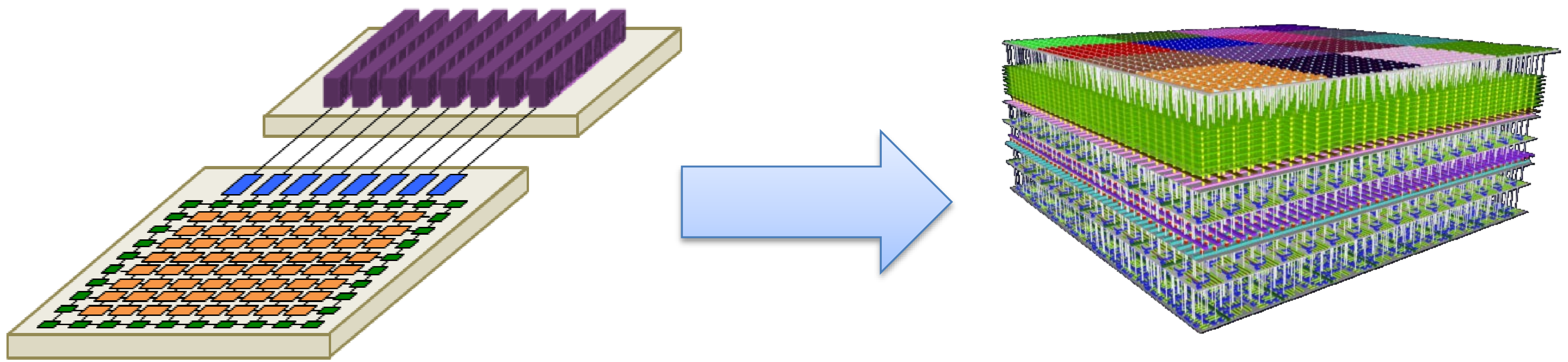
3 Dimensional Monolithic System on a Chip (3DSoC)



Develop novel monolithic 3D fabrication technologies that enable new architectures to drive a >50X improvement in SoC performance at power

Linton Salmon

September 22, 2017

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Agenda

- Motivation for the 3DSoC program
- Program structure
- Program metrics
- Program schedule
- Proposal specifics

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Motivation for 3DSoC

Future
New Materials/Integration

Si Scaling Benefit

Next-Gen Cognitive EW approach, wide band, real time
100W

Curren EW approach, wide band, real time
100W

Current EW approach single channel non-real time
240W

IC

General
Purpose
GPU

General
Purpose
CPU

100000
10000
1000
100
10
1
0

Computational Efficiency (GOPS/W)

130nm
90nm
65nm
40/45nm
28/32nm
20/22nm
14/16nm
10nm

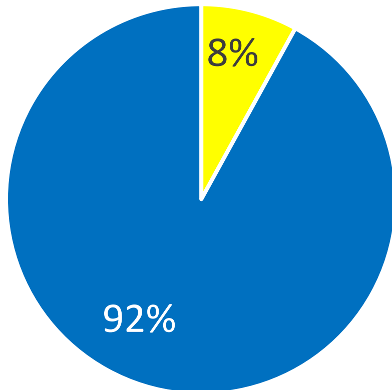
Technology Node

FPGA
DoD Custom IC
Custom IC
GP GPU
GP CPU
Intel CPU

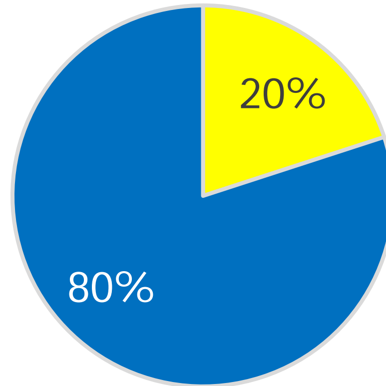
A Major SoC Performance Limitation: Memory Access

