3 Project Description

3.1 Introduction

Since Moore’s first observation in 1965 of the exponential growth in the number of transistors per integrated circuit, there has been an expectation of the continued availability of low-cost circuits of ever-greater complexity. The extraordinary progress charted by “Moore’s Law” has been accomplished through continual, comprehensive and integrated achievements in materials science, device miniaturization, and strategic design at the device through systems level. In recent years, there has been increasing recognition that sustained exponential growth of complex electronic systems will require new breakthroughs in fabrication and assembly with controlled engineering of nanoscale components. The National Nanotechnology Initiative (NNI) has identified Nanoelectronics, one of the “next generation of information technology devices” as one of nine “Grand Challenge” areas because of the potentially significant economic and societal impact of innovations in this area [79]. In the still-early stages of research in nanoelectronic technologies, a variety of candidate device geometries and principles has been proposed; what has been singularly lacking up to now is serious consideration of an accompanying design methodology. Smaller, ultimately nanoscaled electronic devices promise greater information density and hence systems functionality. However, materials properties may change dramatically at the nanoscale, leading to voltage breakdowns and current leakage. Variations in device dimensions of 1-10 nm, perhaps acceptable for circuits with 100nm gate-length devices, result in unacceptable margins for nanoscale devices. At the same time, placing such exacting constraints exceeds present-day manufacturing tolerances. Accordingly, a number of different alternative schemes have been proposed. ‘Molecular Electronics’ is one scheme that replaces the transistor with a molecular switch that can be easily chemically synthesized and electrically controlled. Among the most successful work in this area is arguably the early work reported in [31], and dramatically scaled-up in [66]. Possible control of the size and structure of massively produced carbon nanotubes, together with the possibility of exceptional electron mobility, has generated strong interest in these nanostructures as the basis of a new nanoelectronic technology. A number of simple building-block devices, such as inverters with modest gain, have already been demonstrated [32]. A related strategy is the use of semiconductor nanowires as the building-block devices of a new nanotechnology [33]. This is by no means an exhaustive catalogue of candidates for new nanoelectronic technologies, and the examples described have achieved differing levels of development from single device to integrated system. It is clear, however, that the choice of new nanoscale devices will mandate dramatically different design parameters than apply to even current scaled CMOS devices. We expect different methods of charge control and different metrics of evaluation (e.g., noise margin, threshold voltage, propagation delay, capacitance, parasitics). These differences have obvious profound implications for the operation of the complex circuits that will be constructed from these devices. And yet, ironically, with some few exceptions [31][66], there is very little being done to map possible systems designs onto real candidate nanodevice technologies. Given the very early stages of research in nanoelectronic devices, and the even fewer examples of nanodevice integration, many might argue that it is too early to generate a design methodology for technologies whose performance and properties are still ill-defined. Yet we believe that this is exactly the right time for such an effort. Design methodologies have obtained a tremendous degree of sophistication and predictive value, and have kept pace with shrinking device dimensions and changes in device specifications. We believe that there are tremendous advantages to be gained in using the power of these tools, insights and methodologies to begin to explore and define a context for evaluation of next-generation nanoelectronic technologies. Without such a context, new nanoelectronic devices will continue to be evaluated against the metrics of the existing scaled CMOS technologies: against those metrics, new technologies may always be found wanting. More importantly, without a common context of systems evaluation, it will be difficult to make early assessments of the viability of new nanoelectronic device approaches, nor will it be possible to strategically guide the development of new technologies.

In the latest version of the International Technology Roadmap for Semiconductors (ITRS) [59], most entries associated with the technology nodes of 45nm or below are labeled red, meaning that no manufacturable solution is known. Nanoelectronics design research is in a discovery mode and therefore an ideal candidate for international collaboration because: (i) it has the true nature of pre-competitive research—the results and discoveries in this area will form the foundation for the entire industry to move forward; and (ii) we need to leverage the global investment to share the extremely high cost projected for nanotechnology development and coordinate the research efforts in this area. We believe that actively participating, defining, and leading such an international research effort is the best way to help the U.S. maintain its long-term technology leadership in the future.

This project proposes to establish an International Center on Design for Nanotechnologies (IC-DFN), where we have chosen to collaborate with a group of established researchers from two universities in Taiwan (National Taiwan and
National TsingHua Universities) and three universities in China (Peking, Tshinghua, and Zhejiang Universities). There are several reasons for choosing these universities in Taiwan and China as our international partners:

- Taiwan and China offer a great deal of expertise in semiconductor fabrication and manufacturing. Taiwan has established itself as the leader in the IC foundry business in the past ten years (with over 70% of the worldwide market in 2004 [57]) and has helped to establish the so-called fabless semiconductor industry in the U.S. and worldwide. China is following Taiwan’s lead, and in the past few years has successfully attracted the most new fab openings. We expect that the foundries in Taiwan (such as TSMC and UMC) will be among the first to produce nano-CMOS devices.

- The vast human resources available in China and Taiwan, with their traditional emphasis on engineering education, make them the ideal location for building large design teams. We expect that a large percentage of future designs in nanotechnologies will be carried out in these regions. Most major U.S. semiconductor and IC companies have already set up or are setting up large design centers in China and Taiwan. Indeed, our proposal has received strong support from major U.S. semiconductor, EDA (electronic design automation), and FPGA companies, including Cadence Design Automation, Intel Corp., Magma Design Automation, and Xilinx, Inc. Their support letters (see supplementary documents) have endorsed both the research directions of IC-DFN and its international partnership program with the participating universities in China and Taiwan.

- Similar to the U.S., both Taiwan and China have set up large national investments in nanotechnologies. Our international partners in Taiwan and China are simultaneously applying for funding from the National Science Council (NSC) in Taiwan and the Chinese National Science Foundation (CNSF), respectively, to support the proposed research and education activities of the Center, and are confident that they will get substantial support from these funding agencies. It is therefore mutually beneficial to leverage the investment in these regions for nanotechnology research. Indeed, through the creative leveraging of this international partnership and collaboration, the proposed project will be able to assemble a large research team on design for nanotechnologies with 21 faculty members and over 50 graduate students from the seven universities located in the U.S., Taiwan, and China, with only a modest funding request to NSF ($500K/year). Any research project of this scale would be extremely costly (at minimum, $3-5M/year) if carried out by U.S. researchers alone.

- The two PIs of this proposal have built a very successful international collaboration program in the past five years on giga-scale system-on-a-chip designs, involving 15 faculty members and over 30 graduate students from the U.S., Taiwan, and China sponsored jointly by NSF, NSC, and CNSF, and resulting in significant research output (a more detailed summary is provided in Section 3.5). This proposal is built upon our previous success in international collaboration, yet with a new research scope focusing on nanotechnologies, a strong educational component, and a much expanded and diversified research team including experts in nanodevices, circuit/architecture designs, and design automation.

- The five foreign universities that we selected for partnership are the very top universities in their regions, and the participating faculty members are internationally renowned experts in design and design automation of nanodevices (please refer to their bios attached to the end of this proposal).

- Last but not least, in the past two decades China and Taiwan have been among the fastest growing regions in the world economy, yet their culture is dramatically different from that of the U.S. It is very important for U.S. students and young researchers to witness, first hand, the rapid growth in these regions, to understand the culture and reasons behind such growth, and to prepare themselves as future technology and/or business leaders in a more and more highly coupled global economy.

Each student in this program will have the opportunity to visit one of the five foreign universities, for one to three months each year, participating in relevant research projects with that university’s scientists and students. Faculty members at the host universities have committed themselves to mentoring the visiting students, helping them to appreciate the cultural and technology development of the host country and the linkage of technology to economic development, as well as helping to develop their scientific talents and social interaction skills to prepare them for future participation in this global economy. We have received commitments (see supplementary documents) from the participating foreign universities to provide all local support for this educational plan. In addition, IC-DFN will organize two workshops per year, rotating between the United States, China and Taiwan, to facilitate the exchange of research and education problems, ideas, and results. The workshops will include specific topics and presentations, targeted at students and young researchers to catalyze a deeper understanding of this multidisciplinary field and to develop the cultural and social aspects of international collaboration. A more detailed education plan is presented in Section 3.3.
3.2 Research Plans and Objectives

