graph. The resulting graphs are very sparse, random, and heterogeneous.
- **Actionable knowledge** – data analytics are used to transform the data into actionable knowledge. The analytics operation on graphs. The execution of these analytics have minimal data locality, poor data re-use, fine-grain data movement and data driven parallelism.
- **Fundamental problem** – today's computers solve yesterday's problems – they are not efficient or scalable for the emerging data problems
- ***Need new system-level intelligent mechanisms for moving, accessing, and storing large, random, time-varying data streams, structures, objects, and knowledge***

There are no magic bullets

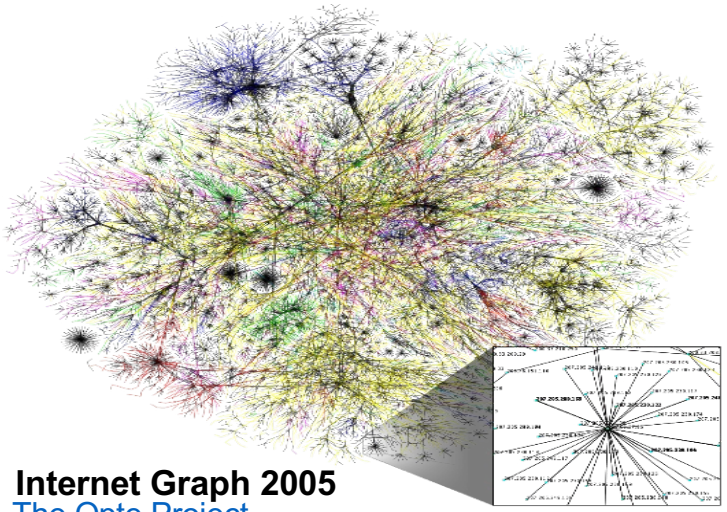
Data Analytics Problem

Relationships between entities are often represented by vertices in a graph and the edges represent the relationships between entities

The graphs are typically sparse (i.e., most entries are null) – there are many entities but very few relationships between the entities

Graphs	Vertices	Edges
Social network	1 Billion	100 Billion
Internet	50 Billion	1 Trillion
Brain	100 Billion	100 Trillion

Technical Report NSA-RD-2013-056002v1, U.S. National Security Agency



Internet Graph 2005
The Opte Project

Extracting Actionable Knowledge Methods

Graph Analytics	Machine Learning	Statistics Methods	Linear Algebra	Data Filtering

"The variety and volume of data collected (today) … far outpace the abilities of current systems to execute complex analytics … and extract meaningful insights." (Buono, D., *Computer*, August 2015)

AGILE Data Problems of Interest

Measurements
Sensors
Sound, Images, Video
Text

Floating Point
Integers
Strings
Bits

Graphs / Matrices
Documents
Tensors
Maps

- As problem sizes grow over time – data is increasingly sparse, random and diverse
- Increasing variety of data sources
- Vertex data & meta-data can vary in size from bytes to giga-bytes and larger

AGILE Challenge: Streaming Data Analytics

reporting the news → predicting the future

Big Data (today)

Large static data sets – can optimize access

Emerging ML capabilities

Data has been previously processed

Queries are based on existing data
– **updates occur on daily basis**

Systems optimized for dense algebra, not graph analytics computations

Streaming Data Analytics

Low data re-usage – random locations

Requires specialized analytics

Data quality and uncertainty are concerns

Stream of queries & updates to persistent data – **time is a factor**

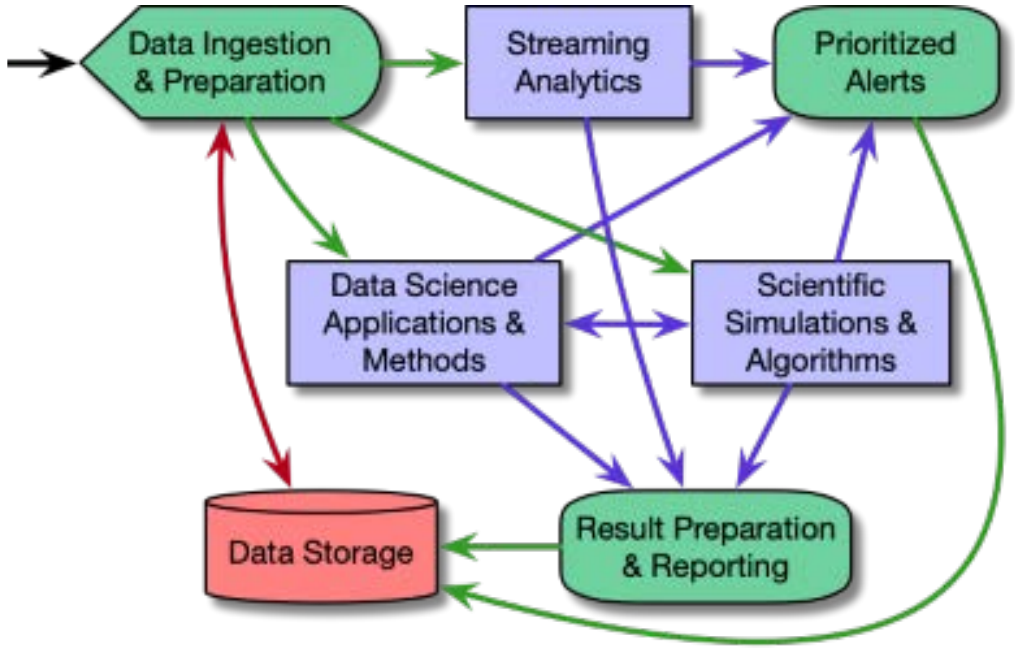
Systems designed for rapidly changing sparse data and processing flow

“Forensic Analysis”

“Predictive Analysis”

AGILE Computational Problems of Interest

- Results required in **near-real-time up to hours**
- Streaming data causes **unpredictable changes** to stored data
- **Extremely fine grain data movement and parallelism**: computations, data are distributed across computer



- Data computation tasks to be performed are typically **determined by the data and streaming queries**
- Tasks have **extremely poor data locality and data reuse**
- **Many graph analytics algorithms can be recast as sparse linear algebra operations**

Sparse Matrix X Vector Algorithm

SpMV *After* A Lot of "Conventional" Optimizing

SpMV = Sparse Matrix times a Vector

Sparse Matrix x Vector algorithm (SpMV)