Data for 7nm instantiation of a state-of-the-art Machine Learning accelerator

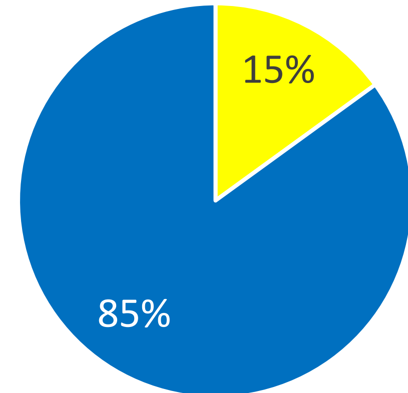
Neural Programmer (LSTM)



ResNet-152 (CNN)



Alex Net (CNN)

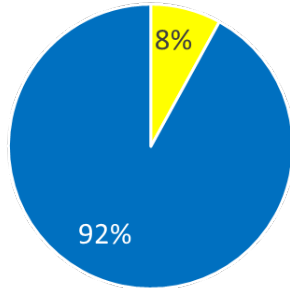


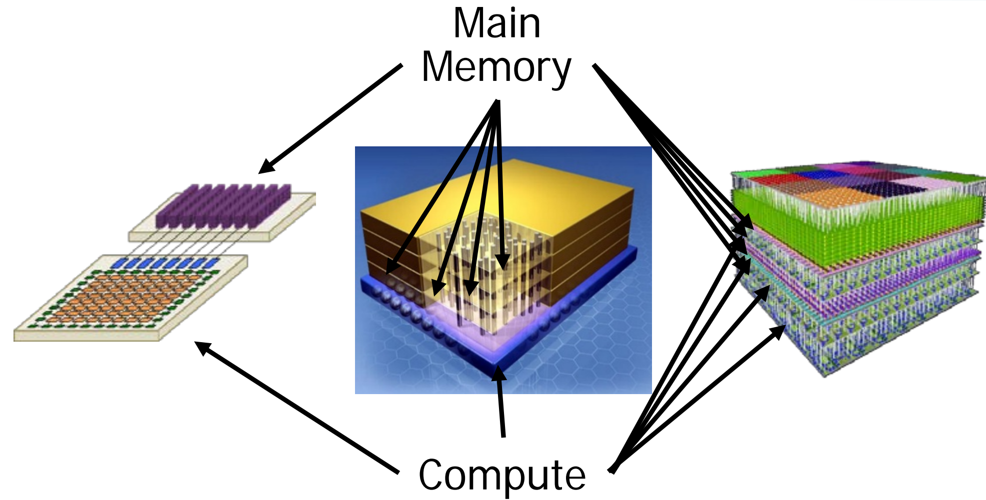
Data from S. Mitra of Stanford

Compute ■ Memory

The memory bottleneck is a major problem for machine learning and other applications requiring computation on large data sets