The proposed research activities address the full hierarchy of the design issues for nanotechnologies, with a special focus on the architectural level and the integration level. The research tasks include investigation and development of efficient design methodologies, algorithms, tools, and microarchitectures for the emerging nanotechnologies, including both nano-CMOS and technology candidates beyond CMOS. The specific DFN research thrusts include: (i) exploring novel techniques to achieve highly reliable and error-resilient designs to overcome likely defects and non-deterministic behavior of new nanodevices, (ii) enabling higher levels of design abstraction to support the extremely high degree of integration enabled by the nanotechnologies, and (iii) developing novel, highly scalable solutions to fundamental design automation problems to cope with the high computational complexity associated with synthesis and verification of complex systems implemented using nanotechnologies. Two technology drivers will be explored. The first one is ultimate CMOS in the 30nm to 10nm generations, where many challenging design issues, such as complexity and reliability, already occur. The second technology driver will be based on one or two technology candidates beyond CMOS, such as nanowires. Moreover, we plan to use nano-FPGA as an architecture/platform candidate to explore various issues in nanotechnology designs. The regularity and reconfigurability makes nano-FPGA an ideal candidate to support the reliability and complexity requirements of designs in nanotechnologies. Finally, a complex system-on-a-chip design with multiple CPU and DSP cores and video codec for multimedia application will be developed as an application driver (testbed) to explore the extremely high capability and complexity associated with designs in nanotechnologies. Fig. 1 illustrates the center organization and research thrusts. The three “horizontal” DFN thrusts will be driven heavily by the two “vertical” technology drivers (“nano-CMOS at 10-30nm” and “beyond CMOS”), as well as the architecture driver (nano-FPGA) and the application driver/testbed (“multi-processor/multi-core SOC”). Every thrust consists of team members from all three regions with relevant expertise to enable close international collaboration. For the application driver which demands tremendous resources, we will leverage the large design teams and fabrication infrastructure currently in place at our collaborating institutes in Taiwan and China (a more detailed summary of task division among participating institutes in three regions is shown in Section 3.4.2). The outcome of this research project will be innovative design methodologies, tools, and algorithms that enable efficient and economic design solutions using the present nano-components and potential new ones as soon as they become available and practical.

3.2.1 Technology and Architecture Drivers

In this section, we shall first discuss nanotechnology candidates/drivers considered in this project, and then propose to use nano-FPGA as an architecture/platform driver, so that we can integrate nanodevices in very large scale to support meaningful computation and reliable digital applications. In this proposal a nano-FPGA refers to a field-programmable gate-array (FPGA) fabricated in nanotechnologies, such as ultimate CMOS or nanowires. It has functionalities similar to those in today’s market that are provided by FPGA companies like Altera and Xilinx, but may have very different circuit-level or logic-level architectures in order to deal with process variation, yield, reliability, complexity, and other issues associated with nanotechnologies. Our reason for using nano-FPGA as an architecture/platform driver is based on the following:

- A nano-FPGA is highly regular. It is easier to come up with a highly optimized tile, with consideration of various manufacturing related issues, and then replicate it many times across the chip to form a nano-FPGA.
- A nano-FPGA is field-programmable, so that we can reconfigure it for yield and/or reliability enhancements.
- A nano-FPGA is a memory dominated device. Given the current trend of nanotechnology development, we expect that memory technology (to be discussed next) will be more mature than logic devices in the near future.

For similar reasons, regular programmable logic arrays have also been considered as candidates by other researchers for exploring the potential of nanotechnologies, such as [34][35][53].

3.2.1.1 Technology Drivers

Future system designs that incorporate new nanodevices into either memories or logic applications should address their variability of properties as well as their control for manufacturability. As the feature size continues to scale
down, the variations of properties for sub 30 nm devices as a result of lack of processing control will become much more significant than today’s 90 nm technology. For example, the number of electrons in the channel will decrease to a single digit and eventually to a single electron, and thus, charge fluctuations in the conducting channel and the gate/channel interface will result in a large variation of the current and threshold voltage. We discuss below the potential new devices for performing system functions.

**Emerging Memory Devices**

As scaling down continues, memory will be extremely important for low power information processing as well as for reliable architecture/platform design, such as nano-FPGAs to be discussed in Section 3.2.1.2. In this section, we focus on the embedded memory drivers because they will be used in system design for computing and reconfigurability.

We categorize the new technologies being developed into high/low speeds, and volatile/nonvolatile memories. For low-speed nonvolatile memories (which are of particular interest for reconfiguration of systems), there are nano floating gate devices, which may incorporate quantum dots. These devices may be fabricated with scaled CMOS technologies. Another class of this type of memory includes molecular, polymer, crossbar CNT and nanowire memories. One example of polymer memory material is 2-amino-5-imidazoleredicarbonitrile, and this type of memory utilizes the effect of electrical bistability in a triple-layer [65][71][86] (Fig. 2(a)(b)). The latter may be fabricated by self assembly after CMOS fabrication.

Potential high-speed emerging memory devices include Insulator Resistance Change (or Phase Transition) Memory, nano negative differential resistance [50][56], and spin-based memories [85][95][106]. The first uses phase transition to change the resistance (Fig. 2(c)). The second one takes the advantage of the nanostructure to exploit tunneling to latch up for memory [105] (Fig. 2(d)). The third class uses spin to control either current flow or uses the electric field to control the storage of information. For the phase transition memory, the mechanism is the reversible formation/annihilation of conducting filaments, and the basic component of this memory is a metal-insulator-metal (MIM) structure, using insulators that show reproducible I-V hysteresis (Cr-doped (Ba, Sr)TiO3 or SrZrO3 and others). We will investigate these emerging candidates for incorporating to the design of nano-FPGA. For example, in the case of nonvolatile polymer memory devices, we will explore the design as of how to use self assembly of these polymer memories after the fabrication of CMOS. Similarly, with the spin based memories, our design will change accordingly.

**Emerging Logic Devices: (1) Ultimate CMOS**

The ITRS roadmap indicates that the scaling of the feature size of CMOS will likely continue for the next decade, and CMOS will evolve to “ultimate CMOS,” beyond which continued scaling of CMOS will reach its fundamental limits [59]. Among the limitations anticipated by ITRS are power dissipation and variability of device characteristics. The evolution of CMOS structures includes silicon on insulator (SOI), dual gate, and surrounding gate structures such as the vertical pillar [102] and FinFET (or Trigate) devices [38][54][104]. Details of the device design are varied for the
improved control of subthreshold behavior to improve short channel effect and the subthreshold performance for low power applications.

In the evolutionary path of scaling, these identified candidates of the advanced CMOS structures can be scaled more aggressively than the classic bulk-Si structure. As shown in Fig. 3, the Trigate (or FIN) device is electrostatically more robust than single-gate MOSFET because three gates are used to control the channel, thus allowing for additional gate length scaling by at least a factor of two. Compared with conventional bulk devices, FinFETs have shown greatly improved short channel characteristics, such as smaller threshold voltage roll off, lower drain-induced barrier lowering and lower subthreshold slope. For vertical structures, illustrated in Fig. 4, the channel can be engineered to include a tunneling structure with a negative differential resistance. This type of structure can be designed into circuits with rich functionalities, as will be discussed in the next section.

**Emerging Logic Devices: (2) Beyond CMOS: Technology Candidate—Nanowire**

ITRS has been continuously examining emerging logic devices. But based on the criteria set forth by ITRS (at least two groups have worked and reported on the structures), and after examining many potential logic devices, nanotubes and nanowires appear to be among the very few that have been identified as the most feasible candidates. Nanowires and nanotubes offer the benefit of higher device performance. The recent development of nanowires/nanotubes provides the control of small feature sizes suitable for nanodevice fabrications.

The additional key motivation is to achieve the high integration that is not possible with conventional top-down approaches, and thus to afford the fabrication of the nano-FPGA in the sublithographic regime. With the emerging bottom-up nanowire synthesis techniques, some programmable logic designs have been demonstrated by using crossed semiconductor nanowires, e.g., molecular switch–based random access memory [74] and nanowire-based logic circuits [5][55]. For example, one promising method for accomplishing the fabrication is the superlattice nanowire pattern transfer (SNAP) approach [77]. This general method can produce ultrahigh-density arrays of aligned metal and semiconductor nanowires and nanowire crossbars, as shown in Fig. 5. Diameters and pitches (center-to-center distances) of nanowires are fabricated in the tens of nanometers. Another technique is nanoimprinting. Nanostructures of around 20 nm have been fabricated. The patterns have been successfully used as a template to produce different kinds of nanowires. Using these techniques, the process can be carried out multiple times to produce programmable cross points at nanowire junctions.

Nanowires circuit architectures are being investigated to achieve higher integration density. For instance, a nanowire NAND logic gate function is illustrated in Fig. 6 with a nanowire logic gate which consists of two parallel semiconductor nanowire arrays with p type field effect transistor (pFET) and nFET functions, respectively. Likewise, nanowires have been used for several applications, for example, PLAs [34]. We plan to develop the use of semiconductor nanowires for implementation of a variety of components for design of nano-FPGA. The nanowire circuits may achieve the desired performance in terms of power consumption, speed, and reliability.