The graph shows the execution speedup for two different sparse datasets

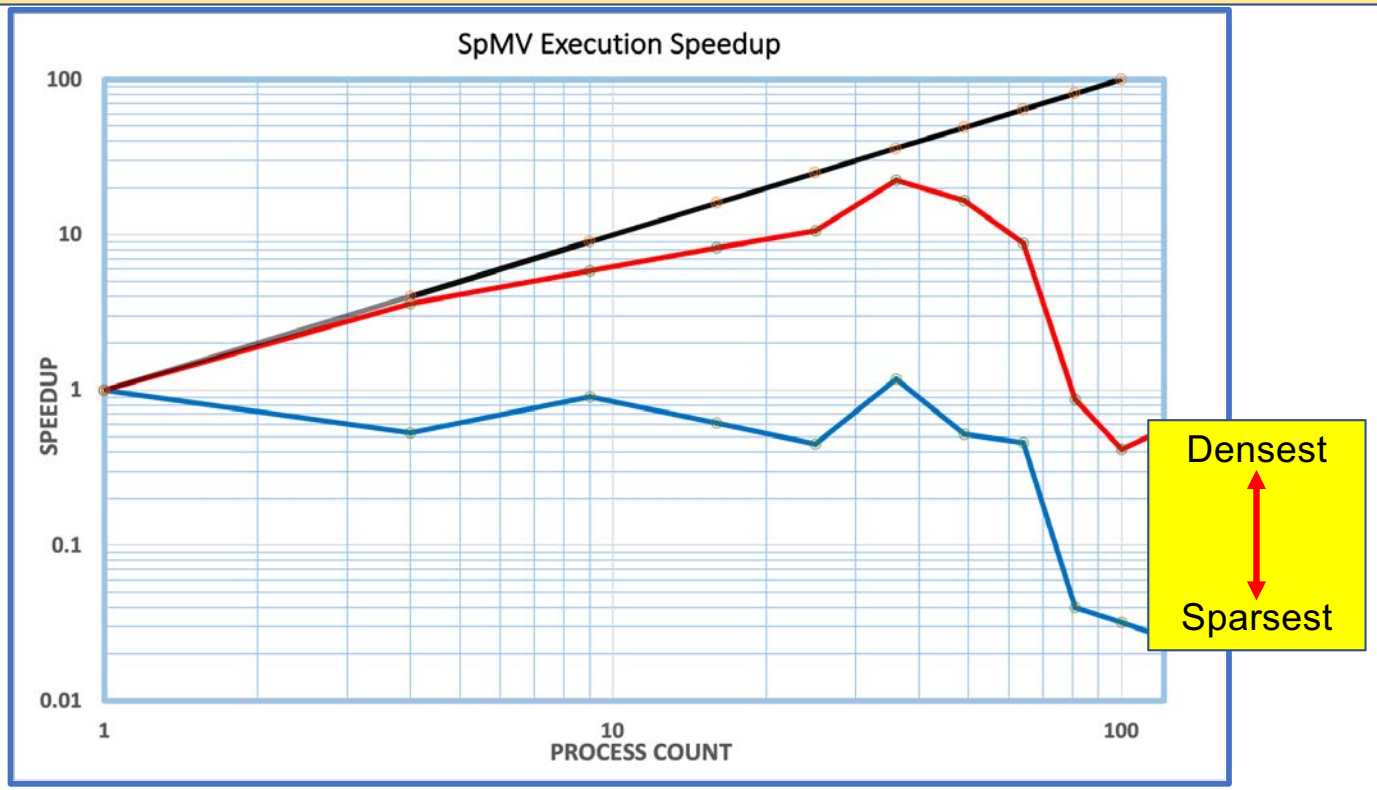


Chart by Brian Page, UND

SpMV is a kernel in many graph analytics algorithms

High Performance Conjugate Gradients (HPCG) Benchmark

Results: Nov 2020

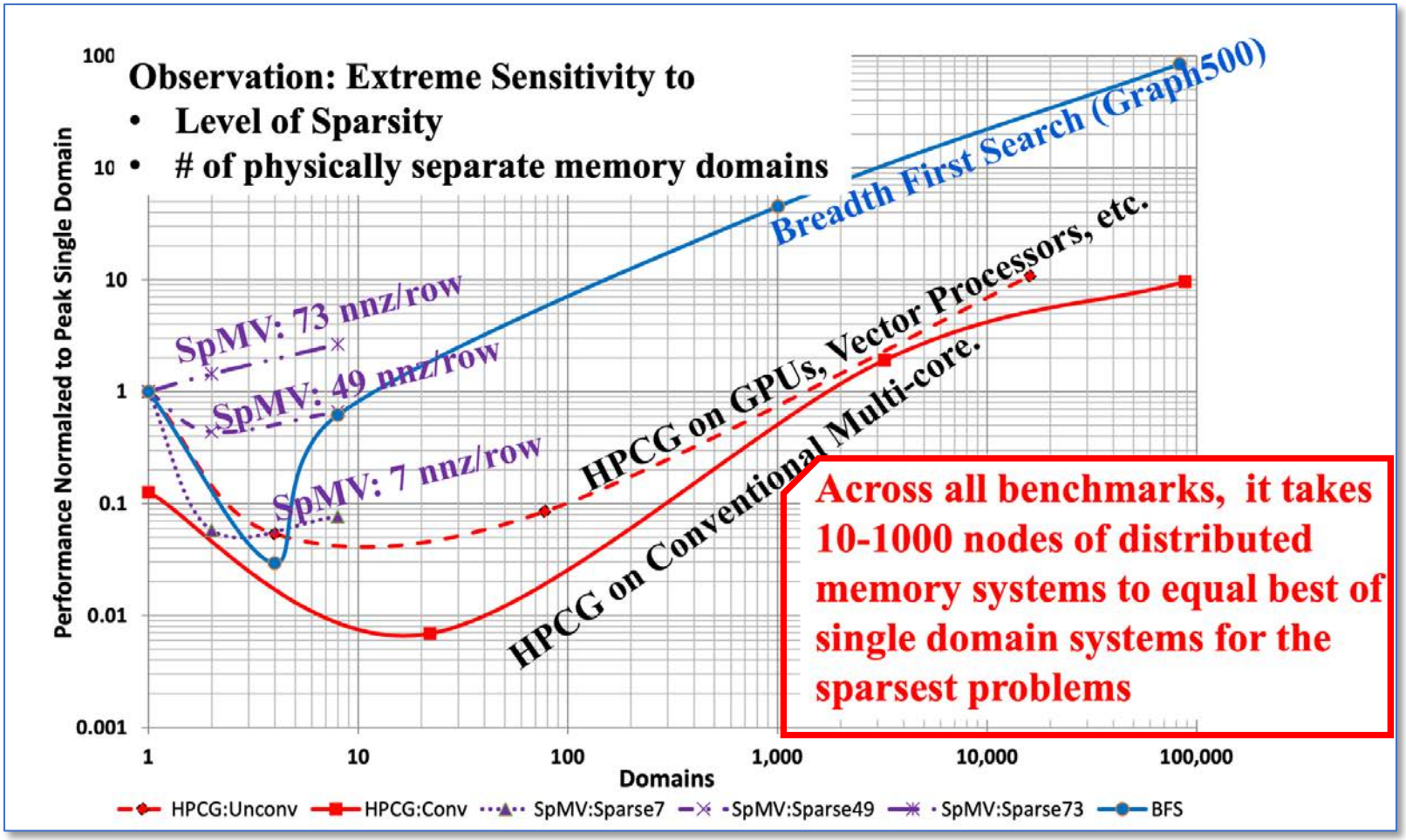
Conjugate Gradient (CG) Algorithm – iterative algorithm for solving sparse linear systems – key operations: Sparse Matrix X Vector (SpMV), global sum & synchronization (https://www.hpcg-benchmark.org)

Rank	Site	System	Cores	TOP 500 Rank	HPL Rmax (TFlop/s)	HPCG (TFlop/s)	% of HPL Rmax
1	RIKEN Center for Computational Science Japan	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu	7,630,848	1	442,010	16,004	3.62%
2	DOE/SC/Oak Ridge National Laboratory Unite States	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM	2,414,592	2	148,600	2	1.97%
3	DOE/NNSA/LLNL United States	Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox					1.90%
4	NVIDIA Corporation United States	Selene - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia	555,520	5	63,460	1,622	2.56%

Dense Algorithm (HPL) 82% of Peak
Sparse Algorithm (HPCG) 3.6% of HPL Rmax

Benchmarks – Graph500 & HPCG

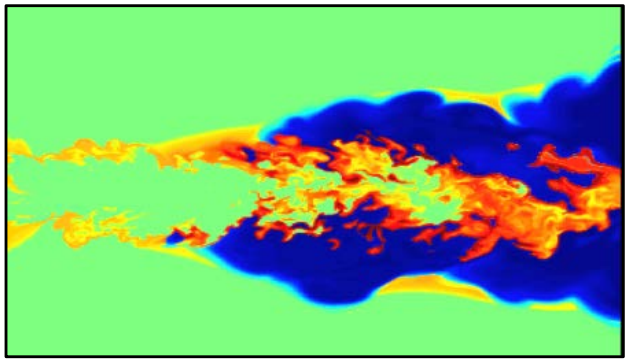
Graph500 / Breadth-first search (**BFS**) is an algorithm for traversing graph structures.



Graph by Peter Kogge: kogge@nd.edu

http://graph500.org

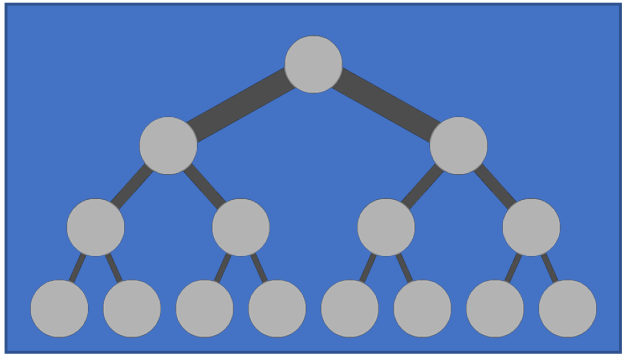
Today's Computers



Multi-Physics Simulation

Jackie Chen, SNL

Accelerators



Conventional Networks

- Designed for yesterday's applications
  - Multi-physics simulations
- Vendors are focused on incremental improvements
  - Accelerators & supporting memory components
  - Focused on processing, not data challenges
- Computers are not computational efficient or scalable for large scale graph analytics problems

Current Systems are Composed of Independently Developed Components

Communication – network that interconnects memory & computation

- ***Standard Interfaces***
- ***Block-oriented data movement***
- ***High overhead / container***

COMMODITY

Mellanox: InfiniBand
Cray: Slingshot
Intel: Omni-Path
Ethernet

Memory – mechanisms for accessing and storing data

- ***Caches - load/store of cache lines***
- ***No local processing***
- ***CPU controls memory accesses***

COMMODITY

DRAM DDR5
HBMnext
GDDR6X
NVRAM

Computation – mechanisms for the execution and flow control of tasks

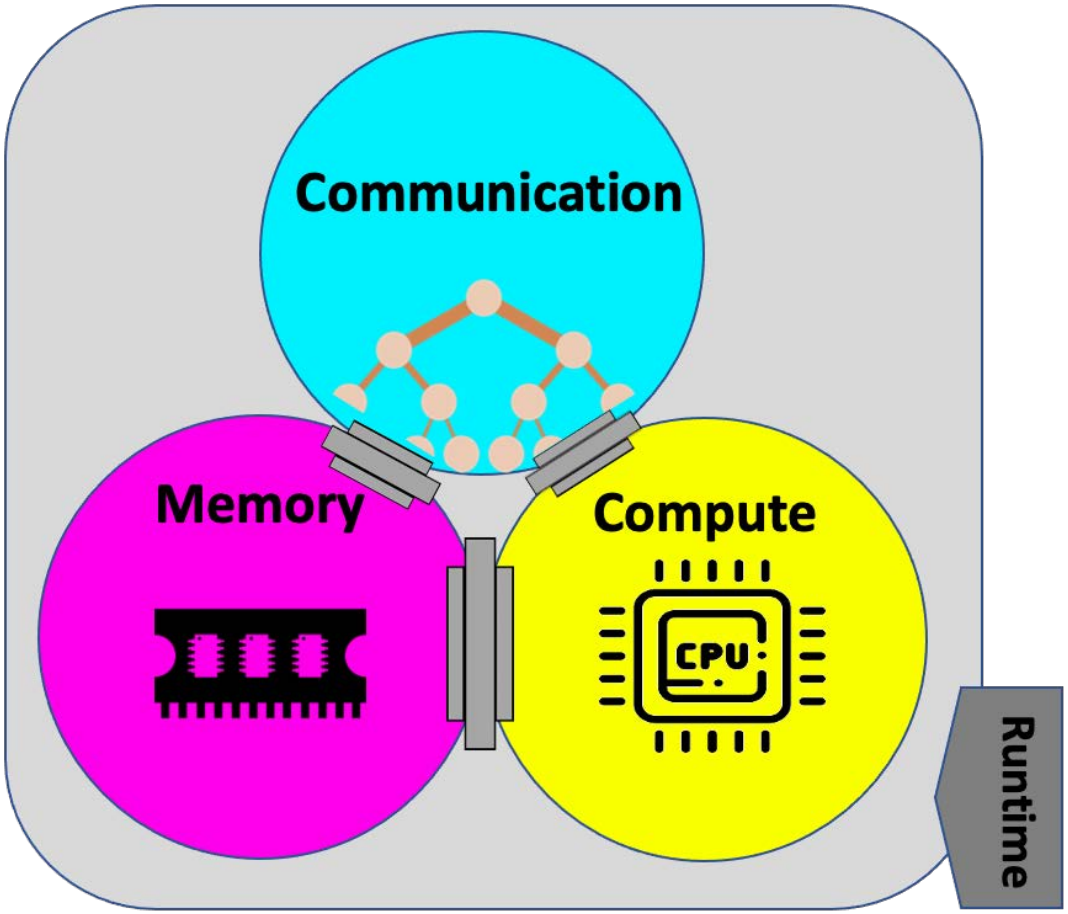
- ***CPU/GPU: von Neumann architectures for 32 / 64-bit FP compute-intensive apps***
- ***CPU drives everything: data, system control, …***

