IDA

INSTITUTE FOR DEFENSE ANALYSES

The Round Table on Computer Performance Metrics for Export Control: Discussions and Results

Alfred E. Brenner, Task Leader
Norman R. Howes

19980303 094

INSTITUTE FOR DEFENSE ANALYSES

The Round Table on Computer Performance Metrics for Export Control: Discussions and Results

Alfred E. Brenner, Task Leader
Norman R. Howes

PREFACE

This document was prepared for the Director, Strategic Policy Directorate, Defense Technology Security Administration, Office of the Under Secretary of Defense for Policy. The work was performed under the task order Technical Analysis of Strategic Impact of Changes in Export Controls Due to Foreign Availability, Rapid Technology Advances, and Foreign Acquisition. The document addresses an objective in the task order, identifying priority information needed to evaluate products or technology subject to rapid technological advances with implications for technology control list changes, foreign assessments/reviews, foreign acquisitions of U.S. companies, or other export-related matters.

This document was reviewed by research staff members at the Institute for Defense Analyses: Dr. Edward A. Feustel, Dr. Richard J. Ivanetich, and Dr. Reginald N. Meeson.

Table of Contents

Preceding Page Blank

EXECUTIVE SUMMARY

At the request of the Director of the Defense Technology Security Administration (DTSA) of the Office of the Undersecretary of Defense for Policy (USD(P)) in coordination with the Bureau of Export Administration (BXA) of the U.S. Department of Commerce (DoC) a Round Table on Computer Performance Metrics for Export Control was convened by the Institute for Defense Analyses (IDA). The purpose of the Round Table was to determine if the current metric, the Composite Theoretical Performance (CTP), used for calculating relative computing performance for purposes of export control, still provides a sufficiently robust measure of the relative performance of current and likely future computer systems in light of current architectural trends. If a new metric was needed, then the Round Table participants were to identify issues and to recommend methods of organizing and conducting a study for a new metric. The participants, who came from industry, government and academia, were selected on the basis of their technical knowledge and/or involvement in the design of computer and software systems.

The Round Table spanned one day and identified a number of issues on which there was general consensus. The key findings were:

- The CTP is still an effective metric for the purposes of export control when applied to a single computing element. Modest refinements could be made to the CTP for systems composed of aggregate computing elements.

- Because of the wide range of architectures in use today, especially with respect to the memory-to-processor integration schemes, there are variances estimated to be about a factor of two in the actual performance of delivered systems relative to the measure given by the CTP calculations. Continuing rapid changes in microelectronic technology may result in yet larger variances in the ratio in the near term future.

- Because of the rapid changes in computer architectures, any export control metric should be reevaluated every two years.

A number of follow-on studies were suggested or implied during the Round Table discussions. These are summarized in an appendix.

BACKGROUND

The Round Table on Computer Performance Metrics for Export Control met on October 15, 1997, in Alexandria, Virginia, at the Institute for Defense Analyses (IDA). The Round Table was sponsored by the Director of the Defense Technology Security Administration (DTSA) of the Office of the Under Secretary of Defense for Policy (USD(P)) in coordination with the Bureau of Export Administration (BXA) of the U.S. Department of Commerce (DoC). The participants came from the major firms involved in the support of or the manufacture of high performance computers, and government agencies or laboratories with major involvement in research or in the use of such computers. The corporate participants were invited as individual technical experts and not as formal representatives of their employers. In addition, a number of observers were invited on the basis of their interest and involvement in the export control of computers. Appendix A of this document lists all attendees.

The purpose of the Round Table was to determine if the current metric, the Composite Theoretical Performance (CTP), used for calculating relative computing performance for purposes of export control, still provides a sufficiently robust measure of the relative performance of current and likely future computer systems in light of current architectural trends. The desired result of the Round Table discussion was a recommendation as to whether the CTP was still sufficient or whether further work on defining a new measure was appropriate. If the discussions indicated a need for a new measure, the Round Table participants would identify the issues and make suggestions on how to organize and conduct a study to determine a new measure.

The CTP was put into effect on September 1, 1991, replacing the then-current metric, the Processing Data Rate (PDR). The PDR was replaced because it did not adequately address the performance variances of modern computer architectures at that time. Its major deficiencies were that it made no explicit provision for pipelines or concurrent operations within a central processing unit (CPU) and that it had no provision for multiple CPU computers with distributed memory.

1

The CTP came much closer to tracking current computer architectures than the PDR did. But in view of the diverse architectural approaches now being used, along with the increasing performance level of commodity microprocessors and the astounding growth in the bandwidth and connectivity of both local and wide area networks (LANs/WANs), it became prudent to reexamine the current suitability of the CTP.

SUMMARY OF THE ROUND TABLE DISCUSSIONS

Peter Sullivan (DTSA) and Tanya Mottley (BXA) presented the purpose for convening the Round Table during the introductory addresses to the Round Table. Mr. Sullivan emphasized that the government was not interested in changing the current metric for another metric of marginal improvement. If a new metric was to be considered, it needed to provide a significant improvement that would justify the effort to change from the existing one. Mr. Sullivan also emphasized that if a new metric was to be introduced, it needed to be put to a practical test to confirm its added value.