**3.2.1.2 Architecture/Platform Driver – nano-FPGA**

The overriding concern in designing a nano-FPGA will be how to overcome the yield and reliability problems that result from the large manufacturing process variation and the much greater likelihood of transient errors due to the extremely small dimensions of nanodevices. We plan to address the yield and reliability issues through smart and efficient use of reconfigurability and redundancy. Reconfigurability
includes the capability for both static reconfiguration and dynamic reconfiguration. The redundancy can be provided at different levels, including circuit level, logic level, and component/sub-system level. In the remainder of this subsection, we shall outline a few possible solutions or directions for use of reconfiguration and reconfigurability.

We plan to provide both static reconfiguration and dynamic reconfiguration capabilities in the proposed nano-FPGA. Static reconfiguration provides one-time programmability and is usually carried out after fabrication to correct faulty elements in the circuit for yield enhancement. One possible way to support static reconfiguration is through the use of metal-to-metal programmable nano vias, similar to the ViaLink® [41] technology used in one-time programmable FPGAs from Quicklogic [84]. Initially, each programmable via is open. When a proper voltage is applied to a programmable via, it can change the state and form a conducting path. The size of a programmable via is comparable to that of a normal via. Such programmable vias can be used between any pair of adjacent metal layers, and thus provide high-density programmable interconnects. We shall explore both the technology and efficient use of such programmable vias in our research project.

Dynamic reconfiguration, on the other hand, supports reprogrammability in the field, and can be used to both repair faulty circuits for yield enhancement and correct transient errors for higher reliability. One common way to support dynamic reconfiguration is through extensive use of the proposed nano memory cells and memory-controlled programmable logic, similar to the SRAM-based FPGAs widely available in the market today. For example, a k-input logic function can be implemented by a lookup-table (LUT) of $2^k$ memory cells, and its functionality can be dynamically changed by writing different values into the LUT. Similarly, a memory-controlled multiplexer enables dynamic change of the input-to-output path by taking different values to the memory cells connected to the selection signals. Fig. 7 shows a typical reprogrammable logic cell, often called basic logic element (BLE), in a SRAM-based FPGA widely used today, which can implement any k-input function, and optionally store the result in DFF. Clearly, compared to programmable nano vias, memory-based reprogrammable cells and interconnects are more costly in terms of area and density. Therefore, it is better to use them for coarser grain reconfiguration. We also notice that there is encouraging progress on the development of reversible programmable nano vias, where one may apply a small voltage and/or energy that results in an order-of-magnitude, nonvolatile change of electrical resistance of the nano via [88][90]. Such a development will lead to much denser dynamically reconfigurable interconnects. We shall further study and explore such technology options by incorporating new nano components, including those described in Section 3.2.1.1, during the course of our proposed research. The design using these nanocomponents is a key part of this study.

With reconfigurability in place, the first natural step is the use of programmable power supplies, which can be very important for combating the large Vt variation in nano-CMOS and possibly in nanowires. For example, Fig. 8 shows a programmable dual-Vdd power supply using two programmable switches. For a low Vt device, we can select the low Vdd (VddL). For a high Vt device, we can select high Vdd (VddH). If both programmable switches are dynamically reconfigurable, they may be both programmed off to achieve power gating to completely eliminate the leakage power for the associated BLE. Such a programmable Vdd concept has recently been proposed in the design and synthesis of low-power FPGAs [17][70]. Obviously, it is costly to deliver programmable Vdds to every single transistor or BLE. So, we shall study the proper granularity to provide programmable Vdds. Fortunately, Vt variation has strong spatial correlation in most cases. For the purpose of leakage power control, it is meaningful to have a large number of devices in a region that shares a common Vdd value.
Furthermore, with reconfigurability in place, we can incorporate different levels of redundancy for yield and reliability enhancement. At the finest level, we can include redundant transistors, and two transistors can be merged using, say programmable vias, as shown in Fig. 9. This ensures that when one transistor is faulty, the pair can still work properly. The merged pair also has a different I-V characteristic—we can merge two high Vt transistors to get a transistor with lower Vt. However, the overhead of redundancy and reconfigurability at the transistor level is very high. The redundancy at this level is likely to be used only for key circuit elements. A more affordable approach is to provide redundancy at the logic cell or logic block level. For example, for a k-input LUT, we may provide more than 2k memory cells so that redundant bits can be used for error detection and correction. Also, it is likely that the nano-FPGA will have a hierarchical structure, as in today’s FPGAs, such as Stratix or Stratix II [2], where a group of identical BLES forms a logic block cluster, also called a configurable logic block (CLB) as named in [8]. Through proper design of the reconfigurable local interconnects within a CLB, we can include one or several redundant BLES, as shown in Fig. 10, so that the CLB is tolerant to one or multiple BLE faults. Such a hierarchy might be repeated for more than one level to provide a higher degree of yield and/or reliability as suggested in [53]. Finally, the redundancy can be provided at the component/sub-system level, such as the use of multiple on-chip processor cores. An important task in our proposed research is the systematic study of various reconfigurability and redundancy schemes in nanotechnology designs for the best density, energy, performance, yield, and reliability trade-off.

We shall employ the quantitative evaluation methodology as used in [8][26][69] to evaluate and optimize the proposed nano-FPGA architecture, both at the circuit level and the logic level. We shall simultaneously develop the associated logic synthesis, mapping, placement, and routing tools for nano-FPGAs in order to carry out such quantitative architecture evaluation. The research teams at UCLA and NTHU have extensive experience in this area (e.g. [15][18][20][28][96][100]). We shall also consider the ArchEvaluator™ tool [4] from Magma for such an evaluation.

### 3.2.2 DFN Thrusts

#### 3.2.2.1 Design for Reliability

A common characteristic of the basic building blocks for computing in future nanotechnologies is their relatively high variability and low reliability. For silicon nanotechnology, the drastic device shrinking, low power-supply levels, and increasing operating speeds significantly reduce the noise margins and increase error rates for both soft-errors and hard-errors. On the other hand, innovations of new devices in nanotechnology seem to be highly prone to errors due to environmental and manufacturing variations as well as the nature and limitations of molecular manufacturing. At this stage, these new innovative devices cannot compete with CMOS in terms of their reliability, scalability, performance, and cost. If we are to continue to design systems using ever-smaller components, down to the atomic scale, we must address the system reliability issues as well as reproducibility for nanomanufacturing.

The scope of this research will address the reliability and reproducibility issues involved in several levels of design hierarchy, including logic, integration, and architectural levels. The goal is to achieve an integrated design methodology supporting self-reconfiguration and error-tolerance for complex systems in the presence of components with relatively high variability and failure rates.

Errors and faults in a system could be either permanent (hard errors) or transient (soft errors). Reconfiguration, done either statistically or dynamically as discussed in Section 3.2.1.2, could be an effective solution to fixing the hard errors. To support reconfigurability, the design must be self-testable and self-diagnosable, to the right level of diagnosis resolution, to automatically identify the failed devices/components before self-reconfiguration can take place. For static reconfiguration, off-line self-test and self-diagnosis will be sufficient. To support dynamic reconfiguration, the design must have on-line self-test and diagnosis capabilities to detect and identify failures when a system is operating. Redundancy, which is a fundamental technique to address soft errors, can be generally categorized into three types: (1) **hardware redundancy** such as N-tuple modular redundancy (NMR) and NAND multiplexing [51], (2) **time redundancy** such as recomputation in time using alternating logic or self-dual logic, and (3) **information redundancy** by adding redundant information to the original data, such as error-correcting codes.

Several research issues ought to be investigated in order to develop a broadly applicable and scalable design methodology to support reliable design built upon reconfiguration and redundancy. In this project we plan to investigate the following subjects, using nano-FPGA as an architectural/platform driver: (1) exploring the tradeoffs between using reconfiguration and redundancy to address the reliability problem for systems built upon unreliable components with various hard and soft failure rates and spatial distributions, (2) functional- and timing-error-tolerant architectural design techniques to incorporate built-in reconfigurability and redundancy, (3) self-test and self-
diagnosis techniques to support built-in reconfiguration, and (4) CAD issues on synthesis, mapping, and routing of reliable design with built-in reconfiguration and redundancy capabilities. The research teams at UCSB, NTU and NTU have extensive experience in verification, test, fault tolerance and FPGA so will jointly address these tasks.