COMMODITY

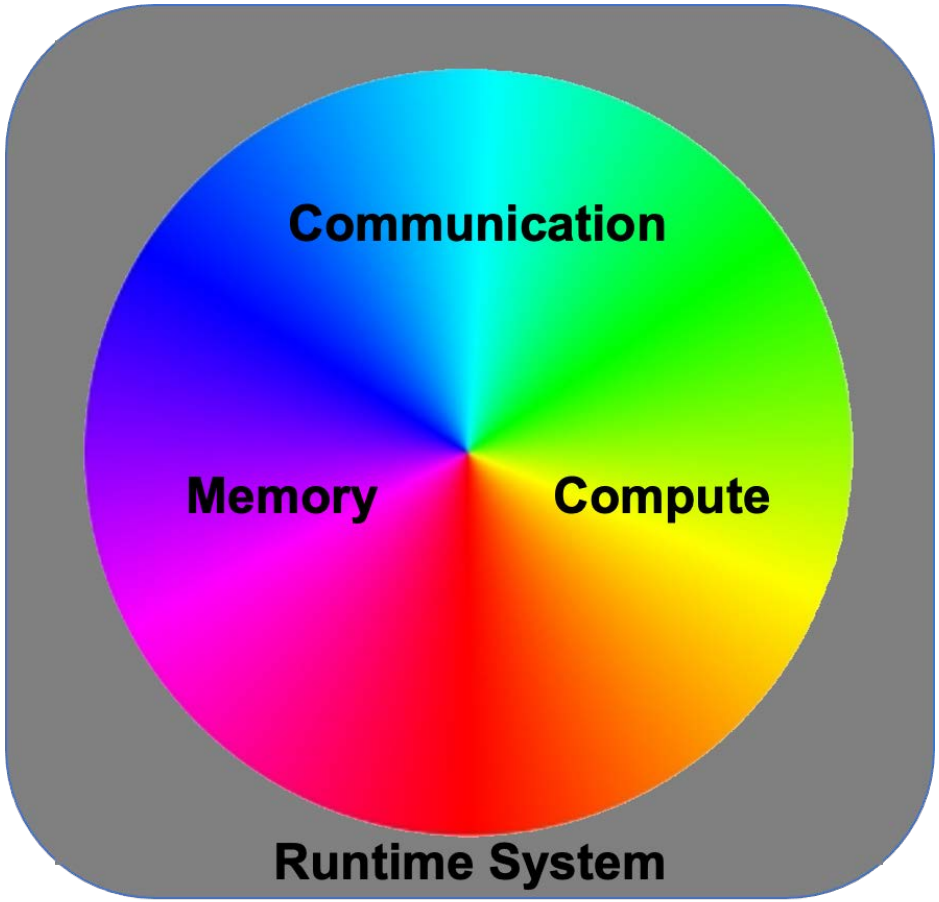
INTEL, AMD, ARM,
NVIDIA, Apple, RISC-V
FPGA
Neuromorphic

… Still focused on processing, not data movement

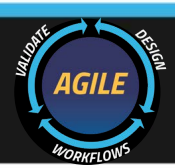
Today's Sub-optimal Computer Designs



AGILE Integrated Computer Designs



Technical Challenges



Designing processing elements to execute efficiently a broad spectrum of data-driven tasks (perhaps unknown until execution time).

- data analytic workloads can have execution patterns with have poor spatial and temporal locality and instruction streams have numerous branch-conditions that cannot be predicted ahead-of-time.
- furthermore, the amount of work executed may be data dependent and not predictable a priori

Support fine-grain, irregular data movement ranging from the processing elements to the system level. Data irregularity will arise in both data layout and compulsory data movement.

- to keep a data-movement pathway fully occupied, there must be enough data-movement requests in progress concurrently that will keep that channel full, thereby hiding the latency.
- these data movement is unpredictable, and data driven alternative approaches to minimizing/hiding latency will be necessary.
- with fine-grained (very small) data transfer requests, the efficiency of data movement can suffer from the overhead cost.

Technical Challenges

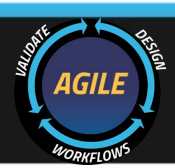
Data-driven parallelism at all levels (fine-to-coarse grain)

- graph processing for irregular, time-varying structures has parallelism that can only be discovered in real time.
- hiding latencies will require processing as many tasks in parallel as possible.
- high percentage of fine-grain tasks may necessitate asynchronous operations.

New memory and storage architectures for random, dynamic irregular data structures. AGILE workflows may have non-regular structures whose number, shape, and size are not known until runtime and may change dynamically during runtime.

- creates challenges for conventional page-based memory management and data layout in memory structures.
- creation and deletion of these data structures may need to happen asynchronously at large scale, which creates huge challenges for memory consistency models at both the node and system scale.
- Since the data properties (the base type) may be highly varied and unknown until runtime, the system and even the processing elements may need to have dynamic introspection.

Technical Challenges



An I/O subsystem capable of ingesting high-velocity streams of data from multiple external sources

- Filesystem semantics make it very challenging to ingest data from multiple data sources into a file interface that has sequential semantics while still maintaining global consistency across all clients.
- software interfaces and layered abstractions to access non-volatile and high-capacity storage are increasingly mis-matched with the underlying technology and should be re-thought from bottom to top.

System security, data integrity, and compliance services to support multiple applications working cooperatively in the same memory space

- For modern high-performance computing systems, multi-tenancy usually involves many different users and applications with different access credentials able to share a common database.
- For AGILE systems, the access control apparatus must allow a common approach to managing authentication and authorization that works across nonvolatile storage and volatile memory.
- Both efficient data transport and end-to-end data integrity protection are necessary and essential.

Summary

AGILE Technical Challenges

1. New processing elements designed to execute efficiently a broad spectrum of data-driven tasks (perhaps unknown until execution time).
2. Fine-grain, irregular data movement at all levels, ranging from the processing elements to the system level.
3. Parallelism at all levels and resolutions (fine-to-coarse grain).
4. New memory and new storage architectures for random, dynamic, irregular data structures.
5. An I/O subsystem capable of ingesting high-velocity streams of data of different types from multiple external sources.
6. System security, data integrity, and compliance services to support multiple applications working cooperatively in the same memory space.

Reversing von Neumann Architecture Heritage

Optimization of ALU/FPU utilization
- Presumes that FPU is most valuable resource

Separation of compute logic and memory
- Referred to as “von Neumann bottleneck”
- Consequence of early disparate enabling technologies

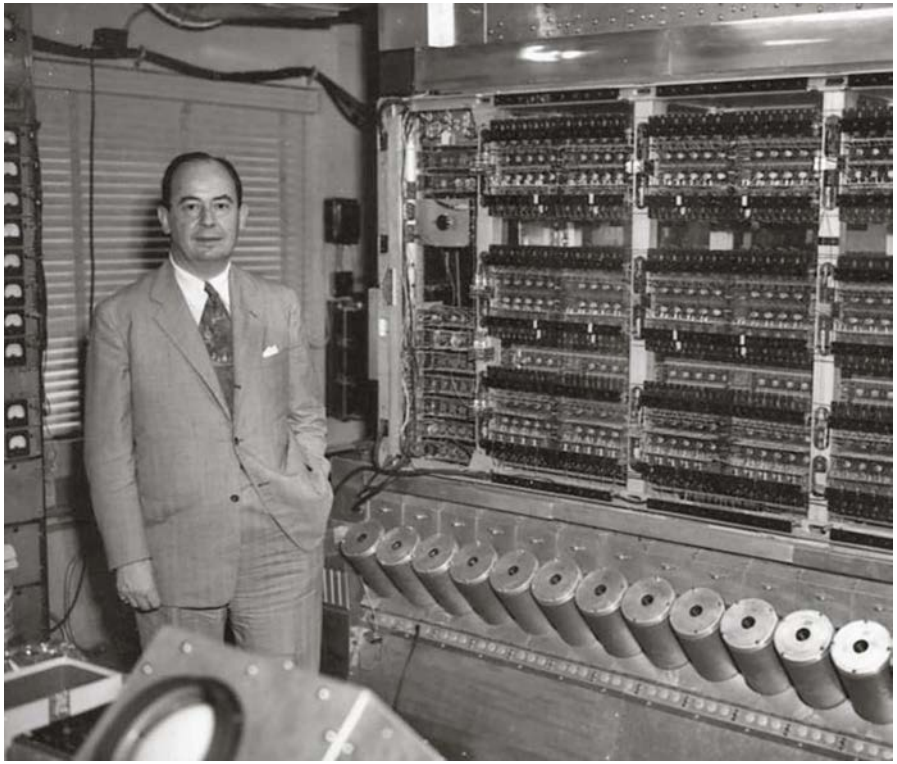
Sequential instruction issue
- Constrains natural parallelism

Sequential consistency memory model
- Narrows ordering of global mutable side-effects
- Imposes cache coherence mechanisms

Registers of isolated local name space
- Load/store operations
- Logically insular
- Causes register overflows
- Fixed length

https://www.nature.com/collections/dhdjceebhg and Thomas Sterling - HPCWire, July 2020

First Draft of a Report on the EDVAC, John von Neumann



Universal machine: John von Neumann and his high-speed computer, circa 1952.
https://medium.com/luteceo-software-chemistry/can-programming-be-liberated-from-the-von-neumann-style-932ba107402b

Innovative Research Strategies

Tightly Integrated Subsystem Design (communication, memory, compute & runtime) resulting in a unified, efficient, and scalable subsystem with transparent access to data and compute resources throughout the system.

