Intel Labs Academic Research Office (ARO)

GRANT PROPOSAL COVERSHEET

Proposal Title:			
3D SPHINX : A Stackable Processor using Heterogeneous Integration and Execution			
University Name / Receiving Organization		Department / Discipline	
Georgia Institute of Technology		Electrical and Computer Engineering Chemical and Biomolecular Engineering	
Representative Authorized To Conduct Grant Administration		Principal Investigator Information	
Contact Name	Siri Melkote	PI Name	Hsien-Hsin S. Lee (ECE)
Mailing Address	777 Atlantic Dr., School of ECE, Georgia Institute of Technology, Atlanta, GA 30332	Mailing Address	266 Ferst Dr., School of ECE, Georgia Institute of Technology, Atlanta, GA 30332
Phone #	(404) 894-7337	Phone #	(404) 894-9483
Fax #	(404) 894-8041	Fax #	(404) 385-3137
E-Mail Address	siri.melkote@ece.gatech.edu	E-Mail Address	leehs@gatech.edu
		Project/PI URL	http://www.ece.gatech.edu/~leehs
		Co-Investigators	Muhannad Bakir (ECE) Abhijit Chatterjee (ECE) Paul A. Kohl (ChBE) Sung Kyu Lim (ECE) Saibal Mukhopadhyay (ECE) Madhavan Swaminathan (ECE) Sudhakar Yalamanchili (ECE)
Amount of Cash Requested	USD $499,041 (year 1) USD $734,804 (year 2) USD $724,984 (year 3)	Includes Overhead? Y/N	No.
Additional Comments			

Response to
Intel RFP on Prototype TSV-based 3-D IC High Performance CPUs Program

3D SPHINX: A Stackable Processor using Heterogeneous Integration and Execution

PI: Hsien-Hsin S. Lee

Co-PIs: Muhannad Bakir
Abhijit Chatterjee
Paul A. Kohl
Sung Kyu Lim
Saibal Mukhopadhyay
Madhavan Swaminathan
Sudhakar Yalamanchili

College of Engineering
Georgia Institute of Technology

Contact Information:
Hsien-Hsin S. Lee
School of Electrical and Computer Engineering
Georgia Institute of Technology,
Atlanta, GA 30332
Email: leehs@gatech.edu
Tel: 1-404-894-9483
FAX: 1-404-385-3137

Total Fund Requested:
USD $499,041 (Year 1)
USD $734,804 (Year 2)
USD $724,984 (Year 3)

3D SPHINX : A Stackable Processor using Heterogeneous INtegration and eXecution

1 Executive Summary

The evolution of 3-D die stacking technology has enabled significant advances in capabilities for process, interconnect, materials, packaging, etc. The industry has now reached a point where product development cycles using 3-D ICs must be aligned with process and design cycle changes. Intel's tick-tock model has provided such a structure for product development of 2D ICs. The proposed program seeks to extend this to 3-D ICs by optimizing the heterogeneous architecture, memory hierarchy, design, interconnect, and testing concepts for 3-D ICs. This will (i) enable the localization of the process and design changes to a single die, and (ii) thereby provide a focused and cost-effective targeting of process and design changes to 3-D subsystems. This would favorably alter the economics of 3-D platforms.

Our technical approach is to formulate several system-level concepts for modular design and a 3-D design process that takes advantage of them. The major system concepts encompassed here include (i) an architecture-independent, scalable data communication system, (ii) a novel modular clock network design for per-die clock domains, (iii) a modular power delivery network, (iv) design for test techniques, and (v) advanced in-house 3-D and packaging prototyping expertise and capabilities. These concepts are supported by a comprehensive suite of measurement and characterization circuits for understanding the consequences of our design decisions, as well as supported by an in-house TSV-based assembly, package substrate fabrication, and die stacking capabilities.

Our objective is to demonstrate the above-stated 3-D system concepts by fabricating, testing, packaging, and evaluating two 3-die stacks over the course of this program. A stack will be comprised of two compute die and one memory die. The compute die will each be comprised of two cores (LEON), two accelerator units, and a shared L2 cache bank. The architecture of the prototype is shown in Figure 1. As described later in this proposal, the prototype will test multiple different system concepts and is not intended to necessarily correspond to a typical deployable configuration. The bulk of the system design effort will be devoted to version 1 (V1) that demonstrates the ability to integrate die from multiple vendors - compute and memory layers are fabricated through different vendors. Version (V2) demonstrates the heterogeneous stacking where the design is the same as V1, but the memory layer die from V1 goes through a process shrink and is implemented in a different CMOS node. The two compute layers feature an interleaved layout of two LEON cores with two architecturally distinct accelerators as shown in Figure 1. The core-accelerator interface is standardized across both core-accelerator pairs to demonstrate the ability to compose architecturally distinct designs (of accelerators) across the same inter-die interface. Moreover, V2 will incorporate advanced prototyping processes and capabilities including modified package and post-CMOS TSV integration at Georgia Tech. The 3-tier design is supported by the necessary important innovations in circuit design, test, and fabrication technologies. In doing so, we seek to validate the proposed concepts of standardized interfaces for cross-generational, system-level integration in heterogeneous 3-D systems.

Collectively, our objective is to demonstrate that our approach to 3-D system design is (i) extensible across multiple tiers, (ii) compatible with future goals of using legacy modules across products, and (iii) greatly improves the ability to compose die from multiple design and process points with consequent productivity and efficiency gains. We make this proposal based on our previous success and existing research programs in each of the relevant areas.

2 Relevance and Impact

This program proposes to design, fabricate, test, and evaluate a three-tier stacked IC prototype called 3D SPHINX or *3D-integrated Stackable Processor using Heterogeneous INtegration and eXecution*. The 3D SPHINX prototype is unique in (i) its composition of die from multiple vendors, (ii) the integration of die at two distinct technology nodes, and iii) composition of two architecturally distinct designs across the same inter-die interface. As illustrated in Figure 1, the 3D SPHINX prototype is a single-package 3-D heterogeneous multi-core system. The bottom two die form the compute tiers and the top die is the memory tier. The proposed program will construct two prototypes. Version 1 (V1) demonstrates the ability to integrate die from multiple vendors — compute and memory tiers are fabricated through different vendors. Version (V2) demonstrates the heterogeneous stacking where the design is the same as V1, but the memory tier die from V1 goes through a process shrink. The two compute tiers feature an interleaved layout of two LEON cores with two architecturally distinct accelerators as shown in Figure 1. The core-accelerator interface is standardized across both distinct core-accelerator pairs to demonstrate the ability to compose architecturally distinct designs (of accelerators) across the same inter-die interface. The 3-tier design is supported by the necessary important innovations in circuit design, test, and fabrication technologies.

Collectively, the two prototype versions demonstrate the issues (and solutions pursued here) for aligning 3-D IC product development cycles with process and design changes as is currently made available in Intel's tick-tock model. Process and design changes can be localized to a die in the stack with major benefits in optimizing the non-recurring-engineering (NRE) costs of 3-D ICs. Importantly, prototype packaging and construction is all done in house and therefore in close collaboration with designers and architects.

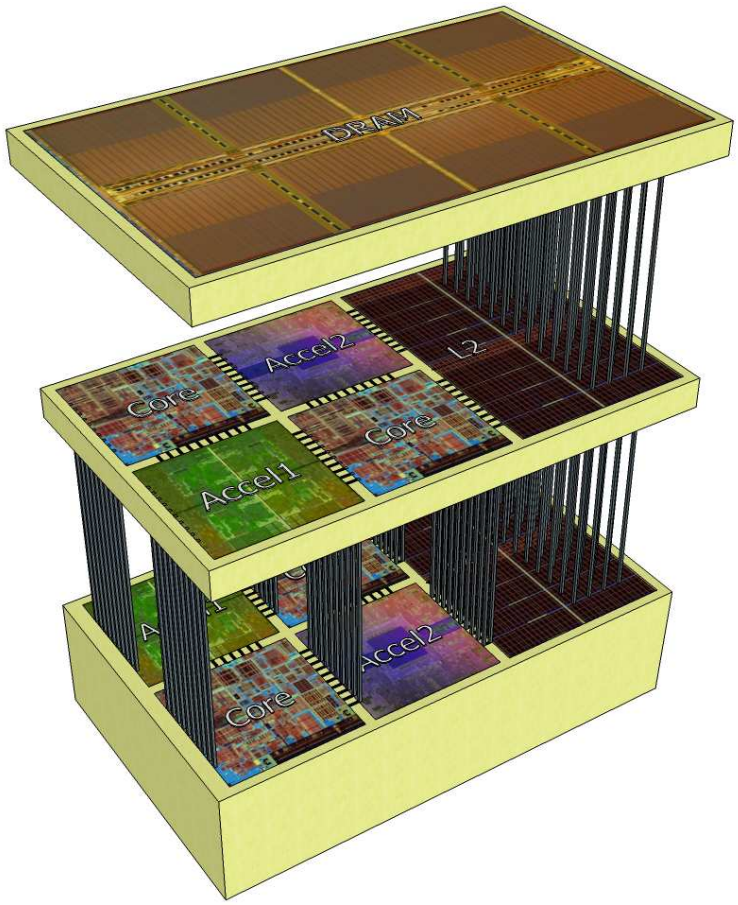


Figure 1: 3D SPHINX Architecture

Localization of design and process changes to each die requires *highly modular design*. Architecturally, the 3D SPHINX prototype seeks to demonstrate multiple extensible stacking concepts. The first is with respect to the cache architecture where the L2 cache is extensible across multiple die. The second is with respect to the use of custom cores (equivalently accelerators) and their composition with cores across die. Such a composition of cache and accelerators across die is made feasible by a configurable packet-based interface. Physical composition of multiple die is demonstrated in-house at Georgia Tech as described in this proposal.

From a design and test perspective, modularity requires innovation across (i) power delivery network (PDN), (ii) clock distribution network (CDN), and (iii) design for test (pre- and post bond). In 3D SPHINX, PDNs and CDNs are designed as independent domains in each die that are in turn extensible across die. For example, each tier has an independent 2D PDN derived from a global 3D PDN using integrated voltage converter and regulator modules (VRM). Each tier is an independent clock domain and the multicore architecture utilizes a standardized interface to a packet-based interconnect that operates across tiers via asynchronous links.

Quantitatively 3D SPHINX is targeted to achieve the goals of 2X to 3X package footprint reduction and 15% to 25% performance improvement over a 2D implementation. To do so, we leverage and integrate several technologies into our 3D SPHINX at the architectural level, circuit level, and package level. The major innovative claims are the following.