**Optimal mixture of reconfigurability and redundancy:** Among recent studies of error-tolerant architectures for large systems built on unreliable components, there are two representative approaches. Teramac [53], a customized configurable computer based on clusters of traditional FPGAs, can implement target designs with up to 75% defective components in the FPGA clusters. The error tolerance is achieved by highly redundant network connectivity, connected via crossbars in a fat-tree network, so that the compiler can avoid the defective parts and routing resources yet still be able to map and route the design using the scattered good components. A self-diagnosis program is required to locate defective components to allow reconfiguration. The second example is the Embryonics project [76], a highly robust integrated circuit which is inspired by the basic processes of molecular biology and the embryonic development of living beings. By adopting certain features of cellular organization, and by transposing them to the two-dimensional world of integrated circuits on silicon, its researchers made some preliminary observations that properties unique to the living world, such as self-replication and self-repair, might also be applicable to artificial objects such as integrated circuits. Using existing FPGA components, the Embryonics project developed VLSI circuits capable of self-repair and self-replication. What we can learn from these examples concerns the feasibility of building reliable computers with imperfect components through reconfiguration by using regular structures with abundant communication bandwidth. However, such reconfiguration techniques are not efficient for the transient errors which are very common for nanodevices. On the other hand, redundancy techniques such as N-tuple modular redundancy (NMR), while efficient for transient errors, require reliable critical components which often are not available for nanodevices. Thus, it is clear that a combination of reconfiguration and redundancy is necessary to support systems with both hard and soft faults. At each level of hierarchy, a key research issue is: what is the right mixture of hardware support to enable adequate reconfiguration and redundancy for a computing system built upon a nanotechnology with known hard- and soft-error rates and error distributions? If sufficient redundancy (in the form of either hardware redundancy, time redundancy or information redundancy) is included in the lower-level components to raise the level of component reliability, then the amount of spare hardware resources required for reconfigurability at a higher level of design hierarchy could be reduced. Similarly, reconfiguration can be used in lower-level components to reduce the requirements for redundancy at higher levels of design hierarchy. Alternatively, at each level of design hierarchy, both reconfiguration and redundancy could be included. Different mixtures and strategies could result in different hardware, performance, and power consumption overheads.

**Architectural design for timing-error-tolerance:** Most process and temperature variations in nano-manufacturing are manifested as delay/performance variations in the logic. We will investigate new architectures to tolerate timing errors through addition of redundant or reconfigurable detection-and-correction modules.

The addition of redundant structures can be used to detect and tolerate timing errors. Fig. 11 gives a simple example to demonstrate the delay tolerance techniques by redundant circuitry. Assume circuit $C_1$ is a domino circuit which implements logic function $F_1 = (a+b+c)d+e = ad+bd+cd+e$. To tolerate delay variations in circuit $C_1$, we add an auxiliary circuit $C_2$ implementing $F_2 = (a+b)d = ad+bd$ and generate a new output $F_1^* = F_1 + F_2$. Because the on-set of $F_1$ apparently covers that of $F_2$ (i.e., $F_1 \supseteq F_2$), the new function $F_1^* = F_1 + F_2$ is identical to the original function $F_1$. Appending circuit $C_2$ does not change the functionality of the original circuit but does increase the tolerance to delay variations. Consider an input $(a,b,c,d,e) = (1,0,0,1,0)$ which induces transitions propagating along critical (highlighted) paths in both $C_1$ and $C_2$. Since both $F_1$ and $F_2$ have only rising transitions and feed into an OR gate, whichever rising signal arrives earlier will dominate the OR gate’s output. Therefore, the earlier arrival of either $F_1$ or $F_2$ determines the output value, not the later one. Delay variation tolerance is consequently achieved because late-arriving signals will not influence the circuit’s delay. We would generalize and extend such technique to the architecture level to tolerate delay errors due to excessive variations.

**Functional error tolerance for nano-FPGA:** Functional error tolerance on regular structures such as FPGA is more feasible than on non-regular structures. In this proposal, we plan to research functional error tolerance design for nano-FPGA. The topics include (1) the error-tolerance design of CLB blocks optimized for a given task,
characterized soft- and hard-error rates, and spatial distributions, (2) self-diagnosis for identification of faulty CLB blocks and routing resources, (3) reconfiguration mechanisms and hardware support for reconfiguration inside the nano-FPGA, and (4) synthesis, mapping, routing, and other CAD and optimization considerations for reconfiguration. We will also investigate effective methods to add redundancy at the architectural level to increase error-tolerance for both nano-FPGA and semi-regular, structured ASIC designs. Fig. 12 illustrates an idea for incorporating a programmable rectification module to tolerate functional errors. The programmable rectification module, possibly an embedded nano-FPGA, is appended to the original implementation. The inputs of the rectification module and the original implementation are connected together pair-wise, and their outputs are pair-wise connected XOR gates. If the output functions of the rectification modules are programmed to constant zeros, the combined module degenerates to the original implementation. When reconfiguration is required, the programmable rectification module can be used to mask the errors in the original module and to accomplish the new implementation. Again, we would extend this idea to the architecture level for functional error tolerance. (Note that all these can be designed using the nano components described in Section 3.2.1.1.)

On-line/off-line self-test and self-diagnosis to support reconfiguration: To support reconfigurability, effective on-line and off-line self-testing and off-line self-diagnosis techniques are required. On-line self-testing allows failure detection in the field when the system is operating, whereas off-line self-testing supports fault detection before the part is shipped out from the manufacturing line before system assembly. After detection, the defective devices/connections causing the failure should be identified through self-diagnosis before reconfiguration can take place. Although self-test techniques for ICs are available today, the assumptions of the available techniques (e.g., the added self-test circuitry has very low failure rates) are not usually valid for nanodevices. There are practically no general self-diagnosis techniques available today. Most of the diagnosis techniques are adaptive and require extensive human intervention, which will not be feasible in the production environment. We therefore propose to investigate effective design for self-diagnosis techniques for nano-FPGA with the assumption that the added self-test and self-diagnosis circuitry will have relatively high failure rates as well. In this project we will advance embedded self-test and self-diagnosis technology for detecting and diagnosing both functional and timing-related defects. We will investigate: (1) how to utilize on-chip reconfigurable, but perhaps unreliable, resources for self-test, (2) how to utilize the regularity of nano-FPGA to facilitate self-test and self-diagnosis, and (3) how to create programmable delivery mechanisms to apply, analyze, compress, and dynamically reconfigure diagnosis tests at the operational speed to locate performance-related defects.

3.2.2.2 Enable Higher Level of Abstractions

The design complexity of integrated circuit systems is outgrowing the capabilities of the existing RTL (register-transfer level) methods, and it is commonly acknowledged that the ultimate solution is to move to the next level of abstraction beyond RTL. For example, in a recent panel at the Design and Verification Conference (DVCon’2005), it was pointed out that a 50-million gate design (feasible in today’s CMOS technology) typically requires 7 million lines of RTL code [46], a complexity that significantly challenges the capacity and ability of human designers. The situation will get much worse as we move into nanotechnologies with over 100x increase of device density on chip. Therefore, we believe it is very important to raise the level of design abstraction to dramatically increase our design productivity in order to explore the full benefits of nanotechnology. Otherwise, the great integration potential implied by the nanotechnology will be wasted. Our research in this area will focus on three directions: (1) Use much more complex building blocks (as opposed to current standard cells) for system-level designs. In particular, we expect that various kinds of processor cores will be important building blocks in future nanotechnology systems that can easily accommodate hundreds and possibly thousands of processor cores in a single chip, with ample capabilities for programmability and reconfiguration. (2) Develop system-level synthesis techniques that can automatically generate heterogeneous multi-core systems with proper software and hardware partitioning. (3) Develop automatic behavior-level synthesis technologies for customized hardware.

Architecture design and exploration of heterogeneous multi-core systems. We expect that future nanotechnology systems will consist of various types of processor cores, programmable fabrics, and customized logic. The processor cores may include those for general-purpose computation (such as those used in chip multiprocessors [81]), as well as specialized cores optimized for certain application domains. Clearly, this presents a
large solution space, and we need an exploration methodology that will allow us to efficiently search for the viable solutions. We plan to use a two-step approach to tackle this problem. First, we will develop an architecture evaluation framework to evaluate and optimize a given multi-core system, where the combination of cores, their layout, and their interconnection are given. We shall consider various optimization techniques for efficient mapping of a given application to such a system, especially using a combination of processor cores and reconfigurable fabric. Once a robust framework with novel architectural optimization techniques has been developed to evaluate any given multi-core heterogeneous systems, we will leverage this framework to explore the solution space by relaxing constraints along specific directions. For instance, we can explore different layouts and interconnections, even for a given heterogeneous mix of cores using nanocomponents and circuits of different technologies. Alternatively, we can compare two different sets of cores and evaluate them under different layout constraints. Such high-level exploration and optimization will be crucial for the design and optimization of high complex systems in nanotechnologies. We shall leverage our recent work on microarchitecture evaluation with physical prototyping [27][61] and related work [72][42]. We shall also adopt interesting ideas from recent work on heterogeneous multi-core processors [67][68].