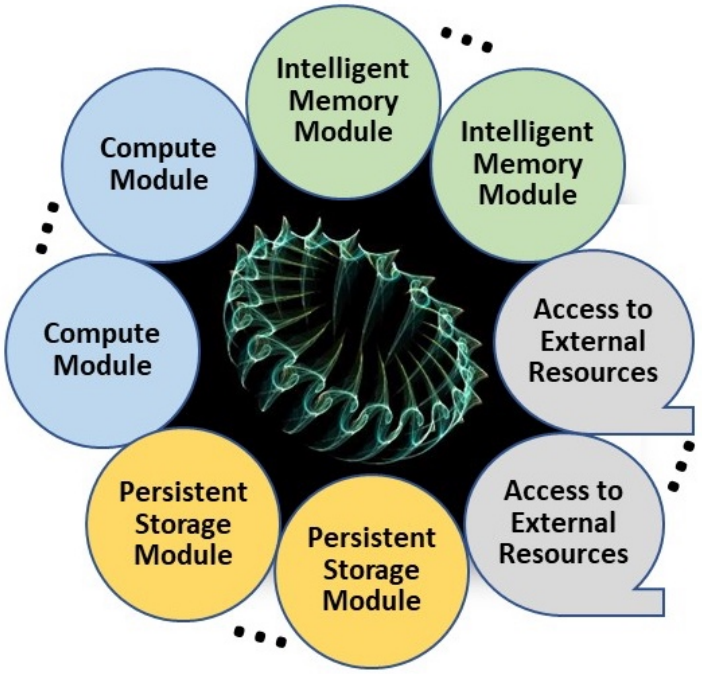
Data-driven compute elements optimized for both data, low-latency, high-BW data access, including moving the compute to the data.

Distributed memory management and security system that efficiently supports fine-grained addressing and protection of objects and data across the system.

System-level intelligent mechanisms for moving, accessing, and storing – large, random, time-varying data streams and structures.

Global Name Space/Global Adaptive Data Transfer mechanism to enable efficient data movement and adaptation of information flows to match complex workflow requirements.

Dynamic adaptive runtime: match activity demands to changing resource availability.



AGILE Modular System Vision

AGILE Program

Program Objectives:

- Enable data analytic problems that involve 10X more data
- Time to solution 10-100 times faster
- Emerging data analytic challenges

Research Teams:

- Funded to perform research on microelectronics / computer architecture projects to develop system-level designs for new AGILE computer architectures.
- **Develop validated designs that achieve or exceed the AGILE Program Target Metrics. These results will be validated by an independent test and evaluation team.**

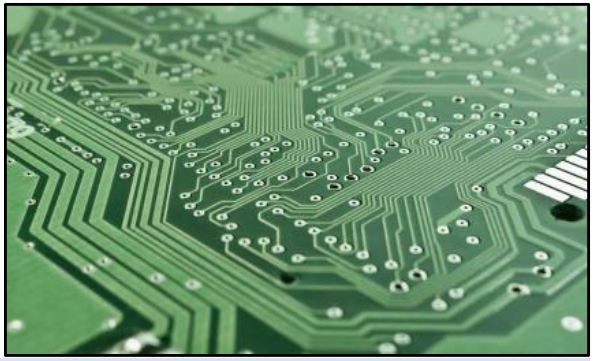
Phases:

- **Phase 1:** Exploratory microelectronics research that results in a high-level behavioral model, with a performance model, of a preliminary system design.
  (duration: 18 months)
- **Phase 2:** Develop a detailed design for proposed AGILE architecture.
  (duration: 18 months)

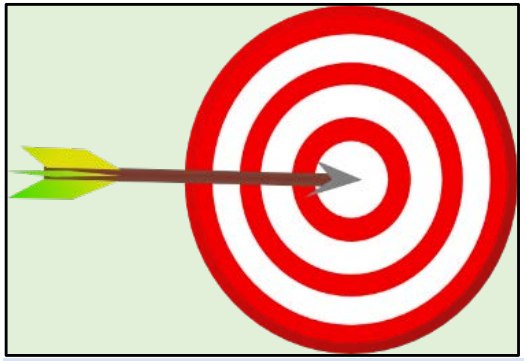
AGILE Design Goals

Cost Effective

Open Designs



Realized in Silicon (2030)



Target Metrics

- Create novel computer architectures and designs that overcome the technical challenges.
- The program will result in the delivery of RTL designs whose performance has been evaluated using an application modeling and simulation environment.
- AGILE architectures must be energy efficient and able to support scaling from a deskside system to large multi-cabinet configurations.

AGILE Program Out of Scope

Designs that don't provide efficient and scalable execution of large-scale data analytics, including streaming analytics.	Design of power distribution and cooling systems for the proposed architecture.
System designs that don't scale from a deskside system to large multi-cabinet configurations.	Research that does not have strong theoretical and experimental foundations or plausible scientific support for the research claims.
Energy inefficient designs.	Approaches that are likely to result in only incremental improvements over the current state-of-the-art.
Resulting system not cost effective.	Approaches that do not leverage AGILE application suite, A-SST, etc.
Designs that cannot be realized in silicon prior to 2030.	Approaches with significant limitations on operating conditions or operational parameters.
Designs that consist of IP that is not open sourced or licensable.	Design of the archival or extended storage system.

AGILE Workflows, Kernels & Benchmarks Codes

The program will provide source for workflows, kernels and benchmarks codes, with datasets, that will be used to drive the research and design AGILE efforts

Workflows

- **Knowledge Graphs –** Groups, Relationships & Interests
- **Detection –** System and Event Patterns
- **Sequence Data –** Identification & Clustering
- **Network –** Cyber-Physical Systems

Kernels

- Based on the workflows that exercise specific features

Benchmarks

- Breadth First Search
- Triangle Counting
- Jaccard Coefficients

AGILE Program Strategy

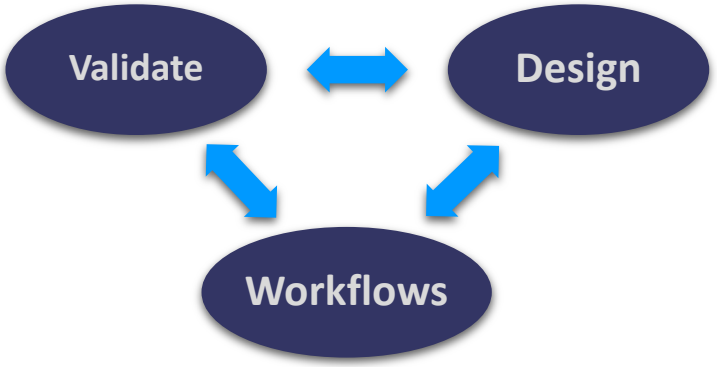
Workflows: Representative data analytic workflows, with realistic datasets

Design: AGILE Research Teams will perform research that will result in an integrated system design for data-driven computations

Utilize a co-design process involving : 1) AGILE workflows, benchmarks and kernels, 2) research/designs, and 3) modeling and simulation

Teams will develop detailed system-level designs, based on RTL designs and Functional models

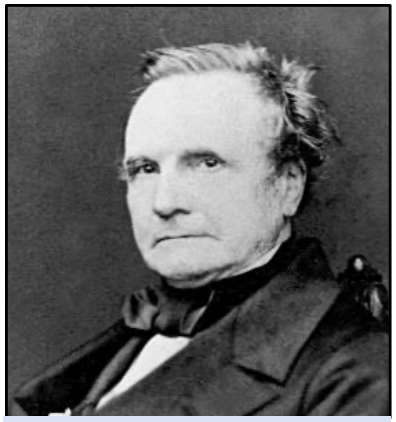
Validate: AGILE enhanced modeling and simulation (A-SST) tool-kit will be utilized to evaluate a team's design characteristics and performance estimates



Output: Open system designs; all non-open IP must be licensable. System designs must enable system security and compliance.

AGILE Architectures

Looking for the next computer pioneer, whose picture will appear below



Charles Babbage
Mechanical Computer

Alan Turing
Turing Machine

John von Neumann
von Neumann Arch.