Dr. Brenner (IDA) chaired the Round Table discussions and gave a short historical introduction, indicating both the technical and procedural issues involved in export control, and reiterated the expected Round Table discussion goals. Dr. Brenner concluded with what he believed to be requirements on any export control metric:

- easy to evaluate
- deterministic
- a good measure of the relative performance for all computer systems, taking cognizance of
    - architectural variations
    - variations in problem characteristics
    - software efficacy variations
    - evolving technology
- meaningful in some range of applicability
- likely to be acceptable in the international export control community

Before the Round Table interactive discussions started, Ballard Troy (BXA) gave a short history of the development of the CTP and a concise review of the definition of the CTP. Appendix B contains the formal definition of the CTP as extracted from the DoC's Export Control Regulations.[1]

---

[1] Export Control Regulations: Technical Note to Category 4, Computers, Supplement No. 1, Part 744.

The Round Table then considered what the form of a performance metric should be if it is to track current architectural trends. After some discussion, it was almost unanimously agreed that a "correct" metric, ψ, should be of the form:

$$\Psi = P \times \alpha_M \times \alpha_C \times \alpha_I \times \alpha_N$$

where P is the peak rate of executing operations and the α_is are scaling parameters, each with values between 0 and 1, that are functions of memory bandwidth (M), cache size (C), network interconnect efficacy (I), and the number of processors (N) in the system.

However, there was also agreement that some of these scaling parameters were not as significant as others and that a study would be appropriate to determine how each of the scaling factors should be evaluated. These factors should be defined to be good measures of the scaling parameters and should not be overly burdensome to evaluate for both vendors and export control personnel.

Upon further discussion, a majority of participants concluded that:

- α_I was probably not significant enough to be included in the metric, considering the amount of work that would be involved in determining the right formula for α_I and it would probably not change the value of the metric much.

- Because both α_M and α_C affect data-to-processor latencies, they might best be combined into a single scaling parameter, $\alpha_{M,C}$, that was a function of the two variables, M and C. In this case a simpler expression for Ψ could be given as:

$$\Psi = P \times \alpha_{M,C} \times \alpha_N$$

- The term $P \times \alpha_N$, which represents the peak instruction execution rate for a system composed of multiple processors, was probably reasonably well represented using the current CTP, thus:

$$\Psi = CTP \times \alpha_{M,C}$$

- Although the CTP still provides a good measure of the peak instruction execution rate for most current system architectures, some refinements could be made with minor adjustments to some of the heuristic parameters in the CTP as now defined. This would give rise to an adjusted CTP_{Adj} calculation that better approximates the composite peak number of operations, the term $P \times \alpha_N$, that could be executed by the collection of processors in the system.

This would leave a possible new metric in the rather simple form:

$$\Psi = CTP_{Adj} \times \alpha_{M,C}$$

Here CTP_{Adj} is the current CTP in form, but has an updated set of coefficients assigned for multiple computing element systems. The new feature of the metric Ψ is contained in the memory term $\alpha_{M,C}$.

The Round Table recognized that the value of the memory bandwidth term α_M—and hence the term $\alpha_{M,C}$—is dependent upon a large number of design and implementation parameters and consequently may be difficult to calculate deterministically from the technical specifications of any given system. Therefore, it appeared that the only viable method to evaluate it would involve a rather simple timing measurement. Participants agreed that this could be done by measuring the time to move a block of data from one segment of memory to another. To eliminate the effects of caching on this measurement, it would be necessary to use a memory segment that was several times as large as the largest cache in the system and then dividing the time by the length of the memory segment. This measurement would most naturally be performed by the vendor of the computer system.

However, as simple as this measurement is to perform by a vendor, the requirement to make a measurement on working equipment rather than to make a calculation based entirely on documented technical specifications of the system changes the dynamics of the metric determinations. Introducing the need for such a measurement would lead to a requirement that the vendors must make this measurement and then certify and publish the results in their technical specification sheets (as they now do for the current parameters that determine the CTP). This would require convincing our international partners of the need for such a dramatic change in approach and would also introduce a number of new thorny issues into the problem. These include questions of the variations of test metrics procedures on different machines and concerns that such measurements may be manipulated by the vendor.

Further discussions revealed that several of the participants were in agreement that the inadequacies of the current CTP—in particular the lack of some measure of the $\alpha_{M,C}$ factor at the current time—might lead to an "unfairness" in the CTP value of up to a factor of two relative to actual performance. (Note that this "factor of two" is a purely subjective estimate of the variances on the part of the participants.) The level of unfairness is, of course, application dependent, and what the CTP gives is an estimate of the peak performance of a system. Because the nature of national security problems spans a wide range of problems, the details of which

are not spelled out[2], using this estimator of the peak performance level makes some sense. However, no user of the system would ever be able to realize this level of performance on a real-world problem.