**System-level synthesis for heterogeneous multi-core systems.** Associated with the architecture design and exploration efforts outlined in the preceding paragraph, we need to develop system-level synthesis capabilities so that we can efficiently compile a complex application in a system-level description (such as SystemC [94]) into a set of tasks to be executed on various processors, a set of functions to be implemented in customized logic, as well as the communication protocols and the interface logic connecting different modules. Such capabilities are part of the so-called electronic system-level (ESL) design automation, which has been identified by Dataquest as the next productivity boost for the semiconductor industry [45][78]. As a very preliminary step, we recently developed the capability to automatically map an application specified in C language to an efficient implementation of an application-specific processor (ASIP). We assume that the base system consists of configurable processors with an extensible instruction set (e.g., as in the Nios processor from Altera [3]) which can be implemented on some programmable fabric. Our tool can automatically determine the set of customized instructions to be generated, the synthesized logic for their implementations in the programmable fabric, and the proper communication protocol. It demonstrated considerable performance gain using the automatically synthesized ASIP versus the baseline processor [23][24]. The promise and success of using ASIPs have also been reported in [52][93][103]. Clearly, much more needs to be researched and developed in the course of this project in order to provide automatic synthesis capabilities for more general heterogeneous multi-core systems.

**Automatic behavior synthesis for customized hardware.** Once certain functions are identified for implementation in customized logic, we hope to be able to automatically synthesize the customized logic from their behavior level description. In general, it has been shown that both the code density and simulation time can be improved by 10X and 100X, respectively, when moved to the behavior level [98][99]. Such an improvement in efficiency is much needed for design in nanotechnologies. Although behavioral synthesis has been a topic of research for almost two decades (e.g., [98][36][49], it has never really been widely accepted by chip designers for multiple reasons, including lack of compelling reason, lack of solid RTL foundation, lack of linkage to lower-level design, and thus, lack of quality results. With the rapid increase in design complexity and the availability of robust RTL-to-GDSII flows in two to three years (finally, by the EDA industry), we think it is time to re-visit behavior synthesis again, this time with full consideration of physical reality. Our goal is to develop novel behavior synthesis technologies to optimize logic, interconnects, performance, and power simultaneously, so that we can improve both design productivity and quality of results. For example, our recently proposed Regular Distributed Register (RDR) micro-architecture and its associated synthesis framework demonstrated the potential of integrating behavior synthesis and physical planning [21][22]. The RDR micro-architecture enables accurate prediction of global interconnect performance at early design stages and efficient support of the multi-cycle (or pipelined) on-chip communication by proper overlapping of computation and communication. The regularity of RDR microarchitecture also enables efficient layout-driven architectural synthesis algorithms. The experimentation on a number of real-life examples demonstrated significant latency improvement over the existing behavioral synthesis tools. While the RDR-based approach has demonstrated success in certain types of examples, we believe that the true potential of behavioral synthesis has yet to be realized. We plan to develop a highly comprehensive synthesis framework that takes behavioral descriptions in high-level languages as input (such as SystemC [94] or SpecC [91]), performs more aggressive high-level synthesis and optimization coupled with physical planning to optimize multiple design objectives, and generates RTL implementations together with physical constraints and timing constraints (e.g., multi-cycle path constraints) which serve as guidelines for the downstream tools.

**System-level property checking and equivalence checking.** Simulation, verification and test with such heterogeneous multi-core systems with automatic system-level and behavior-level synthesis capabilities will be very
challenging. We plan to develop a property-checking and equivalence-checking framework for verifying the correct functional behavior of the specification level using languages like C, SystemC, SpecC and System Verilog. We also plan to further generalize the framework for the equivalence checking task which verifies the equivalence of the RTL implementation to the behavior specification of the design. There are several challenging research issues for developing such a unified verification framework. In particular, proving the equivalence of the behavior model, which is not cycle-accurate, against its cycle-accurate implementation requires the integration of some design knowledge (such as the types of transformations/refinements used in synthesis and optimization) into the verification process. Furthermore, the mapping, abstraction and refinement techniques, used in existing formal verification techniques such as model checking, need to be extended to be applicable at the specification level described in languages like C, SystemC, SpecC and System Verilog.

3.2.2.3 Efficient Solutions to Fundamental Design Automation Problems

The main task of this project is to develop novel, highly scalable solutions to fundamental design automation problems to cope with the high computational complexity associated with synthesis and verification of complex systems using nano-components and nanotechnologies. In addition to UCSB and UCLA, the research teams at THU, ZJU, NTHU and NTU, who have extensive experience in developing efficient and scalable algorithms for design automation problems, will contribute extensively in this thrust. We jointly will focus on the following topics:

**Efficient High-Level Satisfiability Checking (SAT)**

Satisfiability checking is a critical component of EDA tools for synthesis, verification, testing, and diagnosis. To leverage abstractions in high-level models, satisfiability-solvers should be able to efficiently handle the multiple domains of expression present in these high-level models.

We attempt to address the satisfiability checking problem for models specified in high-level descriptions which consist of the following distinct domains of operation: (a) data-path, in which arithmetic and logic operations are performed. The data-path in a high-level description may operate on integers or use bitvector arithmetic; (b) sequential control logic, and (c) memory arrays. Prior research indicates that there is an advantage to solving an overall high-level satisfiability problem in domains other than purely Boolean [10]. Typically, each domain requires a distinct decision theory to solve a sub-problem in that domain. The decision theories for each domain can then be integrated into a whole by combining the individual decision theories under equality, known in the literature as combined decision procedures (CDPs). Several attempts have been made to combine decision theories (e.g., [89][7]). However our initial investigation [60] shows that none of the previous approaches have the robustness required to handle large and complex designs, but there is plenty of room for an approach in this matrix that handles a combination of these domains efficiently. We propose approaching high-level satisfiability problems as a combination of tightly integrated sequential Boolean satisfiability and integer arithmetic satisfiability. Based on our initial investigation, our focus will be to first develop all necessary theories, algorithms and software for complete hybrid SAT solving. Our goal is to achieve at least 10X improvement over today’s SAT solver capacity and capability. We will then publicly release the software, as we have done recently for our sequential Boolean SAT solver [44] which has been used by several semiconductor and CAD companies (including Intel, Cadence, Calypto, Freescale, etc.) and university research groups.

**Multilevel Optimization**

One important technique in dealing with the extremely high complexity inherent in nanotechnology is multilevel optimization. A multilevel algorithm builds a hierarchy from the bottom up by recursive aggregation, then interleaves optimization and disaggregation at every level from the top down. Multilevel methods have strong advantages in scalability, flexibility, and adaptability to new and complex constraints. They have been applied to many areas of scientific computing, such as computational fluid dynamics [92], signal and image processing [19], and large-scale particle simulations [47]. Recent VLSI physical design algorithms use multilevel methods as well, including multilevel partitioning [1][64], placement [12][13], and routing [25]. However, we believe that the full potential of multilevel methods has not been exploited in the field of design automation for micro and nano-systems. For example, many well-known multilevel techniques, such as algebraic multigrid (AMG), have not been used in existing multilevel physical design algorithms. All the existing coarsening schemes are based on strict aggregation (or clustering). Also, we are not aware of any true multilevel design approaches that combine synthesis, physical design, and constraint modeling (e.g., timing, routability, thermal effects) into a single multilevel framework. We will investigate how multilevel optimization techniques can be used to facilitate flexibility and reliability.

**Efficient Solver for Large-Scale Linear Systems**
A number of design automation problems, such as placement, simulation, parasitic extraction and thermal analysis, can be formulated as one for solving a large-scale system of linear equations. The computational effort for solving these linear systems, and the storage needed, have created a bottleneck for high-performance design and have slowed the time-to-market. With the advancements in design technology, linear systems generated from problems of numerical simulation or optimization usually involve thousands or even millions of unknowns. This huge number challenges the general-purpose direct or iterative solvers. We propose to develop novel and highly efficient algorithms for solving large-scale linear systems that are a result of problems related to nano-scale design automation. Initially, we shall investigate in four possible directions: (1) exploit the inherent nature of the problem to get a good approximation of the original linear system, which as a result accelerates the solution manipulation (for example, matrix-vector multiplication), (2) study the Krylov subspace iterative methods and related preconditioning techniques to adapt them to specific problems in design automation, (3) develop specific techniques for a sparse linear system using efficient storage schemes, permutation, reordering and preconditioning, etc., and taking advantage of the sparsity of data in direct or iterative solutions, (4) consider the parallel implementation of the efficient solver for large-scale linear systems.