- The 3D SPHINX prototype demonstrates stackable, scalable, multicore design concepts across heterogeneous die.
- The prototypes employ a variable-width asynchronous communication link design (to enable GALS across 3-D tiers) and standardized interface (across components) that can be composed to realize an inter-die packet based network.
- The proposed prototype design and fabrication activities of the three-tiered 3D SPHINX are aimed at demonstrating the feasibility of the *tick-tock* model employed by Intel. The outcome of this work will provide deep insights into understanding the feasible integration and practical business model issues in 3-D technology, in particular, to improve the NRE cost of chip design, performance, and production yield by adopting heterogeneous 3-D die stacking. The establishment of standard interfaces would allow the stepwise integration of technology innovations to achieve product advances.
- Prototyping activities for a three-tiered die stack will be performed in two phases corresponding to a two-generation tick cycle to implement the technical challenges of heterogeneous die integration, thereby demonstrating the viability of scalable die-stacking technology without major architectural and design changes for future heterogeneous 3-D systems. In addition, our two-in-a-box core-accelerator design strategy will implicitly demonstrate the tock model for integrating heterogeneous architectures.
- The prototype demonstrations of 3-D stacks using die designed from different vendors using disparate CMOS nodes will create new understanding of the 3-D advantages and new challenges of a true, heterogeneous 3-D system.
- The prototype will demonstrate circuit techniques for heterogeneous 3D stacks with different tiers designed in disparate CMOS technologies, having varying V_{dd}/frequency requirements, and operating under different supply noise and thermal conditions. The techniques will include independently controlled power delivery network per-tier with integrated voltage regulator and distributed decap; tunable (with on-chip VCO/DLL) clock delivery network per-tier; and 3D interface circuits for face-to-face vias and back-side TSVs supporting communications across clock domains. Special emphasis will be given to reduce the usage of 3D vias considering their limited yield.
- The prototype will demonstrate methodologies and design-for-test circuits to incrementally test each tier and interface in a heterogeneous 3-D die stack through all stages of the die-attach and package assembly process using a combination of on-chip BIST/DFT and probing (both contact and non-contact) techniques. The detection and diagnosis of timing margins and thermal/supply reliability can

further drive V_{dd} and clock tuning mechanisms thereby maximizing system reliability, and stack yield for heterogeneous 3-D systems.

- The measurement of the PDN impedance in the frequency domain from the probe pads on the die and package will provide insight into their interactions and also provide information into the resonances occurring in the power delivery network. This characterization will enable the validation of tools used for power delivery in 3-D stacked ICs with TSVs for the first time. In addition, this characterization will provide insight into the variability of the voltage droops on each tier of the die stack based on switching activity that can enable the design of future 3-D heterogeneous systems.

Overall, the success of this project will lay the foundation for designing emerging heterogeneous multi-core processors by using 3-D die-stacking technology and seeks to advance 3-D IC design with prototypes that serve both to educate and as well as advance the state of the practice.

3 Technical Rationale and Approach

3.1 3D SPHINX Architecture Concept

To fulfill and demonstrate the technical requirements described in the RFP and address future challenges for 3D-integrated heterogeneous multi-core architectures, we propose **3D SPHINX**— a 3D-integrated *Stackable Processor using Heterogeneous INtegration and eXecution* — a scalable accelerator-based architecture exploiting heterogeneous integration using 3-D die stacking technology. In this implementation, we introduce several extensible design concepts into one single architecture including stacking die tiers fabricated with disparate process technology nodes, integration of heterogeneous execution models using accelerators with standardized interface, and allowing more die to be stacked as a scalable solution in the future.

Figure 1 illustrates the 3D architectural diagram of our proposed 3D SPHINX multi-core architecture. The prototype we plan to design and fabricate will consist of three die tiers including two compute tiers shown at the bottom stacked with one memory tier on the top. The architecture is partitioned in a modular manner in order to enable future stacking of additional compute tiers and/or memory tiers. The following sections will detail the architectural partitioning in 3-D, inter-tier interface, and how does the 3D SPHINX prototype meets the design objectives.

3.1.1 Compute and Memory Tiers

The 3D SPHINX architecture contains two different device tiers: compute tier and memory tier. As depicted in Figure 1, each compute tier is composed of two general-purpose processor cores, two specially tailored accelerators of different functionality, and at least one bank of the level 2 cache. We call the cores and accelerators the *compute plane*. The rest of the memory components are generally referred to as the *uncore* part. As can be seen, the design of the compute tiers are almost symmetric except that their floorplans are different to demonstrate heterogeneous execution to be discussed later in Section 3.1.3. The cores and accelerators across two tiers are intentionally placed in an interleaved pattern. As such, each processor core is paired up with one particular type of accelerator vertically through inter-die vias (assuming face-to-face bonding for these two die tiers). This alternated core-accelerator placement in Figure 1 is the outcome of the 3-D thermal density consideration related to our overall cooling strategy. We assume that the processor core will become idle when its vertically-paired accelerator carries out the computation and vice versa. Therefore, the core and its corresponding accelerator of each vertical pair will not be fully activated simultaneously, thereby reducing both the power and thermal density.

Another consideration of this interleaved placement is with respect to the extensibility of the compute plane. As shown, the 2-D compute plane is organized in an N-by-M grid (N=M=2 is shown) with interleaved cores and accelerators. The same interleaving placement can be extended regularly to future, larger compute plane as N and M are increased when larger die footprint is available as the fabrication process continues to shrink. This layout extends naturally to shared ISA architectures where cores are concurrently active, but the less complex cores (lower power) are similarly strategically placed to manage thermal densities.

For our processor core, we propose to use the synthesizable LEON open core (SPARC V8) from Gaisler Research. We plan to explore different types of accelerators using Berkeley's drawfs model [3] as guidance likely resulting in an ASIC core for typical data-parallel media applications. We will also consider and evaluate the possibility of implementing a more broad-purpose co-processor [35] for our accelerator. The cores, accelerators, and L2 banks will be connected with a 3-D ring network and data coherence will be maintained across the ring. This interconnect structure and the link operation upon which it is based is discussed in the next section.

3.1.2 Interconnection Model and Scalable Interface

The provision of a scalable (size), extensible (across technology generations), adaptable (across distinct IP modules) interconnect fabric requires (i) a standardized interface, (ii) a supporting point-to-point switched network, and (iii) the ability to configure both.

Standardized Module Interface: We propose to develop a standardized interface for all interacting inter-die modules. We have some prior experience with (and past consortium membership in) the Open Core Protocol (OCP) — an industry standard specification for interfacing IP blocks [1] – and plan to build

on that experience. This standardized interface will connect IP blocks such as the LEON core to the 3D SPHINX network and the accelerators. Our design approach will be to start with the base interface protocol based on read/write memory transactions and global address space across the stack. Adapters will translate the interface transactions to 3D SPHINX network packets extending transactions across multiple die and modules on each die. This translation overhead in the network adapters is the price of design re-use and increased design productivity.

Inter-die Network:. The inter-die interconnect is based on a unidirectional, point-to-point, packet-based link. Each link operates asynchronously enabling clock-domain crossings and the packet-based operation enables composition of links to form networks across multiple die. The link model is influenced by several industry standards [1, 6, 28, 29] and is comprised of of three groups of signals: data, control and optional test signals. The control signals are influenced by the *cmd* signals in the OCP protocol as well as their extensions to tags and threads to enable high performance communication. These optimizations support the ease of integration of architecturally distinct components within and across the die. Finally, a channel can be constructed with two links traversing opposite directions. The main attributes of this link are:

- **Variable Width:** Link widths are negotiated at power-up and established in byte-wide increments. This is a well known technique employed in several high speed networking standards.
- **Asynchronous Operation:** Multiple die may operate at different clock speeds. Links are asynchronous with the necessary buffering and synchronization logic for clock domain crossings. Globally asynchronous locally synchronous (GALS) operation will be the operating principle across the stack.
- **Configurable Link Layer Protocol:** All inter-die links use the same physical layer (PHY) protocol. The physical interfaces to the PHYs are intended to be configurable (for future re-use) to enable use with different link layer protocols. For example, the L2 cache interbank network can use the same physical layer as the core-to-accelerator interface. This will make it easier for third-party die to be designed to fit into an existing 3-D platform advancing interoperability across design generations (a "tock" step). Robustness is achieved with link level CRC and automated retry across a link. This program will demonstrate only one link layer protocol.

The packet-based communication enables scalable communication across die. In this prototype we restrict ourselves to 3-D ring topologies for both intra-die and inter-die communication although the packet design will support more complex switched 3-D topologies. Ring topologies capture the necessary components of 3-D scalability with simple, fast routers with ordering properties that simplify coherence and consistency management. The use of a configurable link layer enables the use of future *optional* extensions and optimizations, e.g., virtual channels to preserve ordering and message priorities. In this project, our goal is to keep the physical design as simple as possible, while demonstrating design room for such higher level (microarchitecture) optimizations that can be subsequently explored.

3.1.3 Design Objectives

- **Performance.** To achieve the targeted 15% to 25% performance improvement over a 2-D counterpart design, the 3D SPHINX will exploit the following key benefits enabled by 3D integration: (i) stacking memory directly on processing logic, (ii) reduced wire length using structure folding, and (iii) high-speed and high-density inter-die vias. By simply stacking system memory on top of the logic and integrating the memory controller on die [5, 20], the average memory access latency can be substantially reduced leading to overall performance improvement. To take further advantage of 3-D, structure folding which reduces signal drive can be implemented. In particular, as depicted in Figure 1, we propose to bank-partition the L2 cache across die tiers in our 3D SPHINX architecture. Such a 3-D design strategy partitions an SRAM cache array by folding the bank (also called bank stacking [27, 30]) and cuts down the wire RC delay of a 2-D layout of cache banks. This reduction is mostly due to the small vertical lengths of TSVs, and therefore data transfer between any two tiers in a 3-D design can be achieved within one single hop as studied in [14, 18]. It will be feasible to design a "TSV elevator" which can transport data from the top memory tier to any compute tier below it, and

vice versa, within one clock cycle for the 3D SPHINX contingent on the capacitive load of our TSV. In contrast, the same access needs to traverse across multiple cache banks, I/O pads, off-chip memory controllers, and finally memory in a 2-D system design, inevitably increasing the latency. As shown in [27], folding an SRAM structure can effectively reduce the wire length and improve the access latency for 3-D caches. Moreover, such a partitioning scheme could also reduce the area overhead required for a multi-ported cache by employing smaller 3-D banks.

