

INFORMATION SYSTEMS TECHNICAL ADVISORY COMMITTEE (ISTAC)

January 26, 2011 - 9:00 AM to 4:00 PM

SPAWAR, Bldg. 33 Cloud Room, San Diego CA

MINUTES OF MEETING

Agenda Item Presentations/Discussions:

PUBLIC SESSION (January 26, 2011)

The meeting opened at 9:00 AM. Approximately 35 people were in attendance.

Opening: Jonathan Wise opened the meeting, and Mike Wood of SSC provided a brief overview and welcome. This was followed by introductions and request for comments from the public. There were no public comments.

Working Group Reports: Jonathan Wise reported on the activities of the Category 3A working group; there was no report on Category 3B; Henry Brandt reported on Category 4; Frank Quick reported on Category 5 Part 1; and Roz Thomsen reported on Category 5 Part 2. Key points from the Working Group reports were:

Cat 3A: Current efforts are on cleanup and technical corrections for 3A002 controls on microwave instrumentation.

Cat 4: This group has held teleconferences pertaining to fault-tolerant computing. There was also a presentation on a proposal to delete entry 4A003.a in the November 3, 2010 ISTAC meeting. This group has also arranged for Intel to present its periodic roadmap update (see details below, in these minutes).

Cat 5p1: There are several issues under discussion, but at this time comments are limited to closed session.

Cat 5p2: The recent EAR update (76FR1059) update on executable encryption code derived from publicly available source code is appreciated. Additionally, this group has submitted several proposals for consideration for the 2011 Wassenaar cycle.

Intel Roadmap: This presentation consisted of two parts: Greg Taylor spoke on the technology roadmap, and then Joseph Kim raised an export control question pertaining to MMIC amplifiers at ~60 GHz for emerging short-range communications applications.

The key point of the roadmap is that Intel continues to explore novel technologies to ensure the continuance of Moore's Law. Aspects include manufacturing technologies (current manufacturing technology is at 32 nm, and the roadmap looks to 8 nm in 2017), use of III-V materials, and Performance/Aggregation.

Manufacturing:

Node progression continues on a slope of 0.7x every two years. Intel is currently using immersion lithography, and is not yet using EUV.

The Tick-Tock model continues: a new technology (manufacturing process node) is introduced by migrating a prior (proven) architecture to the new process, and then a new architecture is developed on new process. That is, the focus alternates between new manufacturing technology and new architecture.

The consequence of smaller process nodes is compaction: processor cores become both smaller and more capable. As an example, whereas the Xeon 5500 (45 nm process node) occupied 263 mm² and contained 731M transistors, 4 cores and 8 MB cache, and the Westmere-EP (32 nm process node) occupies 240 mm² and contains 1.17B transistors, 6 cores and 12 MB cache, an improvement of 1.5x in cores and cache and 1.8x in areal transistor density.

III-V Materials:

Intel continues to use a larger fraction of the periodic table (more elements) in the manufacturing process. Examples include GaAs, InAsSb, AlInAsP.

The advantages of III-V materials are higher speeds and lower power, but III-V substrates are impractical for high-volume CMOS production. The challenge, therefore, is to leverage the benefits of III-V materials on a silicon substrate, as a silicon process evolution, not as a revolution to compound semiconductors.

Performance/Aggregation

Estimates are that 64-socket systems will reach 20-25 WT and that even 8-socket systems will reach 2-3 WT by 2013, greatly exceeding the current export-control threshold of 0.75 WT.

HPCs are increasingly being designed using off-the-shelf components. Interconnect speeds and software continue to be significant limiting factors on the practical use of HPCs. Future high-performing multicore processors may rival current HPCs.

GPUs are becoming integral to system performance (more tightly integrated to the CPU).

The Chinese supercomputer Tianhe-1A is rated at 2.5 petaflops. It uses 7,168 Nvidia M2050 GPUs and 14,336 Intel CPUs.

Joseph Kim continued on ~60 GHz applications

There are many emerging applications for MMICs at 60+ GHz, including WLAN/WPAN (~60 GHz), automotive anticollision radar (77 GHz), imaging, both medical and security (>90 GHz). The utility of 60 GHz spectrum is that it can support high bandwidth (and thus high data transfer rates), it is unlicensed, and that it has limited penetration (short range, and relatively quiet/noise-free).

Particularly for the ~60 GHz space, there are two key standards: 802.11ad and 802.15.3c, and an emerging Chinese standard CWPAN.