At the present time, processor chip performance is increasing at about 50% per year, while memory bandwidths are growing approximately 35% per year. Furthermore, as the number of elements on a chip continue to grow at a very high rate[3], major architectural changes are beginning to appear in the design of systems based upon new approaches to integrating memory and processors. These architectural trends in the use of memory may cause additional discrepancies in the unfairness levels of various systems, as estimated by the current CTP, within the next two years. This, of course, may lead to some computer systems being prohibited for export while more effective computers with lower CTPs might be below the cut-off level and hence be exportable.

Finally, with such changes in architecture, and not just in the performance level of CPUs, one might expect changes in real computer performance to occur with a much shorter time constant than heretofore. It was suggested, therefore, that it would be prudent to reevaluate how well the metric continues to reflect actual computer performance every two years.

The recommendation by several of the attendees was that modifying the current CTP metric with factors that make it more closely track current architectural trends would be highly desirable. Such a study should explore how to best calculate the metric in a way that would be simple and straightforward for the computer systems vendors. It was also recommended that any new or modified metric be applied to a number of different types of current high performance computer systems and be compared with values of the current CTP metric. Such a comparison would be necessary information to have in considering whether it would really be worthwhile to change the metric.

---

[2] Over the years that the CTP and the PDR have been in use, the detailed nature of the many classes of problems of national security concerns has not been specified. There is no evidence that this situation will change in the future.

[3] An observation known as Moore's Law, usually quoted as "the number of elements on a microelectronics chip doubles every 18 months."

ADDITIONAL DETAILS OF THE ROUND TABLE DISCUSSIONS

Networks of Workstations and New Communications Technologies

The Round Table technical discussions began with a detailed discussion of the performance capability of networks of workstations. All the participants were fully aware that it is virtually impossible to control sales of inexpensive, commodity personal computers and workstations that can be connected together by someone with a modest understanding of networking into very large networks with tremendous aggregate computational capability. The main concern of most of the participants was to understand how networks of workstations differed from supercomputers. Some problems will run easily and effectively on such networks, while other classes of problems important to national security concerns will not run effectively without a major software redesign effort. For many problems no amount of software redesign will allow networks of workstations to compete with appropriately designed high performance computers.

Initially not everyone understood that even if a "rogue state" assembled such a large network of workstations by legitimately acquiring large numbers of commodity processors, the actual effort to produce the software necessary to realize the full potential of such an aggregate system would take several years. During this time, the state of the art of computational technology would have increased by approximately an order of magnitude. After considerable discussion, most of the participants were in agreement that there was a fundamental difference between a system designed by a single vendor that was built as an aggregate of many commodity processors and included the software to enable these processors to cooperatively work on solving single problems of national concern, and a large collection of commodity processors not subject to export control that are externally networked together.

A related discussion followed regarding the difficulty of controlling new very high-speed networking and interconnect products using new communications technologies. It was agreed that controlling such devices was almost as difficult as controlling commodity processors. But it was also noted that at this time these are not easy to install and operate effectively without a great deal of expertise, understanding, and effort on the part of the end

7

user. Furthermore, it was generally agreed that, if necessary, performance metrics might be developed to take account of various methods of interconnecting large numbers of processors by various bus or network technologies that may be employed to construct such "scalable parallel" systems.

But in the end it was agreed that aggregates of commodity processors and high-speed networking hardware technology were beyond the scope of discussions for a computer performance metric for export control. With this agreement, the Round Table was then able to focus on a fairly well-defined technology domain for its considerations of the adequacy of the existing CTP metric.

New Architectures and the Current CTP Metric

When discussions got underway in earnest about the current CTP metric, it became apparent that many of the participants were of the opinion that the current metric did not reflect relative performance very accurately while some were of the opinion that the difference was not significant enough to worry about. It was generally agreed that two computer systems with the same calculated CTP could have up to a factor of two difference in the real-world performance that a user of the systems would be able to realize. But the issue is very complicated and is very dependent upon the application and the class of problems for which the manufacturer has tuned the machine. However, it was agreed that the primary factor giving rise to these discrepancies was related to the memory-to-processor bandwidth. Consequently, if one could improve the estimator for this factor, the variance in the value of the metric for equivalent machines of different architecture might be reduced.

An extensive discussion followed, centering around current architectural trends that have developed since the current CTP metric was put in place in 1991. These include the use of larger caches, the introduction of secondary caches, and the trend toward much more fully integrated memory and processor elements on the same chip. The range of architectural variations currently being explored include smart memories, processor in memory, and memory on processor chips. All of these approaches have become viable product options because of the very high number of elements capable of being mass produced on a single chip today.