Another 3-D feature to exploit for performance is the via density. It is known that off-chip memory bandwidth will eventually be limited by the available package pin count and its corresponding I/O power consumption. By exploiting 3-D TSVs, this limitation can easily be circumvented due to their much higher density. More than 2x performance gain has been demonstrated in [36] by re-designing the DRAM interface and using a wide TSV bus to fetch data at a much larger granularity than the size of a regular cache line. The design essentially leverages the basic property of (i) TSV bandwidth to eliminate trailing-edge effect and (ii) short TSV depth to transfer data at very high speed. We will investigate the trade-off among TSV density, area overhead, and design alteration for the 3D SPHINX to achieve our performance goal. On the other hand, the same wide TSV bus design can be applied to the interface between each processor core and its corresponding accelerator across die tiers to further expedite the data transfer during the acceleration mode [11, 35].

- **Tick Model Demonstration.** To facilitate Intel's *tick model* allowing disparate process technologies to be independently applied to each die tier for a 3D-stacked design, our 3D SPHINX will demonstrate this viability with two features in our prototype. Both require a *scalable interface design* across die tiers (Section 3.1.2). The first demonstration is via the memory interface standardization, in particular, the system memory and each L2 bank placed on each individual compute tier. Our goal is to define the interface to enable seamless stacking of future system memory tiers, either fabricated using process shrink or new memory technologies (*e.g.,* PCM or STT-RAM) with no design change. For this project, we plan to demonstrate two designs by stacking memory tiers fabricated at two process nodes. Section 3.4 will elaborate the details.

  Secondly, we will demonstrate the tick model across the compute tiers in 3D SPHINX using a globally asynchronous and locally synchronous (GALS) execution model. In our 3-tier prototype, the bottom two compute tiers will be designed and clocked at different frequencies to demonstrate heterogeneous integration. The heterogeneity could either come from die tiers fabricated by disparate process technologies or from two speed bins. Such integration will increase the overall post-bond yield by allowing stacking of compute tiers running at different speeds. Similarly, to enable this integration style we require a uniform and scalable interface design to synchronize the communication between the processor core and the accelerator, in particular, synchronization buffers and different drivers need to be implemented to synchronize communication between two compute tiers (Section 3.1.2.

- **Tock Model Demonstration.** For Intel's *tock model* demonstration, the 3D SPHINX prototype uses two heterogeneous accelerator designs, each stacked with a common core ISA placed in an interleaved fashion in 3-D. This two-in-a-box design choice is aimed at inherently showcasing the tock model with architectural changes without taping out another brand new design. Again, to be able to stack different types of accelerators on top of each processor core, the interface should be defined scalable and broad-purpose to facilitate different types of accelerator integration. Operationally, the acceleration function can be treated like an API call or an ISA plug-in employed in DSP/embedded systems. The standardized interface for these accelerators includes commands and data for architectural scalability. Our goal is to investigate and design such a unified interface for the two accelerators in the 3D SPHINX prototype in order to implicitly realize the tock model within one single design.

3.2 Three-Tier Design for 3D SPHINX

Our 3D SPHINX architecture contains three tiers, the first two based on a Tezzaron 2-tier IC, and the second based on a third party 2-D IC. These two are then bonded with TSVs by our assembly team in Georgia Tech. We will use semi-custom design process for our 3-tier design. 2-D CAD tools available from Cadence, Synopsis, and Mentor will be used for synthesis and layout of the logic blocks (core, accelerator,

and network). SRAM arrays in the 2-tier IC and the 3rd tier will be designed using custom components (including custom cells if memory compilers are not available). We further have the unique advantage of using in-house 3-D CAD tools, specifically developed for 2-tier 3-D IC design based on these vendor tools and enhanced with custom point tools to handle TSVs and the 2-tier stack [4,8,12,13,15,21,25,38,40]. The CAD tool chain will be enhanced (under existing projects) to support the custom circuits (described later) and design-for-testability (DFT) components. We will build a bridge between our chip-level CAD tools and package-level tools (from CST, Agilent, Ansys and in house custom tools) to co-design 3D SPHINX and its package for better power and signal delivery and for sign-off. Section 5 (Fabrication and Assembly Facilities) will provide further details on our chip and package level CAD tool capabilities. The custom circuit design efforts, discussed here, will focus on three principal components of our prototype: (i) power delivery, (ii) clocking, and (iii) 3-D interface circuits. The other custom design efforts are related to DFT circuits and discussed in Section 3.3.

3.2.1 Design of the Power Delivery Network

In a 3-D stack, the power delivery networks (PDNs) of different tiers have different distance to P/G bumps (or wirebonds) and hence, different R, L, and C characteristics and resonance frequency. Therefore, same architectural blocks (e.g. cores or accelerators in 3D SPHINX) with identical current profile when placed in a different tier will lead to different supply noise profiles (*e.g.*, 2nd or 3rd droop, supply settling time, and IR drops etc.). This is a critical challenge for heterogeneous 3-D structures. Therefore, our first design objective for the PDN network in 3D SPHINX is to ensure that the same component when placed in different tiers experiences a similar noise profile even if the R-L-C property of the physical network changes. The second objective is to *ensure robustness of the PDN of each tier through on-chip voltage regulation instead of using a large number of 3-D vias*. Limiting the number of 3-D vias is important considering their limited yield (specifically for TSVs). Reducing the dependence of power supply robustness on the number of 3-D vias also helps pre-bond testability.

We will design an independent 2-D power delivery network for each tier which will be derived from a global 3-D PDN using integrated voltage converter and regulator modules (VRM) [9, 16, 24, 39]. The independent voltage domains for the tiers help provide power to the tiers during pre-bond test. Further, it also allows different voltage for each tier. The design options to be explored here are either a few complex centralized converters or simpler distributed converters. The control circuits for VRMs will be built to provide options for on-chip voltage scaling. Power gating transistors will be included in the core and memory modules to drive active power management. The power gating transistors will also allow us to control the power demand during testing to limit the Ldi/dt and IR drop. We will use regularly placed distributed decaps [10] in each tier to minimize the voltage droop. The interaction of the on-chip VRM based power delivery network with physical design issues such as number, placement, and sizing of the 3-D P/G vias (both face-to-face 3-D vias and back-side TSVs) will be assessed during the design process. We will consider the interaction of chip-level and package-level PDN for our prototype by quantifying the chip-package resonance and its effect on signal integrity through on-die and package measurements. Two major issues we need to consider associated with the transmission of the signals through the TSVs related to the PDN are: (i) RC effect due to the capacitance and conductance to the P/G TSVs and (ii) slow wave effect causing pulse broadening and coupling to signal TSVs far away even with P/G TSVs interspersed in between. Both these effects will be quantified through measurements to construct models for sign-off. A low impedance path from the voltage regulator on the package to the 3-D stack will be designed considering the role of the die and package in generating the first, second and third droops in the power supply noise.

3.2.2 Clocking for the 3D SPHINX

Similar to PDN design, minimizing the usage of 3-D vias in the clock network is also important from yield and testability standpoint. Further, the use of GALS architecture also allows different tiers of 3D SPHINX to operate in independent clock domains with relaxed constraints on clock skew between the tiers. The asynchronous communication between the tiers provides robustness against across tier skew variations. Our clocking strategy will be to design independent clock networks for each tier. We will use off-chip source and on-chip Voltage-Controlled-Oscillator (VCO) with integrated Delay-Locked-Loop (DLL) based clock

source embedded in each tier. The de-skewing technique will be used while distributing clocks in each tier. Synchronizing the two DLLs will be a problem if we plan to run the tiers at the same frequency (particularly tier 1 and tier 2) and use synchronous clocking. This is because of the unpredictability of the voltage and thermal conditions of each tier. We will evaluate the option of spine based as well as H-tree type distribution architecture. The independent clock tree for each tier will also help testability of individual tiers so that each tier can be tested without other tiers. Separating the power delivery network for the clock buffers will be considered for safer design if pad allocation permits.

3.2.3 3-D Interface Circuits

The 3-D interface in our design includes the micro-bumps in the face-to-face bonded 3-D Tezzaron stack, TSVs and solder-bumps in the connection between Tezzaron die-stack and memory die (3rd tier), and the TSVs or wire-bonds in the final system. Circuit design considerations for the interface will focus on robust and low-latency signal communication. We envision that the signal delay through the face-to-face connections will be minimal and may not require additional considerations. The challenge will be the communication through the back-side TSV and solder-bumps that will connect our Tezzaron stack and MOSIS die. The design trade-off is reliability (use of a large number of redundant TSVs) versus speed (increased capacitance). We also need to consider whether ESD protection circuits will be required at the interfaces to prevent gate damage during pre-bond probing. Further, these 3-D interfaces need to support asynchronous data communication between tiers operating at different frequencies as well and hence, require circuit support for handshaking and storage queues. Level shifters also need to be used across the interface to communicate across voltage domains. Both data and clock can be transferred through the 3-D interface for synchronization. Our plan is to include in-built signal recovery [7] and slew correction to improve signal integrity. Duty-cycle correction may be necessary for the clock. The interface characterization methods (see Section 3.3) will help estimate the safe data rate through the TSVs in our prototype. The capacitive loads of the vias can clearly limit the achievable performance and power dissipation. We plan to explore low-swing signaling to partially alleviate this challenge depending on the resistance of the bonding material. We will evaluate whether a serial 3-D interface (at least for the back-side TSVs) can be used to minimize the number of 3-D vias. Note limiting the number of 3-D interfaces will be beneficial from yield standpoint. We can focus on improving integrity of signal (data and clock) through the TSVs even incorporating complex circuits instead of simply increasing the number of TSVs.

3.3 Testing 3D SPHINX and Design for Testability

3-D stack testing techniques will be employed that allow rapid testing of the die processor/logic/memory elements and vertical through-silicon vias in 3-D chip stacks using DFT/BIST structures located in the stacked die. The incorporated test structures will allow testing of individual (thinned) die before assembly as well as full incremental test of the 3-D chip stack as it is assembled and at each stage of stack assembly. It will be possible to test partially assembled chip stacks as well as fully assembled ones using hierarchical boundary scan based test techniques [31, 32]. Specially designed series/parallel boundary scan structures activate lower level scan cell chains that stimulate incoming and outgoing thru-via signal paths as well as inter-die interconnections while minimizing test time and maximizing test coverage [26]. In addition, due to the inability to close signal paths in partially assembled stacks, the scan mechanisms will be used to activate at-speed BIST procedures to stimulate thru-via stubs individually or in predetermined patterns. In this context, we will demonstrate how the integrity of such (stimulated) through-vias can be determined by picking up and amplifying the resulting signals via capacitively coupled non-contact sensors separated by air dielectric from the through-via stubs. For comparison, we will also use direct probing of exposed-surface vias to perform testing of through-vias for the purpose of via characterization. The purpose of the comparison will be to explore the kinds of via defects that can be detected/diagnosed using capacitively coupled vs. direct probing techniques and the corresponding limitations and benefits. Incremental continuity/resistance testing of through-via power/ground interconnections will be performed using simplified AC tests [34] generated by DFT circuitry and supported by non-contact/contact probing methods. The developed test techniques will be supported by algorithms that minimize incremental test time while providing accurate failure diagnostics.