There are also two major consortia, WiGig and WirelessHD.

The 57-64 GHz spectrum has been allocated and unlicensed by many authorities: 57-64 GHz by US, CA, KR; 57-66 GHz by EU; 59-64 GHz by CN; and 59-66 GHz by JP.

Export controls that might apply are: 3A002b2 (MMIC amplifiers), 3E001/5E001d (technology for MMIC amplifiers), 3A002d (signal generators related to microwave radio), 3E001/5E001c (technology for microwave radio), 5A001d/5E001 (phased array antennae and related technology), 6A008 (low power radar).

The export control concern is that existing controls might not adequately address ~60 GHz devices, and that clarification might be needed to provide decontrol mechanisms for them.

It was agreed that the ISTAC will take this up as a topic in 2011. Jonathan Wise will coordinate with Joseph Kim and Dave Robertson, and also seek to engage others in the industry. This topic will tentatively be discussed further during the July 2011 ISTAC meeting.

Godson Microprocessor and Suggestions for Cat 4 Controls: Chuck Moore of AMD made a two-part presentation: first, a report on the Godson microprocessor, and second some suggestions for approaches and modifications to the Cat 4 export controls.

The Chinese Godson (Longsoon) microprocessor project began in 2001 with university funding; the goal was to develop concepts, skills and confidence. It culminated with demonstration of a 1 GHz, 4-way OOO superscalar, 15 W, MIPs architecture-compatible core.

Funding levels are significant. For 2006-2020, funding is US \$10B each for “CPU and OS” and “VLSI technology”, plus US \$10B each for 14 other projects that make use of advanced computing technology. That is, “Desktop, Server and HPC products based on a domestic-designed CPU.”

The next goals are: Significantly increased levels of integration (16-core CPU; CPU+GPU single-chip solution); Architectural extensions for vector processing; Architectural extensions for efficient x86 architecture emulation; Scalable coherent interconnect to build larger SMP nodes; Improved Reliability/Availability/ Serviceability characteristics; Scalable cluster interconnect to build large multi-node supercomputers (with intent to build a 10 PF HPC system by 2013/2014).

Summarizing, key retrospective points are 1) Domestic microprocessor design capability in China is rapidly catching up to US capability, with potential intercept of US processor/system capability by the middle of this decade; 2) Impressive investments in the use of computers for science and technology; 3) Advanced computing capabilities are likely to be available worldwide despite the Wassenaar export controls.

The second part of the presentation addressed AMD’s perspective on export controls on HPCs.

AMD’s Accelerated Processing Unit (APU) strategy is integration of multiple CPU and GPU onto a single chip. This enables high performance for traditional (double-precision) workloads and also high performance processing for 3D graphics acceleration. It also enables augmentations for data parallel compute offload to the GPU (which in this case is acting as a vector processor).

APU technology is the current stage in the sequence of drivers for computing performance. The 1990s was the single-core era, in which there were steady performance increases due to increasing clock frequency. The 2000s was the multi-core era, in which there was reduction in per-core performance but increases due to doubling of the number of cores every few years. The 2010s are

the heterogeneous era, in which increasing core count offers diminishing returns and instead there is a step increase followed by steady increase due to data parallel compute offload.

APU technology results in low-cost commodity hardware. Export control definitions and control thresholds will need to be adjusted to address APU.

AMD estimates that by 2015, single chip performance will extrapolate to 0.4 WT for discrete CPU, to 4.0 WT for discrete GPU, to 0.1 WT for client CPU, and to 3.0 WT for high-end APU.

AMD continued with three suggestions pertaining to existing or proposed export controls.

First, they recommend modifying existing APP Aggregation Note 4 as follows: “Do not include processors that are limited to input/output and peripheral functions (e.g., disk drive, communications, graphics, audio processing, image processing, video processing, visualization and video display) when calculating APP.”

Greg Taylor opined that expanding the list of exclusions might not help, because the note contains the qualifying phrase “limited to”, which sets a high standard that few functionalities achieve. Specifically, if a GPU has open programmability, then it cannot be said to be “limited to” any particular functionality.

Dave Robertson added that the list is illustrative.

Chuck Moore agreed that some chips do have open programmability, and agreed that that is of concern.

It was decided that no action was needed on this.