John Atanasoff
1st Electric Computer

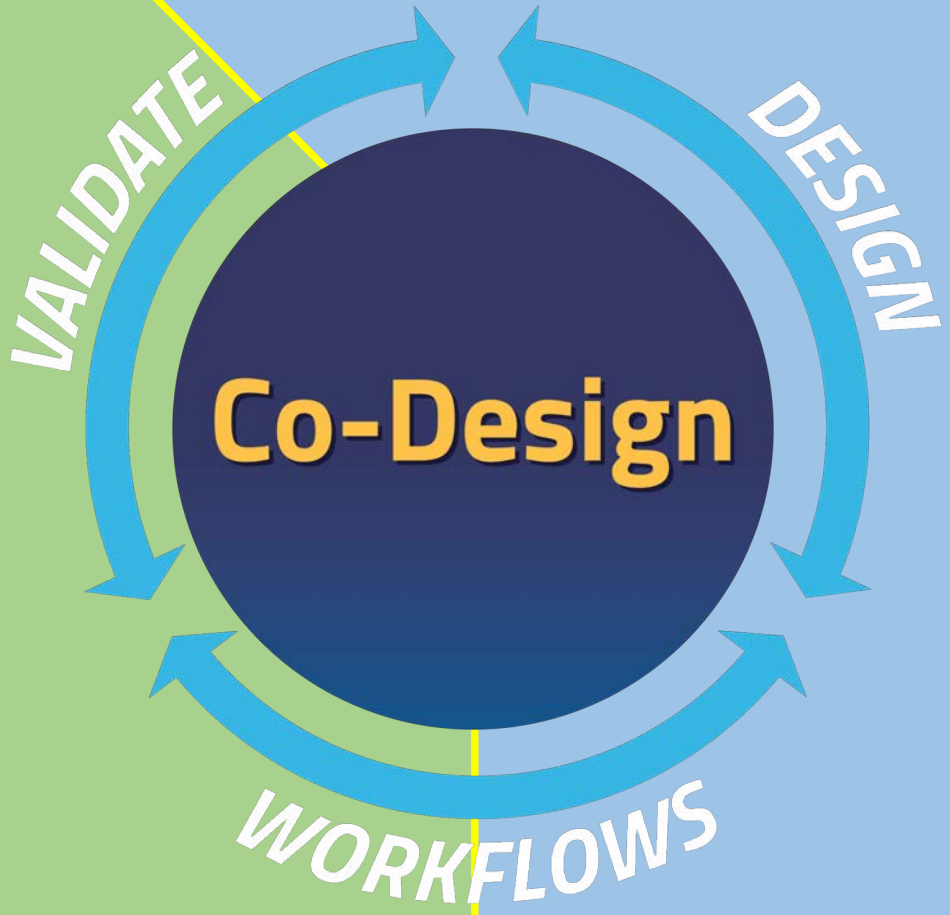
Grace Hopper
Programming

- Architectures that enable scalable, efficient execution of data intensive applications
- System-level intelligent mechanisms for moving, accessing and storing large, random, time-varying data streams, structures, objects, and knowledge
- Dynamic adaptive runtime systems to match activity demands to changing resource availability supported by hardware capabilities
- Declarative interfaces for intelligent programming environments that provide an intelligent determination of efficient and scalable data and computation operations

AGILE Research Teams: Overview

- Large, multi-disciplinary teams with R&D expertise in Communication, Memory, Computation, Runtime subsystems, including experts in the emerging data analytics
  - Multiple teams funded to perform microelectronics & architectural research that results in system design
  - Utilize co-design process to develop the design, optimize application codes and develop modules and system model for A-SST
  - Develop application framework that supports the AGILE workflows & are optimized for the AGILE system design
  - Validate the performance of the designs using the A-SST
  - Demonstrate that the design can achieve the Target Metrics

Research Teams



Test and Evaluation Team

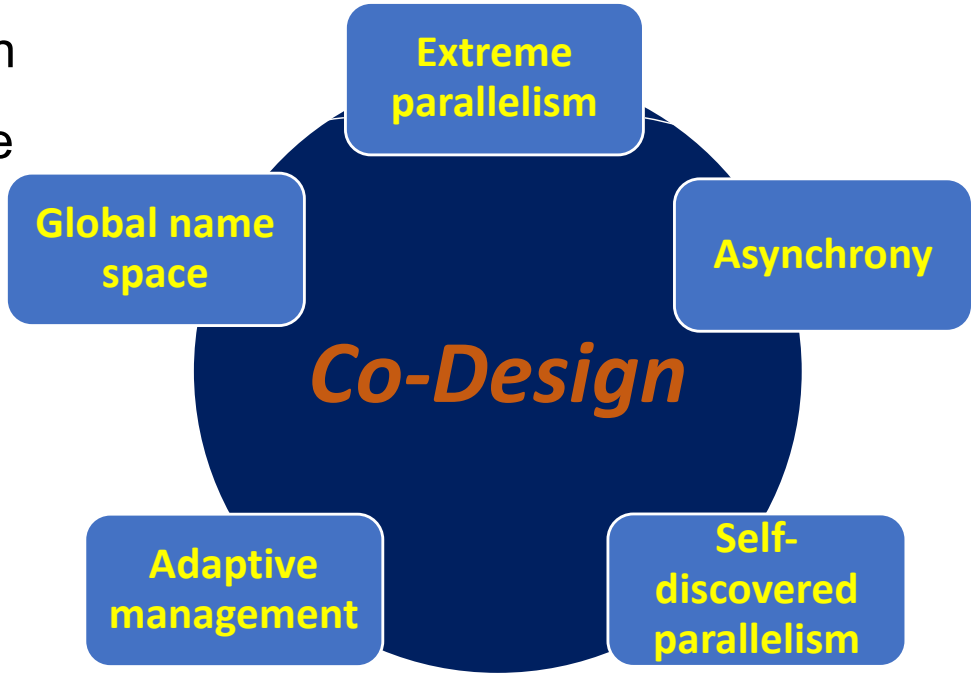
Attributes of an AGILE Runtime

The design of the runtime is an abstraction of computing system software structure and operation for a specific system design

It provides a conceptual framework for the co-design of technology: architecture, applications, and system software.

Design enables

- autonomous execution of data-driven computations.
- dynamically reconfigurable environment.
- performance & efficiency objective functions
- understanding of what are the physical first-class objects that are “run.”
- real-time adaptive control of resources.



System Performance Factors

Drives the research of the integrated system design

Starvation	• Poor utilization of work recourses • Impacts scalability & resource efficiency
Latency	• Added time to complete work when accessing remote resource • Impacts resource efficiency
Overhead	• Added time for managing work • Impacts scalability & resource efficiency
Waiting	• Delay due to simultaneous accesses of a specific resource • Impacts system efficiency

work = parallel tasks executed by the resources: communication, memory, computation and/or runtime.

Research Must Focus on Avoiding S, L, O, W to Achieve AGILE Goals

Impact of Data Analytics on S, L, O, W

Starvation

- Limited parallelism per thread due to limited operations per data
- Under utilization of the capabilities of network resources

Latency

- Limited benefit received when overlapping communication and computation operations
- Inefficient data paths for small packets with random locations

Overhead

- Time to setup data transfer exceeds the time to transfer the data involving small data packets
- Time spent on efforts not directly impacting thread execution

Waiting

- Large number of data requests to a remote memory, causing conflicts for the memory device
- Network switching resource conflicts caused by excessive number of small packets & different nodes

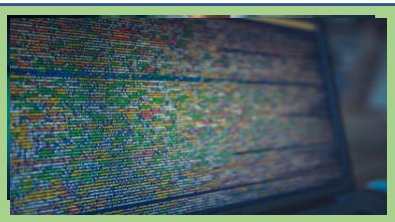
Mission Driven Applications

Four workflows have been selected that represent challenging problems of interest to the IC.

The Proposed Workflows are:

- **Knowledge Graphs – Groups, Relationships, and Interests**
- **Detection – System and Event Patterns**
- **Sequence Data – Identification and Clustering**
- **Network – Cyber-Physical Systems**

Knowledge - Groups, Relationships & Interests



Detection – System and Event Patterns

Sequence Data – Identification & Clustering

Network – Cyber-Physical Systems

Knowledge Graphs

Knowledge - Groups, Relationships & Interest

Knowledge Graph workflows operate on stored entities, their semantic types, properties, meta-data, provenance, and relationships; may be time varying.

AGILE goal: reduce transformation of data to knowledge from days to minutes; generate new discoveries & predictions with high-confidence evidence pathways.

Metric	Today	AGILE Target
Data ingestion rate	0.1 B per second	10 B per second
Embedding learning	1,440 minutes	30 minutes
Vertex classification with graph neural networks	> 1,440 minutes	30 minutes
Link prediction with graph neural networks	> 1,440 minutes	30 minutes
Multi-hop reasoning with graph neural networks	1 – 2 hops (exact) in 30 minutes	3 – 5 hops (fuzzy) in a minute

Detection

Detection – System and Event Patterns

Detection workflow have entities with state that may transition to new states when properties, etc. change. Workflow identifies events, event requirements, temporal evolution, precursors, key event drivers/influences.

AGILE goal: discover multi-day, multi-location patterns; issue prioritized alerts and actions in hours.

Metric	Today	AGILE Target
Size of graph	0.01 PB	10 PB
Number of data streams	1	3
Insert/Delete/Modify vertices and edges	0.1 B / second (batched)	10 B / second (continuous)
Pattern Detection per minute	Single event, linear paths, exact;	Multiple events, branches, prioritized fuzzy matching
Incremental analysis	NOT DONE	Never recompute from scratch
Multiple Day / Multiple Location queries	NOT DONE	Completed in minutes

Sequence Data

Sequence Data – Identification & Clustering

Sequence Data workflow seeks to extract common structures, inheritance, and functionality from massive temporal sequences of data. Workflows characterize agents/groups, recognize associated activities, and identify potential threats.

AGILE goal: increase number of sequences analyzed per unit time by 1,000x; reduce time to results from month(s) to hours-days; estimate uncertainties/prioritize alerts (not done today).

Metric	Today	AGILE Target
Ingest and prepare sequence data from multiple data lakes	**100 hours**	**4 hours**
Construct similarity graphs using metrics such as Jaccard index	**400 hours**	**6 hours**
Perform graph clustering of similarity networks	**500 hours**	**10 hours**
Predict labels (functions) of new sequences	**200 hours**	**4 hours**
Estimate uncertainties of labels and functions to prioritize alerts	**NOT DONE**	**Completed**

Network - Cyber

Network – Cyber-Physical Systems

Network workflow models complex interconnected networks and their defense. Examples include power grid, comms & control, social networks, etc.