Addressing the Memory Limitation





Production 2D | Development 3D TSV Package | DARPA 3DSoC

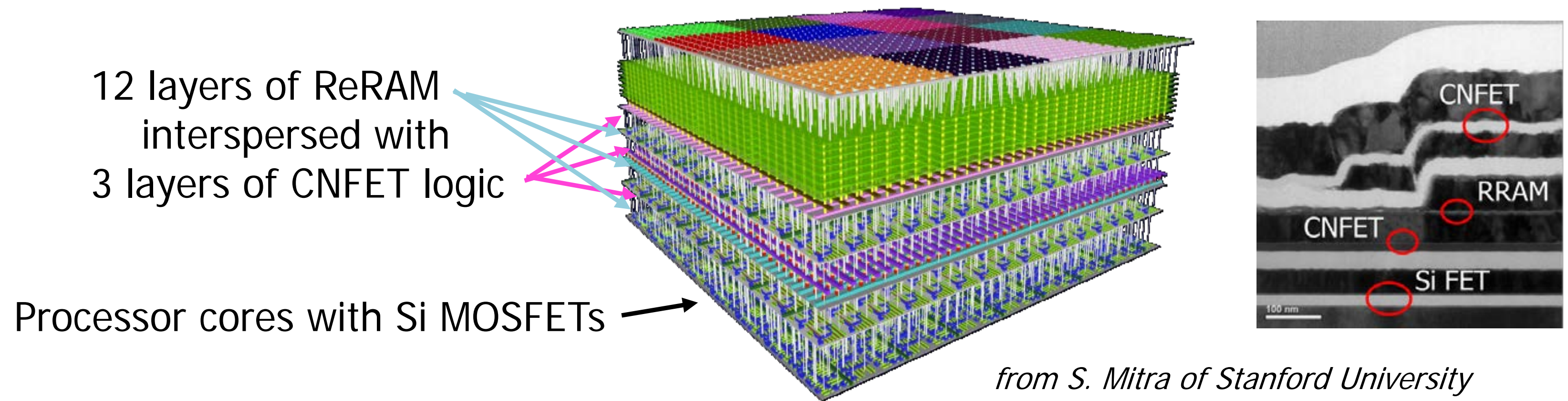
Memory Access Parameter	2D		3D TSV Package		3DSoC
Total I/O	512		1K		33K
Max Bandwidth (Gb/s)	400		1K		46K
Memory access energy (pJ/bit)	52		32		1.5
VDD (Volts System)	1.6		1.2		0.6

3DSoC increases the IO count and bandwidth by >50X from current 2D fabrication architectures

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

An Integrated, Monolithic SoC (3DSoC) Solution

An example of an integrated flow that fabricates 3D logic and memory on a single die



from S. Mitra of Stanford University

Note: This is an example only. Other technical approaches are expected.

Critical characteristics for a monolithic solution

- Must permit new architectures that leverage fast, configurable access to non-volatile main memory
- Stackable 3D logic and memory functions that allow new architectures
  - Low temperature formation
  - Logic AND memory
  - High density of memory – at least 4GB (Giga-Byte)/die
- Possible to fabricate in existing domestic, commercial, high-yielding infrastructure
  - 90nm on 200mm wafers
  - High yield on large SoCs

Stanford Simulation Results for Machine Learning

LSTM Network	Model Size	Training/Inference	2D at 7nm	
			Benefit 3DSoC at 7nm	**Benefit 3DSoC at 90nm**
Language Model	2.5 Gbytes	Training	645X	75X
		Inference	626X	73X
Neural Programmer	1 Gbyte	Training	359X	40X
		Inference	493X	55X
Image Captioning	150MByte	Training	367X	41X
		Inference	323X	35X

- 2D vs 3DSoC comparison
  - 2D: 7nm technology for accelerator and 4GB of off-chip DRAM main memory
  - 3DSoC: 90nm Carbon Nanotube FET (CNFET) for accelerator and 4GB of on-chip ReRAM (non-volatile) memory
- Uses published traces from an accelerator SoC and the LSTM algorithm
- Benefit = $(E^*t)_{2D} / (E^*t)_{3D}$ **
- Benefits would be enhanced if design targeted at 3DSoC technology

** E=Energy of computation t=execution time of computation

3D SoC Benefits Beyond Deep Learning

Benefits of 3DSoC at same node over 2D Baseline
Energy × execution time for large datasets >50GB

Benefits

80x
60x
40x
20x
0

851x 656x 400x 510x 700x 970x

PageRank
SSSP
Connected Components
BFS
Logistic Regression
Linear Regression

Energy
Execution Time

From Deep Dive & IBM Graph

The benefits of 3D SoC technology extend beyond DNNs to large graph analytic and statistical applications

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

3DSoC Target Characteristics

Taken from the BAA

- Capability of > 50X the performance at power when compared with 7nm 2D CMOS technology.
- Interconnect densities > 3K interconnects/mm and 9M interconnects/mm^2 between 3D layers.
- Interconnect bandwidth > 50Tb/s between 3D layers.
- Memory access energy < 2pJ/bit.
- Inclusion of > 4GB of non-volatile memory in a monolithic SoC that has a 2D footprint of no more than 200mm^2 and dissipates < 500mW of average operating power.
- Provision for logic densities > 1M gates/mm^2 (as measured in a 2D projection) across multiple logic layers in the 3D stack.
  - Note that the process temperatures required for the 3D layers must be low enough not to compromise earlier logic or memory layers in the 3D stack.

