Increasing Performance in Computer Systems
How is it done & Why does it matter

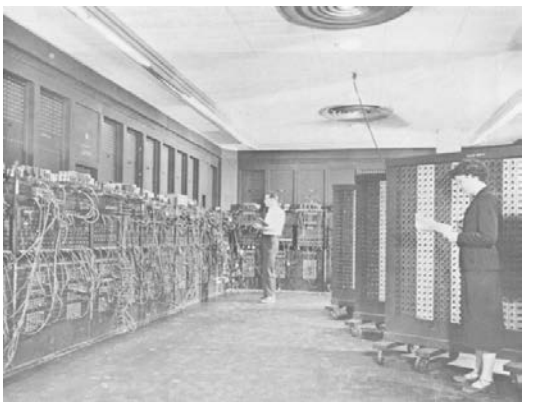
Throughout the last 50 years, computer technology has progressed rapidly. What once would take an entire room and twenty men can now be accomplished with a device we carry in our pockets or on our wrists. This rapid pace has benefitted America and the world. In the past, simple methods were used to speed up systems, but technical issues mean these methods will not work forever, and there are different ways to proceed.

There are many applications in industry and research that are 'compute-bound,' meaning that they would go faster if they were given a faster, more capable computer. These include:

- Weather simulation for climate modeling
- Drug interaction simulation for pharmaceutical research

More powerful computers equate to faster analysis, which can lead to an edge in the global market.

Slowing of the pace of computer development could also lead to a decrease in consumer and business demand for new computers.



The room-sized ENIAC was built in 1946 for calculating missile firing tables. In 1997, researchers recreated the computer in the form of a quarter-inch square silicon chip.

Why can't this rapid pace continue?

Several problems crop up as computer performance increases. These include:

- Power issues
  Chips take much more power to run, which leads to a great deal of waste heat which must be removed.

- Quantum effects
  Parts of computer chips are getting small enough that physical laws that only come into play on such small scales are causing problems in fabrication.

Following are summaries of two possible approaches, one traditional and evolutionary and one highly quite different from other solutions.

Research related to clusters yields widely applicable approaches to accelerating real world problems. While quantum computing shows promise, it is many years from practical usage.

Presenter Contact information:
Sean Riddle . swriddle@ucdavis.edu / Computer Science / One Shields Avenue / Davis, CA 95616

Ways of Improving Computer Performance

Large Scale Clusters / The Cloud

Clusters are formed of many interconnected computers that are used to collectively solve a problem. Cloud computing, in which companies sell their product as a service run on the company's systems, is also becoming increasingly popular.

There is much active research in computer science on the many facets of efficiently coordinating computers that may be on different sides of the world.

This also needs complementary insights from electrical engineers to deal with power consumption, cooling issues, etc.

Pros

- Uses existing manufacturing infrastructure and computer science expertise
- Provides other benefits over increased performance, like redundancy

Cons

- Will provide gradual improvement instead of revolutionary change
- A lot of power is needed to run many computers

The Cloud Tradeoff: Security & Convenience

Security

Computers often store personal data that we want to remain private (eg., medical information). In the same way that people can either keep their money stuffed in their mattress or deposited at a bank, they can store their data on their own computer or in the cloud.

- Upside of the cloud: You don't have to worry about someone breaking into your computer and stealing your data
- Downside of the cloud: If the cloud service is hacked, all data for all clients can be compromised

Convenience

The Internet has already filled the niche of exchanging small files that was filled by the floppy disk.

As Internet access grows even more ubiquitous, it will become a viable way to exchange larger files that has advantages over the current standard, USB flash drives:

- Easily exchange data over large distances - the physical drive does not have to be present
- Nothing to forget
- No fixed capacity - The cloud provider can add resources at any time

State of the art

- Frameworks like Google's Map/Reduce system allow large clusters to easily work on a problem that can be split into chunks.
- Distributed cloud systems let computation and data storage be sold as a service that consumers can purchase only when needed.

Quantum Computing

Quantum computing is a nascent field, quite different from other approaches to computing. The theoretical underpinnings were discovered in 1982, but practical applications are still years away.

Pros

- Accelerates some problems substantially

Cons

- Drastically different model from normal computing would require massive workforce retraining
- Only accelerates a limited group of problems
- Requires a great deal of physics research before it can be scaled to a useful size

State of the art

- D-Wave Systems makes commercially available specialized quantum computers ($10,000,000).
- The Joint Quantum Institute (among other research institutions) is performing basic research.