AGILE goal: incremental cyber analysis and defense of network of network systems, returning prioritized actions in < 1 hour for graphs with 10 billion vertices and 100 billion edges

Metric	Today	AGILE Target
Graph construction through game theoretic modeling	120 minutes	2 minutes
Identification of top *K* influential nodes	60 minutes	1 minute
Enhanced models for identifying top *K*	600 minutes	30 minutes
Belief propagation	120 minutes	2 minutes
Incremental analysis	NOT DONE	Never recompute from scratch
Ranked courses of action	NOT DONE	Returned in less than an hour

Graph Benchmark Target Metrics

Assumption:

- System has 1+ PB of storage for graph data
- A core represents a basic computing element within the system, capable of executing graph algorithms and loading data from a memory

Graph Data

- Uses the Graph500 Benchmark datasets (http://graph500.org/)
- Scale 36: 2^{36} vertices & storage 17.6 TB
- Scale 42: 2^{42} vertices & storage 1.1 PB

Benchmark	Target - Scale 36	Target – Scale 42	Report
Breadth First Search (BFS)	1 GTEPS/core	1 GTEPS/core	Traversed edges per second (TEPS)
Triangle Counting	1M triangles/sec/core	1M triangles/sec/core	Number of triangles & number of 2-paths
Jaccard Coefficients	1M coefficients/sec/core	1M coefficients/sec/core	Number of Jaccard Coefficients & number of Jaccard Coefficients per second

Performance Evaluation: Summary

Objectives

BAA: These objectives will motivate the Workflow-driven R&D plans specified in the proposals

AGILE Program: An aggregate estimate, will be determined, based on performance estimates of the Workflows and Benchmarks

Workflows

BAA: Will describe Workflow components and target performance metrics. Proposers must describe how their approach will provide the required improvements and details concerning the estimated target metrics

AGILE Program: Performers will have source code for the Workflows to be used in the Co-Design process

Benchmarks

BAA: Proposers will report the performance estimates and justification

AGILE Program: Performers will utilize the Benchmark Codes in the Co-Design process and report performance estimates

AGILE Major Deliverables

Phase 1

1. Preliminary Design Review
2. Hardware & Application Test Plans, with test suites
3. Modified Application Software
4. **High-level Behavioral Models**

Phase 2

1. Critical Design Review
2. Hardware & Application Test Plans, with test suites
3. Modified Application Software
4. **Detailed Design**

High-level Behavioral Models – model behavior and performance of node components at the instruction and transaction level – do not provide accurate circuit timing but provide functionally correct results and approximate cycle timing. Must include node-level abstractions and parameterized multi-node simulations.

Detailed Design – a system-level model consisting of: 1) RTL design for AGILE integrated circuits (plus functional models for any COTS components); 2) precision node abstraction models that accurately describe the interactions between the components on a node, including an accurate performance model for the node; and 3) precision parameterized multi-node models for accurate system-level simulations.

Test & Evaluation Process

Government Test and Evaluation **(T&E) Team will reproduce (validate)** significant performer results and validate all projections, experimental results and test suites.

Government Funded Equipment (GFE): ***A-SST Environment*** is a computer modeling and simulation environment based on the Sandia National Laboratory SST toolset (https://github.com/sstsimulator).

Performers will submit their models/designs for validation and performance estimation using A-SST.

AGILE performer designs may consist of a mixture of A-SST-provided modules and their own AGILE custom modules. The **T&E Team will assist the performers** in ensuring that their custom modules can be incorporated into the A-SST environment.

Offerors may thus propose their own methods for augmenting the A-SST environment to achieve their proposed architectures' performance estimates. However, this methodology must be validated by the T&E Team.

Achieving program **target metrics is one factor** considered in deciding which teams move forward to later phase of the program

Proposals should be written with T&E in mind.

Testing protocols as written are preliminary and will be finalized by program kickoff meeting.

T&E Evaluation Efforts

Design V&V

Access COTS IP & validate functional models

Validate Performers' Hardware Test Plans

Validate Performers' Application Test Plans

Evaluate Performers' models/designs for correctness & completeness

Validate the results generated using A-SST

A-SST

Assist Performers with the integration of their models/designs into the A-SST* Environment

Validate Performers' models/designs for use in the A-SST Environment

Using A-SST, provide performance estimates of the Performers' models/designs

*AGILE-enhanced Structural Simulation Toolkit (Modeling and Simulation Environment)

Application Codes

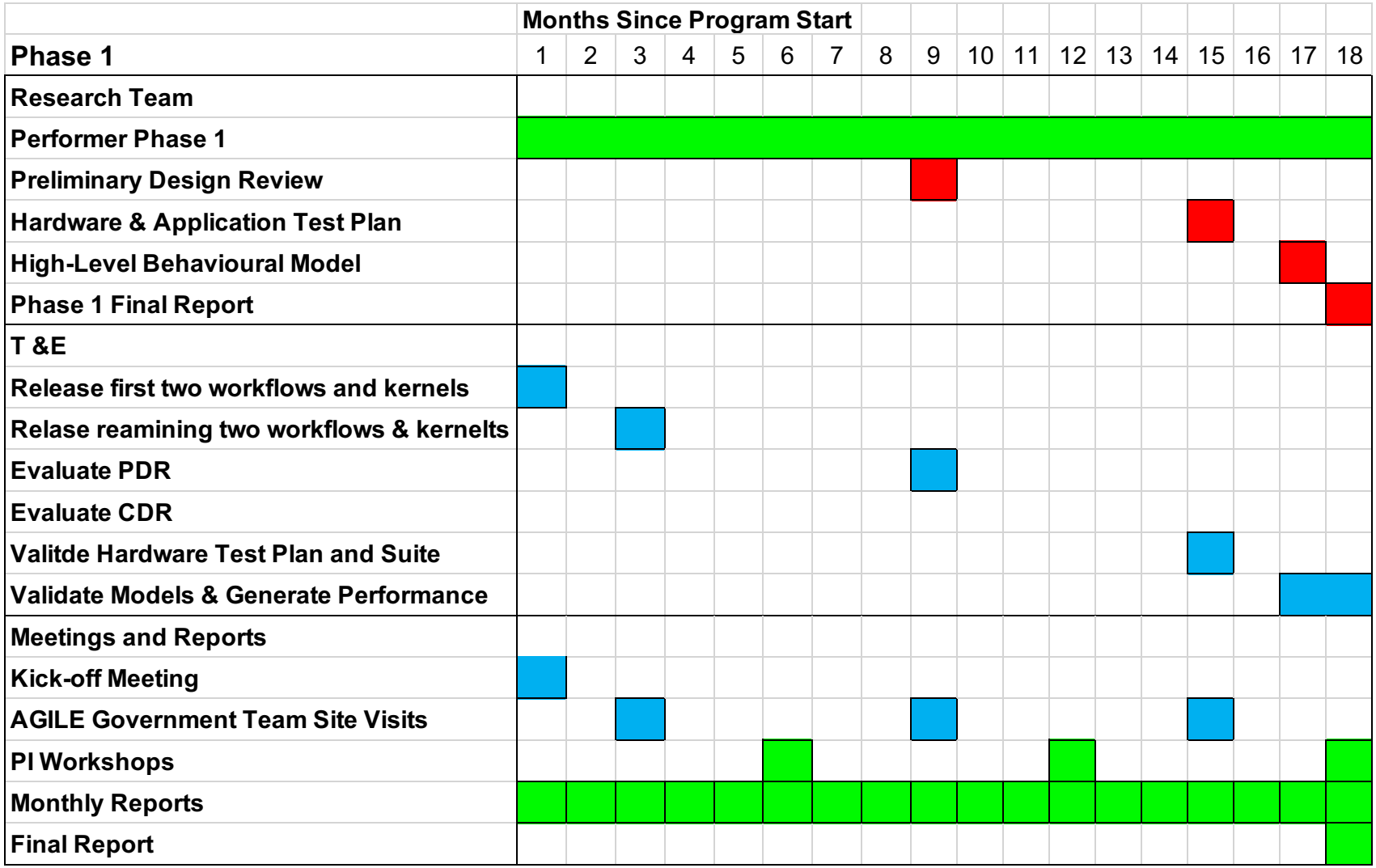
Develop AGILE Workflows and kernels

Generate performance of AGILE Workflows and kernels on Baseline Configuration qualified systems

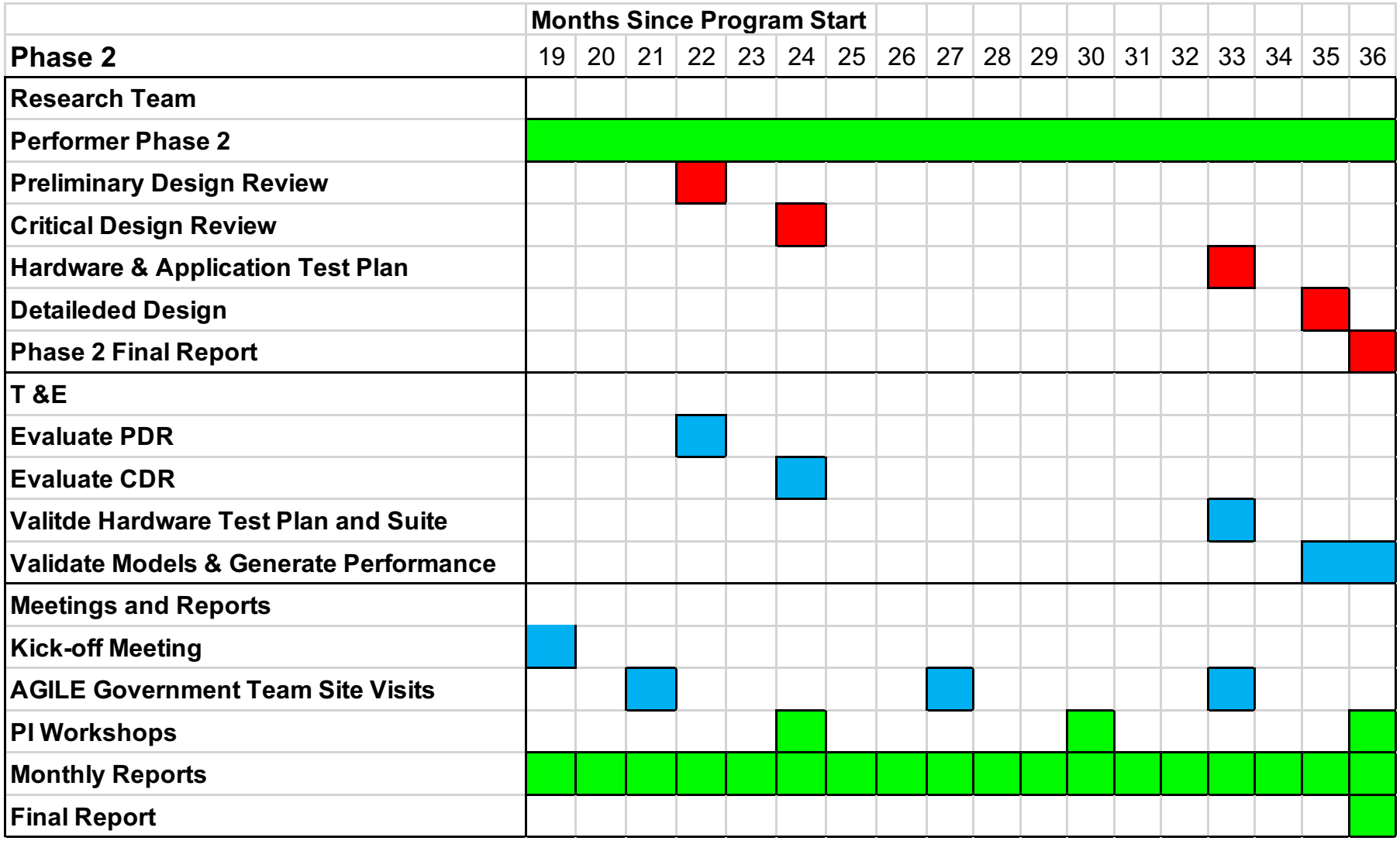
Assist the Performers with understanding and implementing the application codes