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

3DSoC Target Characteristics (Continued)

Taken from the BAA

- 3D transistor performance approximately equal to or to exceed that of standard 90nm CMOS transistors:
  - NFET: I_{on}=640μA/μm at I_{off}=5nA/μm, f_t=100GHz, Subthreshold Slope=90mV/dec.
  - PFET: I_{on}=280μA/μm at I_{off}=5nA/μm, Subthreshold Slope=90mV/dec.
- Projected fabrication cost comparable to 7nm 2D technology.
- Capability to fabricate, at reasonable yield, 3DSoCs with > 4GB (Giga-Byte) of memory and > 50M logic gates.
- Capability to be fabricated as an on-demand foundry service for multiple DoD and commercial customers using a U.S. fabrication facility.

Program Structure

Taken from the BAA

- Technical Area 1 (TA-1) will develop the 3D monolithic manufacturing process that will be utilized to build DoD and commercial SOCs.
  - End goal of this technical area is to be prepared to offer the 3D monolithic process as a commercial foundry offering to external users, including defense contractors, for fabrication of relevant large-scale compute/memory systems.
- Technical Area 2 (TA-2) will focus on developing a test chip (DEC), including an SoC to facilitate process development and drive yield improvement.
  - End goal of this technical area is to develop a test chip with test structures and an SoC that will facilitate process development, drive yield improvement, and exercise the design elements required to design a 3DSoC
- Technical Area 3 (TA-3) will develop the EDA design tools required to design large-scale compute/memory systems that utilize 3DSoC technology.
  - End goal of this technical area is develop the EDA design tools required to enable design of large-scale compute/memory systems that utilize 3D monolithic processes to achieve the performance at power goals of the program

Technical Area – 1 Milestones

Taken from the BAA

Task	Phase 1	Phase 2	Phase 3
Process Technology	1. Process module description 2. Initial integrated process description	1. Demonstration of integrated process flow 2. Cost of fabrication model for the 3DSoC technology	1. Utilization of integrated, high-yielding process flow to successfully fabricate multiple 3DSoC designs
Yield Improvement	1. Yield improvement plan in place 2. Measurement test plan in place	1. Successful fabrication of Development Evaluation Circuit (DEC) 2. Successful fabrication of initial 3DSoC demonstration circuits	1. Successful fabrication of multiple 3DSoC designs
Technology Enablement	1. Delivery of a V0.5 (Beta) PDK that is hardware-based and includes at a minimum; a. Transistor and memory cell models b. Parasitic extraction models c. Design rules	1. Delivery of a V1.0 (Design Ready) PDK that is hardware-verified 2. Development of the enablement infrastructure required for initial 3DSoC demonstration circuit design	1. Refinement of PDK and other enablement infrastructure
Foundry Service	1. Prepare for MPW runs in Phase 2	1. Establish the infrastructure required for MPW participation 2. Schedule and fabricate 2 MPW runs	1. Refinement of MPW infrastructure 2. Schedule and fabricate 2 MPW runs
System Performance Evaluation	1. Provide hardware performance comparison based on 3DSoC simulations	1. Provide hardware performance comparison based on 3DSoC test chip (DEC) hardware	1. Provide hardware performance comparison based on 3DSoC demonstration hardware