3.3.1 Characterization of Supply Noise

The supply noise behavior of our prototype system is critical to characterize [2, 33]. We aim to design high-speed (∼100s MHz) but low-resolution ADCs to characterize high-frequency supply variations while low-speed but higher-resolution ones for low-frequency (transient IR drop) variations. The second option will be to use delay based sensors for supply noise characterization to simplify the design. Our main goal is to characterize how the supply noise profile varies from tier to tier and whether noise propagates across tiers. We will perform frequency domain post-processing of the data to understand how 3-D integration modulates the supply noise behavior (*e.g.*, principal frequency of first, second, and third droop, how the resonant frequency changes from tier-to-tier) and whether the supply noise coupling exists across tiers.

3.3.2 Characterization of Die-to-Die Interface

The design of the interface circuits will also include specific circuits for interface characterization. The first objective is to characterize the RC parameters of interfaces. We will exploit our recent work to perform DC test for resistance variation. The resistance variation will be captured in terms voltage and converted to digital output either using ADCs or VCO (and counter). The RC property can be characterized by incorporating the interface either as a capacitive load or as a serial part (both R and C) of the oscillator. We next plan to characterize the maximum data rate that can be reliably sustained across a 3-D interface. The goal will be to drive the interface through specialized ring-oscillators with controllable frequency and monitor (frequency-to-digital-converters) signal at the other end. The interface circuits drive at-speed logic tests (or delay tests) by designing in-built circuits to drive required test vectors at each interface point.

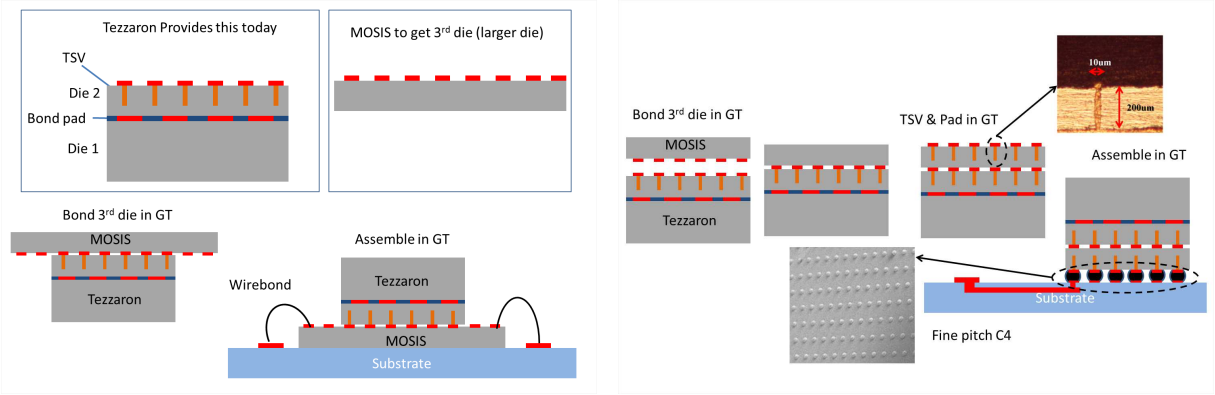
3.3.3 Delay and Temperature Sensor

We also incorporate in-built delay sensor to localize post-fabrication performance bottlenecks to understand what determines the 3-D prototypes performance — interface or compute tiers or memory tiers. The delay sensing will be performed using a replica path approach (with minimal invasion to the original layout of the core). The delay sensors will be correlated to at-speed test to understand whether the physical property or the logical considerations limit the performance. We will also design on-chip temperature sensors [19, 23, 37]. The delay sensor will be extended to temperature sensing. The differential delay sensor (a temperature sensitive and one temperature insensitive delay sensor)can be used to improve the accuracy of temperature sensing. Signal post-processing (low-pass filter) can be used to improve the sensing accuracy.

3.4 3-D Technology Prototyping and Fabrication for 3D SPHINX

Full 3-D prototyping capabilities will be exercised in a step-wise manner in this program with each step enabling to a generational advance in system memory technology and performance. The prototyping will be a combination of in-house routine and innovative processes, and third-party capabilities for those components or processes which are too costly, complex, or time consuming to perform internally given time and budget constraints. The third-party prototyping includes chip fabrication. The internal, Georgia Tech 3-D prototyping will include all aspects of 3-D chip-stacking including TSV interconnect, flip-chip bonding, and assembly. TSVs will be fabricated externally for the logic die since it is a standard part of the Tezzaron process flow. The prototyping innovations we will demonstrate include (i) in-house stacking of third-party CMOS die manufactured at different foundries using different CMOS technology nodes (heterogeneous die integration), (ii) in-house fabrication of TSVs using a post-CMOS process (combining routine fabrication steps and Georgia Tech's innovation), (iii) in-house fabrication of novel, low-capacitance TSVs featuring low-k liners (leading to high-speed and lower energy signal links), (iv) path-finding demonstration of fine pitch chip bonding, and (v) path-finding demonstrations for the integration of novel cooling within the 3-D stack. For the most part, the in-house prototyping innovations [22] have already been created in other programs. We will leverage an enormous amount of existing investment in facilities and expertise, and use the funds from this program to explore the integration for the prototype.

The prototyping activities for the 3-die stack will be performed in two phases corresponding to two tick-tock cycles. We will also outline the generational changes which would occur on additional cycles (which are beyond the time and budget anticipated in this program). The first phase as shown in Figure 2(a),



(a) Packaging V1 (b) Packaging V2

Figure 2: Bonding and Packaging for 3D SPHINX

provides a low-risk, rapid turn-around, 3-die chip stack (V1) to demonstrate the architectural benefits and benchmarking of 3D SPHINX system. The prototype will feature a 2-die stack using the Tezzaron multi-project-wafer (MPW) run with 130nm CMOS technology. The third die will be fabricated separately in the same technology node through a different vendor (MOSIS). The latter is a critical part of the novelty we seek to demonstrate in this proposal, namely, the stacking of CMOS chips formed at different foundries. The three-die stack will have Tezzaron TSVs in one die; these TSVs will be approximately 1.2μm in diameter. In order to form the interconnection between the 2-die Tezzaron stack and the MOSIS die,a face-to-back flip-chip attachment will to performed at Georgia Tech.

The first phase prototyping will demonstrate two essential features: flip-chip technology (for chip-to-chip stacking and stack-to-package assembly), and ability to perform wafer-level processes on single chips. Normally, wafer level processing would be performed to facilitate TSV bonding, and flip-chip operations. Since we will not be able to procure full wafers, we will need to expend extra effort to transform single chips into pseudo-wafers. Individual die will be temporarily mounted into a recess in a silicon wafer so as to make a planar, wafer-like structure which can be handled and processed in the wafer-level format. Flip-chip bonding of the die will use C4 solder bumps. Solder can be evaporated (such as in the original C4 controlled collapse configuration) or electroplated in a plating mold, which is our preference in this program since we will have little control over the chip metallization. Under-bump metallization can be customized to each type of chip, including zincating aluminum I/O pads. The target for the bump pitch will be from 70μm pitch down to sub-10μm in this program. In V1, the first-level interconnect technology between the chip stack and the package substrate is through wire-bond from the third die to the stack substrate. The third die has a larger footprint than the Tezzaron stack allowing access to the bond pads. The stack substrate will be fully designed and fabricated at Georgia Tech. This first phase prototype technology will be adequate for the system power, speed, and cooling requirements, yet demonstrates our ability to stack heterogeneous chips fabricated in disparate technology nodes. It also paves the way for the introduction of additional speed, power, and cooling in future generations, shown in Figure 2(b). The substrate will be fabricated at Georgia Tech using a two-layer, thin-film, inorganic technology (one or two layers of fine-line copper or gold wiring on a silicon substrate) so that the 3-D stack can be powered and tested.

The second version prototype (V2) anticipates higher speed, power, and density connections between the 3-D stack and the substrate. Importantly, the memory die will be fabricated at a different technology node (90nm or 65nm). As depicted in Figure 2(b), post-CMOS TSVs will be fabricated at Georgia Tech in the third die. As was the case in V1, a separate processing jig will be designed to transform the single chip into a wafer-like structure because we will unlikely be able to procure chips in wafer-form due to budget. Depending on the die thickness, we may back-grind the silicon die (or etch it) to thin down the MOSIS die to a thickness that enables TSV processing; sub 200μm thickness is the target. If the die is too thin, we may run into silicon cracking during handling or bonding. Please note that our post-process team routinely processes sub-200μm thick wafers. We have even processed 1μm thick silicon on some occasions (*i.e.*, free-

standing device layer of SOI wafers). Following die thinning, bosch process will be used to etch vias in the back side of the wafer (alignment of features will be made using IR camera). Next, either low-temperature PECVD oxide or room-temperature parylene will be deposited on the sidewalls of the vias as a sidewall liner (dielectric). Next, using a directional plasma etch, the dielectric will be removed from the base of the via to explore underlying chip-on-chip interconnects. A seed layer deposition followed by copper filling occurs next (CMP will be used to polish the overburden). We will also explore electroless copper TSVs as well as annular copper vias (which are easier to process). Target TSV diameter for this demonstration is around 50μm , although we will try as small as 10μm diameters. Demonstrations will be made for higher performance TSV technologies with low-k liners for future high-speed vertical interconnect.

Using processes similar to above, we will next form C4 bumps to bond the MOSIS die to the 2-die Tezzaron stack. Next, the 3-die stack will be flip-chip bonded to an in-house made inorganic substrate (similar to above prototype). The first-level C4 pitch will be on the order of 70μm .

Future phases of process advancements can be envisioned and will be outlined in this program with physical demonstrations. The advances include: higher density, scalable TSVs, advances in the density and layers in the first-level package substrate, and inter-layer cooling. These advances will be planned for 3-D proliferation for higher clock speed, chip power, and cooling needs to coincide with future technology/architecture tick-tock cycles.

3.5 Post-Silicon Validation and Benchmarking for 3D SPHINX

After the 3D SPHINX silicon stack is returned, we will then perform post-silicon validation and execute benchmark programs for the 3-D chip stack. The task will be done in two phases: (1) probe-based bare die testing prior to packaging; (2) post-packaging chip validation and benchmarking. The probe-based testing (prior to packaging) will primarily focus on the analysis of the electrical parameters (including on-chip characterization of supply integrity, temperature, interface, delay, leakage/power etc.) and the minimal functionality of 3D SPHINX architecture with test-vectors for structural and functional tests. The at-speed tests will also be aimed. The detailed testing mechanisms and logic were described in Section 3.3. In the second phase, we plan to validate the architectural features, measure and characterize system performance, and demonstrate a functional 3D SPHINX design by running real benchmark applications. We next discuss the validation plan for our prototype.