**Multi-Space Search and Search Space Smoothing**

**Multi-space search.** Many optimization problems in design automation can be seen as one of solution space searching. An arbitrary optimization problem is composed of value space, variable space, objective space, constraint space, parameter space, and other kinds of space. We want to find a search path from the initial solution point given by a heuristic guess, or a random solution to the final highly optimized solution point. Traditional optimization methods always search solutions only in their value space (i.e., they change only variable values). It is hard to reach the optimal point for complex problems in this way. If we make search adjustments in different spaces (i.e., when the search encounters an obstacle in one search space, it jumps into another space), the search may traverse through multiple spaces and finally reach the “optimal” point. This idea is derived from the non-equilibrium thermodynamics evolution, particularly the principle that structural changes are more fundamental than quantitative changes. This idea has been exploited in global routing [62] and standard cell placement [73] with encouraging preliminary results.

**Search space smoothing.** Most optimization problems in design automation are of multi-objective and multi-constraint types. There are many local minimal points in the rugged solution space. Local search approaches can reach a local minimal point in general. Although stochastic methods, such as simulated annealing and genetic algorithms, can climb a hill to get into another valley, it is hard to reach precisely the global minimal point. Our idea is to reconstruct a sequence of problems dynamically, in which any two adjacent problems are very similar. The first problem is very smooth, with an easy-to-reach optimal solution. The final one is the original problem. The main idea is to smooth the original problem-solution space gradually and proceed to solve the easiest problem with the most smoothing solution space, continuing to the final, original problem. The approach has achieved promising preliminary results in the Traveling Salesman Problem (TSP) [48] and floorplanning [37].

For research tasks outlined in this section, we plan to involve experts in mathematics, theoretical computer science, and operations research. In particular, we plan to hold several joint activities with the NSF-funded Institute of Pure and Applied Mathematics (IPAM) at UCLA [58] and the Center of Mathematical Sciences (CMS) at Zhejiang University [11], include joint seminars and workshops, faculty and student exchanges, etc., so that the researchers in those related areas can help address the significant challenges in design for nanotechnologies. We have received supporting letters from the directors of IPAM and CMS (included as supplementary documents) which show strong endorsement to this proposal and strong interest for collaboration.

### 3.2.3 Testbed — Multi-Core Heterogeneous SOC Design in Nanotechnologies

It is not at all easy to conceive and construct a design or application at this stage to fully utilize the potential offered by nanotechnologies (with tens or hundreds of billions of transistors on a single chip). In order to explore the nanotechnology potentials and practice the DFN methodologies proposed in this research, we plan to develop a testbed (design driver) of multi-core heterogeneous system-on-a-chip (SOC) design, intended for multimedia applications, including multiple RISC CPU cores, DSP cores, and customized logic, such as an H.264 advanced video codec. The system-level diagram of the testbed is shown in Fig. 13.

The RISC cores will be designed by our collaborators from Peking University (PKU), which has gained extensive experience in designing RISC CPUs for Net-PCs in the past five years [83]. The initial design will target a performance range up to 1,000 Dhrystone MIPS, with 800+ MHz clock frequency, to be validated on 0.13µm
foundry processes. This RISC core will became one of the UniCore Processor families which include 8-bit, 16-bit, 32-bit, and 64-bit embedded processors. The heterogeneous multi-core SoC platform provides a framework for the fast creation and delivery of digital SoC devices based on the UniCore Processor families and many key peripheral IPs, including DDR SDRAM, PCI 2.2/PCI Express, 1000M Ethernet Mac and so on. This platform will provide automatic SoC IP integration flow and associated hardware, software, verification, and back-end methodologies. The DSP core will be developed by our collaborators from Zhejiang University (ZJU) [82][101]. It will provide full performance support to multiple standards in areas of video, audio, image, and voice coding/decoding. The native data type is 16-bit (integer and fixed-point fractional), and the internal data type in MAC is 40-bit (8 guard bits and 32 accumulation bits). Scalable motion estimation accelerator and 3D accelerator are also planned for integration with the core itself. The initial design validation is planned for the CMOS 0.13um process and then retargeted for 90nm and 65nm and nanotechnologies.

The H.264 Video Codec will be designed by our collaborators from National Tsinghua University [16][63]. It will support the main profile/CIF format at 30 fps. Our emphasis is on extreme low power. Fast verification for conformation to the standard is a key research topic. The initial design will consist of about one million gates and a large number of memory macros, which provides a challenge for memory BIST and SOC testing research.

The resource level required to develop such a complex SOC is very high. We will leverage the large design teams currently in place at NTHU, PKU, and ZJU. In fact, both PIs have had very productive collaborations with these teams in the last five years through another NSF-funded international collaboration effort on giga-scale SOC designs (see Section 3.5). Such a complex testbed will be very useful for practicing and validating the DFN methodologies proposed in Section 3.2.2, such as system-level synthesis for heterogeneous multi-core systems, system-level simulation and verification, and design for reliability. In fact, the multiple instances of CPUs and DSPs planned for the testbed provide additional opportunity for reliability design at the system level. We also will investigate how to design and implement such a complex SOC on the proposed nano-FPGA platform as discussed in Section 3.2.1.2.

### 3.3 Educational Plan and Objectives

#### 3.3.1 Objectives

The students participating in the IC-DFN will have first-hand experience in a truly exceptional environment. They will be pioneers in defining a design methodology for a nascent field of electronic systems that will incorporate new nanoscale device principles. As such, they must be exposed to a multi-disciplinary environment that spans the spectrum of electrical engineering, computer science, materials science, physical sciences and mathematics. Moreover, the environment of IC-DFN will also bridge simulation with experimental validation. We expect these innovations to have a profound influence on the deployment of new information technologies that will be made available in the marketplace, and we want our students to be cognizant of these aspects of their research. Finally, the research enterprise has historically been an international endeavor, and today’s economic enterprise must take place within a global context. Through IC-DFN, our students will have the chance to observe the collaborative and complementary elements of three major forces in research, technology and the global economy in the development of powerful new strategies for future information systems. IC-DFN education will foster a cross-cultural understanding that bridges areas of research activities and lays the foundation of understanding for different cultural/societal contexts, educational and economic strengths. Thus, our intention is to form an IC-DFN education program that will “produce young people who are more adaptable and flexible, as well as technically proficient,” as recommended by the National Academy of Sciences report on “Reshaping the Graduate Education of Scientists and Engineers” [87].

#### 3.3.2 Plan

**Globally-Aware Undergraduate Engineers: Aligning with the UC Education Abroad Program.** The University of California Education Abroad Program (EAP) is a distinctive and distinguished international program that provides academic exchanges and integrates international curricula and learning opportunities into UC’s degree programs. Through institutional partnerships and initiatives, EAP expands the context and content of learning by exposing students and faculty to the challenges of diverse languages and intellectual traditions, alternative approaches to knowledge, and different cultural assumptions. For a variety of reasons (including numerous detailed graduation requirements), engineering and science students do not frequently avail themselves of the opportunities provided by EAP. Yet in many ways, it is the engineering students, who will be working in a globally competitive and rapidly changing environment, who would most profit from the EAP program. We propose to attract and recruit top undergraduate science and engineering students to this program by linking EAP participation with a research internship. Students who take a prior intensive language course under the UC-Peking Joint Center for International Studies will have part of their study costs paid for by the IC-DFN, and will, on their return, join an IC-DFN research
Team-based research projects, semi-annual workshops and web seminars. IC-DFN graduate and undergraduate students will have the chance to experience team-based research projects under the mentorship of several faculty members. The semi-annual center workshops will be systematically hosted in locations rotating among the U.S., mainland China and Taiwan. We intend that the workshops will include highlight oral and poster presentations by students and young researchers. We also anticipate special sessions (perhaps panel discussions) that will focus on the social context of research and education. Topical themes might include subjects such as “Most-Valued Technical Career Options: a Cross-Cultural Dialogue.” Participation could include industrial, academic and student panelists. Other themes might include sharing insights on such topics as “Ideal Educational Programs,” “The Role of Technology in Society,” or “Funding of Science: By Whom and How Much?” We also plan to establish regular web-based seminars and teleconferences to connect students, faculty, and researchers from China, Taiwan, and the United States to discuss technical, social and cultural issues. Despite the 16-hour difference in time zones, we can coordinate web-based meetings that match late-afternoon meetings in the U.S. with morning meetings in China and Taiwan.