The expected architectural changes emanating from this continuing memory/processor integration on a single chip was considered by the group to be the most important factor in

future computer performance and in issues related to the export control of computers. This will change the relevancy and character of cache and latency, and will revolutionize computer chip and board architecture designs in the very near future in even more profound ways. One recent paper by a group of researchers from the University of Wisconsin with which several of the participants were familiar was cited during the discussion. This paper summarized what is currently happening as follows:

> Today's technological trends point to a widening gap between the rate at which a processing unit can consume operands and the rate at which the memory system can supply them. Present designs are addressing this trend by introducing one or two levels of on-chip cache. While this on-chip memory effectively reduces memory access latency, the delay incurred when it is necessary to go off-chip is high. As a consequence, processors extrapolated from current designs will be more and more frequently stalled waiting for operands.[4]

The architectural changes beginning to appear are designed to attain high levels of *memory bandwidth* and *memory efficiency* (a term defined in the Wisconsin research papers). They are making the role of cache much less relevant in tolerating off-chip memory fetches. This may result in the current CTP metric reflecting even less well the relative real-world performance of several equivalent systems because it gives only an estimate of the peak CPU power of the computer system. Consequently, unless the export control metric takes this dependency on memory bandwidth and efficiency into consideration, the metric may become even less reliable in the near future.

---

[4] *The Declining Effectiveness of Dynamic Caching for General-Purpose Microprocessors*, Douglas C. Burger, James R. Goodman, and Alain Kägi; University of Wisconsin-Madison Computer Sciences Dept. Tech. Report 1261, January, 1995. This group has also published a number of other relevant papers on this issue, some of which are available on the World Wide Web at http://www.cs.wisc.edu/galileo. For example, see *System-Level Implications of Processor/Memory Integration*, Douglas C. Burger, which was presented at the Mixed Logic/DRAM Workshop at the 24th International Symposium on Computer Architecture (ISCA), June, 1997.

RESULTS

The Round Table participants concluded their discussions with consensus on a number of issues. The key findings are summarized here.

1. The CTP is an effective metric for the purposes of export control. It provides a well-defined and easily evaluated measure of the peak instruction rate for a single computing element.

2. For systems composed of aggregations of computing elements, some modest refinements might be made to the heuristic assignments made in the CTP definition for the weighting factor coefficients.

3. Due primarily to the wide range of architectures of systems in use today, especially with respect to the memory-to-processor integration schemes, there are variances estimated to be about a factor of two in the actual performance of delivered systems relative to the measure given by the CTP calculation.

4. Rapid changes in microelectronics technology are likely to further affect memory latency as new architectures emerge to take advantage of these technological changes. This may result in yet larger variances in the ratio of actual performance relative to the CTP evaluations.

5. With the continuing rapid evolution of the semiconductor industry and the resulting effects on computer architectures, reevaluation of the metric used for export control should be made every two years.

As is well understood by the export control community, the CTP is an approximate measure of the relative performance of computer systems. The Round Table concluded that there is no clearly better metric to replace it today. But it did warn that with rapidly changing technology it is important to track the effectiveness of the metric on a continuous basis. A list of additional studies that would address many of the issues raised during the course of the Round Table is given in Appendix C.

APPENDIX A.
ROUND TABLE ATTENDEES

The attendees at the Round Table on Computer Performance Metrics for Export Control that met on October 15, 1997, at the Institute for Defense Analyses were:

Participants:

Greg Astfalk	Hewlett-Packard Company
Ronald Boisvert	National Institute of Standards and Technology
Mike Booth	Silicon Graphics, Inc./Cray Research
Henry Brandt	IBM
William Carlson	Center for Computing Sciences
Hank Dardy	Naval Research Laboratory
Tom Gannon	Digital Equipment Corporation
Roger Golliver	Intel
Gary Koob	Defense Advanced Research Projects Agency
Doug Martin	National Security Agency
John McCalpin	Silicon Graphics, Inc./Cray Research
Dave Powers	National Security Agency
Jeff Rulifson	Sun Microsystems
Margaret Simmons	San Diego Supercomputer Center
Horst Simon	Lawrence Berkley National Laboratory
Ballard Troy	Bureau of Export Administration
Steve Wallach	Centerpoint Ventures

Hosts:

Tanya Mottley	Bureau of Export Administration, DoC
Peter Sullivan	Defense Technology Security Administration, DoD

IDA:

Alfred Brenner, Chair	IDA
Norm Howes	IDA

Preceding Page Blank

Observers:

Gordon Boezer	IDA
Ed Feustel	IDA
David Hoger	Intel
Paul Koenig	Defense Technology Security Administration, DoD
Alex Marusak	Los Alamos National Laboratory
Oksana Nesterczuk	Defense Technology Security Administration, DoD
Dale Nielsen	Lawrence Livermore National Laboratory
Kenneth Pocek	Intel
Jim Ramsbotham	IDA
Joe Young	Bureau of Export Administration, DoC

APPENDIX B.
COMPOSITE THEORETICAL PERFORMANCE
TECHNICAL NOTE

§740.7

COMPUTERS (CTP)

(a) Scope

License Exception CTP authorizes exports and reexports of computers and specially designed components therefor, exported or reexported separately or as part of a system for consumption in Computer Tier countries as provided by this section. (Related equipment controlled under 4A003.d, .f, and .g is authorized under this License Exception, only when exported or reexported with these computers as part of a system.) You may not use this License Exception to export or reexport items that you know will be used to enhance the CTP beyond the eligibility limit allowed to your country of destination. When evaluating your computer to determine License Exception CTP eligibility, use the CTP parameter to the exclusion of other technical parameters for computers classified under ECCN 4A003.a, .b and .c, *except* of parameters specified as Missile Technology (MT) concerns or 4A003.e (equipment performing analog-to-digital conversions exceeding the limits in ECCN 3A001.a.5.a). This License Exception does not authorize the export or reexport of graphic accelerators or coprocessors, or of computers controlled for MT reasons.

