# Introduction

## Moore’s Law

**Doubles Every 2 years:** Moore's law is the observation that over the history of computing hardware, **the number of transistors** on integrated circuits doubles approximately every 2 years.

**Speed Increase Halted at 2004:** Initially processor frequency increased consistently over the years until 2004. Then processor speed increase halted.



**More Cores Added Since:** Processor computing power increased by introducing more and more cores to the processor.

Moore’s law continue to hold true. But since 2004, the increase comes with multicores.



## Three Walls of single processors

Traditional single processor designs, like the x86 architecture, have reached their maximum performance potential.

Three major limiters have been identified for contemporary processors:

1. **The power wall:**

**Higher clock frequency:** Until recently, increasing performance of processors meant increasing the clock frequency.

**Heating problem:** Higher frequencies require more power, which generates heat that needs to be dissipated or the processor will simply melt.

**Two processors cancelled:** The industry was stunned when Intel cancelled not one but two premier processor designs in May of 2004.

**Too Hot, Abandoned**: The Tejas had been projected to run 7 GHz. It never did. When microprocessors get too hot, they quit working and sometimes blow up.

**Direction Change:** So, Intel quickly changed direction, slowed down their processors, and announced dual-core/multicore.

1. **The memory wall:** The latency of main memory has been a bottleneck for some time, because memory technology has not been able to keep up with advances in processor technology.

**Processors forced to wait idle:** If data cannot be fed to processors fast enough, the processors are forced to idle while waiting for data.

**Secondary Solutions not enough:** Fast local cache and better branch prediction have been able to reduce this problem, but these techniques have reached their limits.

**Slow Memories:** Memory which can match the speed of modern processors are currently not available, and with memory latency approaching 1000 clock cycles many CPU clock cycles are often wasted on waiting for data.

1. **ILP wall (Instruction-level Parallelism):** There is increasing difficulty to find enough parallelism in the instructions stream of a single process to keep higher performance processor cores busy.

**Multicores Everywhere**

Today many PCs have quadcore or more.

Many smart phones have quadcore CPU.

PS4 has 8 cores (2 quadcore). A GPU with 18x64=1152 cores.

## Bigger Problems

In 2004, the U.S. Office of Science and Technology Policy released a report titled “Federal Plan for High-End Computing.”

This report lists four broad application areas:

* climate and weather,
* nanoscale science and technology,
* life sciences,
* aerospace vehicle design

with problems requiring massive amounts of computation, that can and do benefit from parallel computing.

Problems in other areas, such as astrophysics, mathematics, games, animation, are also attacked using parallel computing.

### Weather forecasting:

**Hurricane Katrina:** In August 2005, Hurricane Katrina devastated the U.S. Gulf Coast, flooding the city of New Orleans, killing more than 1,800 people, and causing $100 billion in damage.

**WRF Model:** Computer models of the atmosphere, such as the WRF Model, can predict:

* a storm’s track and
* intensity (wind speed)

**Dividing atmosphere into 3D cells:** The WRF program takes a portion of the earth’s atmosphere—a given region of the earth’s surface, up to a given height above the surface—and divides this three-dimensional region into many small 3-D cells.

**Calculating each cell conditions:** The WRF program then uses physical models to calculate atmospheric conditions in each cell, as influenced by the neighboring cells, and advances the forecast in a series of many small time steps.

**WRF Uses Parallel Computing:** The WRF program uses parallel computing to handle the calculations for large numbers of cells and time steps.

**Accurate prediction is vital:** Accurate hurricane track and intensity forecasts can help officials decide how to prepare for a storm and where to evacuate if necessary.

**Current forecasting not all that accurate:** However, current hurricane forecasting programs are not all that accurate. With current models:

* track forecast errors can be as large as 70 km, and
* wind speed forecast errors can be as large as 37 kilometers per hour.

**Less Error in prediction is vital:** A wind-speed shift of 37 kmph can change a mere tropical storm to a Category 3 major hurricane. A track error of 70 km could cause officials to evacuate Boca Raton, Florida when they should have evacuated Miami.

**Smaller cells needed:** To get more accurate predictions, the cells and the time steps in the model must be made smaller. This means that:

* more cells and
* more time steps

need to be calculated.

**Cell size reduction:** For example, if the cell’s dimensions are decreased by a factor of 2, the number of cells must increase by a factor of 8 to cover the same 3-D region.

**Time step reduction:** If the time step size is also decreased by a factor of 2, the number of time steps must increase by a factor of 2 to cover the same time span.

**More computing power needed:** the total amount of computation goes up by a factor of 16. This in turn means that even more powerful parallel computers and parallel programs will be needed to calculate the models.

### Climate modeling.

**Rain season prediction:** On what date will the rainy season begin in Brazil this year, so farmers will know when to plant their crops?

**Climate Change Analysis:** Why is rainfall decreasing in the Indian subcontinent—could it be caused by pollution from burning wood for fuel and cooking?