International internships. IC-DFN graduate students will spend at least one summer at a foreign university (either Taiwan or China) participating in meaningful research projects with that university’s researchers. If the need arises, IC-DFN will also arrange for intensive language programs to enhance the quality of their experience. Similarly, IC-DFN researchers in the U.S. will host graduate students from Taiwan or China for an extended stay of three to six months. Faculty members and their graduate students at the host universities will commit themselves to mentoring these visiting students, helping them to appreciate the culture of the host country, as well as developing their scientific talents and social interaction skills to prepare for the future global economy. We have received written commitments from all five foreign participating universities to support this activity (included as supplemental documents). Alternatively, DFN graduate students will have the opportunity to work as interns on-site in China or Taiwan, at the R&D facility of a U.S. company. This program, in addition to the typical benefits of internships, will provide students with a unique opportunity to discover the technical and business world in a very different environment, understand new working rules, and appreciate another culture without being burdened too much with learning a new language. We have received commitments to offer such internship positions from several companies in support of this program (included as supplemental documents). We will approach more U.S. high-tech companies with R&D centers based in China and Taiwan, and we expect to secure more internship positions for this activity.

Nanotechnology and the global context. IC-DFN students at both UCLA and UCSB will have access to an ongoing vital intellectual environment through weekly seminars at both locations, in addition to shared seminars as part of SIA’s Focus Center on Functional Engineered Nano Architectonics, FENA (http://www.fena.org/), which is based in UCLA with UCSB participation. IC-DFN will enhance this by providing invited speakers who will address issues of technology, society and the global context. IC-DFN will help fund guest speakers to present talks on topics such as “The Taiwan-to-San Jose Commute: Companies Straddling the Pacific,” “Brain-Drains, Visas and Technological Resources,” and “New Markets and New Legislation.”

Outreach. We will work closely with organizations such as WiSE (Women in Science and Engineering) and MESA (Mathematics, Engineering, Science Achievement) in the University of California to reach, recruit and mentor students from under-represented groups. California Nanosystems Institute (CNSI) also administers several programs that coordinate with high schools and community colleges to recruit and retain talented students from under-represented groups into science and engineering disciplines (www.cnsci.ucsb.edu/education/education.html): In the past three years INSET (Internships in Nanosystems Science, Engineering and Technology) has provided research mentorships and experiences for about 50 community college students. EPSEM (Expanding Pathways to Science, Engineering and Mathematics) provides a multi-tiered, articulated academic mentorship for talented and motivated students who, under typical circumstances, would not be matriculating to the University of California immediately after high school. Activities take place year round, and an intensive summer residential program brings new students onto campus to learn new strategies for academic success, work on team design/problem solving, develop communications, career and leadership skills, and become acquainted with research programs. IC-DFN will participate with these programs through mentorship, research, and seminar opportunities to attract a wider, more diverse group of students. IC-DFN will identify and fund speakers to present talks in the workshops and meetings organized by WiSE and MESA on nanotechnologies, as well as topics that could enhance a student’s understanding of the cultural, economic, and technical developments in Asia and the Pacific Rim.
3.4 Management Plan, International Coordination and Logistics

3.4.1 Research Team by Institutions

The center will be co-directed by Professor Jason Cong from UCLA and Professor Tim Cheng from UCSB. It will involve 21 faculty members from the U.S., Taiwan, and China. Fig. 14 shows the organizational chart.

United States:
- UCLA: Professor Jason Cong (Center co-director) and Professor Kang Wang
- UCSB: Professor Tim Cheng (Center co-director) and Professor Evelyn Hu

Taiwan:
- National Tsinghua University (NTHU): Professors Shih-Chieh Chang (Taiwan coordinator), Cheng-Chung Chi, Shi-Yu Huang, Tingting Hwang, Youn-Long Lin, C. L. Liu, and Cheng-Wen Wu
- National Taiwan University (NTU): Professors Yao-Wen Chang, Juin-Lang Huang, Chien-Mo Li, and Ric Huang

China:
- Tsinghua University (THU): Professors Jinian Bian and Xianlong Hong (China coordinator)
- Peking University (PKU): Professors Xu Cheng and Ru Huang
- Zhejiang University (ZJU): Professors Xiaolang Yan and Zhizhen Ye

Fig. 14. Organizational view of IC-DFN

3.4.2 Division of Research and Educational Tasks

To address the multiple challenges within the scope of DFN, we have structured the center along three research areas with a total of six interlocking research thrusts, each led by one or two thrust leaders. The education activities will be integrated into the research thrusts. This subsection details the division of research tasks among different participating institutions and faculty members.

- Technology and Architecture Drivers
  - Technology Driver -- Thrust leaders: Kang Wang (UCLA) and Evelyn Hu (UCSB); Team members: Cheng-Chung Chi (NTHU), Ru Huang (PKU), and Zhizhen Ye (ZJU)
    - Technology characterization, in terms of reliability, process variation, etc. (UCSB, ZJU)
    - Design of circuit blocks, e.g. memory cells, logic gates, and multiplexors (UCLA, PKU)
    - Bottom-up assembly techniques (UCSB, NTHU)
  - Architecture/Platform Driver: nano-FPGA -- Thrust leaders: Jason Cong and Kang Wang (UCLA); Team members: Yao-Wen Chang (NTU), Xianlong Hong (THU), Ru Huang (PKU) and Tingting Hwang (NTHU)
    - Exploration of reconfiguration technologies (PKU and UCLA)
    - Circuit-level design of nano-FPGA (PKU and UCLA)
Design for Nanotechnologies Thrusts
- Design for Robustness -- Thrust leader: Tim Cheng (UCSB); Team members: Shih-Chieh Chang (NTHU), Yao-Wen Chang (NTU), Tingting Hwang (NTHU), Shi-Yu Huang (NTHU), Cheng-Wen Wu (NTHU), and Xiaolang Yan (ZJU)
  - Exploring tradeoffs between reconfiguration and redundancy for reliable design (UCSB)
  - On-line/off-line self-test and self-diagnosis to support reconfiguration (UCSB)
  - Architectural design for timing-error-tolerance (NTHU, NTU, and UCSB)
  - Functional error tolerance for nano-FPGA and nano-structured ASIC (NTHU)
  - CAD issues on synthesis, mapping, and routing of reliable design with built-in reconfiguration and redundancy capabilities (NTHU, NTU, and ZJU)
- Enable Higher Level of Abstractions -- Thrust leader: Jason Cong (UCLA); Team members: Jinian Bian (THU) and Tim Cheng (UCSB)
  - System-level performance modeling and estimation (UCLA)
  - System-level and behavior-level synthesis (UCLA and THU)
  - System-level and behavior-level property check and equivalence checking (UCSB)
- Efficient Solutions to Fundamental Design Automation Problems -- Thrust leaders: Xianlong Hong (THU) and C.L. Liu (NTHU); Team members: Yao-Wen Chang (NTU), Tim Cheng (UCSB), and Jason Cong (UCLA)
  - Efficient high-level satisfiability checking (UCSB)
  - Multilevel optimization (UCLA and NTU)
  - Efficient solver for large-scale linear systems (THU and NTHU)
  - Multi-space search and search space smoothing (THU)
- Design Driver -- Leader: Youn-Long Lin (NTHU); Team members: Xu Cheng (PKU) and Xiaolang Yan (ZJU)
  - CPU core designs (PKU), DSP core designs (ZJU), and video codec designs (NTHU)
  - On-chip interconnect structure design (NTHU)
- Education -- Thrust leaders: Tim Cheng, Evelyn Hu (UCSB) and Jason Cong (UCLA); Team members: all
  - Semi-annual workshops (locations rotating among the U.S., mainland China and Taiwan)
  - Web seminars
  - Establishment of international internships
  - Partnership with UC EAP program, joint activities with IPAM (UCLA) and CMS (ZJU)
  - Providing seminars to WiSE and/or MESA and participate in CNSI's INSET and EPSEM programs

3.4.3 Schedule
The following are the major milestones that IC-DFN expects to achieve during the course of this project:

**Years 1-2:**
- Technology characterization, in terms of reliability, process variation, etc. (UCSB, ZJU)
- Design of circuit blocks, such as memory cells, logic gates, and multiplexors (UCLA, PKU)
- System-level performance modeling and estimation (UCLA)
- Studies of tradeoffs between reconfiguration and redundancy on system reliability (UCSB)
- Algorithm and tool development of high-level satisfiability checking (UCSB)
- Architectural-level design technique for timing-error-tolerance (NTHU, NTU, and UCSB)
- Component-level design, key technology research and performance evaluation on high-performance low-power CPU cores, DSP cores, and video codec needed for the testbed (NTHU, PKU, and ZJU)

**Years 3-4:**
- Exploration of bottom-up assembly techniques (UCSB, NTHU)
- Circuit-level and logic-level design of nano-FPGA (PKU and UCLA)
- Basic design flow for mapping gate-level circuits to nano-FPGAs (UCLA, NTU and THU)
- Design technique for functional error tolerance for nano-FPGA (UCSB and NTHU)
- System-level and behavior-level synthesis (UCLA and THU)
- System-level and behavior-level property check and equivalence checking (UCSB)
- Validation of testbed components through component-level tape-outs at 90nm technologies, and heterogeneous multi-core SoC platform development (NTHU, PKU, and ZJU)
Year 5:

- Completion of nano-FPGA design (UCLA, NTU and THU)
- Completion of heterogeneous multi-core testbed (NTHU, PKU, and ZJU)
- Integration of DFN design technologies into a prototyping design flow for the nano-FPGA architecture and demonstration of the results on the proposed testbed (all)
- Final project report (all)

Each thrust will have a quarterly conference call to discuss the research progress. There will be two center-wide workshops each year, rotating between China, US, and Taiwan, to review research progress, exchange results, and facilitate international education program. One of the PI/co-PIs will visit the NSF once a year to report on the overall progress made in the center for both research and education programs.