(b) Computer Tier 1

(1) Eligible countries. The countries that are eligible to receive exports and reexports under this License Exception are Australia, Austria, Belgium, Denmark, Finland, France, Germany, Greece, the Holy See, Iceland, Ireland, Italy, Japan, Liechtenstein, Luxembourg, Mexico, Monaco, Netherlands, New Zealand, Norway, Portugal, San Marino, Spain, Sweden, Switzerland, Turkey, and the United Kingdom.

(2) Eligible Computers. The computers eligible for License Exception CTP to Tier 1 destinations are those with a CTP greater than 2,000 Mtops.

(c) Computer Tier 2

(1) Eligible countries. The countries that are eligible to receive exports under this License Exception include Antigua and Barbuda, Argentina, Bahamas, Barbados, Bangladesh, Belize, Benin, Bhutan, Bolivia, Botswana, Brazil, Brunei, Burkina Faso, Burma, Burundi, Cameroon, Cape Verde, Central Africa, Chad, Chile, Colombia, Congo, Costa Rica, Cote d'Ivoire, Cyprus, Czech Republic, Dominica, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Eritrea, Ethiopia, Fiji, Gabon, Gambia (The), Ghana, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong, Hungary, Indonesia, Jamaica, Kenya, Kiribati, Korea (Republic of), Lesotho, Liberia, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritius, Micronesia (Federated States of), Mozambique, Namibia, Nauru, Nepal, Nicaragua, Niger, Nigeria, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Rwanda, St. Kitts & Nevis, St. Lucia, St. Vincent and Grenadines, Sao Tome & Principe, Senegal, Seychelles, Sierra Leone, Singapore, Slovak Republic, Slovenia, Solomon Islands, Somalia, South Africa, Sri Lanka, Surinam, Swaziland, Taiwan, Tanzania, Togo, Tonga, Thailand, Trinidad and Tobago, Tuvalu, Uganda, Uruguay, Venezuela, Western Sahara, Western Samoa, Zaire, Zambia, and Zimbabwe.

(2) Eligible computers. The computers eligible for License Exception CTP to Tier 2 destinations are those having a Composite Theoretical Performance (CTP) greater than 2,000, but equal to or less than 10,000 Millions of Theoretical Operations Per Second (Mtops).

Preceding Page Blank

(d) Computer Tier 3

(1) Eligible countries. The countries that are eligible to receive exports and reexports under this License Exception are Afghanistan, Albania, Algeria, Andorra, Angola, Armenia, Azerbaijan, Bahrain, Belarus, Bosnia & Herzegovina, Bulgaria, Cambodia, China (People's Republic of), Comoros, Croatia, Djibouti, Egypt, Estonia, Georgia, India, Israel, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lithuania, Macedonia (The Former Yugoslav Republic of), Mauritania, Moldova, Mongolia, Morocco, Oman, Pakistan, Qatar, Romania, Russia, Saudi Arabia, Serbia & Montenegro, Tajikistan, Tunisia, Turkmenistan, Ukraine, United Arab Emirates, Uzbekistan, Vanuatu, Vietnam, and Yemen.

(2) Eligible computers. The computers eligible for License Exception CTP to Tier 3 destinations are those having a Composite Theoretical Performance (CTP) greater than 2,000 Millions of Theoretical Operations Per Second (Mtops), but less than or equal to 7,000 Mtops.

(3) Eligible exports. Only exports and reexports to permitted end-users and end-uses located in countries in Computer Tier 3. License Exception CTP does not authorize exports and reexports to Computer Tier 3 for military end-users and end-uses and nuclear, chemical, biological, or missile end-users and end-uses defined in part 744 of the EAR. Exports and reexports under this License Exception may not be made to known military end-users or to known military end-uses or known proliferation end-uses or end-users defined in part 744 of the EAR. Such exports and reexports will continue to require a license and will be considered on a case-by-case basis. Retransfers to military end-users or end-uses and defined proliferation end-users and end-uses in eligible countries are strictly prohibited without prior authorization.

(e) Restrictions

(1) Computers eligible for License Exception CTP may not be accessed either physically or computationally by nationals of Cuba, Iran, Iraq, Libya, North Korea, Sudan or Syria, except commercial consignees described in Supplement No. 3 to part 742 of the EAR are prohibited only from giving such nationals user-accessible programmability.

(2) Computers eligible for License Exception CTP may not be reexported/retransferred without prior authorization from BXA i.e., a license, a permissive reexport, another License Exception, or "No License Required". This restriction must be conveyed to the consignee, via the Destination Control Statement, see §758.6(a)(ii) of the EAR.