Validate changes to the Performers' versions of the AGILE Workflows and kernels (optimized for their systems)

Program Schedule – Phase 1



Program Schedule – Phase 2



Summary

- The chasm between the *DEMANDS* of today's escalating data-intensive problems and the *CAPABILITIES* of yesterday's computing systems is unbridgeable
- Evolutionary improvements are now providing diminishing returns
- AGILE is the first program in decades to offer a clean slate for completely re-thinking system-level computing architectures
- AGILE systems will enable new areas of data analytic applications that turn chaos into order
- AGILE offers you the opportunity to pioneer a new class of high-performance systems that will change the face of computing

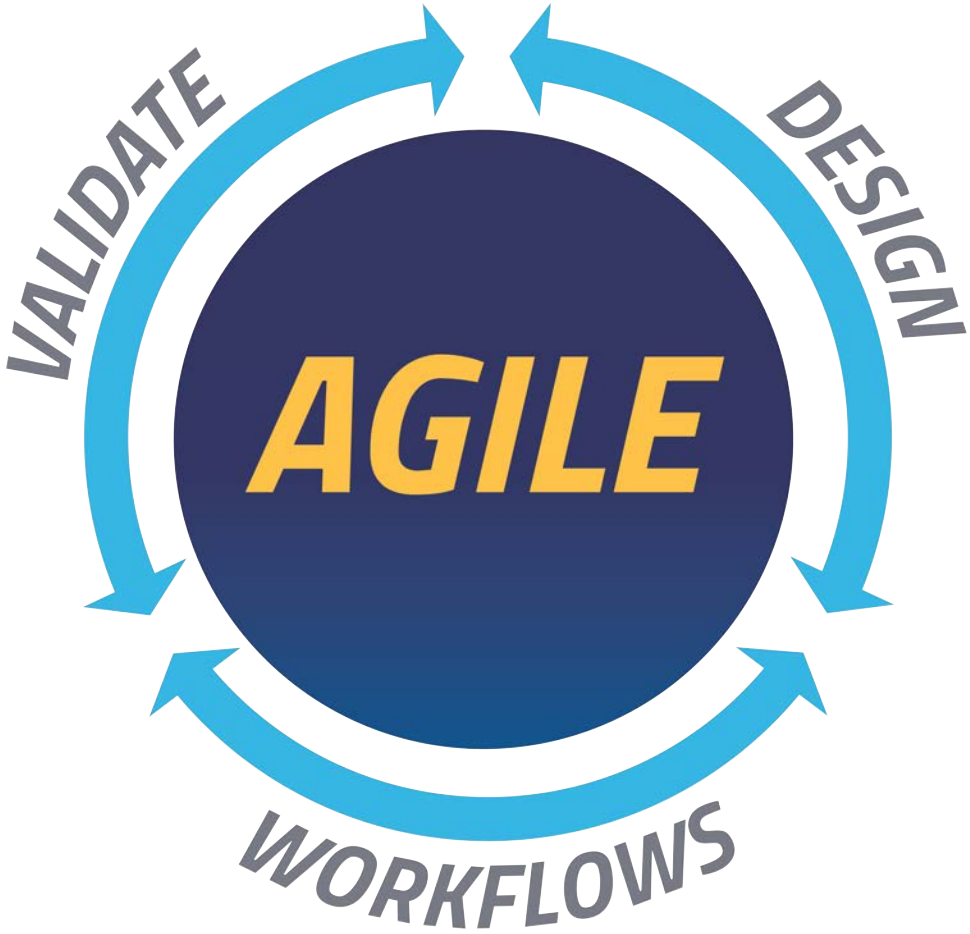
Point of Contact

Dr. William (Bill) Harrod
Program Manager
Office of the Director of National Intelligence
Intelligence Advanced Research Projects Activity (IARPA)
Washington, DC 20511
Phone: (301) 243-1814

Electronic mail: dni-iarpa-BAA-21-01@iarpa.gov
include IARPA-BAA-21-01 in the Subject Line

Website: https://www.iarpa.gov/index.php/research-programs/agile

THANK YOU!



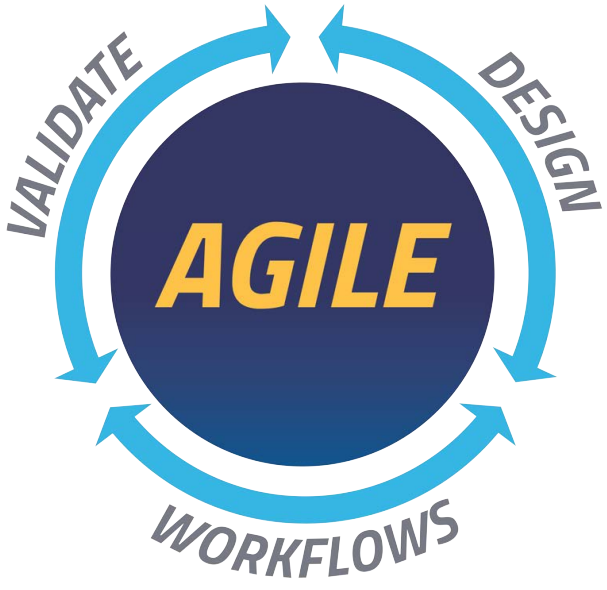
Agenda

Time	Topic	Speaker
12:00 PM – 12:10 PM	Welcome, Logistics, Proposer's Day Goals	Dr. William Harrod Program Manager, IARPA
12:10 PM – 12:20 PM	IARPA Overview	Dr. Pedro Espina, Director, Office of Collections
12:20 PM – 1:20 PM	AGILE Program Overview	Dr. William Harrod
1:20 PM – 1:35 PM	Break to formulate Questions	
1:35 PM – 2:00 PM	Doing Business with IARPA	Chris Fox, IARPA Contracting Officer
2:00 PM – 2:30 PM	Break Questions after 2:15 PM will not be addressed	
2:30 PM – 3:30 PM	AGILE Questions & Answer	Dr. William Harrod

Note: All times EST (Washington, DC Time)

Break to Formulate Questions

Doing Business with IARPA begins at 1:35 pm EST



Doing Business with IARPA
December 22, 2020 (FY21)

Chris Fox, IARPA Contracting Officer

Office of the Director of National Intelligence

IARPA

BE THE FUTURE

General

Broad Agency Announcement (BAA)

Questions and Answers

Eligible Applicants

Preparing the Proposal

Submitting the proposal

Evaluation and award process

Other

- Organizational Conflicts of Interest
- Intellectual Property
- Pre-Publication Review
- Academic Institution Acknowledgement
- Multiple Proposal Submissions
- Contract Type

- Disclaimer

BAA

- IARPA uses BAA type solicitations conducted under FAR Part 35, Research and Development Contracting
- BAAs will be posted to beta.SAM.gov
- We typically allow 45 – 60 days for proposals
- All the information needed to submit a proposal will be in the BAA.

Questions and Answers

- The BAA will have a Q&A period during which prospective offerors can submit questions
- The email for questions will be provided in the BAA
- Q&As will be posted to beta.SAM.gov so be sure to check regularly
- No answers will go directly to offeror nor shall questions be sent to other than the email designated in the BAA
- Note that your question will be posted so be careful not to reveal information that you don't want made public..

Eligible Applicants

Collaborations and teaming are generally encouraged by IARPA

- team formation is the responsibility of Offerors

Foreign organizations and/or individuals

- This is program dependent, the BAA will specify if there are any limitations
- Regardless of eligibity, must comply with:
  - Any contract security clauses or requirements
  - Export Control Laws (ITAR, EAR) and implementing contract clauses

Eligible Applicants, cont.