During the first phase, we will directly probe the bare die. Along with the I/O pads, our designs (2-D MOSIS dies and the two-tier Tezzaron chip) will also have specifically designed probe-pads to facilitate pre-bond and post-bond probing. First, we will perform pre-bond probing of the 2-D MOSIS chip before and after the formation of the back-side TSVs. The I/O pads and probe-pads for pre-bond tests will be probed. We will next probe the 2-tier Tezzaron stack and the post-bond 3-tier die stack using probe cards. These tests will be performed using a 200mm probe-stations (currently being purchased through an NSF grant) and the probe-cards specifically designed for the pad configurations of our test chips. The tests will be controlled using LabView, PXI-based measurement equipment setup, and a high-speed oscilloscope.

In the second phase, we will test the packaged die stack. We will evaluate two options for this phase. The first option is to perform the packaging (using wire-bonding) and design of a simple PCB in-house at Georgia Tech. The designed PCB will be fabricated externally. We will integrate the 3D SPHINX chip in the board internally at Georgia Tech. The second option is based on previous experiences of our 3D-MAPS design (see Section 6.1.1). We will employ a contractor to design a daughter card for 3D SPHINX chip. The board can either be tested by interfacing with LabView and loading the benchmark applications or by using an FPGA developer's board. In the latter case, 3D SPHINX daughter card will be connected with an FPGA developer's board (*e.g.,* Xilinx' MLxxx series) through the standard connectors such as PCI Mezzanine Connectors (PMC) or FPGA Mezzanine Connectors (FMC). We will then program the FPGA board and upload our benchmark program bitstream into the memory of 3D SPHINX (on the daughter card) through scan chain for functional validation and performance measurement.

4 Statement of Work and Schedule

4.1 Schedule and Milestones

To achieve the goals of this research we anticipate following an overall schedule of activities and measuring progress by milestones as presented in Figure 3. There are 7 major research activities, each with several sub-activities listed as shown. Significant progress milestones are shown on the schedule.

4.2 Major Milestones

The program will demonstrate the ability of the 3D SPHINX design concepts to create a modular architecture for 3-D systems that supports composition of modules across technology generations (a tick) and designs within a technology generation (a tock). The major activities in support of this demonstration are the following.

1. A 3-tiered die stack (version 1 or V1) comprised of two compute layers and one memory layer wire bonded to the package. All three die will be fabricated in the same technology node (130 nm) and placed in a test set-up providing functional and performance data.

2. A 3-tiered die stack (version 2 or V2) using the same die for the compute layers but integrated and packaged with a memory die fabricated at a different technology node (90nm or 65 nm). The package will be delivered in a test set-up to provide functional and performance data.

3. All software (for design, test, and simulation) will be made available open source. Note that in particular much of this software development (especially the CAD extensions) has been developed by parallel and independently funded efforts.

4. All hardware designs will be made available as open source modules.

5. Writings documenting new processes, techniques, and insights will be published in conference and journal proceedings and/or as GT technical reports.

The major activity milestones marking technical progress at the granularity of a quarter are presented in the schedule in Section 4.1 (Figure 3).

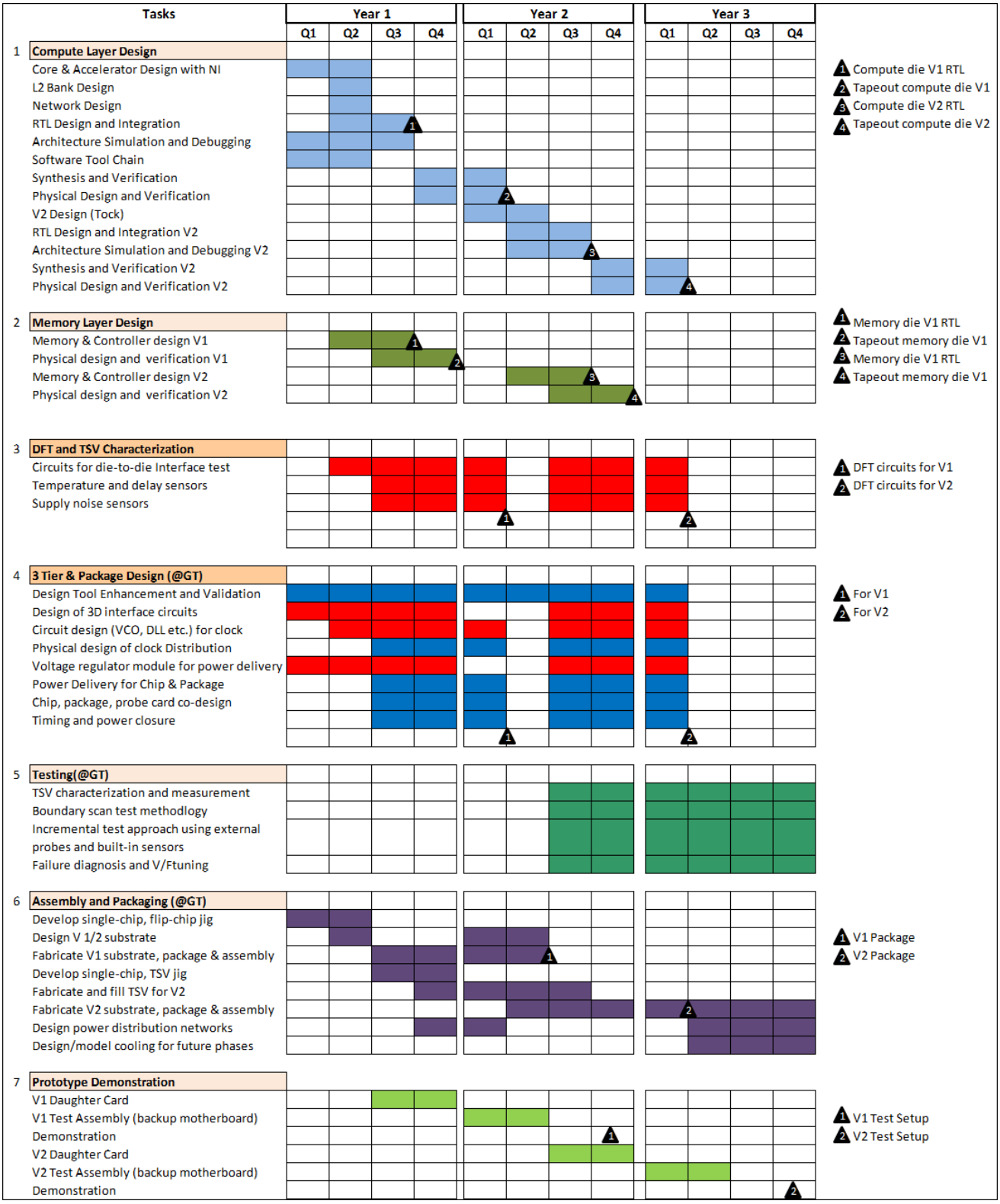


Figure 3: Schedule and Milestones

5 Fabrication and Assembly Facilities

5.1 Georgia Tech In-House Capabilities

In this program, we will leverage a lot of capabilities including design and analysis tools, fabrication, assembly, package capability, and 3-D design knowhow already being developed or under development using existing research fund for our 3D SPHINX prototyping. The following sections briefly overviews our in-house capabilities.

5.1.1 Microelectronics Fabrication and Packaging Facility

Process Step	Tools in GT Cleanroom	Chemicals Used	Comments
Via etch	PlasmaTherm ICP and STS ICP	SF6 and C4F8	etch time, gas flow rate/cycles, and power settings function of diameter and density of TSVs
Via passivation	Uniaxis PECVD	silane and oxygen plasma for silicon dioxide	quality and thickness of sidewall passivation will be tested on 'dummy' wafers
Seed layer coating	Unifilm sputtering system	-	we will ensure smooth via coverage
Copper plating	In-house plating setup	electrolytic acid copper solution	we have developed processes for void-free high aspect ratio TSV filling
Flip-chip assembly	Finetech Lambda	-	System has alignment accuracy of 0.5μm
CMP	Logitech Polisher	CMP slurry	we will ensure smooth surface for bonding

Table 1: TSV process and Assembly using Georgia Tech Microelectronics Facilities

The Georgia Tech Nanotechnology Research Center (GT NRC) is an open user facility and consists of the Pettit Microelectronics Building and the newly completed Marcus Nanotechnology Building. These two buildings maintain over 140 research tools. The facilities include the following relevant capabilities for the proposed research: optical lithography (spin-coaters and mask aligners), thin film deposition (thermal growth, low-pressure chemical vapor deposition, plasma enhanced chemical vapor deposition, and atomic layer deposition), reactive ion etching, inductively coupled plasma etching, electron-beam evaporators, direct current sputtering, packaging facilities (dicing saws, probe stations, lapping tools, polishing tools, wire-bonders, and flip-chip aligners), micro-scale printing tools, and a host of metrology tools. The specific tools we will use for the processes required to perform the TSVs and assembly are summarized in Table 1.

In addition, several of the PIs on this project are involved in 3-D integration as part of the Interconnect Focus Center (IFC) and Interconnect and Packaging Center (IPC). This collective expertise in the areas of design, CAD, characterization, bonding, power delivery and others will be leveraged as part of this project.

5.1.2 2-D/3-D CAD Tool and Test Capabilities

We will leverage our CAD tool capabilities to design the layout for both the 2-tier 3D IC and 1-tier 2-D IC, separately. The 2-D IC design can be done with readily available tools from Cadence, Synopsis, and Mentor. The 2-tier 3-D IC design can also be done with our existing CAD tools based on these vendor's tools, enhanced with custom point tools to handle TSVs and 2-tier stack that is shown in Figure 4, which we have already used to design our earlier prototype described in Section 6.1.1. In addition to the above, we will also align 3D SPHINX prototyping with other on-going tasks **supported by existing 3-D projects** as follows:

- *Tool enhancement.* Our in-house 3-D sign-off analysis tools will be enhanced to deal with both face-to-face and face-to-back bonding in 3D SPHINX. The entire set of tools including full-stack extraction, timing, clock, power/signal integrity, thermal, and design verification need to be revisited. Layout optimization tools including buffering and sizing for timing and power closure will also be extended.

