

*Defense Science Board
Task Force*

On

*HIGH PERFORMANCE
MICROCHIP SUPPLY*



February 2005

**Office of the Under Secretary of Defense
For Acquisition, Technology, and Logistics
Washington, D.C. 20301-3140**

APPENDIX D. FUTURE TECHNOLOGY DEVELOPMENT

The Department needs to secure continued “Moore’s Law” improvements in processing capacity that will enable it to maximize the advantages inherent in its superior sources of information and the superiority of the algorithms and networks that are used to process and benefit from them. Even a “level playing field” in which “standard” programmable microchips of ever increasing capacity are available to both the DOD and its adversaries works to U.S. advantage, especially if the suppliers of those parts are U.S. based. In contrast, if the microchips available are not powerful enough to take full advantage of superior sensor, actuator, algorithm, networking and systems capabilities, U.S. superiority will rapidly erode.

Historically, the rapid rate of growth in U.S. microchip capability resulted from a robust national portfolio of long-term research that incorporated both incremental and revolutionary components. Industry excelled in evolutionary technology developments that resulted in reduced costs, higher quality and reliability and vastly improved performance. DOD now is no longer perceived as being seriously involved in – or even taking steps to ensure that others are conducting – research to enable the embedded processing proficiency on which its strategic advantage depends. This withdrawal has created a vacuum where no part of the U.S. government is able to exert leadership, especially with respect to the revolutionary component of the research portfolio.⁴³ The problem, for DOD, the IT

43. This development is partly explained by historic circumstances. Since World War II, the DOD has been the primary supporter of research in university Electrical Engineering and Computer Science (EECS) departments, with NSF contributing some funds towards basic research. From the early 1960’s through the 1980’s, one tremendously successful aspect of the DOD’s funding in the information technology space came from DARPA’s unique approach to the funding of Applied Research (6.2 funding), which hybridized university and industry research through a process that envisioned revolutionary new capabilities, identified barriers to their realization, focused the best minds in the field on new approaches to overcome those barriers and fostered rapid commercialization and DOD adoption. The hybridization of university and industry researchers was a crucial element; it kept the best and the brightest in the

industry and the nation as a whole, is that no effective leadership structure has been substituted. Instead, research in these fields is managed through a hodge-podge of programs spawning numerous government agencies. The President's budget includes "cross-cuts" of the government's Nanotechnology (the NNI) and IT (the NITRD) research investments, each of which is stitched together by committees representing the participating agencies. However, there is no unified source of leadership that can mount revolutionary programs, let alone ensure that the DOD's future requirements for programmable microchips will be met.

While it is tempting to believe that the future capabilities DOD will require will emerge solely from the private sector, most commercial entities can only engage in relatively short-term incremental R&D on their own. Although the industry has pooled its resources to support a limited degree of long-term university research (e.g., through the SRC), this is far from sufficient to meet DOD's needs. Similarly, it may be tempting to believe that NSF funding will be sufficient to sustain information superiority. Although NSF funding has risen and the NSF does fund more longer term incremental work than industry, the NSF does not typically support development of revolutionary capabilities, along the lines of those achieved by DARPA's 6.2 programs.

university sector well informed of defense issues and the university researchers acted as useful "prods" to the defense contractors, making it impossible for them to dismiss revolutionary concepts whose feasibility was demonstrated by university-based 6.2 efforts that produced convincing "proof of concept" prototypes. As EECS grew in scale and its scope extended beyond DOD applications, a "success disaster" ensued in that EECS essentially "outgrew" the ability of the DOD to be its primary source of directional influence, let alone funding. Furthermore, DOD never developed a strategy to deal with this transition. With pressures to fund developments are unique to the Defense (e.g., military aircraft, tanks, artillery, etc.), the DOD withdrew its EECS research leadership. Recently, DARPA has further limited university participation, especially as prime contractors, in its Computer Science 6.2 programs, which were by far its most significant investments in university research (vastly outstripping 6.1 funding). These limitations have come in a number of ways, including non-fiscal limitations, such as the classification of work in areas that were previously unclassified, precluding university submission as prime contractors on certain solicitations, and reducing the periods of performance to 18-24 months.

Past investments in research would appear to secure Moore's Law scaling through 2009 or 2011. However, beyond that, the microchip industry faces three key challenges, each of which puts DOD's information superiority at risk, but also poses opportunities for DOD to tilt the economics of non-standard (e.g., low volume) microchip manufacturing in its favor:

- There is a significant technology gap in the 2013-2019 time frame, as transistor critical dimensions shrink below 10-12 nm. While the continued use of charge-based devices (devices based on the movement of electrons) remains feasible during this period, there is an urgent need for revolutionary approaches to development new operating principles and materials for use in those devices. Surprisingly, very little of the U.S. government's National Nanotechnology Initiative (NNI) is addressing this issue – yet the window for the timely conduct and insertion of such research is rapidly closing. The need for new technologies for use in ten years may also present an opportunity to contain the capital costs associated with microchip manufacturing. For example, one alternative that has been suggested involves the use of chemical processes to create nano-wire / nano-dot devices in bulk and to use self-assembly techniques to create small clusters of “pre-wired” devices and precisely position them on a microchip substrate. Taking this line of reasoning a step further, it may be possible to use what would then be “legacy” 10-20 nm lithography technology to wire together these self-assembled clusters. Since the “legacy” lithography and associated mask-making equipment would by then have been substantially depreciated, the capital equipment costs associated with such a manufacturing approach might not be as prohibitive to the DOD as they are today. Direct imprint lithography is another alternative that is being

investigated, though its viability at the finest feature sizes remains uncertain.

- New approaches to the architecture of programmable microchips and to their programming concepts are required to extract benefit from the huge numbers of transistors made available by Moore's Law advances. The design of super-complex chips is being influenced by three transitions, each of which presents risks and opportunities. First, standard CPUs and DSPs are evolving from single core (i.e., processor) devices to multi-core devices.⁴⁴ Secondly, standard Programmable Gate Arrays (PGAs) are undergoing a transition in which PGA fabrics are combined on the same chip with CPUs and DSPs. This process, which is to the DOD's advantage, will likely intensify with the advent of multi-core processor architectures described above. Finally, the sustained use of sub-wavelength lithography over coming generations of transistors will push designers of complex logic towards regular transistor patterns. This may reduce the gap between PGAs and other forms of logic, further accelerating the combination of PGAs with many-core processors.
- In the early 2020's, thermal noise limitations suggest a need to transition beyond charge-coupled devices

44. Initially, this will be a relatively mild transition as the components of dual processor systems are integrated onto a single chip. However, as the number of cores scales up to hundreds and then thousands of cores per die, the impact of this transition will be far more pronounced. Going forward, the use of parallel independent computational elements (vs. frequency scaling, pipelining, etc.) will be the means of extracting increased performance from these microchips - not just for supercomputers, but for desktops, handhelds and embedded systems as well.. **Arguably, this is the computer industry's largest transition since the adoption of the microchip - and possibly since the invention of the stored program computer.** Key challenges arising from this transition have to do with: the on-chip networks interconnecting the cores; Memory and I/O bandwidth to feed these microchips; and the programming environments that will allow their capacity to be harnessed by large numbers of programmers / users.

to those based on other physical principles, such as spin. Since these new technologies are unlikely to instantaneously transition to high volume manufacturing, it is likely that there will be some period during which their economics could favor lower volume applications, presenting another opportunity for the DOD to re-assert leadership and gain strategic advantage. Conversely, ceding leadership in this transition to nations that are potential adversaries would certainly undermine our information superiority – and may also undermine a key economic component of our national security.