

**DARPA Tech, DARPA's 25th Systems and Technology Symposium
August 8, 2007**

Anaheim, California

**Teleprompter Script for Dr. Chuck Morefield, Deputy Director,
Information Processing Technology Office**

Extreme Computing

» **CHUCK MOREFIELD:**

In 1956 the early thinkers in artificial intelligence, including Oliver Selfridge, Marvin Minsky, and others, met at Dartmouth.

This influential group made a large bet, turning their backs on analog cybernetics and jumping on the digital train in earnest as they started a long quest for human levels of machine reasoning.

From our vantage point 50 years later, it appears they placed a pretty good bet, since not many years from now small and cheap digital systems will have the memory and processing capacity of a human brain.

A few decades after that, we will likely find the processing capacity of all of humanity contained in a single computer.

IPTO is following a number of threads as we wend our way up this steep ramp in raw performance.

We like to refer to our thoughts in this area as "Extreme Computing".

As I give you a glimpse of what we mean by Extreme Computing, I also

want to solicit your help.

This will be a bit like an Easter Egg hunt:
my job is to place some eggs along the path, and your job is to find them.

Don't worry,
it won't be that hard!

Along the way,
we will touch on productive computing, Moore's law,
the "memory wall",
novel computing architectures, AI, and software complexity.

As you find things of interest, please check in with Bill Harrod,
our manager for Extreme Computing, and others of us from IPTO to
discuss your technical thoughts.

As Charlie mentioned, DARPA is vitally interested in the hardware
aspects of computing performance.

DARPA's largest foundational hardware program is
High-Productivity Computing Systems (HPCS for short).

HPCS leverages switched architecture, linking many homogenous multi-
core components into a single integrated computing bundle.

The goal of this program is to bring online usable general-purpose
systems by the year 2010.

To ensure that these high-end systems meet critical national security

needs, the HPCS program has established a strong collaboration with key government user agencies.

Nonetheless, an important requirement for HPCS is to achieve *commercial viability*.

Custom small production stove-pipes were not allowed, and current performers have received a lot of encouragement to develop scalable systems that serve a spectrum of commercial markets, markets that range from the Ito calculus of computational finance to the vast compute farms of Web 2.0.

HPCS is now in its final phase, which will culminate in prototype petascale systems that will undergo significant acceptance testing by stressing applications.

If successful, the vendors will begin selling their machines to the DoD and in commercial marketplaces.

IPTO is also looking beyond petascale, toward the exascale horizon.

The HPCS systems almost in the grasp of the user community will not scale directly to the next level.

Several technical roadblocks stand in our way, including power, size, and especially programmability.

So moving from petascale to exascale requires some more innovation.

At exascales,

we are focused less now on immediate commercial or military transition, and much more on enabling technology, particularly in the areas of power and programmability.

As Moore's paradigm hits the power wall, we can no longer continue to turn to increasing clock rates for more cycles.

Our sister office MTO is looking for solutions to this problem at the chip level.

One way around the power wall can be found in variable precision arithmetic units –components that require the application *itself* to select run-time precision.

We are also looking at techniques that kick off computation before all data is received,

and at stochastic processing that relaxes constraints on precision.

Meanwhile, data movement is not keeping up with processing speed, so we are also hitting a “memory wall”.

It will be important to find new techniques that make access to data rapid and transparent,

and that minimize data movement within and without the chip.

IPTO continues to explore novel microprocessor architectures.

Our most recent program in this area is called Polymorphous Computing Architectures,

PCA for short.

This program aims at deployed devices with very limited power and space, for example embedded in the webbing of a soldier on patrol.

PCA architectures can morph in milliseconds –their architectural resources changing configuration at run-time, adapting hardware to the mission as it unfolds.

Our goal is high performance in small packages that give users in the field immediate local access to powerful computing.

These fractional forms of extreme computing will be in great demand as technology such as IPTO's hand held translators move into widescale use.

Beyond the current PCA program, revolutionary processing architectures and approaches are sought that address key exascale hardware issues.