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Technical Area – 1 Metrics

Taken from the BAA

Task	Phase 1	Phase 2	Phase 3
Process Technology	1. >90% of all process modules demonstrated 2. Initial integrated fabrication lots complete	1. 100% of all process modules demonstrated 2. Full integrated process demonstrated	1. Full integrated process demonstrated on 3DSoC designs
Yield Improvement	1. > 60% yield on representaive 3DSoC circuit building blocks 2. Sufficient yield on test structures to predict 30% yield at final DEC circuit complexity	1. > 30% yield on the Development Evaluation Circuit (DEC) 2. Successful fabrication of initial 3DSoC demonstration circuits	1. > 60% yield on the Development Evaluation Circuit (DEC) 2. > 30% yield on initial 3DSoC demonstration circuits
Technology Enablement	1. < 10% difference between measured test structure results and PDK prediction	1. < 5% difference between measured test structure results and PDK prediction	1. < 2% difference between measured test structure results and PDK prediction
Foundry Service	1. Full MPW infrastructure in place	1. < 6 month MPW turn-around time from design delivery to chip delivery	1. < 4 month MPW turn-around time from design delivery to chip delivery
System Performance Evaluation *	1. > 50X compute/memory performance at power improvement over 2D 7nm technology predicted through simulation	1. > 10X compute/memory performance at power improvement over 2D 7nm technology demonstrated using the DEC	1. > 50X compute/memory performance at power improvement over 2D 7nm technology demonstrated with 3DSoC demonstration circuits

Technical Area -2

Taken from the BAA

Task/Metric	Phase 1	Phase 2
DEC Design Metrics	1. > 100MB of memory on 3 or more 3DSoC layers 2. > 20M gates of logic on 3 or more 3DSoC layers 3. At least 1 microprocessor core of the complexity of a RISC-V or ARM Axx 4. Implementation of a complex logic function such as DSP or DNN	1. > 1GB of memory on 3 or more 3DSoC layers 2. > 50M gates of logic on 3 or more 3DSoC layers 3. At least 1 microprocessor core of the complexity of a RISC-V or ARM Axx 4. Implementation of a complex logic function such as DSP or DNN
DEC Design Task	1. Complete initial DEC design using the V0.1 (initial) PDK	1. Complete DEC design using the V0.5 (beta) PDK

Technical Area -3 Milestones

Taken from the BAA

Task	Phase 1	Phase 2	Phase 3
3DSoC Logic EDA Tool Development	1. Define an EDA flow that will place and route logic monolithically across 3DSoC layers 2. Distribute the EDA tools to a single sophisticated user	1. Provide an EDA flow that can be used by multiple external 3DSoC design teams	1. Provide an EDA flow that is generally available for external use, specifically by DoD users
3DSoC Memory EDA Tool Development	1. Define an EDA flow that will compile, place, and connect memories with logic and across 3DSoC layers 2. Distribute the EDA tools to a single sophisticated user	1. Provide an EDA flow that can be used by multiple external 3DSoC design teams	1. Provide an EDA flow that is generally available for external use, specifically by DoD users
Monolithic 3DSoC Design EDA Tool Development	1. Define an EDA flow that will design an entire 3DSoC that spans across the entire 3DSoC stack 2. Distribute the EDA tools to a single sophisticated user	1. Provide an EDA flow that can be used by multiple external 3DSoC design teams	1. Provide an EDA flow that is generally available for external use, specifically by DoD users