3.5 Results from Prior NSF Support


Both PIs have very relevant experience in managing a large international research program. They have been directing the NSF “Giga-Scale System-on-a-Chip Design” program, which began in 2000 and is now in its final year (after receiving a no-cost extension). The scope of this research project includes: (i) investigation and development of efficient system-on-a-chip (SOC) synthesis tools and methodologies, (ii) investigation and development of SOC verification, test, and diagnostic technologies, (iii) use and/or development of one or several SOC design drivers to motivate and validate various synthesis, verification, and test techniques developed during the course of this research project. It also has a strong international collaboration component which led to the establishment of the International Research Center on Giga-Scale System-On-A-Chip Design (ICSOC), and which also includes three collaborating universities from China (Peking University, Tsinghua University, and Zhejiang University) under the support of the Natural Science Foundation of China and two collaborating universities from Taiwan (National Tsinghua University and National Chiao Tung University) under support of the National Science Council of Taiwan. During the course of this ICSOC project, 15 faculty members from 7 universities and over 30 graduate students were involved in research. Significant progress has been made in both scientific research and international collaboration. The center has produced a total of 293 publications to date, including 73 publications involving both PIs from U.S. The complete publication list is available from the Center website at http://ballade.cs.ucla.edu/icsoc/. We shall highlight some of the research results in the following.

Architectures and Synthesis Systems for Multi-Cycle On-Chip Communication: As part of our effort in developing a communication-centric design methodology, we developed the Regular Distributed Register (RDR) micro-architecture which offers high regularity and direct support of multi-cycle communication [22] and the associated synthesis methodology that integrates behavior-level optimization with physical design [22]. Experimental results demonstrated remarkable improvement on the clock frequency and total latency. This work has generated significant interest from the research community and the industry. The PI gave invited talks or keynote speeches on this topic at a number of conferences and workshops, including Int’l Symp. on Physical Design, (April 2003), Int’l Rapid System Prototyping Workshop (June 2003), Intel Physical Design Research Symp. (Aug. 2003), and Int’l Conf. on Hardware/Software Codesign and System Synthesis (October 2003).

Optimality and Scalability of Circuit Placement Algorithms: We made significant progress on the optimality study of existing placement algorithms for integrated circuits. Our result, published in [14], showed for the first time that leading placement tools from both industry and academia are 70% to 150% (!) away from the optimal solutions in terms of the total wirelength using a set of placement examples with known optimal (PEKO examples) constructed in their research. After this significant result, we further extended the optimality studies to develop more general benchmarks (the PEKU examples, reported in [29]) and for timing optimization (T-PEKO examples, reported in [30]). These results have generated great interest from both the research community and the industry. These studies were covered in the EE Times multiple times on February 5, 2003, April 4, 2003, and Nov. 13, 2003. PEKO/PEKU/T-PEKO examples have been made available on the Internet and downloaded by researchers and engineers from over 250 companies and universities worldwide. They have now been used in almost every placement paper submitted or published since the summer of 2003.

CSAT - A Circuit SAT Solver for Functional Verification: In this project, we have developed an entirely different SAT solver design concept that is circuit-based and employs Signal Correlation Guided Learning (SCGL). Our solver, named CSAT, is able to utilize circuit topological information and signal correlations to enforce a decision ordering that is substantially more efficient for solving circuit-based SAT problem instances. In particular, for unsatisfiable circuit examples, which are very common in functional verification applications, our solver is able
to achieve 10x-100x speedup over state-of-the-art SAT solvers. This has been validated in many real industrial examples by the industrial users of CSAT. The source code of CSAT has been publicly released under the BSD license model. Several other universities and research institutes have also used it for their research in verification. Intel has successfully integrated the software into its FV formal verification system for production use. Cadence and Freescale are including the package in their verification systems as well. A start-up company, Calypto, has also adopted CSAT as their core solver for the verification system they are developing.

**A Statistical Delay Testing Framework for DSM Defects:** We have developed a new statistical timing analyzer that can be used for analyzing the circuit performance in the presence of delay faults caused by manufacturing defects, excessive noise, process variations, modeling errors. The software can be applied for vectorless static analysis as well as for dynamic simulation with given input vectors. We have successfully demonstrated how to use our statistical timing analyzer as a core engine for selecting critical paths considering power supply noise effects on the path propagation delays. We were the first group to apply such statistical analysis to path delay testing and have produced results for more intelligent path selection and better test generation strategy under the statistical timing models. The source code of the statistical timing analyzer has been publicly released. A paper reporting partial results on this project won the Best Paper Award in Test Category at the 2003 Design and Test Conference in Europe (DATE 2003).

We have also made great progress on establishing close international collaboration. We held ten workshops during the course of this project in United States, China and Taiwan which attracted many local participants including graduate students and young researchers. Each workshop includes 1 to 2 days of technical presentations followed by another day of one-on-one discussions interleaved with a social program. Over 30 students from UCLA and UCSB attended these workshops, and a significant portion attended our workshops overseas. The workshop presentations are posted on the center website. Moreover, over 10 students and faculty members from UCLA and UCSB visited the foreign universities for research collaboration. We also hosted the visits of over 10 students and faculty members from the foreign participating universities at UCLA and UCSB.


The Internships in Nanosystems Science, Engineering and Technology (INSET) program brings science and engineering community college undergraduates to the UC Santa Barbara campus for a summer research experience. Interns gain first-hand experience in scientific investigation in a dynamic, collaborative research environment. They are matched individually with UCSB faculty and graduate student lab mentors who provide training and support. Interns attend weekly meetings, special seminars, and have the opportunity to develop their presentation skills throughout the summer. Strong and continuous interaction with community college faculty and administrators assures effective recruitment of students and ongoing communications during the full academic year. Community college staff will participate in the group meetings during the summer. Over the first three years of this program, nearly 50 student internships were awarded, involving students from Santa Barbara City College, Ventura College, Allan Hancock College, Pasadena City College, Contra Costa College and Santa Rosa Junior College. We recently found that the program has been renewed for another three-year term. Additional information may be found by accessing: [http://education.cnis.ucsb.edu/inset.html](http://education.cnis.ucsb.edu/inset.html).


Photonic crystal device elements have a high potential for structuring the optical systems fabric at the nanoscale through novel control of optical emitters, low-loss waveguiding, and high-resolution filtering of wavelengths in tailored optical dispersion properties. To form truly effective building blocks for real optical systems, such photonic crystal device elements have to be characterized, optimized and integrated within the fabric of large optical systems. Our work comprised one part of a multi-faceted photonic integration approach. Professor Hu’s group, in close collaboration with Professor Blumenthal, carried out the initial development of the simulation, fabrication and characterization of photonic crystal waveguides in InGaAsP. Major achievements of the program: (1) development of appropriate 2D and 3D simulation tools to predictively model the performance of photonic crystal waveguides, (2) development and optimization of photonic crystal fabrication, and fine-tuning the process to accommodate photonic crystal waveguide integration into existing device platforms, (3) initial characterizations of transmission and coupling characteristics of photonic crystal waveguides and 1- to 3-line defect waveguides, (4) design and implementation of different tapering structures in the photonic crystal waveguides to achieve optimal matching between those structures and more conventional ridge waveguides, and (5) setting the infrastructure of understanding in place to look at highly dispersive transmission of optical signals at the band-edge of the photonic waveguide (work that is currently being undertaken).