Candidate technologies that will be considered include, for example, novel approaches to input/output and storage, or non-Von Neumann processing architectures that greatly increase key metrics such as FLOPS per watt.

Of course, increased raw power constantly entices us to build more complex software.

The list of power users that initially motivated the DARPA petascale program are concerned about domains like weather forecasting, biomolecular modeling, or the simulation of large physical platforms.

But other power users are emerging, many of whom are concerned with the architecture of the brain.

As Dan Oblinger will tell you momentarily, we want to explore AI programming applied from the "bottom up", in addition to building good old fashioned expert systems from the "top down".

We are particularly interested in modeling the sub-symbolic "instruction

set" of the brain.

Adding multiple layers of learning to these low level models, we can move progressively upward to complex abilities such as prediction and planning for robotics or command and control.

Success here would provide alternate approaches to perception, reasoning, and language that compliment those of IPTO's symbolic AI work under Dave Gunning.

All of this work consumes increasing amounts of computing power.

Even more computationally intense modeling problems lie just ahead.

The world is watching as we face an elusive foe in concurrent, multi-faceted conflicts.

As retired General Barry McCaffrey recently noted, for these conflicts the political and economic struggle for power has become the actual field of battle.

For this reason, as you will hear later in Sean's talk, we want culturally sensitive models of large asymmetric nets of individual actors and the macro economic and political forces through which they interact.

These models will involve huge numbers of reasoning agents, scaling up Dan's individual AI models by many orders of magnitude.

Beneath *all* these complex applications, whether big physics or models of the human terrain, sit very large multi-core bundles of computing

power.

The evolution of the ARPAnet has ensured that numbers of these multi-core systems will be attached to wide area nets.

And today's commercial technology already allows each core of a netted multi-core system to itself contain multiple virtual machines (which are of course netted themselves!)

As we move along this continuum, our computing fabric already has started to exhibit the classic features of mathematical complexity.

Managing this complexity is a key issue for IPTO.

For lack of a more exciting word, we often use the word "productivity" to describe our work in managing complexity.

Complexity (or "productivity" if you will)

is the major bottleneck to exploiting the full capacity of machines already in use.

For future machines,

it will become an even more dominant issue.

Removing the bottlenecks associated with complexity will require new scalable and adaptive software tools with enough internal intelligence to

hide vast quantities of complex software development

dog work beneath their surface.

Extreme Computing must face head on the problem of how to program this vast increase in computing power.

We need to reduce the cost, time, and especially the expertise required to build software for large multicores and specialized computing devices.

It is no exaggeration to say new software advances are key to our children's future, given the economic and military competition we face, and given the likely size of our skilled programmer base.

Extreme Computing must focus on cognitive support to software developers,

as our systems evolve toward rich combinations of processing elements from the chip to the system level.

Today, using these tremendous architectural riches requires highly sophisticated developers.

But even sophisticated developers face challenges, and under-achieve on such systems, which as a consequence have become increasingly expensive to end users.

This calls for us to
re-think our development technology,
for example with new **adaptive compilers**.

Today, production-quality compilers are expensive, and are unique to their target platform.

They are usually designed to encompass a huge portfolio of applications and system resources, and as a result impose needless overhead on development.

This prevents developers from extracting all the available cycles from the system, and requires extensive developer expertise to ensure run-time modules of varying provenance can co-exist in the same universe without crashing.

IPTO is interested in your thoughts for new compilers such as these, compilers that transparently adapt to the underlying platform, and hide complexity from less-sophisticated developers.

So.....

as petascale computing technology becomes a reality with HPCS,

IPTO has reached another critical crossroads,

and we want to “kick it up a notch” with Extreme Computing.

Thanks for your attention to our basket of
Easter eggs.

We want your best ideas as we ramp up performance and deal with software complexity.

I've given you a top-down view of some IPTO thoughts, but we need ***your*** fresh ideas...

from the bottom up...

to start the next revolution.

We need your creativity, and we want to hear from your best minds.

Please: don't be shy.

Now I would like to introduce Dan Oblinger, who will speak to you about IPTO's programs in Learning and Reasoning.