(f) Recordkeeping requirements

In addition to the recordkeeping requirements in part 762 of the EAR, you must keep records of each export under License Exception CTP. These records will be made available to the U.S. Government on request. The records must include the following information:

(1) Date of shipment;

(2) Name and address of the end-user and each intermediate consignee;

(3) CTP of each computer in shipment;

(4) Volume of computers in shipment;

(5) Dollar value of shipment; and

(6) End-use.

Information on How to Calculate "Composite Theoretical Performance" ("CTP"):

Technical Note: "COMPOSITE THEORETICAL PERFORMANCE" (CTP).

Abbreviations used in this **Technical Note:**

CE   "computing element" (typically an arithmetic logical unit)

FP   floating point

XP   fixed point

t   execution time

XOR exclusive OR

CPU central processing unit

TP   theoretical performance (of a single CE)

CTP "composite theoretical performance" (multiple CEs)

R   effective calculating rate

WL   word length

L   word length adjustment

*   multiply

Execution time 't' is expressed in microseconds, TP and "CTP" are expressed in Mtops (millions of theoretical operations per second) and WL is expressed in bits.

Outline of "CTP" calculation method:

"CTP" is a measure of computational performance given in millions of theoretical operations per second (Mtops). In calculating the
"Composite Theoretical Performance" ("CTP") of an aggregation of "Computing Elements" ("CEs"), the following three steps are required:

1. Calculate the effective calculating rate (R) for each "computing element" ("CE");

2. Apply the word length adjustment (L) to the effective calculating rate (R), resulting in a Theoretical Performance (TP) for each "computing element" ("CE");

3. If there is more than one "computing element" ("CE"), combine the Theoretical Performances (TPs), resulting in a "Composite Theoretical Performance" ("CTP") for the aggregation.

Details for these steps are given in the following section.

NOTE 1: For aggregations of multiple "computing elements" ("CEs") that have both shared and unshared memory subsystems, the calculation of "CTP" is completed hierarchically, in two steps: first, aggregate the group of "computing elements" ("CEs") sharing memory, second calculate the "CTP" of the groups using the calculation method for multiple "computing elements" ("CEs") not sharing memory.

NOTE 2: "Computing elements" ("CEs") that are limited to input/output and peripheral functions (e.g., disk drive, communication and video display controllers) are not aggregated into the "CTP" calculation.

The following table shows the method of calculating the "Effective Calculating Rate" (R) for each "Computing Element" ("CE"):

Step 1: *The effective calculating rate R.*
For Computing Elements (CEs) Implementing: Effective calculating Rate, R

Note: Every "CE" must be evaluated independently

XP only (R_{xp}) $1 / [3 * (t_{xp\ add})]$

If no add is implemented use:

$1 / (t_{xp\ mult})$

If neither add nor multiply is implemented use the fastest available arithmetic operation as follows:

$1 / (3 \times t_{xp})$

See Notes X and Y

FP only (R_{fp})   Max $1 / t_{fp\ add}$, $1 / t_{fp\ mult}$

See Notes X and Y

Both FP and XP (R). Calculate both R_{xp}, R_{fp}.

For simple logic processors not implementing any of the specified arithmetic operations.

$1 / (3 \times t_{log})$

Where t_{log} is the execute time of the XOR, or for logic hardware not implementing the XOR, the fastest simple logic operation.

See Notes X and Z

For special logic processors not using any of the specified arithmetic or logic operations.

$$R = R^t \times WL/64$$

Where R is the number of results per second, WL is the number of bits upon which the logic operation occurs, and 64 is a factor to normalize to a 64 bit operation.

NOTE W: For a pipelined "CE" capable of executing up to one arithmetic or logic operation every clock cycle after the pipeline is full, a pipelined rate can be established. The effective calculating rate (R) for such a "CE" is the faster of the pipelined rate or non-pipelined execution rate.

NOTE X: For a "CE" that performs multiple operations of a specific type in a single cycle (e.g., two additions per cycle or two identical logic operations per cycle), the execution time t is given by:

t = cycle time / (the number of arithmetic operations per machine cycle)

"Computing elements" ("CEs") that perform different types of arithmetic or logic operations in a single machine cycle are to be treated as multiple separate "computing elements" ("CEs") performing simultaneously (e.g., a "CE" performing an addition and a multiplication in one cycle is to be treated as two "CEs", the first performing an addition in one cycle and the second performing a multiplication in one cycle).

If a single "Computing element" ("CE") has both scalar function and vector function, use the shorter execution time value.

NOTE Y: For the "CE" that does not implement FP add or FP multiply, but that performs FP divide:

$$R_{tp} = 1 \: / \: t_{fp\ divide;}$$

If the "CE" implements FP reciprocal, but not FP add, FP multiply or FP divide, then:

$$R_{fp} = 1 \: / \: t_{fp\ reciprocal.}$$

If the divide is not implemented, the fp reciprocal should be used.

If none of the specified instructions is implemented, the effective floating point (FP) rate is 0.

