Thoughts on Challenges for Molecular Modeling

Paul Saxe
CEO, Materials Design
psaxe@materialsdesign.com

materials design®

Challenges

- Automation
- Forcefields
- Making the tools accessible and usable
  - Why are we doing this?
  - Who is the target audience?
  - Guaranteeing the integrity of results
  - Providing estimates of the accuracy, i.e. error bars

Molecular modeling is a tool, just as a spectrometer, diffractometer, fracture test or other experimental method is. Therefore it needs to act like other tools: be as simple and easy to use as possible, and work with other tools in the environment.

0th Challenge: Nomenclature ☺

Molecular modeling? Atomistic Modeling? Computational Materials Science? Computational Chemistry? Computational Molecular Biology?

Modeling? Simulation?

Molecule? Crystal? System? Model?

Forcefield? Force field? Potential?

Harris Slash 4

4-bit bipolar microprocessor!

System Size Reduced

The system CPU, which incorporates six 4-bit bipolar microprocessor chips, a programmable read-only memory (Prom) bootstrap and eight priority interrupts, is packaged on a single multilayer printed circuit board, the firm said.

Single-bit error correction using 5-bit code is a standard feature, the firm said, and memory is expandable in 48K-byte increments to 768K bytes. The 24-bit word is used throughout the Slash series.

Effective processor cycle time is said to be 600 nsec.

600 ns cycle time!

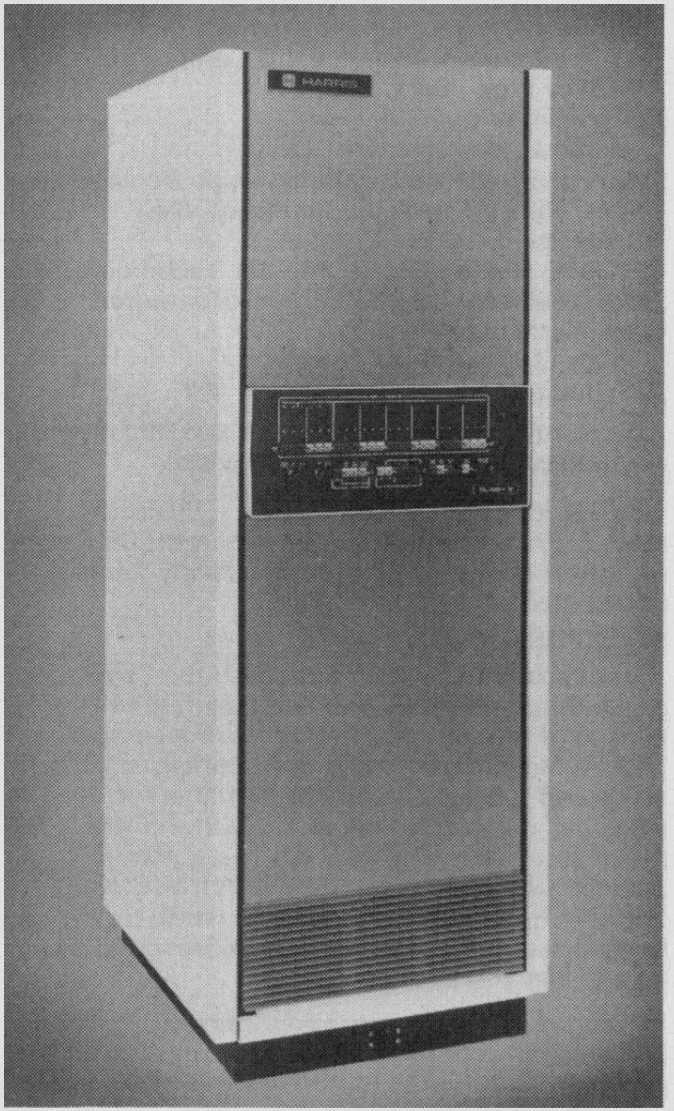
Up to 768 kB memory!

24-bit word / 48-bit double precision!

(with optional Scientific unit)

HARRIS COMPUTER. A medium-scale addition to the line of Slash computers, designed for high-performance and real-time applications, the Slash 6 will fit between and is software compatible with the recently-introduced Slash 7 SuperMini and the Slash 4. Designed around a 4-bit bipolar microprocessor and featuring direct memory access channels for high-speed peripherals, the Slash 6 also permits sharing of peripherals in multiple CPU configurations.
BOOTH NUMBER 2217.

June 1976



CDC-6600

64k 60-bit words

CDC 6600's Five year Reign

The CDC 6600 toppled speed records in 1964. Control Data sold about 100 of them, penetrating markets beyond the usual government and military customers.