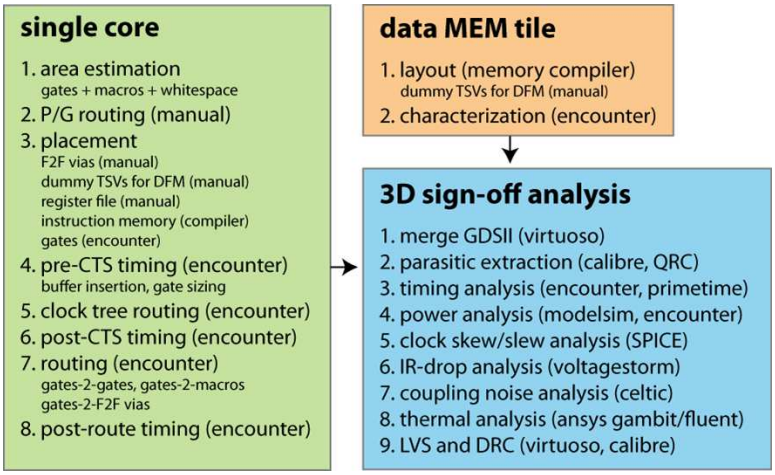


Figure 4: 3-D IC Design Flow

- *New tool addition.* Full-stack mechanical stress and electro-migration analysis caused by TSVs based on Ansys, Abaqus, and Comsol will be added. We plan to extend the tools to design and verify full-stack clock and power delivery networks and scan chains for pre/post-bond testing.
- *Circuit design support.* Our design team will deploy a variety of DFT circuit elements to enable pre/post-silicon and pre/post-bond testing for 3D SPHINX. Our CAD tools will need to be tailored to accommodate these DFT elements and provide feedback in terms of overhead.
- *Chip/package tool integration.* The packaging team requires its own set of tools from CST, Agilent, Ansys for analysis and optimization. We will build a bridge between our chip-based and package-based tool chains to co-design and co-optimize 3D SPHINX and its package for better power and signal delivery.

In addition, the test work will leverage results/prototypes from SRC and NSF research funding at Georgia Tech on testable and tunable multi-cores. Test equipments such as probe station, LabView software, oscilloscope from currently funded projects will be used to carry out our validation plan.

5.2 Potential Foundry Partners

We are exploring the collaboration opportunities of our foundry partners for fabrication. Here are two likely foundry/fabrication partners we will team up with.

5.2.1 Tezzaron

Our preferred choice of fabrication is to use the 3D process of Tezzaron and their foundry partner (GLOBALFOUNDRIES), whom our team members Lee and Lim have worked with in their prior 3D-MAPS prototyping project through DARPA's 3D MPW program, and is currently planning to tape out their second design 3D-MAPS V2 using their MOSIS-3DIC MPW run in 2011. (More details are given in Section 6.1.1.) Most importantly, Tezzaron offers technology way ahead of the curve in terms of the following reasons.

- Easy, inexpensive access to the fab line. They offer MPW runs with MOSIS twice a year at relatively low cost (around \$1500 per 1mm^2 as of spring 2011).
- Small size of TSVs: 1.2μm diameter and 6μm height, which is several times smaller compared with IBM and IMEC (5-10um diameter and 20-50um height), giving advantage of low capacitive load for signaling.
- Very high yield according to the CEO Bob Patti.

- Good track record: have manufactured several successful 3D ICs.
- Roadmap: have plans up update device technology from 130nm to 65nm and offer 3 and 4 tier stacking in late 2011, according to the CEO.

In this program, we anticipate to continue our relationship with Tezzaron and seek for using their future 3-D fabrication service for manufacturing our two-die stack.

5.2.2 ITRI and TSMC

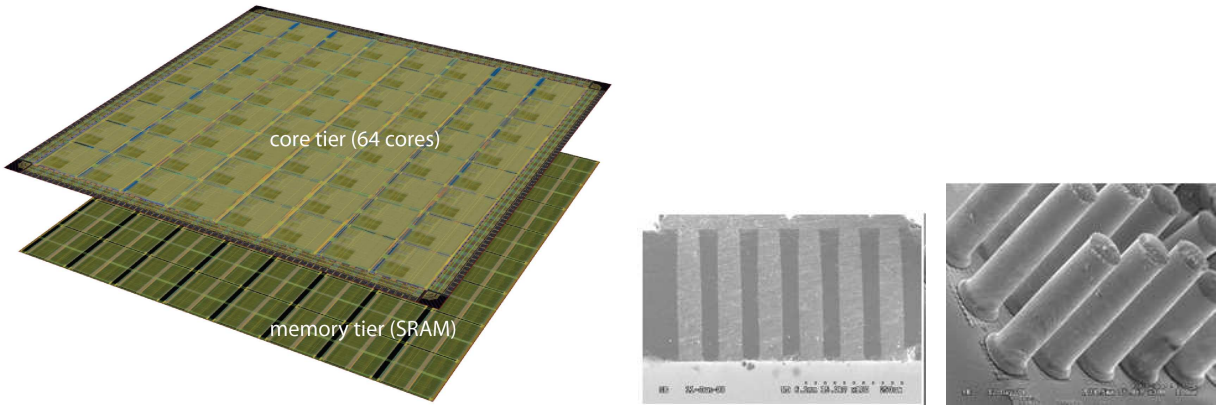
In addition to the first option, we are also assessing a backup option of using manufacturing and integration service provided by Industrial Technology Research Institute (ITRI). ITRI, a foundation and technology incubator based in Hsin-Chu, Taiwan, has made verbal commitment as our foundry and integration potential partner. They are also working with our team member Hsien-Hsin Lee on a 3-D IC design to be taped out in 2012. ITRI provides 90nm CMOS shuttle service from Taiwan Semiconductor Manufacturing Company (TSMC) for their 3D-IC research and development. The die stacking and bonding (currently limited to two die layers face-to-back), TSV fabrication, and post processing will be performed inside ITRI. ITRI will also provide TSMC IP including embedded DRAM (logic-process eDRAM) for satisfying our design requirement. Moreover, based on their partnership with a commercial CAD tool company, ITRI can provide us their established, complete 3D design flow which they used to design their current 3D ICs According to their current projection, ITRI will be able to provide 65nm process (TSMC shuttle service) with three die stack integration around the time we tape out our design.

6 Proposal Team

Our team includes eight members from Georgia Tech specialized in computer architecture, physical design, CAD tools, testing, power distribution, fabrication, and packaging. All of them have been heavily engaged in 3-D stacked ICs research for many years, and had done physical design, prototyping, fabrication, and packaging in their respective 3-D IC projects. Before their qualifications are described, we first highlight critical 3-D IC projects related to this Prototype RFP.

6.1 Highlight of Prior 3-D IC Projects from Team Members

6.1.1 3D-MAPS Many-Core Prototype



(a) 3D-MAPS 64-core processor (b) 50μm TSV (c) TSV pillars without Si

Figure 5: Highlight of Prior 3-D IC Projects at Georgia Tech

Our team members Hsien-Hsin Lee and Sung Kyu Lim led a team who designed and taped out a two-tier 64-core processor called *3D-MAPS* [11, 17] in March 2010 using GLOBALFOUNDRIES (Chartered) 130nm process and TSV process and bonding provided by Tezzaron through the DARPA MPW run. The two-tier 3-D chip contains 64 cores on one die and 256KB scratchpad SRAM on another, totaling 33M transistors. The taped-out two-tier GDSII is shown in Figure 5(a). Around 50,000 face-to-face vias were employed for power and signals to achieve 71.9 GB/s bandwidth at 277MHz core frequency. It also uses 50,000 TSV, each 1.2μ in diameter, for off-chip power delivery and signal I/O. At 1.5V, the 5mmx5mm 3-D chip stack consumes up to 6.2W. The power density is 24W/cm^2, and a special cooling mechanism is employed in the package. The RTL-to-GDSII tool chain is based on commercial tools from Synopsys, Cadence, and Mentor Graphics. Since these tools can only handle 2D ICs, we have developed plug-ins to handle TSVs and 3-D stacking. The chip is being bonded by Tezzaron and will be packaged by Amkor. The same team is currently working on *3D-MAPS V2* featuring 128 cores and 2Gbit DRAM, a 5-die stack to be taped out in May 2011.

6.1.2 Packaging and Cooling Technologies

Muhannad Bakir and Paul Kohl have extensive experience in 3-D IC technology demonstration, packaging, and cooling. They have demonstrated numerous TSV technologies with diameters ranging from 50 μm down to 100 nm. Figure 5(b) and Figure 5(c) are the SEM images showing TSV manufactured by them using Georgia Tech's facility. Moreover, they have demonstrated fine-pitch solder and copper-copper chip-to-chip bonding. In the 3-D prototype we propose, both TSVs and fine pitch bonding and assembly are critical elements. The PIs also have extensive experience with integration of novel on-chip cooling in 3-D ICs and have won numerous paper awards on their research.

6.2 Team Profile

Hsien-Hsin S. Lee (PI). Hsien-Hsin Sean Lee received his Ph.D. from University of Michigan, Ann Arbor. He was a senior processor architect and researcher at Intel (1995-2001) and managed the architecture team at Agere/Motorola's StarCore DSP Center (2001-2002). He started 3-D IC research in 2004 at Georgia Tech and will be the lead Principal Investigator for the 3D SPHINX project. Sponsored by FCRP's GSRC and C2S2 centers, Department of Defense, and National Science Foundation, Lee had led several research projects in 3-D architecture, physical design, design-for-test, and prototyping. In the 3D-MAPS project mentioned above, Lee's responsibility is the architecture definition and design, 3-D partitioning, RTL development, DFT, test plan, software and simulator, and post-silicon validation. Currently, Lee and his team are working with ITRI, Taiwan, to design a two-tier 3-D chip to be fabricated using TSMC's shuttle service (90nm) and ITRI's TSV process. Lee received 3 Best Paper Awards in MICRO-33, CASES-04, IBM PAC2-05, one IEEE MICRO Top Picks, and another 3 papers nominated for Best Papers (HPEC-07, FPL-07, ICCAD-09). Lee will be responsible for the 3D SPHINX architecture and physical design.

Muhannad Bakir. Muhannad S. Bakirs areas of interest include three-dimensional (3-D) electronic system integration, advanced cooling and power delivery for 3-D systems, biosensors and their integration with CMOS circuitry, carbon based interconnect networks, and nanofabrication technology. He is the editor of a book entitled Integrated Interconnect Technologies for 3-D Nanoelectronic Systems (Artech House, 2009) and is the author/coauthor of more than 70 journal publications and conference proceedings, 5 book chapters, 12 US patents. Dr. Bakir received an Intel Ph.D Fellowship (2002) and is the Associate Director of the Georgia Tech Interconnect and Packaging Center (IPC). His projects include multi-mode I/O for 3-D and chip-to-substrate attachment. In particular, significant advances have been made in the co-design of inter-tier chip cooling and tier-to-tier stacking for 3-D. Bakir will focus on integration and packaging in this project.

