Grand Challenge Themes
Information Technology R&D
Shankar Sastry

Architectures for a Billion Transistor Chip

Huge advances in photonics, MEMS, sensors, actuators, optoelectronics, digital, analog and mixed signal processing and new fabrication technology have enabled us to conceive of integrated designs in 3-D on a chip with billions of transistors. Clock delays across the chip will be so large that it will be critical to design new and modular architectures for these chips. While we are redesigning architectures with attendant considerations of fault tolerance and software reconfigurability, we will also be able to integrate security into the architecture using a number of promising new ideas that had been discarded as having too much overhead in the past. A critical features of DoD investment in this grand challenge is a current strategy to meet DoD's embedded computing needs on a number of computationally-intensive platforms, while harnessing market forces that work on commercially-driven processor needs. Programmability of these new architectures may include novel approaches, such as amorphous computing methods.

Beyond the Silicon CMOS Barrier

While the exact dates when we will hit physical limitations for 2-D electronics CMOS technology are a matter of conjecture, it is clear that we will hit them soon, Thus, it is important to begin research in new computational substrates, such as quantum, biological substrate, moletronics, smart fabric described below:

- *Quantum computing, communication, and information science* holds the potential for:
  - Ultrasecure communication over optical backbone
  - Exponential speed up in algorithms for unsorted data base search, factoring large numbers
  - Quantum computers that can give detailed and faithful simulations of string theoretic phenomena and molecular processes

- *Biological substrate: DNA/RNA/Protein computing* holds the potential for:
  - High Volume content addressable storage
  - Solutions to computationally hard problems

Self assembly of nano-structures using DNA/RNA tiling. The nano-structures in turn could be used for nanoscience, such as moletronics.

- *Smart fabric*. By using textile technology to interweave battery, fiber optic cable, and metal connectors, we can produce fabric that can be embossed with processors to provide on-person processing on the order of 10s of teraflops (the size of today's larger super computers).

- *Moletronics*. Using computation at the molecular scale we can conceive of extremely fast and extremely high density processing power for the next-generation of strategic computing for the warfighter.

Bio Computation

- *Biomimetic architectures and secure computation*. The raw processing power of computers is approaching that of the neuronal system. However, the architecture of the brain is very different from the von Neumann model.
  - Take for example the vision system that accounts for between 50 and 80 % of the mammalian neo-cortex. Processing from about 10 million neurons on the retina compresses the information to 1 million optical fibres, but the visual cortex stores over 100 filtered versions of the image. Processors in memory and the tradeoff of massively redundant storage are key features of the visual system that enable visual processing.
  - The human neural system is remarkably secure and fault tolerant. By building in a "secure" architecture it is very difficult for "worms" or "viruses" to infiltrate our memory or processor. In addition performance degrades gracefully.
  - DNA/RNA/protein is a programming paradigm for biological systems. Software of the future will have the characteristics of DNA to allow for composability, adapatability, scalability, and fault tolerance on a truly biological scale.

We will build with algorithms and architectures computers with 10-100s of petaflops capability that will have the functioning of neuronal systems.

- *Humanoid robots*. Humanoid robots that can interact with us in natural ways including gesture, speech and vision as well as the ability to perform with minimal intervention. Applications including repairs in space and possibly combat.

- *Rapid bug-to-drug*. Computational models to rapidly identify functioning of new strains of viruses or pathogens and to formulate targets for the selection of therapeutics based on mathematically quantitative control models of the functioning of gene-protein circuits, signaling mechanisms, immune responses. This part of the bug-to-drug agenda would identify mechanisms for infection and targets. The *Human on a Chip* research agenda described below would complete the process.

Human on a Chip

It will be important to develop a hybrid biological, electronic chip that captures enough properties of human responses to the environment so that drug testing/screening can be done much more inexpensively. Thus, the effects of long-term low-level exposure to toxic substances can be determined without expensive and inconclusive in-vitro or in-vivo toxicity studies. Public health, pharmaceutical industry and DoD's Chem/Bio programs would benefit from having this tool to determine side effects of drugs. The research could be expanded to account for genetic differences in humans and determine genetic variance effects.

When combined with computational biological methods, such as Biospice, for the discovery of mechanisms for infections and targets for therapeutics, this technology will greatly speed up the fielding of new drugs

Software for the Real World (Embedded Software)

There is a great need for *embedded software*, which is software operating with and controlling the physical world. The problem is hard because commercial B-to-B and enterprise software has only an idealized model of the real world. When one studies why there are cost overruns on every new procurement, e.g., from FCS to UCAV, TBMD to NMD, and F-22 to JSF, it is largely due to the vast under-appreciation of the cost required to design, verify, validate, and certify the embedded software. The problem is especially acute for aging military platforms that must be updated to maintain their superiority over long lifetimes (for example B-52s are expected to last 70+ years as airframes). While the DoD is a key stakeholder for embedded software, tremendous commercial innovations and spin-offs are also necessary for competitiveness in

1. Commercial avionics and automotive electronics (where it is predicted that the cost of the computers and embedded software will exceed the drive train, body, etc.by early 2003)
2. Consumer electronics such as PDAs, cell phones
3. Copier/printer and FAX machines,
4. Televison and
5. Process control for chemical and industrial manufacturing processes.

It is extremely important for the DoD to pay close attention to the consumer electronics market sector. In particular, as the IT market continues to be commoditized all over the world (i.e., as functionality grows and prices plummets), the essential source of DoD and US superiority will come from increased functionality of the embedded software on board. Thus, as hardware gets commoditized, we must stay ahead by more rapidly introducing new embedded software functionality that exploits the hardware.