Technical Area -3 Metrics

Taken from the BAA

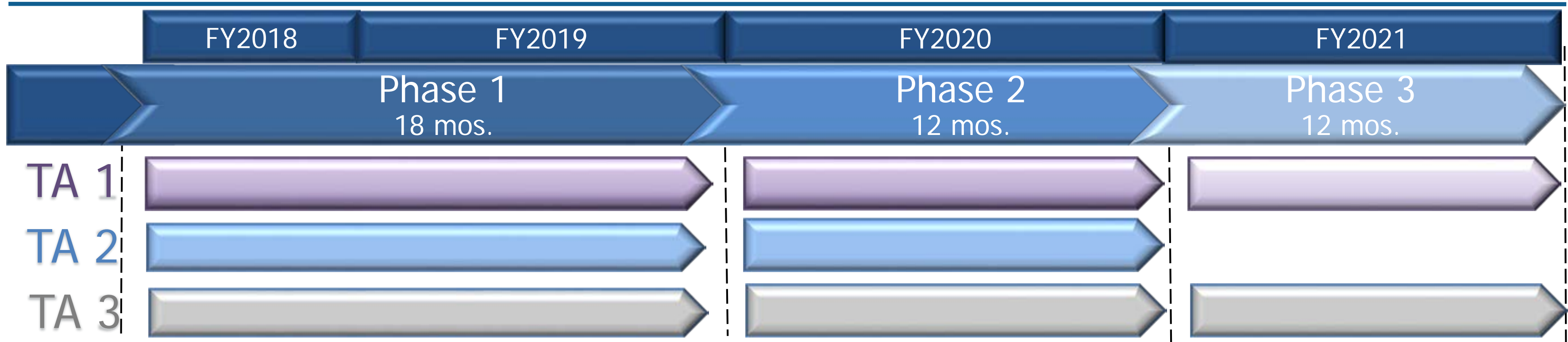
Task	Phase 1	Phase 2	Phase 3
3DSoC Logic EDA Tool Development	EDA flow capable of designing the DEC chip	EDA flow capable of designing a 500M gate 3DSoC design	EDA flow that is sufficiently robust to be used by multiple external users
3DSoC Memory EDA Tool Development	EDA flow capable of designing the DEC chip	EDA flow capable of designing 4GB of memory in a 3DSoC design	EDA flow that is sufficiently robust to be used by multiple external users
Monolithic 3DSoC Design EDA Tool Development	EDA flow capable of designing the DEC chip	EDA flow capable of designing a full 3DSoC design with 4GB of memory and 500M logic gates	EDA flow that is sufficiently robust to be used by multiple external users

Out-of-Scope Technical Areas

Taken from the BAA

1. Extensions of 2D CMOS technology
2. Packaging technologies that do not meet the interconnect bandwidth and density goals of the program
3. Sensor fusion

3DSoC Program Schedule



Phase 1 Outcomes

- Initial 3DSoC process defined
- PDK V0.5 defined
- 3DSoC technology benefits simulated
- First pass DEC fabricated and tested
- Initial 3DSoC EDA tools released

Phase 2 Outcomes

- 3DSoC process demonstrated
- PDK V1.0 released for design
- 3DSoC benefits demonstrated
- Final DEC design fabricated
- 3DSoC EDA tools released for targeted designs

Phase 3 Outcomes

- 3DSoC process used for external designs
- Final PDK released for design
- MPW runs successfully yielded
- 3DSoC EDA tools released for general designs

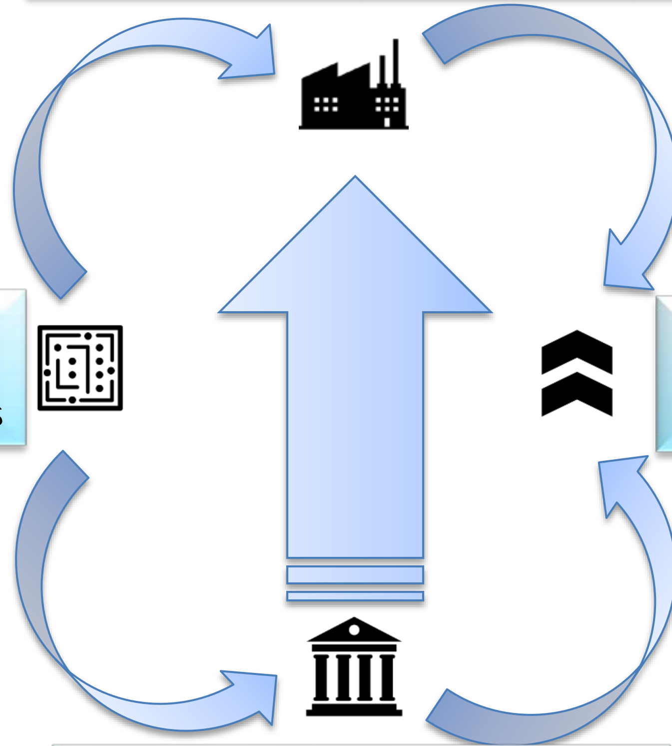
- TA-1: Developing the 3DSoC fabrication process
  - Establish unit processes and flow integration
  - Define the 3DSoC technology PDK
- TA-2: Designing and Implementing the DEC
  - Design 1st and 2nd pass DEC design
  - Foster use of the DEC to drive development and yield
- TA-3: Developing the 3DSoC EDA design flow
  - Develop EDA tools for 3DSoC compute/memory designs
  - Support tools for advanced 3DSoC designs