Abhibit Chatterjee. Abhijit Chatterjee is a Professor in the School of Electrical and Computer Engineering at Georgia Tech and a Fellow of the IEEE. He received his Ph.D in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1990. Chatterjee received the NSF Research Initiation Award in 1993 and the NSF CAREER Award in 1995. He has received four Best Paper Awards and three Best Paper Award nominations. In 1996, he received the Outstanding Faculty for Research Award from the Georgia Tech Packaging Research Center, and in 2000, he received the Outstanding Faculty for Technology Transfer Award, also given by the Packaging Research Center. Chatterjee has published over 350 papers in refereed journals and meetings and has 12 patents. He co-founded Ardext Technologies Inc, a test solutions company and served as Chairman and Chief Scientist from 2000 - 2002. He is currently directing research at Georgia Tech in mixed-signal/RF design and test funded by NSF, SRC, MARCO-DARPA and industry. He will be responsible for DFT and test plan in this project.

Paul A. Kohl. Paul Kohl received a Ph.D. from the University of Texas in 1978 and was employed at AT&T Bell Labs until 1989. He held a number of positions at Bell Labs including the creation and supervision of an Advanced Packaging Group to explore high-density memory-processor modules on silicon. He is currently a Regents Professor at Georgia Tech, Director of the Interconnect Focus Center, and founding Director in the SRC Interconnect and Packaging Center of Excellent at Georgia Tech. His research interests include materials and structures for electronic packaging and interconnect. He holds some of the original patents on porous low-k dielectrics created from thermal decomposition of thermally labile groups and received the Callinan Award from the ECS for the work. He has made numerous contribution of low-loss transmission line structures for chip-to-chip communications. Dr. Kohl is Vice President of the Electrochemical Society (2011-2014) and President elect (2014-2015). Recent work has focused on high-bandwidth, low-loss, air-clad transmission line structure for chip-to-chip communication, novel cooling methods for electronic systems, and power. Kohl will focus on integration and packaging in this project.

Sung Kyu Lim. Sung Kyu Lim has been working on design and CAD tool development for 3-D ICs since 2001. He is currently involved in 8 ongoing works all related to 3-D IC with his 10 PhD students. His group developed sign-off level physical design, analysis, and verification tools for 3-D ICs, which is

primarily based on commercial tools for 2D IC and enhanced with various in-house tools. This tool is use to tape-out 3D MAPS chips V1 and V2 mentioned earlier. His work on 3-D IC is nominated for the Best Paper Award at ISPD'06, ICCAD'09, CICC'10, and DAC'11. He is a member of the 3-D IC Design International Technology Working Group of the International Technology Roadmap for Semiconductors (ITRS). He has been leading the Cross-center Theme on 3-D Integration for the Focus Center Research Program (FCRP), Semiconductor Research Corporation (SRC), since 2010. Lim will be responsible for physical design, CAD tools, DFT, and power distribution network in this project.

Saibal Mukhopadhyay. Saibal Mukhopadhyay have been involved in successful tape-out of four test-chips from University (as a lead designer in two and as a faculty supervisor in two). Further he has designed one test-chip during his industrial tenure at IBM T. J. Watson Center. Results from his test-chips were published in premier circuit conferences and journals including ISSCC, VLSI Circuits, CICC, and JSSC. His particular custom circuit design experience includes: SRAM, on-chip sensors (leakage, device mismatch), logic circuits, analog bias generators (body-bias, voltage reference), and circuits for on-chip clocking (VCO, clock delivery, etc., 200MHz-1GHz). He has designed/supervised test-chip design in following technologies: 130nm RF-CMOS, 150nm FD/SOI (MIT), 180nm CMOS, and sub-45nm SOI. Mukhopadhyay will be involved in defining and supervising the circuit design and design-for-testability problems.

Madhavan Swaminathan. Madhavan Swaminathan received his PhD in Electrical Engineering from Syracuse University in 1991. He is currently the Joseph M. Pettit Professor in Electronics in the School of Electrical and Computer Engineering and Director of the Interconnect and Packaging Center, an SRC Center of Excellence at Georgia Tech. Prior to joining Georgia Tech he was with IBM designing packages for super computers. His areas of interest include power delivery, CAD, high frequency measurements, test and package integration. He has over 325 publications and 22 patents in this area and is the principal author of the book "Power Integrity Modeling and Design for Semiconductors and Systems", Prentice Hall, 2007 and co-editor of "Introduction to System on Package - Miniaturization of the Entire System", McGraw Hill, 2008. He has won many awards including the Technical Excellence Award from SRC in 2007 for his work on power delivery. He is an IEEE Fellow (for contributions to power delivery in mixed signal systems). He will be focusing on power distribution, TSV characterization and package integration in this proposal.

Sudhakar Yalamanchili. Sudhakar Yalamanchili has been involved in the development of high performance multiprocessor interconnection networks since the early 1990's when his group taped out Ariadne - a fully self-timed, reliable wormhole router. Since then this work has grown to include reliable and reconfigurable communication and computation where he has been the PI on projects funded by NSF, DARPA, DOE and ONR. His is currently Co-PI of a project funded by the FCRP's Interconnect Focus Center to explore the limiting impact of 3-D interconnects on multicore architectures. In the last decade he has led the development of compiler and optimization infrastructures for heterogeneous computing notably as Co-PI for DARPA's XMONARCH TRIX compiler and PI for the current CPU-GPU targeted Ocelot infrastructure supported by IBM, NVIDIA, LogicBlox and NSF. He will be responsible for the development of the configurable link and interconnection network for 3D SPHINX, and the software tool chain.

7 Cost Proposal

	TITLE: 3D SPHINX Stackable Processor using Heterogeneous INtegration and eXecution					
	PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR	No. of Months	12	12	12	
A.	SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates					
	(List each separately with title, A.7. show number in brackets)		Period 1	Period 2	Period 3	TOTAL
	1. Post Doc		$0	$60,000	$61,200	$121,200
	7. (1) TOTAL SENIOR PERSONNEL (1-6)		$0	$60,000	$61,200	$121,200
B.	OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
	1. ( ) POST DOCTORAL ASSOCIATES		$0	$0	$0	$0
	2. ( ) OTHER PROFESSIONALS		$0	$0	$0	$0
	3. (10-12) GRADUATE STUDENTS (Full-Time GRA is 20 hours/week)		$284,400	$290,088	$295,890	$870,378
	4. ( ) UNDERGRADUATE STUDENTS		$0	$0	$0	$0
	5. ( ) SECRETARIAL-CLERICAL		$0	$0	$0	$0
	6. ( ) OTHER		$0	$0	$0	$0
	TOTAL SALARIES AND WAGES (A+B)		$284,400	$350,088	$357,090	$991,578
	Graduate Student Health Insurance		$2,275	$2,321	$2,367	$6,963
C.	FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)		$0	$15,660	$15,973	$31,633
	TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A+B+C)		$286,675	$368,069	$375,430	$1,030,174
D.	PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $1,000)					
			$0	$0	$0	$0
	TOTAL PERMANENT EQUIPMENT		$0	$0	$0	$0
E.	TRAVEL	1. DOMESTIC (INCL. CANADA AND U.S.	$6,000	$6,000	$6,000	$18,000
		2. FOREIGN	$0			$0
F	OTHER DIRECT COSTS					
	1. MATERIALS AND SUPPLIES		$ 30,000	$ 172,500	$ 142,500	$345,000
	2. PUBLICATION COSTS/DOCUMENT DISSEMINATION					$0
	3. CONSULTANT SERVICES					$0
	4. CLEANROOM ACCESS FEES (2 students)		$28,000	$28,000	$28,000	$84,000
	5. SUBCONTRACTS					$0
		<25K				$0
		<25K				$0
		>25K	$0	$0	$0	$0
		>25K	$0	$0	$0	$0
	TOTAL Subcontract		$0	$0	$0	$0
						$0
	6. OTHER	Tuition cost	$148,366	$160,235	$173,054	$481,656
G.	TOTAL OTHER DIRECT COSTS		$206,366	$360,735	$343,554	$910,656
	TOTAL DIRECT COSTS (A THROUGH G)		$499,041	$734,804	$724,984	$1,958,829
H.	INDIRECT COSTS (57.1 % overhead)	BASE:	$350,675	$574,569	$551,930	$1,477,174
	Excluding tuition charges, equipment, and subcontract costs over $25K each					
	TOTAL INDIRECT COSTS		$0	$0	$0	$0
I.	TOTAL DIRECT AND INDIRECT COSTS		$499,041	$734,804	$724,984	$1,958,829

Figure 6: Anticipated Budget

As presented as one of the options in the RFP, we plan to place all intellectual property developed in this program in the public domain including offering software developed under this program under an open source license. Accordingly, and as per the guidance in the RFP (Ref: Section on Intellectual Property) the anticipated budget (without overhead) is provided in Figure 6. Note, such a grant/gift does not carry overhead and cannot contain contractual items. Alternatively, if the work is to be performed as part of a sponsored research agreement, overhead is assessed at a rate of 57.1% of several budget categories (such as personnel costs and travel expenses) and specific terms can be negotiated as desired with Georgia Tech's Office of Sponsored Programs.

7.1 Budget Justification

The specific elements of the budget are described in more detail below.

1. **Personnel:** We have budgeted for 12 graduate research assistants engaged in the project. We also plan to hire one postodoctoral research assistant for years 2 and 3 to lead several of the design efforts and assist in a continuous and smooth coordination of activities with Intel.

2. **Travel:** The budget includes approximately 5 person trips ($1200/trip) each year for travel to Intel sites for interaction with Intel engineers.

3. **Cleanroom Costs:** The cleanroom costs are budgeted at approximately $14K per student per year for two graduate students. This cost includes access and use of cleanroom facilities, the bonder, two boxes of dummy silicon wafers and related packaging materials and supplies.

4. **Tuition:** As required by the Institute, tuition remission is assessed at $1030.32 per month for each graduate student.

5. **Materials and Supplies:** This category includes several items.

   (a) Two Die Stack: We have budgeted $37,500 for each two die stack based on anticipated costs. Exact expense formulas cannot be released at this time due to some confidentiality issues. However, we are confident that the estimates are reliable.

   (b) Third Die Fabrication: We have proposed an aggressive strategy for fabricating the third die. Costs are difficult to estimate exactly at this time. We have budgeted $100,000 per die (one die each for V1 and V2).

   (c) Daughter Cards: We have budgeted $10K for sets of daughter cards upon which to mount the packaged 3 die stack. This cost will most likely drop as we plan to design the cards in house.