Four thrusts are critical to success in embedded software:

- *Automated design, verification, and validation*. Current embedded software design practices are stove-piped, with different engineers and software designers working sequentially in different domains. For instance, in the avionics domain a weapons software engineer works on networked fires, sensors, counter measures, etc., a guidance and navigation control engineer works on the flight dynamics, and a propulsion engineer works on the engine software. Each engineer has specific domain expertise, but seldom a clear understanding of hardware, operating systems, and networking issues traditionally implemented by the computer scientists and IT software development teams on the project. We need design practices that allow simultaneous design and propagation of constraints among these different domain specific design teams, which enable

1. *Verified design*, in a mathematical or formal sense
2. *Validated design*, in an engineering sense, and
3. *Certifiable design*, to allow regulatory agencies to certify.

- *High confidence systems*. These are systems for human-centered automation, such as flight control systems, combat systems, and early warning national missile defense systems. A key concern with these types of mission-critical systems is the fragility of their software and their ability to be compromised by security breaches and denial of service attacks. Important challenge areas in high confidence systems and software include:

4. *Narrow-waisted middleware*. The tremendous success of the Internet was the standardization of IP protocols, which allowed for large variability in the underlying transmission physical layers (optical, ATM, Ethernet, etc.) and diverse application layers, which used the same abstractions of the network because of the abstraction of the IP layer. We need to create narrow-waist middleware to allow for a diversity of lower level operating systems and networking protocols to present stable abstractions to higher level application and service layers. The middleware should address multiple considerations in qualities of service, qualities of information assurance, etc.
5. *Security and composable operating systems*. We need operating systems of varied size footprints to support a wide spectrum of applications, ranging from PDAs to routers to servers, with modularity and assurance of multiple levels of security.
6. *Tamper-proof software*. One way of protecting hardware is to make it tamperproof. If superiority is encapsulated in embedded software, it must be made tamper proof as well.

- *Generative programming*. A fundamental difficulty in creating software for embedded systems is the large number of interdependent design concerns, constraints, and the massive amount of details that influence the structure and composition of the code. Even if the vast amount information used for the design, verification, and validation of embedded systems is captured in the form of models by design automation tools , the current relationship between the models and the code of the embedded software would be only loose and indirect. Generative programming is a new software paradigm that automatically manufactures highly optimized code from elementary, reusable implementation components using high-level design models by means of domain-specific configuration knowledge. A central issue in generative programming is the specification and synthesis of generators, i.e. programs that take high-level design models and produce efficient and correct implementations.

- *Intelligent Microsystems*. Intelligent Microsystems are a new class of highly adaptable, highly integrated components (micro-systems) with the ability to self-assess and adapt in real-time, optimizing their micro-level performance and providing new levels of macro-level functionalities to meet the needs of next generation of military sensor and weapon systems. Conceptually, intelligent microsystems can be thought of the inorganic equivalent of higher level living organism. These organism have two levels of intelligence.

1. At the microsystem level, the interaction with environmental factors is autonomous, i.e., decision-making is done with no interaction from higher level thinking (examples include pulling away from hot surfaces, adjusting pupil size to light conditions, or spiking hormone levels in response to fearful situation). In these cases the technical objective is to have the module or sub-system respond to and control its operation in the face of varying conditions (temperature, noise, available power, signal strengths etc).
2. At the macrosystem level, how can modules and sub-systems be integrated together to create distributed functions (computing, sensors, actuators etc). Such a capability involves understanding how to reliably integrate and control multiple units of multiple types under a wide variety of operational conditions. This would not only allow complex, distributed systems, but would provide adaptability and redundancy (in event of non-functional units). The "intelligence" under such situations is analogous to the higher order thinking that is done to not only collect data and respond locally when necessary, but to analyze the data and make changes based not only on what has happened, but what may happen and to adjust accordingly.

Ubiquitous Computation

In the post-PC era we are evolving into a world of ubiquitous or pervasive computation and sensing. In this era, we are surrounded by computational elements and sensors embedded in the environment around us. We are in the computational medium and surrounded by it. When this trend is fully implemented and deployed, it will have a dramatic impact on our military doctrine, which is currently based on decision making in the face of poor or incomplete information. Key areas to be addressed in being able to harness this computational power are:

- *Oceanic databases*. In a world of distributed sensing and computation and data storage, it is critical to provide users with consistent and current views of the world. It is important that queries not be directed to specific locations but to ask for information and the "oceanic data base" flow appropriately to the query.
- *Secure collaborations*. Different levels of trust need to be dynamically determined and the ability to seamlessly collaborate across these coalitions is critical.
- *Natural user interfaces* including speech, gesture, vision and other modalities to allow for natural interaction with a pervasive computational environment.

Augmented Cognition

While the speed, bandwidth and processing power of the computational world is growing exponentially around us, the cognitive ability of humans is not. Phrases like "information overload," non-availability of "human cycles" are now routine. These problems are occurring not only because of poor user interfaces. A research agenda is needed to use computation to enhance the cognitive abilities of humans. In some sense, we are all "computationally disadvantaged" or handicapped. Some immediate agenda items pertaining to augmented cognition include:

- *Augmenting human memory* using artificially generated pointers to trigger the correct retrieval.
- *Real-time visual processing* of video data to store compact versions of complex scenes for later recall on demand by information content.
- *Use of multiple modalities* to prevent warfighter, pilot, commander work flow overload.
- *Use of analysts* doing over the shoulder situation analysis and ready to provide suggestions when called for.