NOTE Z: In simple logic operations, a single instruction performs a single logic manipulation of no more than two operands of given lengths. In complex logic operations, a single instruction performs multiple logic manipulations to produce one or more results from two or more operands.

Rates should be calculated for all supported operand lengths considering both pipelined operations (if supported), and non-pipelined operations, using the fastest executing instruction for each operand length based on:

1. Pipelined or register-to-register operations. Exclude extraordinarily short execution times generated for operations on a predetermined operand or operands (for example, multiplication by 0 or 1). If no register-to-register operations are implemented, continue with (2).

2. The faster of register-to-memory or memory-to-register operations; if these also do not exist, then continue with (3).

3. Memory-to-memory.

In each case above, use the shortest execution time certified by the manufacturer.

Step 2: *TP for each supported operand length WL:*

Adjust the effective rate R (or Rt) by the word length adjustment L as follows:

$TP = R \times L$, where $L = (1/3 + WL/96)$.

Note: The word length WL used in these calculations is the operand length in bits. (If an operation uses operands of different lengths, select the largest word length.)

The combination of a mantissa ALU and an exponent ALU of a floating point processor or unit is considered to be one "computing Element" ("CE") with a Word Length (WL) equal to the number of bits in the data representation (typically 32 or 64) for purposes of the "Composite Theoretical Performance" ("CTP") calculations.

This adjustment is not applied to specialized logic processors that do not use XOR instructions. In this case TP = R.

Select the maximum resulting value of TP for:

Each XP-only "CE" (R_{xp});

Each FP-only "CE" (R_{fp});

Each combined FP and XP "CE" (R);

Each simple logic processor not implementing any of the specified arithmetic operations; and

Each special logic processor not using any of the specified
arithmetic or logic operations.

Step 3: *"CTP" for aggregations of "CEs", including CPU's:*

For a CPU with a single "CE", "CTP" = TP (for CEs performing both fixed and floating point operations, $TP = \max (TP_{fp}, TP_{xp})$).

"CTP" for aggregations of multiple "CEs" operating simultaneously is calculated as follows:

NOTE 1: For aggregations that do not allow all of the "CEs" to run simultaneously, the possible combination of "CEs" that provides the largest "CTP" should be used. The TP of each contributing "CE" is to be calculated at its maximum value theoretically possible before the "CTP" of the combination is derived.

N.B.: To determine the possible combinations of simultaneously operating "CEs", generate an instruction sequence that initiates operations in multiple "CEs", beginning with the slowest "CE" (the one needing the largest number of cycles to complete its operation) and ending with the fastest "CE". At each cycle of the sequence, the combination of "CEs" that are in operation during that cycle is a possible combination. The instruction sequence must take into account all hardware and/or architectural constraints on overlapping operations.

NOTE 2: A single integrated circuit chip or board assembly may contain multiple "CEs".

NOTE 3: Simultaneous operations are assumed to exist when the computer manufacturer claims concurrent, parallel or simultaneous operation or execution in a manual or brochure for the computer.

NOTE 4: "CTP" values are not to be aggregated for "CE"-combinations (inter)connected by "Local Area Networks", Wide Area Networks, Input/Output shared connections/devices, I/O controllers and any communication interconnection implemented by "software".

NOTE 5: "CTP" values must be aggregated for multiple "CEs" specially designed to enhance performance by aggregation, operating simultaneously and sharing memory,- or multiple memory/ "CE"- combinations operating simultaneously utilizing specially designed hardware. This aggregation does not apply to "electronic assemblies" controlled by 4A003.c.

"CTP" $= TP_1 + C_2 * TP_2 + ... + C_n * TP_n$, where the TPs are ordered by value, with TP_1, being the highest, TP_2 being the second highest, ... and TP_n being the lowest. C_i is a coefficient determined by the strength of the interconnection between "CEs", as follows:

For multiple "CEs" operating simultaneously and sharing memory:

$C_2 = C_3 = C_4 = ... = C_n = 0.75$.

NOTE 1: When the "CTP" calculated by the above method does not exceed 194 Mtops, the following formula may be used to calculate C_i:

$C_i = 0.75 / m^{0.5}$ (i = 2, ... n)

where m = the number of "CEs" or groups of "CEs" sharing access.

Provided:

1. The TP_i of each "CE" or group of "CEs" does not exceed 30 Mtops;

2. The "CEs" or groups of "CEs" share access to main memory (excluding cache memory) over a single channel; and

3. Only one "CE" or group of "CEs" can have use of the channel at any given time.

N.B.: This does not apply to items controlled under Category 3.

NOTE 2: "CEs" share memory if they access a common segment of solid state memory. This memory may include cache memory, main memory, or other internal memory. Peripheral memory devices such as disk drives, tape drives, or RAM disks are not included.