**Effects of increased carbon dioxide**: What effect will increased levels of atmospheric carbon dioxide have on the world’s climate—none at all, or drastic warming that will melt the polar ice caps and inundate coastal cities?

**Computer models are used for prediction:** Computer-based climate models can answer these questions (and fan the controversies).

**The Community Climate System Model (CCSM),** for example, models the atmosphere, ocean, sea ice, and land surface using a three-dimensional grid of cells like the WRF model.

**Simulating whole earth climate for thousands of years**: The CCSM program runs on a parallel computer to simulate the earth’s climate over the entire globe for time spans of thousands of years.

**More details, more computing power needed:** Because there is no end to the number of climatological features and the level of detail that can be included in climate simulation programs, such programs will continue to need the power of parallel computers well into the future.

### Protein sequence matching.

**A new protein:** Imagine you are a biochemist. You have isolated a new protein from the creature you are studying, but you have no idea what the protein does—could it be an anti-cancer agent? Or is it just a digestive enzyme?

**Match your protein against other proteins:** One way to get a clue to the protein’s function is to match your protein against other proteins; if your protein closely matches proteins of known function, chances are your protein does something similar to the matching proteins.

**Protein sequence (a string from 20 char alphabet):** Chemically, a protein is a group of amino acids linked together into a long string. Twenty different amino acids are found in proteins, so a protein can be represented as a string of letters from a 20-character alphabet; this string is called the protein’s “sequence.”

**Protein sequence databases:** Protein sequence databases collect information about proteins, including their sequences and functions.

**The Swiss-Prot database**, for example, contains well over 385,000 protein sequences ranging in length from 2 to 35,000 amino acids, with a median length of around 300 amino acids.

**Matching against the database:** You can determine your new protein’s sequence and match it against the protein sequences in the database. However, doing so is more complicated than looking up a credit card number in a financial database.

**Inexact match search:** Rather than finding a single, exact match, you are looking for multiple, inexact but close matches.

**The Basic Local Alignment Search Tool** (BLAST) program is at present the premier tool for solving the preceding protein sequence matching problem.

**local alignment:** The BLAST program combines a “local alignment” algorithm, which matches a piece of one protein sequence against a piece of another protein sequence, with a search of all protein sequences in the database.

**local alignment and parallel computing:** Because local alignment is a computationally intensive algorithm and because the databases are large, parallel versions of BLAST are used to speed up the searches.

### Star cluster simulation.

**Star cluster or galaxy evolution:** Astrophysicists are interested in the evolution of star clusters and even entire galaxies. How does the shape of the star cluster or galaxy change over time as the stars move under the influence of their mutual gravitational attraction?

**Supermassive black hole movement prediction:** Many galaxies, including our own Milky Way, are believed to have a supermassive black hole (SMBH) at the center. How does the SMBH move as the comparatively much-lighter stars orbit the galactic center?

**Galaxy interaction prediction:** What happens when two galaxies collide? The galaxies as a whole might merge, or they might pass through each other but with altered shapes, or certain stars might be ejected to voyage alone through intergalactic space.

**Computational astrophysics:** Astrophysicists have turned to observing the evolution of star clusters or galaxies simulated in the computer. In recent years, “computational astrophysics” has revolutionized the field and revealed a host of new phenomena for theorists to puzzle over.

**Direct N-body methods**: The most general and powerful methods for simulating stellar dynamics, the so-called “direct N-body methods,” require enormous amounts of computation.

**Time steps:** The simulation proceeds as a series of time steps.

**Gravitational force calculation for each star:** At each time step, the gravitational force on each star from all the other stars is calculated, each star’s position and velocity are advanced as determined by the force, and the cycle repeats.

**Computing requirements:** A system of N stars requires O(N) calculations to determine the forces for one star and O(N2) calculations to determine the forces for N stars.

**Calculations for one step:** To simulate, say, one million stars requires 1012 force calculations—on each and every time step; and one simulation may run for thousands or millions of time steps.

**Special purpose hardware:** To run their simulations, computational astrophysicists may use special purpose hardware or parallel computing.

**GRAPE-6:** One example is the GRAPE-6 processor, developed by Junichiro Makino and his colleagues at the University of Tokyo. (GRAPE stands for GRAvity piPE.)

**Special Purpose Processor:** The GRAPE-6 processor is a parallel processor that does only gravitational force calculations, but does them much, much faster than even the speediest general-purpose computer.

**Gravitational supercomputer:** Multiple GRAPE-6 processors are then combined with multiple general-purpose host processors to form a massively parallel gravitational supercomputer.

**gravitySimulator** is another special purpose system built by David Merritt and his colleagues at the Rochester Institute of Technology.

### Mersenne primes

**Mersenne prime:** Mersenne numbers are numbers of the form 2n–1. If a Mersenne number is also a prime number, it is called a Mersenne prime.