Second, they disagree with the proposed APP Aggregation Note 8 (clarification of treatment of GPUs for APP aggregation) on the basis that neither “integration onto a single die” nor “mutually coherent or unified caches/memory” are required to gain access to data parallel computing capabilities in an attached device. Rather, these are merely optimizations designed to reduce overhead and/or improve programmability, and should not be defined as feature thresholds for the more general use of attached devices for computing.

Henry Brandt responded by agreeing that these features are only proxies for the actual parameters of concern, but that they were chosen pragmatically for ease of use in a regulatory context.

Third, AMD suggested relaxations to the APP control thresholds, as follows:

2011: 3.0 WT if GPUs are aggregated, otherwise 0.5 WT

2012: 4.0 WT if GPUs are aggregated, otherwise 1.0 WT

2015: 10 WT if GPUs are aggregated, otherwise 3.0 WT

Additionally, they questioned the weighting factors in the APP formula (currently 0.3 for CPU and 0.9 for GPU/VPU), noting that either kind of system can achieve 0.85-0.9 on linpack, but that real-world performance on both kinds of systems is generally much lower.

David Lindsay commented that these values (0.3 and 0.9) were initially suggested by DoD, based on actual performance data.

Al Courduff explained that these values were chosen 5 years ago based on coherent vector systems, and are not necessarily appropriate for modern CPU-GPU systems.

Chuck Moore added that memory bandwidth is at least as important as peak compute value when considering real workloads.

Henry Brandt explained that there is widespread agreement that memory bandwidth is the relevant technical parameter, but that despite many years of trying, nobody has yet figured out how to codify it as an export control regulatory parameter. The challenge is that determination of memory bandwidth tends to require complex proprietary data that is not available in datasheets.

Chuck continued by noting that programmability challenges are important for GPU/VPU and that Amdahl's Law plays a significant role in real workloads. The GPU/VPU does well on computation kernels but poorly on most other workloads.

Chuck suggested that consideration of CAGR (Compound Annual Growth Rate) would be useful in export controls, given that computing capability continues to increase and that there is a long lead time to implement changes to the export regulations. Henry Brandt noted that indexing of the control thresholds has been suggested many times, but has never been accepted.

Finally, Chuck suggested that software might eventually become more important than computing node performance.

Summarizing, the key points were: 1) Global competitors are catching up to US computing capability; 2) The computing industry is approaching a step function (increase) in APP following by new sustainable CAGR. 3) It is recommended that the APP control thresholds be raised.

No specific actions were needed as follow-up to this presentation. AMD will consider whether and how memory bandwidth might be used as a control parameter.

EDA Overview: Larry Disenhof and Dan Page, representing the EDA Consortium, gave an overview of the EDA Consortium, of EDA software, and made some comments responding to a proposal for entry 3D3.

Electronic Design Automation (EDA) is the software, IP and tools used to design electronic products.

The EDA Consortium is an association of EDA and IP vendors, founded in 1989. Its mission is to work together to resolve issues of common concern, where cooperation benefits both vendors and customers. Some key areas of interest are: export controls, industry OS roadmap, anti-piracy solutions, industry trade shows.

With regard to export control, in 2003 the EDA Consortium initiated a proposal for modification to entry 3D3, which resulted in the controls that exist today (controls on physics-based modeling software). The existing 3D3 control language is

considered by the EDA Consortium to be good, because engineers understand it and can readily determine what is and what is not controlled.

Dan Page continued by providing an overview of Optical Proximity Correction (OPC). OPC modifies shapes in the layout pattern so that the mask compensates for dimensional distortions that occur in wafer lithography and processing. OPC is model-based and iterative (that is, go through several successive rounds of iteration until predicted pattern matches well enough with the desired pattern). The reason that the process is iterative is pragmatic: with typically hundreds of millions of transistors, this is a huge computational problem (several days worth of computation may be required for each mask).

David Lindsay asked whether OPC uses any libraries of patterns. Dan responded that model-based libraries have been considered theoretically, but that they have remained impractical and have never been used.

Functionally, OPC refines a mask design to conform to fab tool capabilities, but does not alter the basic capabilities of those tools. This is fundamentally the same situation as in 2003, except that the refinement is more complex now than in 2003 because of smaller feature sizes (process nodes) now in use.

From a worldwide perspective, OPC is quite common. It is researched and taught in many academic and commercial settings, in both Wassenaar and non-Wassenaar countries (e.g., Taiwan, India, China).

The key conclusion is that because OPC is widely available and that it was explicitly decontrolled by the 2003 changes in 3D3, it would be ineffective and counterproductive to seek to recontrol it now.