Metric	Goal
3DSoC Capability	> 50X 7nm 2D PaP
Hardware Accuracy	< 2% deviation from3DSoC technology targets
Yield	> 30% for full 3DSoC designs
EDA Tools	Successful use of EDA flow for a > 500M gate/4GB memory design

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Need for Close Collaboration

Commercial SoC Manufacturers
Manufacturing process development
Process integration at high complexity



CAD Companies
Novel SoC design tool development
Design assistance to Defense teams

Defense Industry
Knowledge of defense applications
3DSoC demonstrations

Academia
Fundamental process development
Initial SoC simulations

Providing a DoD and national manufacturing capability will require a strong cooperative effort between commercial SoC manufacturers, academia, commercial CAD companies, and defense contractors.

Other Items

Taken from the BAA

Proposal Schedule

- BAA posting date: September 13, 2017
- FAQ submission deadline: October 23, 2017 at 1:00 PM
- Proposal due date: November 6, 2017 at 1:00 PM
- Estimated period of performance start: April, 2018

Proposers may propose to:

- TA-1 and TA-2, or
- TA-3, or
- TA-1, TA-2, and TA-3

Cost Sharing

- Cost sharing is not required; however, it will be carefully considered where there is an applicable statutory condition relating to the selected funding instrument. Cost sharing is encouraged where there is a reasonable probability of a potential commercial application related to the proposed research and development effort.

Other Transaction Agreements (OTA)

- All proposers requesting an OT must include a detailed list of payment milestones. Each payment milestone must include the following:
  - Milestone description (Do not include proprietary data)
  - Completion criteria (Do not include proprietary data)
  - Due date
  - Dollar Amount (to include, if cost share is proposed, awardee and Government share)

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Evaluation Criteria

Taken from the BAA

1. Overall Scientific and Technical Merit

   The proposed technical approach is innovative, feasible, achievable, and complete. The proposed technical team has the expertise and experience to accomplish the proposed tasks. The proposal identifies major technical risks and planned mitigation efforts are clearly defined and feasible.

2. Potential Contribution and Relevance to the DARPA Mission

   The potential contributions of the proposed effort are relevant to the national technology base. The proposed transition plan clearly demonstrates a comprehensive understanding of competing technical approaches on the market and how this new technology is expected to compare.

3. Impact on the Overall Electronics Landscape

   The technology represents a radical departure from current trends in the electronics industry to broadly impact the semiconductor industry in the 2025 to 2030 timeframe.

4. Cost Realism

   The proposed costs are realistic for the technical and management approach.

Proposal Components

Taken from the BAA

A. **Executive Summary (no more than 2 pages)**
Summarize the technical approach, and the technology transition plan.

B. **Technical Approach**
This section is the centerpieces of the proposal and should succinctly summarize the innovative claims for the proposed research.

C. **Statement of Work (SOW)**
In plain English, clearly define the technical tasks/subtasks to be performed, their durations, and dependencies among them.

D. **Results and Technology Transfer**
Proposers should describe their initial hypothesis on how this technology will transition from lab to market and DoD impact.

E. **Ongoing Research (no more than 2 pages)**
Comparison with other ongoing research.

F. **Proposer Accomplishments (no more than 2 pages)**
Proposer's previous accomplishments and work in closely related research areas.

G. **Facilities (no more than 1 page)**
Description of the facilities that would be used for the proposed effort.

H. **Teaming (no more than 1 page)**
Description of the formal teaming agreements.

I. **Schedules and measurable milestones**
Schedules and measurable milestones for the proposed research.

www.darpa.mil