**The first few Mersenne primes** are 22–1, 23–1, 25–1, 27–1, 213–1.

**The largest known Mersenne prime** is 243,112,609–1, with nearly 13 million decimal digits, discovered in August 2008 by the Great Internet Mersenne Prime Search (GIMPS) project.

**Starting in 1996**, the GIMPS project has been testing Mersenne numbers to find ever-larger Mersenne primes.

**The GIMPS project** uses a “virtual parallel computer” to test candidate Mersenne numbers for primality in parallel. The GIMPS parallel computer consists of PCs and workstations contributed by volunteers around the globe.

**Volunteered PCs:** Each volunteer downloads and installs the GIMPS client program on his or her computer.

**GIMPS client**: The client runs as a lowest-priority process, and thus uses CPU cycles only when the computer is otherwise idle. The client contacts the GIMPS server over the Internet, obtains a candidate Mersenne number to work on, subjects the candidate to an array of tests to determine whether the candidate is prime, and reports the result back to the server.

**Long computing times:** Because the candidate Mersenne numbers are so large, the primality tests can take days to weeks of computation on a typical PC.

**GIMPS Discoveries:** Since commencing operation, the GIMPS project has found 12 previously unknown Mersenne primes with exponents ranging from 1,398,269 to the aforementioned 43,112,609.

**SETI and others:** Search for extraterrestrial intelligence (SETI) and many other projects use the same approach as GIMPS to run programs in volunteer computers.

**BOINC:** Berkeley Open Infrastructure for Network Computing (BOINC), a general framework for Internet-based client-server parallel computation.

**GIMPS virtual supercomputer**: As of March 2012, GIMPS has a sustained throughput of approximately 86.1 teraflops, theoretically earning the GIMPS virtual computer 153rd place among the TOP500 most powerful known computer systems in the world.

### Chess

**On May 11, 1997**, Garry Kasparov, chess grandmaster and then world chess champion, sat down opposite IBM chess computer Deep Blue for a six-game exhibition match.

**15 months Earlier Kasparov prevailed:** Kasparov and Deep Blue had met 15 months earlier, on February 10, 1996, and at that time Kasparov prevailed with three wins, two draws, and one loss.

**This Time DeepBlue won:** After extensive upgrades, Deep Blue was ready for a rematch. This time, the results were two wins for Deep Blue, one win for Kasparov, and three draws, for an overall score of Deep Blue 3.5, Kasparov 2.5.

**A first for a computer:** It was the first time a computer had won a match against a reigning human world chess champion. After his defeat, Kasparov demanded a rematch, but IBM refused and retired the machine.

**In a museum:** Deep Blue’s two equipment racks are now gathering dust at the National Museum of American History and the Computer History Museum.

**Deep Blue was a massively parallel, special-purpose supercomputer**, consisting of 30 IBM RS/6000 SP server nodes augmented with 480 specialized VLSI chess chips.

**It chose chess moves by brute force**, analyzing plays at a rate of 200 million chess positions per second. It also had an extensive database of opening positions and endgames.

**Software focus:** While Deep Blue is gone, computer chess research continues. The focus has shifted away from specialized parallel hardware to software programs running on commodity parallel machines.

**The Deep Fritz chess program,** in November 2006, a parallel program running on a PC with two Intel dual-core CPU chips, beat then world chess champion Vladimir Kramnik with two wins and four draws.

**End of the story:** Many people now believe that in the realm of chess, human dominance over computers is at its end.

**Watson on Jeopardy:**

* IBM computer to answer jeopardy questions (Riziko in Turkish).
* Watson competed two former winners and won in 2011.
* The system has 2,880 POWER7 processor cores and has 16 terabytes of RAM.

### Animated films

**First Animated Film:** In 1937, Walt Disney made motion picture history with the animated feature film Snow White and the Seven Dwarfs.

**Each frame is drawn by hand:** In those days each frame of the film was laboriously drawn and colored by hand on celluloid.

**First computer-animated film:** All that would change in 1995, when Pixar Animation Studios and Walt Disney Pictures released **Toy Story**, the world’s first feature-length computer-animated film.

**Many animated films:** Since then, scarcely a year has gone by without several new feature-length computer-animated film releases.

**Early works for animated films:** During the early stages of production on a computer-animated film, the artists and designers work mostly with individual high-end graphics workstations.

**Rendering and render farms:** But when the time comes to “render” each frame of the final film, adding realistic surface textures, skin tones, hair, fur, lighting, and so on, the computation shifts to the “render farm”—a parallel computer with typically several thousand nodes, each node a multicore server.

**For Toy Story**, the render farm had to compute a 77-minute film, with 24 frames per second and 1,536×922 pixels per frame—more than 157 billion pixels altogether.

**Longer Rendering Times:** Despite the render farm’s enormous computational power, it still takes hours or even days to render a single frame.

**Larger parallel computers needed:** As movie audiences come to expect ever-more-realistic computer-animated films, the render farms will continue to require increasingly larger parallel computers running increasingly sophisticated parallel rendering programs.