The following are generally **not** eligible to submit proposals for IARPA research programs or participate as team members under proposals submitted by eligible entities

- Other Government Agencies,
- Federally Funded Research and Development Centers (FFRDCs),
- University Affiliated Research Centers (UARCs)
  - An entity of which only a portion has been designated a UARC may be eligible to submit subject to an OCI review if stated in the BAA
- Any organizations that have a special relationship with the Government that would give them
  - access to privileged and/or proprietary information
  - access to Government equipment or real property

Preparing the Proposal

- The BAA contains proposal preparation instructions such as:
  - Due date and time
  - Page limitations and format
  - Information to be addressed in the proposal (e.g., technical, cost and administrative)
  - Templates for required proposal attachments (e.g., Cover sheets, OCI notification, Academic Institution Acknowledgement, IP/Data Rights Assertions, Cost breakdown)
- The BAA also contains the evaluation factors for award including the technical evaluation criteria (e.g., technical approach, relevance to IARPA, work plan, experience, key personnel, resource realism, etc.)
- The BAA describes the method of evaluation and selection
- IARPA may only request the Technical Volume initially with the detailed Cost volume requested after selection

Submitting the Proposal

Proposals must be submitted through IARPA's IDEAS system

- Interested Offerors must register electronically IAW instructions on: **https://iarpa-ideas.gov.** (will be available after BAA is posted)
- Interested Offerors are strongly encouraged to register in IDEAS at least one week prior to proposal "Due Date"
- Offerors must ensure the version submitted to IDEAS is the "Final Version"
- For Classified proposals, the BAA will contain separate delivery instructions

The BAA will have instructions for how to respond if there are system problems with IDEAS

If the Cost Volume is not requested until after selection, it will be directly submitted to the contracting officer, not through IDEAS

Evaluation and Award Process

Each BAA will detail the method for evaluation and selection but IARPA generally follows a two-step process:

- Fist step is evaluation and selection for negotiations. This is conducted through a scientific/peer review process after which offerors are notified of selection
- Second step is negotiation and contract award conducted by the contracting officer

- Proposals will be reviewed individually against the BAA requirements in accordance with FAR 35 and not against each other

Organizational Conflicts of Interest

IARPA follows FAR Part 9 regarding Organizational Conflicts of Interest (OCIs). The main principles being:

- preventing conflicting roles that might bias a contractor's judgement
- preventing an unfair competitive advantage

The BAA will describe how offerors are to identify and disclose all facts relevant to potential OCIs for the offeror as well as any proposed team members

OCI disclosures may require a mitigation plan describing the actions the offeror will take or intends to take to prevent the conflict

IARPA generally prohibits contractors from concurrently providing System Engineering Technical Assistance (SETA) and T&E support while being a technical R&D performer due to OCI concerns. Each case will be determined individually.

Intellectual Property

The Government needs to be able to effectively manage the program and evaluate the output and deliverables, communicate the information across Government organizations and support further use and development of program results

Offerors will address their IP Rights assertions in their proposal. The Government may request additional information as may be necessary to evaluate

The Government will evaluate the IP rights being offered and whether they are in the Government's best interests.

Pre-Publication Review

IARPA encourages publication of **UNCLASSIFIED** IARPA-funded research in peer-reviewed journals, presentation at conferences and publication in conference proceedings.

Prior to public release of any work submitted for publication, the Performer will:

- Communicate results to be publicly released with the IARPA Program Manager to discuss any sensitivities (e.g., security, speculation on IC use cases, etc.)
- Provide advance courtesy copies to the IARPA PM and Contracting Officer Representative (COR/COTR)

Academic Institution Acknowledgement

According to Executive Order 12333, contracts or arrangements with academic institutions may be undertaken only with the consent of appropriate officials of the institution.

An Academic Institution Acknowledgement letter is required for offerors that are academic institutions and for any proposed teammate that is an academic institution.

A template for this letter will be included in the BAA. Each letter must be signed by a senior official of the institution (e.g. President, Chancellor, Provost or other appropriately designated individual).

IARPA requires this letter before entering into negotiations and/or awarding a contract. It is highly advised that it be submitted with the proposal.

Multiple Proposal Submissions

Proposal Submissions to other entities:

- Typically, the BAA asks offerors to name, in their proposal, other federal, state or local agencies and/or other parties receiving the proposal (or substantially the same proposal) or funding the proposed effort.

- If the offeror has submitted the same or substantially the same proposal to other entities, it may impact IARPA's decision to select and fund the effort.

Multiple Proposal Submissions to IARPA:

- BAAs usually allow an entity to participate in multiple submissions as a prime or subcontractor. If allowed by the BAA, multiple submissions which include a common team member shall not receive duplicative funding for the same work (i.e., no one entity can be paid twice for the same work).

Contract Type

Cost or Cost-Plus-Fixed-Fee type contracts are typically awarded due to the nature of the R&D work. IARPA may, in some instances, consider other contract types, such as Firm Fixed Price, as well as non-FAR based agreements such as Other Transactions.

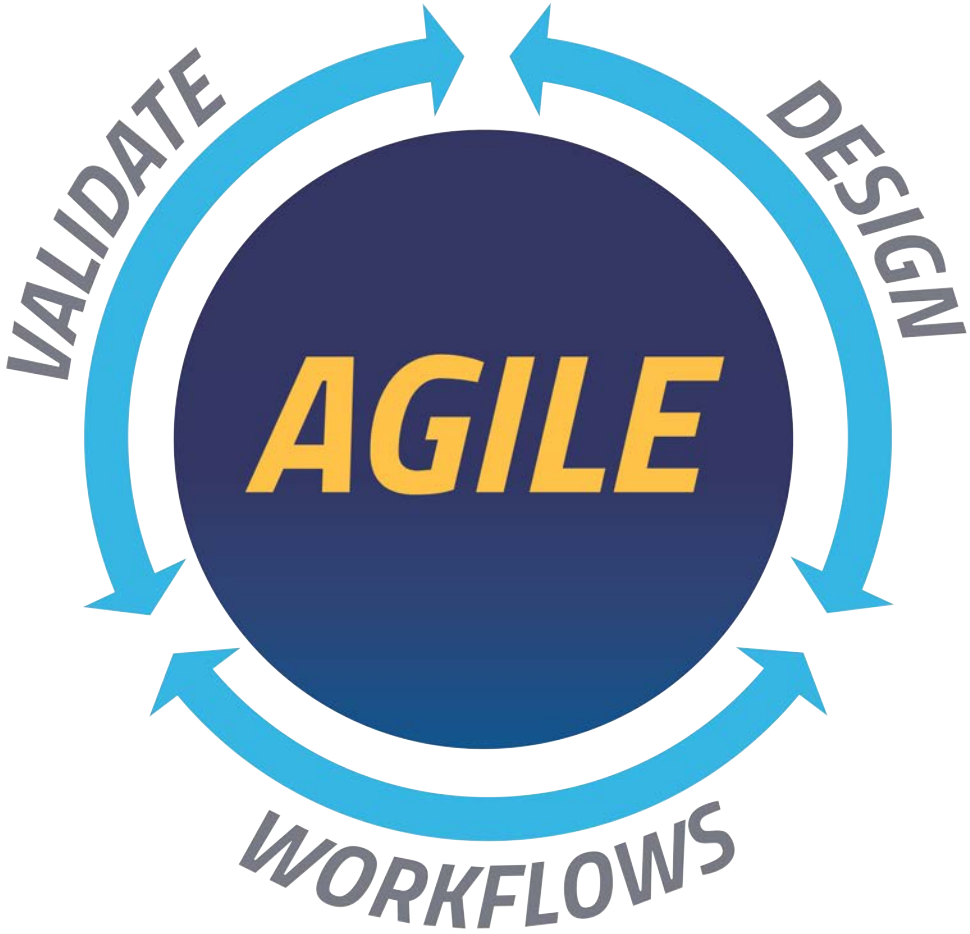
The types of contracts and agreements that will be considered and the conditions for such consideration (e.g., small business, start-ups, commercial, foreign entities, etc.) will be addressed in the BAA.

Disclaimer

The information conveyed in this brief is for planning and general information purposes and is subject to change.

Please carefully read the final BAA and adhere to its requirements which may differ from what has been presented in this briefing.

THANK YOU!



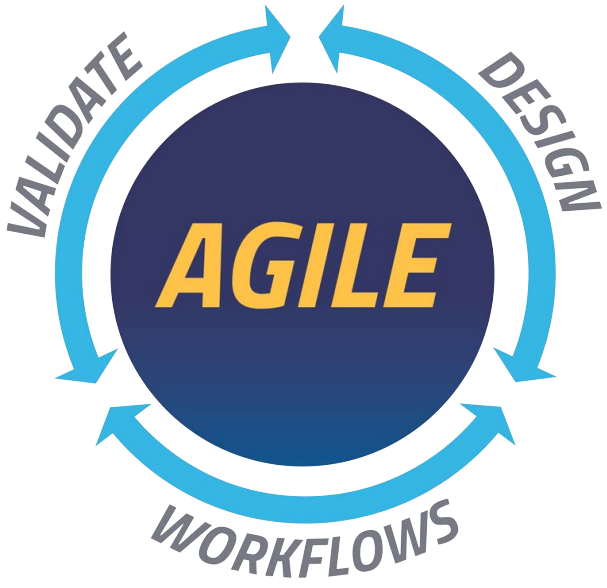
Agenda

Time	Topic	Speaker
12:00 PM – 12:10 PM	Welcome, Logistics, Proposer's Day Goals	Dr. William Harrod Program Manager, IARPA
12:10 PM – 12:20 PM	IARPA Overview	Dr. Pedro Espina, Director, Office of Collections
12:20 PM – 1:20 PM	AGILE Program Overview	Dr. William Harrod
1:20 PM – 1:35 PM	Break to formulate Questions	
1:35 PM – 2:00 PM	Doing Business with IARPA	Chris Fox, IARPA Contracting Officer
2:00 PM – 2:30 PM	Break Questions after 2:15 PM will not be addressed	
2:30 PM – 3:30 PM	AGILE Questions & Answer	Dr. William Harrod

Note: All times EST (Washington, DC Time)

Submit Questions by 2:15pm

Answers to Questions Begin at 2:30 pm EST



Answers to Questions