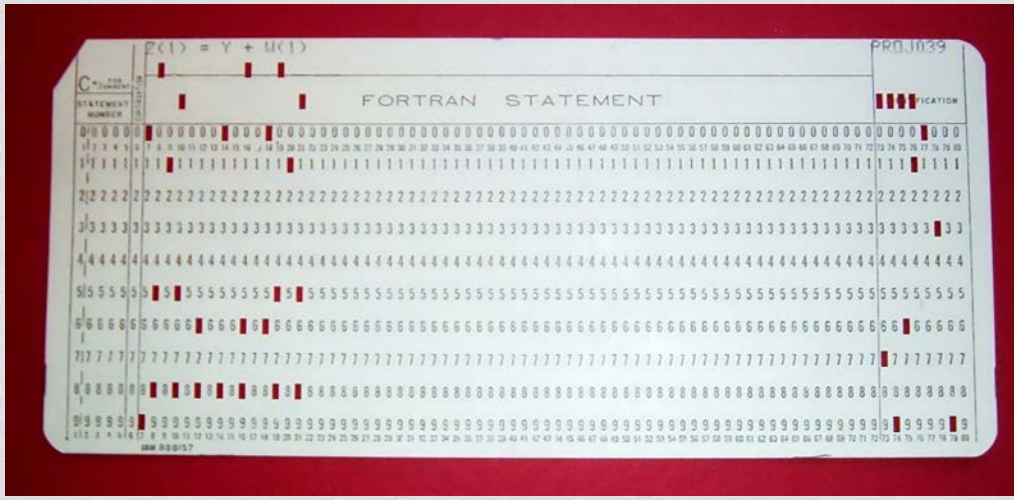
The 6600 executed a dizzying 3,000,000 instructions per second. Designed by supercomputer pioneer Seymour Cray, it wasn't surpassed until the CDC 7600 in 1969—another Cray design.

CDC-6600

The system used a 10 MHz clock, but used a four-phase signal, so the system could at times effectively operate at 40 MHz. A floating-point multiplication took ten cycles, a division took 29, and the overall performance, taking into account memory delays and other issues, was about 3 MFLOPS. Using the best available compilers, late in the machine's history, FORTRAN programs could expect to maintain about 0.5 MFLOPS.

Automation!

- Good old days
  - Assembly code
  - No virtual memory
  - No timesharing on some machines
  - Card decks / UPDATE
  - Overlays
  - Slooooooooow! 0.1-1 MFLOP



What does speed mean?

- 1 MFLOP : 1 month
- 1 GFLOP : 43.2 min
- 1 TFLOP : 2.6 s
- 1 PFLOP : 2.6 ms
- 1 EFLOP : 2.6 us

- Small changes are quantitative
- Large changes are qualitative!

Need to automate – reduce human intervention!

Forcefields (Potentials)

- AMBER/AMBER
- CHARMM/CHARMM
- MM2/MM2
- Discover/CVFF,PCFF,COMPASS
- LAMMPS/?

- Gaussian/Pople basis sets
- VASP/PAW potentials
- MOPAC/PM6,PM7

Forcefields

you seem to be neglecting how little transferable most (empirical) potentials are. unless somebody has parameterized a potential for a specific application that is equal of very similar to what you want to do, you are likely to get bad results. you could practically pick the parameters for any other element and get the same quality of result (i.e. nonsense).

-- Axel Kohlmeyer

Forcefields

- Who is developing wide coverage forcefields?
- Can we provide “forcefields-on-demand”

LAMMPS = Formula 1 car

Forcefields = fuel

Are we using crude oil in our Formula 1 car?

Making the tools accessible and usable

Why are we doing this?

A large majority of us want to solve scientific and engineering problems in materials science, molecular biology and related areas

A small group is most interested in developing the tools themselves

A very small group is interested in the mathematics, physics, etc. of the methods themselves.

➔ Materials scientists, molecular biologists, etc. comprise the target audience

How do we make the tools accessible and usable?

- Focus on the science and engineering enabled by the tools, not the tools per se
- Provide the best tool or combination of tools
- Ensure interoperability
- Provide the complete ecosystem
  - *Builders*
  - *Compute Engines*
  - *Analysis Tools*
  - *Data management and handling*
  - *Compute management and handling*
- Provide guidance, help, documentation
- Education e.g. at Universities, i.e. insert these methods in the curriculum

Challenges

- Automation
- Forcefields
- → Who are the end users!!!

Validation

Ease-of-use

Reliability

Error bars

Human time vs computer time

materials design®

psaxe@materialsdesign.com