   (d) Motherboards: We have budgeted $5K for a motherboard for each version. This motherboard will host the daughter card and FPGA-based interface. Should a PCIe based daughter card solution serve our testing and demonstration needs, we will not need to have these motherboards designed and fabricated.

   (e) Probe Cards: We have budgeted $20K each for two probe cards for testing the two-die stack and memory die with our probe station.

   (f) Packaging: No packaging costs are incurred as these activities are being performed in-house.

Note that the M & S purchases are spread across project years. For example, in year 1 we will acquire a probe card and the daughter cards for the 3-die stacks for SPHINX V1 while the remaining fabrication and assembly costs for V1 are incurred in year 2.

References

[1] Open core protocol. http://www.ocpip.org/.

[2] E. Alon, V. Stojanovic, and M. Horowitz. Circuits and techniques for high-resolution measurement of on-chip power supply noise. *Solid-State Circuits, IEEE Journal of*, 40(4):820–828, 2005.

[3] K. Asanovic, R. Bodik, J. Demmel, T. Keaveny, K. Keutzer, J. Kubiatowicz, N. Morgan, D. Patterson, K. Sen, J. Wawrzynek, et al. A View of the Parallel Computing Landscape. *Communications of the ACM*, 52(10):56–67, 2009.

[4] K. Athikulwongse, A. Chakraborty, J.-S. Yang, D. Z. Pan, and S. K. Lim. Stress-Driven 3D-IC Placement with TSV Keep-Out Zone and Regularity Study. In *Proc. IEEE International Conf. on Computer-Aided Design*, pages 669–674, Nov. 2010.

[5] B. Black, M. Annavaram, N. Brekelbaum, J. DeVale, L. Jiang, G. H. Loh, D. McCauley, P. Morrow, D. W. Nelson, D. Pantuso, P. Reed, J. Rupley, S. Shankar, J. Shen, and C. Webb. Die Stacking (3D) Microarchitecture. In *Proceedings of the 39th International Symposium on Microarchitecture*, 2006.

[6] Brian Holden and Jay Trodden and Don Anderson . *HyperTransport 3.1 Interconnect Technology*. Mindshare Inc., first edition, 2008.

[7] M. Cho, C. Liu, D. Kim, S. Lim, and S. Mukhopadhyay. Design method and test structure to characterize and repair TSV defect induced signal degradation in 3D system. In *Computer-Aided Design (ICCAD), 2010 IEEE/ACM International Conference on*, pages 694–697. IEEE.

[8] M. Cho, C. Liu, D. H. Kim, S. K. Lim, and S. Mukhopadhyay. Design Method and Test Structure to Characterize and Repair TSV Defect Induced Signal Degradation in 3D System. In *Proc. IEEE International Conf. on Computer-Aided Design*, pages 694–697, Nov. 2010.

[9] J. Gjanci and M. Chowdhury. A Hybrid Scheme for On-Chip Voltage Regulation in System-On-a-Chip (SOC). *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, (99):1–11.

[10] J. Gu, R. Harjani, and C. Kim. Distributed active decoupling capacitors for on-chip supply noise cancellation in digital VLSI circuits. In *VLSI Circuits, 2006. Digest of Technical Papers. 2006 Symposium on*, pages 216–217. IEEE, 2006.

[11] M. B. Healy, K. Athikulwongse, R. Goel, M. M. Hossain, D. H. Kim, Y. J. Lee, D. L. Lewis, T.-W. Lin, C. Liu, M. Jung, B. Ouellette, M. Pathak, H. Sane, G. Shen, D. H. Woo, X. Zhao, G. H. Loh, H.-H. S. Lee, and S. K. Lim. Design and analysis of 3D-MAPS: A many-core 3D processor with stacked memory. In *2010 IEEE Custom Integrated Circuits Conference (CICC)*, 2010.

[12] M. Jung, J. Mitra, D. Pan, and S. K. Lim. TSV Stress-aware Full-Chip Mechanical Reliability Analysis and Optimization for 3D IC. In *Proc. ACM Design Automation Conf.*, June 2011.

[13] D. H. Kim, K. Athikulwongse, and S. K. Lim. A Study of Through-Silicon-Via Impact on the 3D Stacked IC Layout. In *Proc. IEEE International Conf. on Computer-Aided Design*, pages 674–680, Nov. 2009.

[14] J. Kim, C. Nicopoulos, D. Park, R. Das, Y. Xie, V. Narayanan, M. Yousif, and C. Das. A Novel Dimensionally-Decomposed Router for On-Chip Communication in 3D Architectures. In *Proceedings of the International Symposium on Computer Architecture*, 2007.

[15] Y.-J. Lee, R. Goel, and S. K. Lim. Multi-functional Interconnect Co-optimization for Fast and Reliable 3D Stacked ICs. In *Proc. IEEE International Conf. on Computer-Aided Design*, pages 645–651, Nov. 2009.

[16] K. Leung, Y. Mai, and P. Mok. A chip-area efficient voltage regulator for VLSI systems. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 18(12):1757–1762, 2010.

[17] D. L. Lewis, M. B. Healy, M. Hossain, T.-W. Lin, M. Pathak, H. Sane, S. K. Lim, G. Loh, and H.-H. S. Lee. Design and Test of 3D-MAPS, a 3D Die-Stack Many-Core Processor. In *IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits*, 2010.

[18] F. Li, C. Nicopoulos, T. Richardson, Y. Xie, N. Vijaykrishnan, and M. Kandemir. Design and Management of 3D Chip Multiprocessors using Network-in-Memory. In *Proceedings of the International Symposium on Computer Architecture*, 2006.

[19] Y. Li, H. Lakdawala, A. Raychowdhury, G. Taylor, and K. Soumyanath. A 1.05 V 1.6 mW 0.45°C 3σ-resolution $\Delta\Sigma$-based temperature sensor with parasitic-resistance compensation in 32nm CMOS. In *Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 340–341. IEEE, 2009.

[20] C. Liu, I. Ganusov, M. Burtscher, and S. Tiwari. Bridging the Processor-Memory Performance Gap with 3D IC Technology. *IEEE Design & Test of Computers*, 22(6):556–564, 2005.

[21] C. Liu, T. Song, J. Cho, J. Kim, J. Kim, and S. K. Lim. Full-Chip TSV-to-TSV Coupling Analysis and Optimization in 3D IC. In *Proc. ACM Design Automation Conf.*, June 2011.

[22] M. Bakir and J. D. Meindl. *Integrated Interconnect Technologies for 3D nanoelectronic Systems*. Artech House, 2009.

[23] K. Makinwa and M. Snoeij. A CMOS temperature-to-frequency converter with an inaccuracy of less than$\pm$0.5°C (3a) from-40°C to 105°C. *IEEE journal of solid-state circuits*, 41(12):2992–2997, 2006.

[24] R. Milliken, J. Silva-Martinez, and E. Sanchez-Sinencio. Full on-chip CMOS low-dropout voltage regulator. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 54(9):1879–1890, 2007.

[25] M. Pathak, Y.-J. Lee, T. Moon, and S. K. Lim. Through Silicon Via Management during 3D Physical Design: When to Add and How Many? In *Proc. IEEE International Conf. on Computer-Aided Design*, pages 387–394, Nov. 2010.

[26] R. Pendurkar, A. Chatterjee, and Y. Zorian. Switching Activity Generation with Automated BIST synthesis for Performance Testing of Interconnects. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 20(9):1143–1158, 2001.

[27] K. Puttaswamy and G. H. Loh. 3d-integrated sram components for high-performance microprocessors. *IEEE Transactions on Computers*, 58(10):1369–1381, 2009.

[28] R. Maddox and G. Singh and R. Safranek and R. Colwell . *Weaving High Performance Multiprocessor Fabric: Architectural Insights to the Intel QuickPath Interconnect*. Intel Press, first edition, 2010.

[29] Ravi Budruk and Don Anderson and Tom Shanley . *PCI Express System Architecture*. Mindshare Inc., first edition, 2008.

[30] P. Reed, G. Yeung, and B. Black. Design Aspects of a Microprocessor Data Cache using 3D Die Interconnect Technology. In *Proceedings of the 2005 International Conference on Integrated Circuit and Technology*, 2006.

[31] K. Sasidhar, L. Alkalai, and A. Chatterjee. Testing NASAs 3-D stack MCM Space Flight Compute. *IEEE Design and Test of Computers*, 15(3):538–545, 1998.

[32] K. Sasidhar and A. Chatterjee. Hierarchical Diagnosis of Identical Units in a Systems. *IEEE Transactions on Computers*, 50(2):186–191, 2001.

[33] T. Sato, A. Inoue, T. Shiota, T. Inoue, Y. Kawabe, T. Hashimoto, T. Imamura, Y. Murasaka, M. Nagata, and A. Iwata. On-Die Supply-Voltage Noise Sensor with Real-Time Sampling Mode for Low-Power Processor Applications. In *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, pages 290–603. IEEE, 2007.

[34] P. Variyam and A. Chatterjee. Specification Driven Test Generation for Analog Circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 19(10):1189–1201, 2000.

[35] D. H. Woo, J. B. Fryman, A. D. Knies, M. Eng, and H.-H. S. Lee. POD: a 3D-integrated Broad-purpose Acceleration Layer. *IEEE MICRO special issue on Accelerator Architectures*, 28(4):28–40, July/August 2008.

[36] D. H. Woo, N. H. Seong, D. L. Lewis, and H.-H. S. Lee. An Optimized 3D-Stacked Memory Architecture by Exploring Excessive, High-Density TSV Bandwidth. In *Proceedings of the 16th International Symposium on High-Performance Computer Architecture*, pages 429–440, 2010.

[37] K. Woo, S. Meninger, T. Xanthopoulos, E. Crain, D. Ha, and D. Ham. Dual-DLL-based CMOS all-digital temperature sensor for microprocessor thermal monitoring. In *Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 68–69. IEEE, 2009.

[38] J.-S. Yang, K. Athikulwongse, Y.-J. Lee, S. K. Lim, and D. Z. Pan. TSV Stress Aware Timing Analysis with Applications to 3D-IC Layout Optimization. In *Proc. ACM Design Automation Conf.*, pages 803–806, June 2010.

[39] Z. Zeng, X. Ye, Z. Feng, and P. Li. Tradeoff analysis and optimization of power delivery networks with on-chip voltage regulation. In *Proceedings of the 47th Design Automation Conference*, pages 831–836. ACM, 2010.

[40] X. Zhao, D. Lewis, H.-H. S. Lee, and S. K. Lim. Pre-bond Testable Low-Power Clock Tree Design for 3D Stacked ICs. In *Proc. IEEE International Conf. on Computer-Aided Design*, pages 184–190, Nov. 2009.