EDA tools were decontrolled ~20 years ago, because they are not a chokepoint. They are not tied to specific designs or applications. An appropriate control point may still be certain tools for advanced manufacturing processes, because there are a limited number of suppliers and because the tools are large and easily tracked.

Dave Robertson summarized the situation by stating that OPC is an empirical process, not a physics-based modeling process: OPC is a pre-distortion technique, but the valuable physics models are elsewhere. Henry Brandt suggested that the empirical nature of OPC is a useful criterion for concluding that OPC need not be export-controlled.

Action: The ISTAC will continue to express the industry position that 3D3 should not be extended to cover OPC.

Cloud Computing: Michael Angelo of NetIQ spoke on Cloud Computing, including taxonomy of clouds, export control and security.

The presentation began with a discussion of cloud taxonomy: a “Private Cloud” is within the user’s firewall (e.g., within a corporation) whereas a “Public Cloud” is composed of unknown computers outside of the user’s control. This distinction prompted comments from the audience to the effect that a “Private Cloud” is really a “Private Network”, and that the entire concept of “Cloud” implies the use of computers outside of one’s control

With regard to export control, the starting point for this presentation is the BIS Advisory Opinion “Application of EAR to Grid and Cloud Computing Services” dated Jan 13, 2009

(http://www.bis.doc.gov/policiesandregulations/advisoryopinions/jan13_2009_ao_on_cloud_grid_computing.pdf). This Advisory Opinion is written from the perspective of the Cloud service provider; the essence of this Opinion is that a Cloud service provider is not the exporter. This prompted a lively discussion.

There was broad consensus that the Advisory Opinion provides excellent guidance and clarification of the obligations of the Cloud service provider, but is silent with regard to the Cloud user. There was also broad consensus that the Cloud user should take appropriate precautions when using the Cloud, complying with the EAR and applying the concept of caveat emptor.

Jonathan Wise suggested that an advisory opinion clarifying the obligations of the Cloud user might be helpful.

Dave Robertson commented that Clouds should be considered insecure and that users should be careful about releasing data.

Roz Thomsen reminded the audience that the aforementioned Advisory Opinion and opined that it was well written.

Jonathan Wise asked what obligations one has with regard to interception of transmissions. For example, if one sends a hardcopy letter domestically within the US, does the sender have any obligation to ensure that the letter does not transit other countries en route? And if it does transit other countries, does that constitute an export? More generally, if one makes a transmission (letter or email) to a destination that is authorized under the EAR (whether NLR, license exception, or license), what obligation does the sender have with regard to the knowledge/control of the transmission channel? This was thought to be an interesting and subtle question, probably an emerging issue, and possibly of increasing importance. There was some uncertainty as to whether this matter is within ISTAC purview; the matter was tabled, for possible reconsideration at a future meeting.

Returning to the technical question of Cloud taxonomy, Henry Brandt added that the common terms may be “grid” and “cloud”. A “grid” is well known, and the user controls its configuration. A “cloud” is amorphous and unknown.

No decisions were made on further action; the issue may be reconsidered at a future ISTAC meeting.

Export License Case Review: Joe Meinhardt, of the Allied Signal Kansas City Plant, spoke briefly on export license case review, with focus on 3B1 (fab equipment).

The two general cases are 1) application for hardware export, and 2) deemed export.

Specific details are critical to a good end use statement. For example the statement “to conduct research” is not a priori wrong, but it is not very informative. A better statement is, for example, “to conduct research on doping of GaAs for use in LEDs”.

These specific details also help to match the end use to the end user, allowing the reviewer to assess questions such as: Is the end user known to do work in that field? Is this on the end user's roadmap?

Precise information on the make and model of the equipment proposed for export are also important, as these are relevant to the process technology node. For newer equipment, which the reviewers may be less familiar with, it is often helpful to provide greater detail (more specifications) than for older equipment, which the reviewers may understand better.

For deemed export licenses, applications should include a detailed job description, to define the scope of the job and the applicable ECCN(s). The application should also contain details about the foreign national (resume, background, credentials) to demonstrate that they are a good match for the job.

It is also useful to include a Technology Transfer Control Plan to explain how exactly the hiring company will protect export-controlled data. The absence of a TTCP in the application package is somewhat of a red flag.

As this presentation was informational in nature, no follow-up actions were proposed or required.

The open session was adjourned at 3:20pm.