For multiple "CEs" or groups of "CEs" not sharing memory, interconnected by one or more data channels:

$C_i = 0.75 * k_i$ (i = 2, ..., 32)
      (see NOTE on k_i factor)
    $= 0.60 * k_i$ (i = 33, ..., 64)
    $= 0.45 * k_i$ (i = 65, ..., 256)
    $= 0.30 * k_i$ (i > 256)

The value of C_i is based on the number of "CEs", not the number of nodes.

where $k_i = \min (S_i/K_r, 1)$, and
K_r = normalizing factor of 20 MByte/s.
S_i = sum of the maximum data rates (in units of MBytes/s) for all data channels connected to the i_{th} "CE" or group of "CEs" sharing memory.

When calculating a C_i for a group of "CEs", the number of the first "CE" in a group determines the proper limit for C_i. For example, in an aggregation of groups consisting of 3 "CEs" each, the 22nd group will contain "CE"$_{64}$, "CE"$_{65}$ and "CE"$_{66}$. The proper limit for C_i for this group is 0.60.

Aggregation (of "CEs" or groups of "CEs") should be from the fastest-to-slowest; i.e.;

$TP_1 \geq TP_2 \geq TP_3$, and

in the case of $TP_i = TP_{i+1}$, from the largest to smallest; i.e.:

$C_i \geq C_{i+1}$

Note: The k_i factor is not to be applied to "CEs" to 2 to 12 if the TP_i / of the "CE" or group of "CEs" is more than 50 Mtops;

i.e., C_i for "CEs" 2 to 12 is 0.75.

APPENDIX C.
SUGGESTED FURTHER STUDIES ON
EXPORT CONTROL METRICS

During the course of the Round Table, a number of suggestions were made or implied for additional studies that might clarify understanding of some of the issues relevant to export control metrics and/or lead to refinements in the CTP or its replacement. These include:

- Explore options to refine α_N giving rise to the adjusted CTP_{Adj}.

- Analyze possible approaches to evaluate a new metric Ψ as discussed by the Round Table. In particular:

  - explore options to evaluate $\alpha_{M,C}$
  - evaluate and compare results with the existing CTP

- In light of new architectural issues, explore new technical approaches for an alternate metric.

- Analyze further the potential effects of increasing memory-processor integration on export control issues.

- Evaluate further the subjective "factor of 2" in unfairness measure of the CTP.

- Analyze the consequences of the rapid changes in architectural approaches in the design of new systems expected by the Round Table. If two years is the time constant for major changes, how does it affect the Wassenaar Arrangement[1] process?

- Analyze the potential effects of new high-speed networking products on the control of high performance computers.

- Pursue extension of the Round Table discussions with Japanese and European partners.

- Examine the implications for export control for emerging new military applications using high-end distributed computer systems.

- Develop a process for gaining concurrence on a new metric in the international community.

---

[1] The Wassenaar Arrangement is an export control organization established in 1996 that replaced the Cold War's export control organization COCOM (Coordinating Committee). The Wassenaar Arrangement has a broader organization and narrower scope than COCOM.

Preceding Page Blank

LIST OF ACRONYMS

BXA	Bureau of Export Administration
COCOM	Coordinating Committee
CPU	central processing unit
CTP	Composite Theoretical Performance
DoC	Department of Commerce
DTSA	Defense Technology Security Administration
IDA	Institute for Defense Analyses
LAN	local area network
PDR	Processing Data Rate
USD(P)	Under Secretary of Defense for Policy
WAN	wide area network

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December 1997	3. REPORT TYPE AND DATES COVERED Final

4. TITLE AND SUBTITLE
The Round Table on Computer Performance Metrics for Export Control: Discussions and Results

5. FUNDING NUMBERS
DASW01-94-C-0054

Task Order T-K2-1351

6. AUTHOR(S)
Alfred E. Brenner, Norman R. Howes

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses (IDA)
1801 N. Beauregard St.
Alexandria, VA 22311-1772

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Document D-2116

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Oksana Nesterczuk
Defense Technology Security Administration
400 Army Navy Drive, Suite 300
Arlington, VA 22202

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, unlimited distribution: January 28, 1998.

12b. DISTRIBUTION CODE
2A

13. ABSTRACT (Maximum 200 words)

Rapidly changing microelectronics technologies and their effects on computer architectures have required that the U.S. government evaluate its measures of computer performance for export control. Both the Departments of Defense and Commerce realized the need for determining whether the current metric, the Composite Theoretical Performance (CTP), still provided the necessary measurements for current and future computer systems, particularly in light of architectural trends. A Round Table on Computer Performance Metrics for Export Control was held on October 15, 1997, at IDA. Attending were participants and observers from major firms involved with the manufacturing or support of high performance computers, and representatives from government organizations, including the laboratories. The Round Table found that the CTP was still a relatively effective metric when applied to a single computing element, and that this metric could be refined for measuring systems composed of aggregate computing elements. But advances in microelectronics technology will very likely lead to changes in the design of computers, which may diminish the effectiveness of the CTP. Consequently, the Round Table participants suggested that any export control metric should be reevaluated every two years. Additional findings and issues were identified.

14. SUBJECT TERMS
Export control, computer performance metrics, computer architectures, Composite Theoretical Performance (CTP), microelectronics.

15. NUMBER OF PAGES
46

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL