Information, Logistics, and 21st Century Research

http://www.esp.org/rjr/ccaf.pdf

Robert J. Robbins
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue North, LV-101
Seattle, Washington 98109

rrobbins@fhcrc.org
(206) 667 2920

Abstract

Over the next few years, the relentless exponential effect of Moore's Law will profoundly affect nearly all areas of science and technology. By 2005, analytical power previously available only at supercomputer centers will exist on every desktop and the volume of electronic data will be enormous. Even now, a standard Intel computer delivers more computational power than the first supercomputer and GenBank acquires more data every ten weeks than it did in its first ten years.

Advanced biomedical research of the 21st Century will require an adequate information infrastructure and a mastery of logistics. Those with access will participate in the transformation of biomedical science; those without may become irrelevant. If public support for information infrastructure is inadequate, some types of research may only be possible in the private sector.

Topics

- Cancer is a pathology of the genetic system.

Topics

- Cancer is a pathology of the genetic system.
- Moore’s Law constantly transforms IT (and everything else).

Topics

- Cancer is a pathology of the genetic system.
- Moore's Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology, especially genetics and genomics.

Topics

- Cancer is a pathology of the genetic system.
- Moore's Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology, especially genetics and genomics.
- 21st-Century biology will be based on bioinformatics and powered by logistics skills.

Topics

- Cancer is a pathology of the genetic system.
- Moore’s Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology, especially genetics and genomics.
- 21st-Century biology will be based on bioinformatics and powered by logistics skills.
- Bioinformatics is emerging as an independent discipline.

Topics

- Cancer is a pathology of the genetic system.
- Moore's Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology, especially genetics and genomics.
- 21st-Century biology will be based on bioinformatics and powered by logistics skills.
- Bioinformatics is emerging as an independent discipline.
- Currently, support for public bio-information infrastructure seems inadequate.

Topics

- Cancer is a pathology of the genetic system.
- Moore's Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology, especially genetics and genomics.
- 21st-Century biology will be based on bioinformatics and powered by logistics skills.
- Bioinformatics is emerging as an independent discipline.
- Currently, support for public bio-information infrastructure seems inadequate.
- In the future, much current public research may move into the private sector.

Cancer as Pathology of Genetic System

Moore's Law

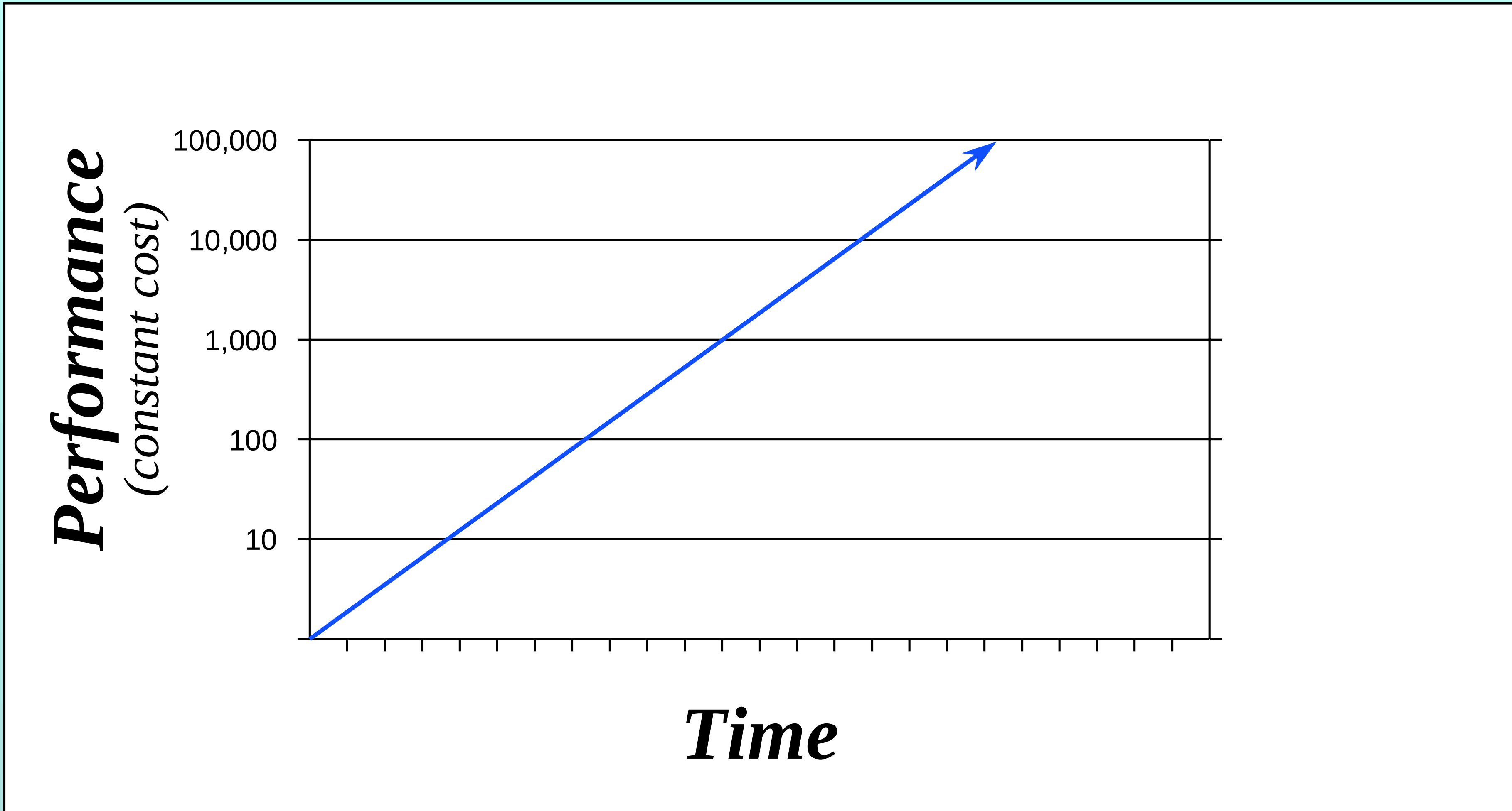
Transforms InfoTech (and everything else)

Moore's Law: *The Statement*

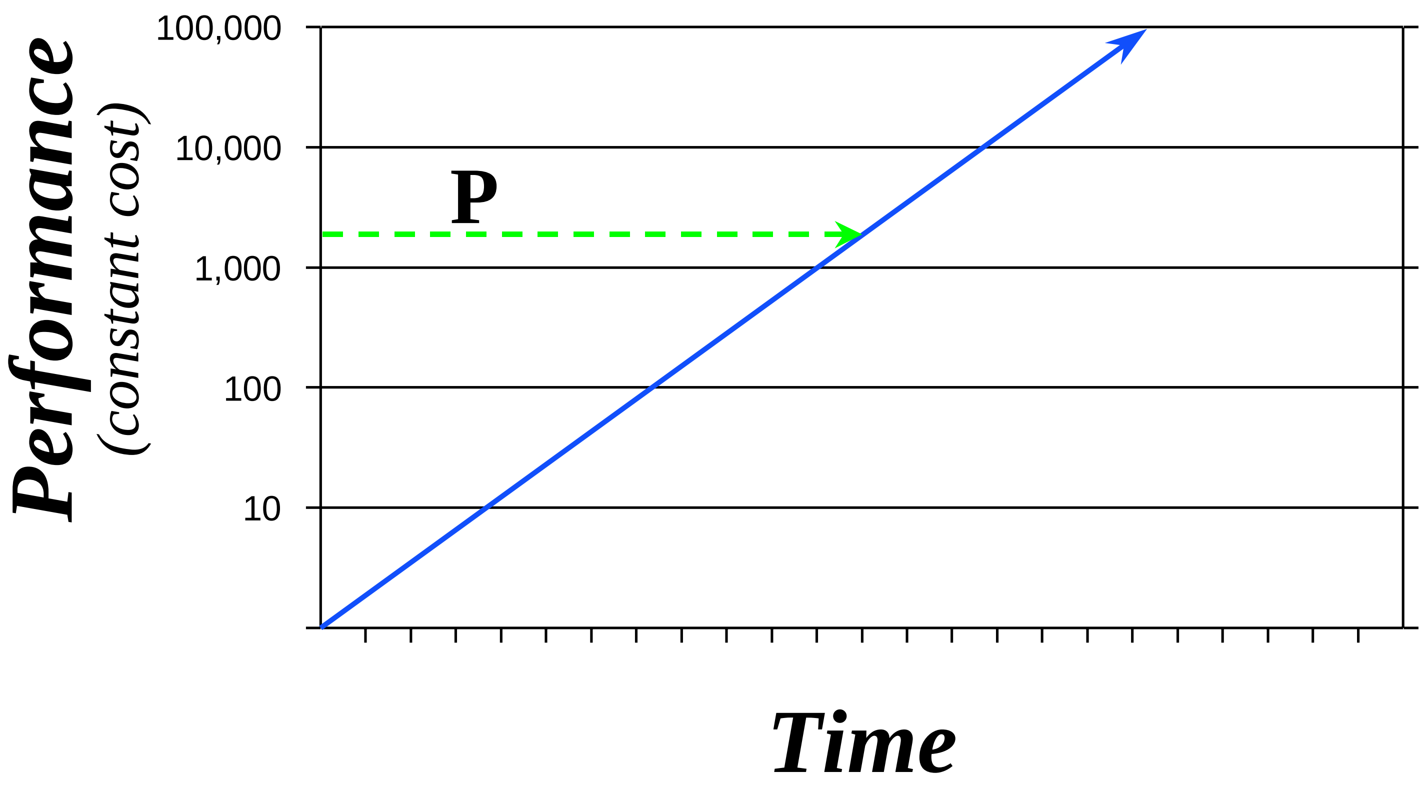
Every eighteen months, the number of transistors that can be placed on a chip doubles.

Gordon Moore, co-founder of Intel...

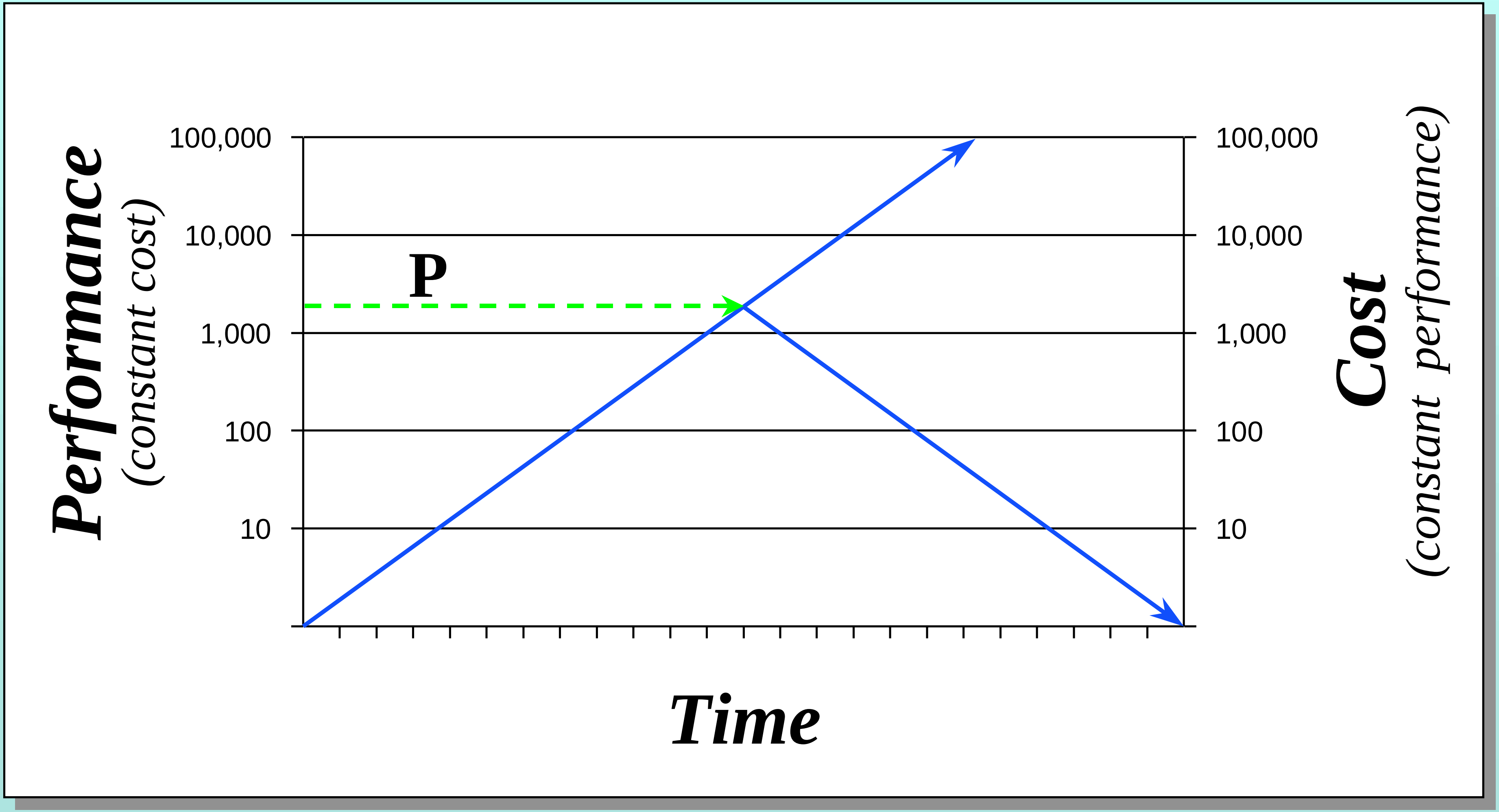
Moore's Law: *The Effect*



Moore's Law: *The Effect*



Moore's Law: *The Effect*



Moore's Law: *The Effect*

Three Phases of Novel IT Applications

- It's Impossible

Moore's Law: *The Effect*

Three Phases of Novel IT Applications

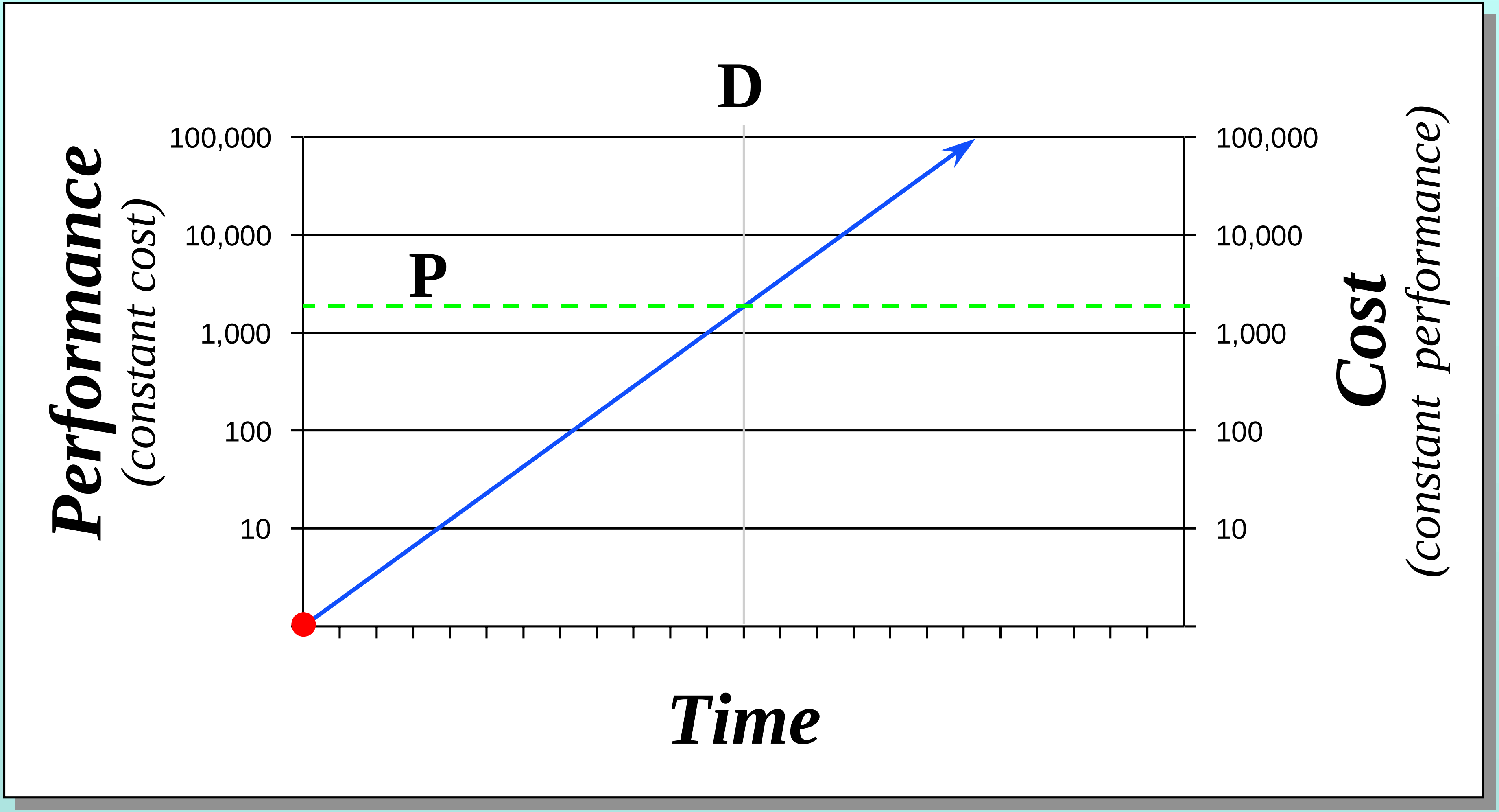
- It's Impossible
- It's Impractical

Moore's Law: *The Effect*

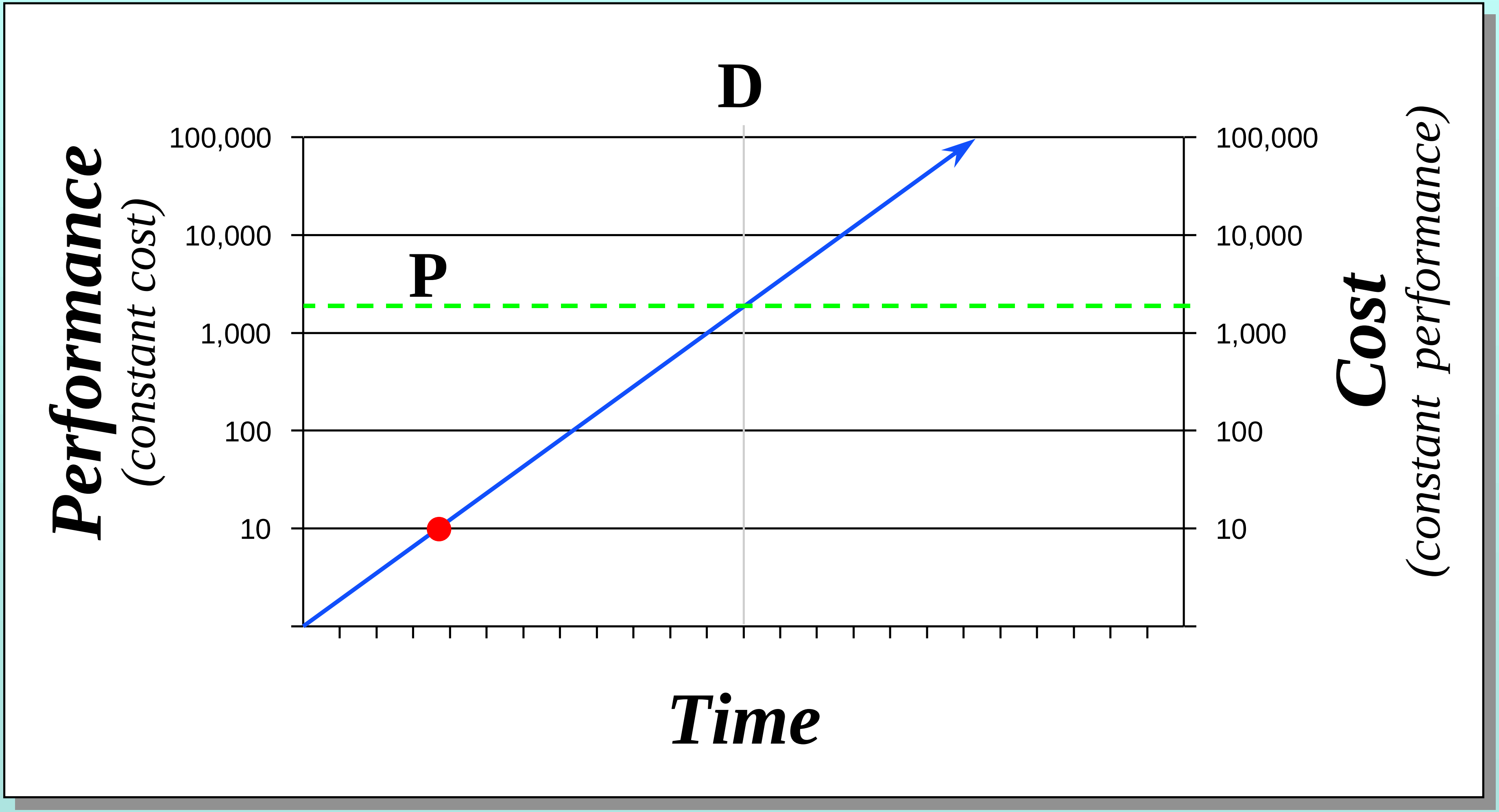
Three Phases of Novel IT Applications

- It's Impossible
- It's Impractical
- It's Overdue

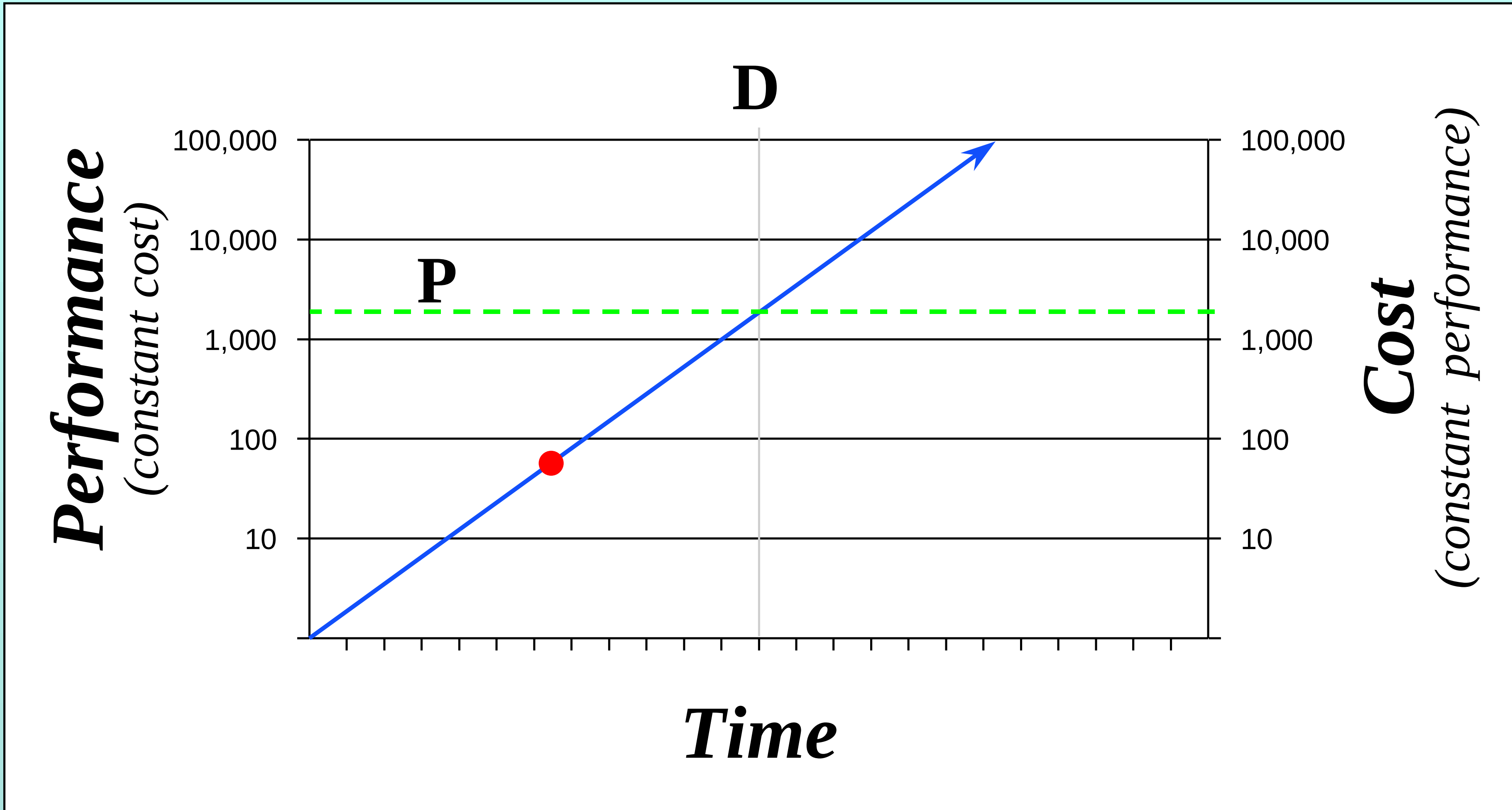
Moore's Law: *The Effect*



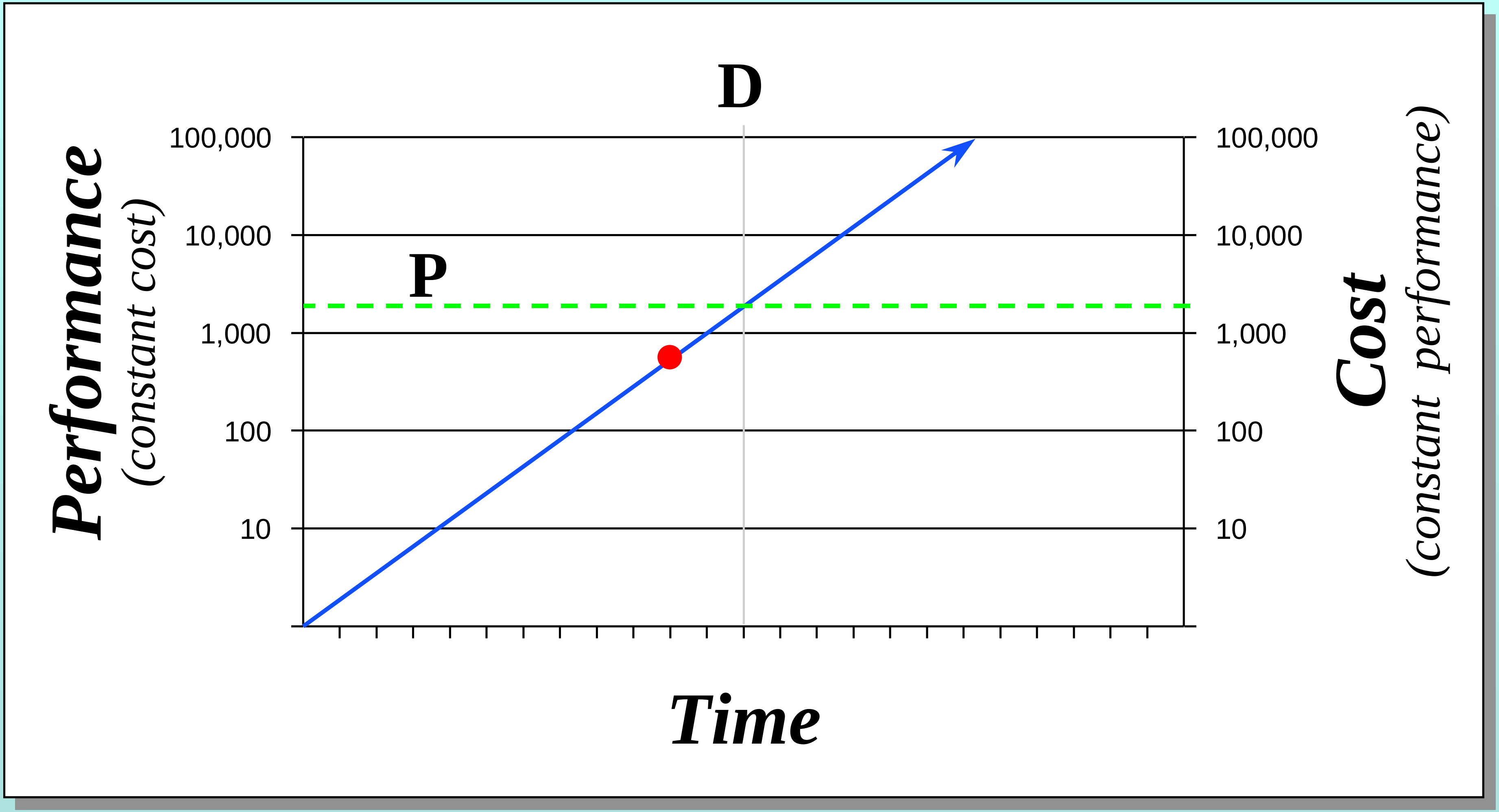
Moore's Law: *The Effect*



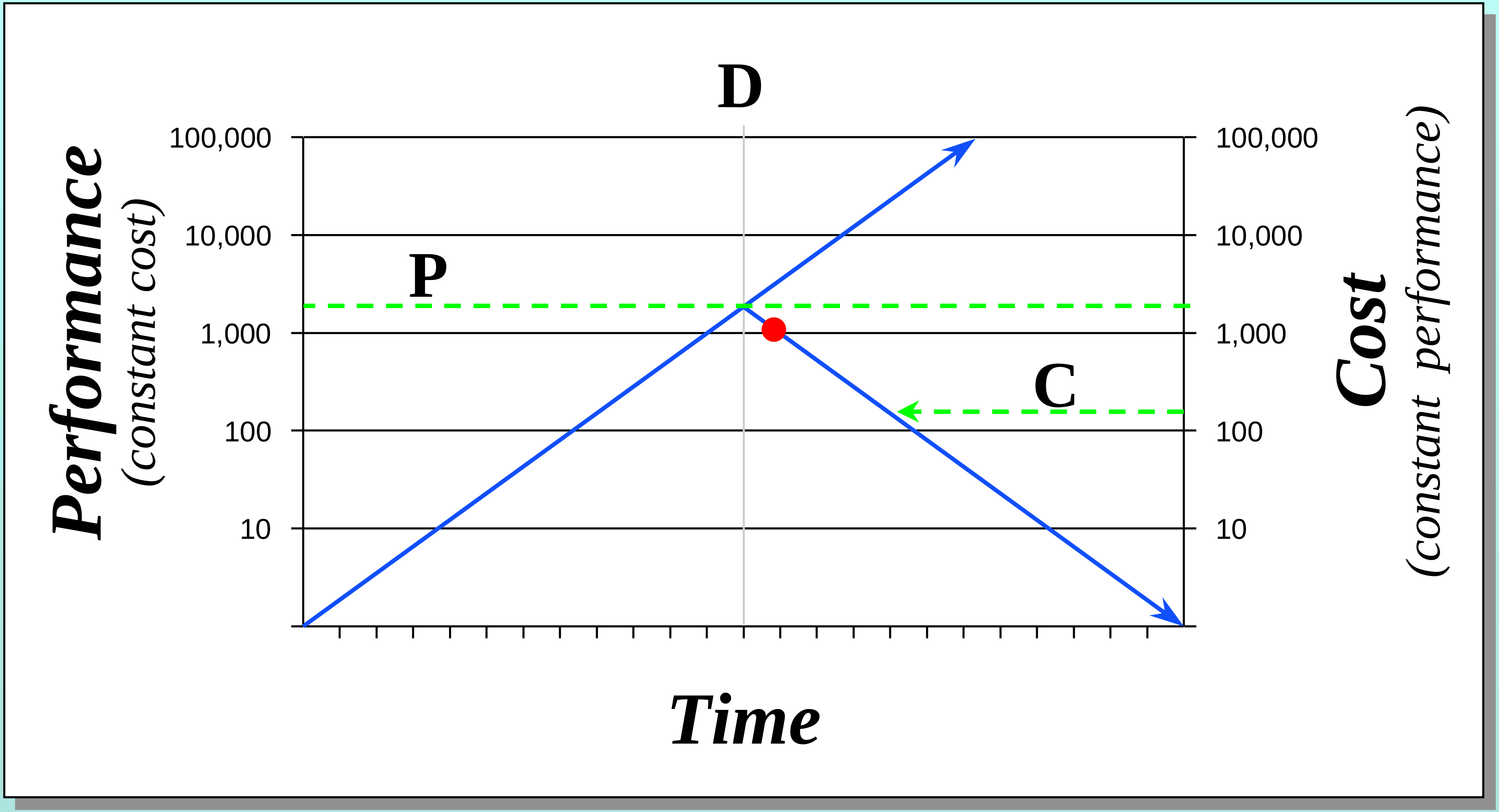
Moore's Law: *The Effect*



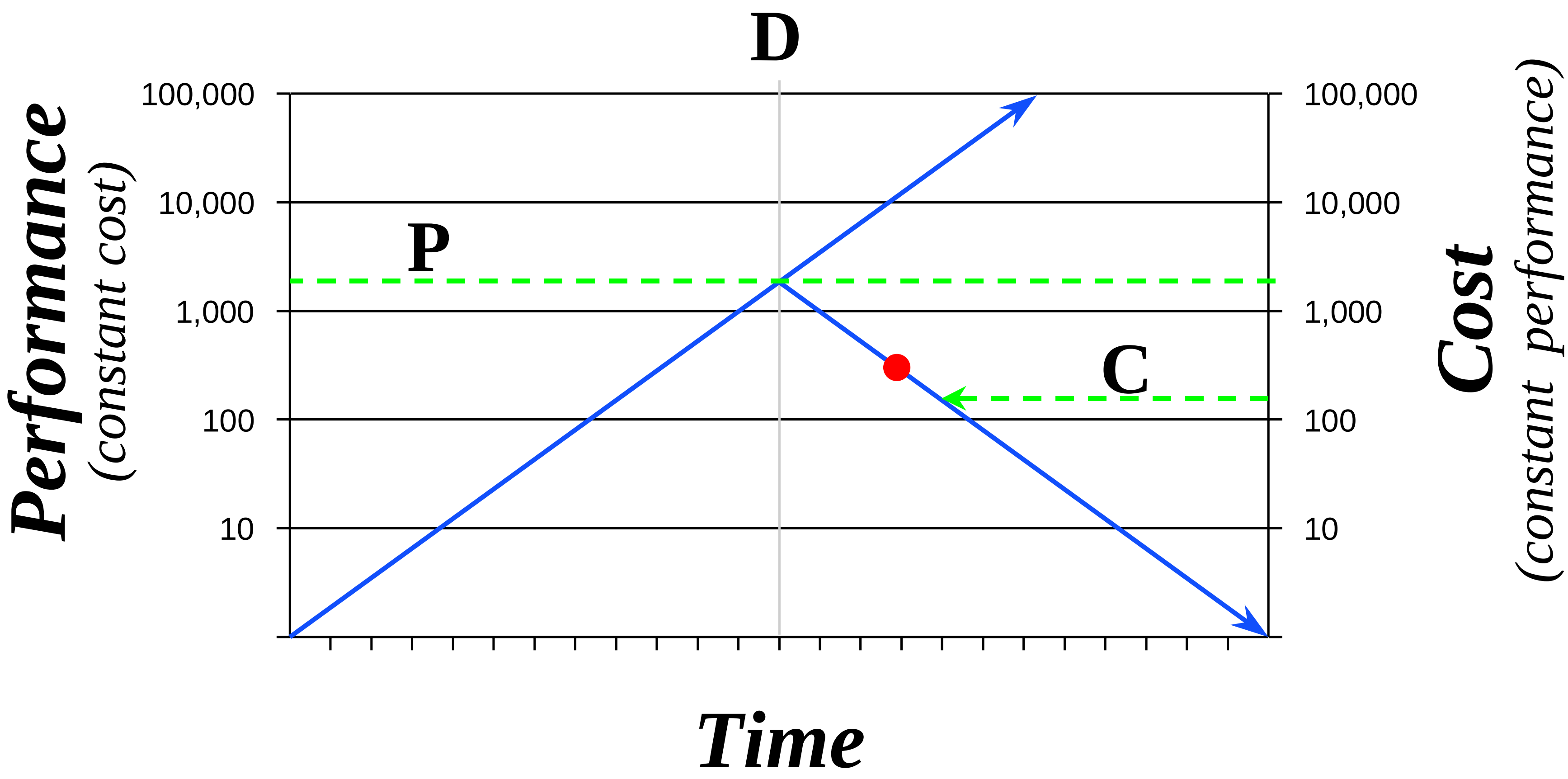
Moore's Law: *The Effect*



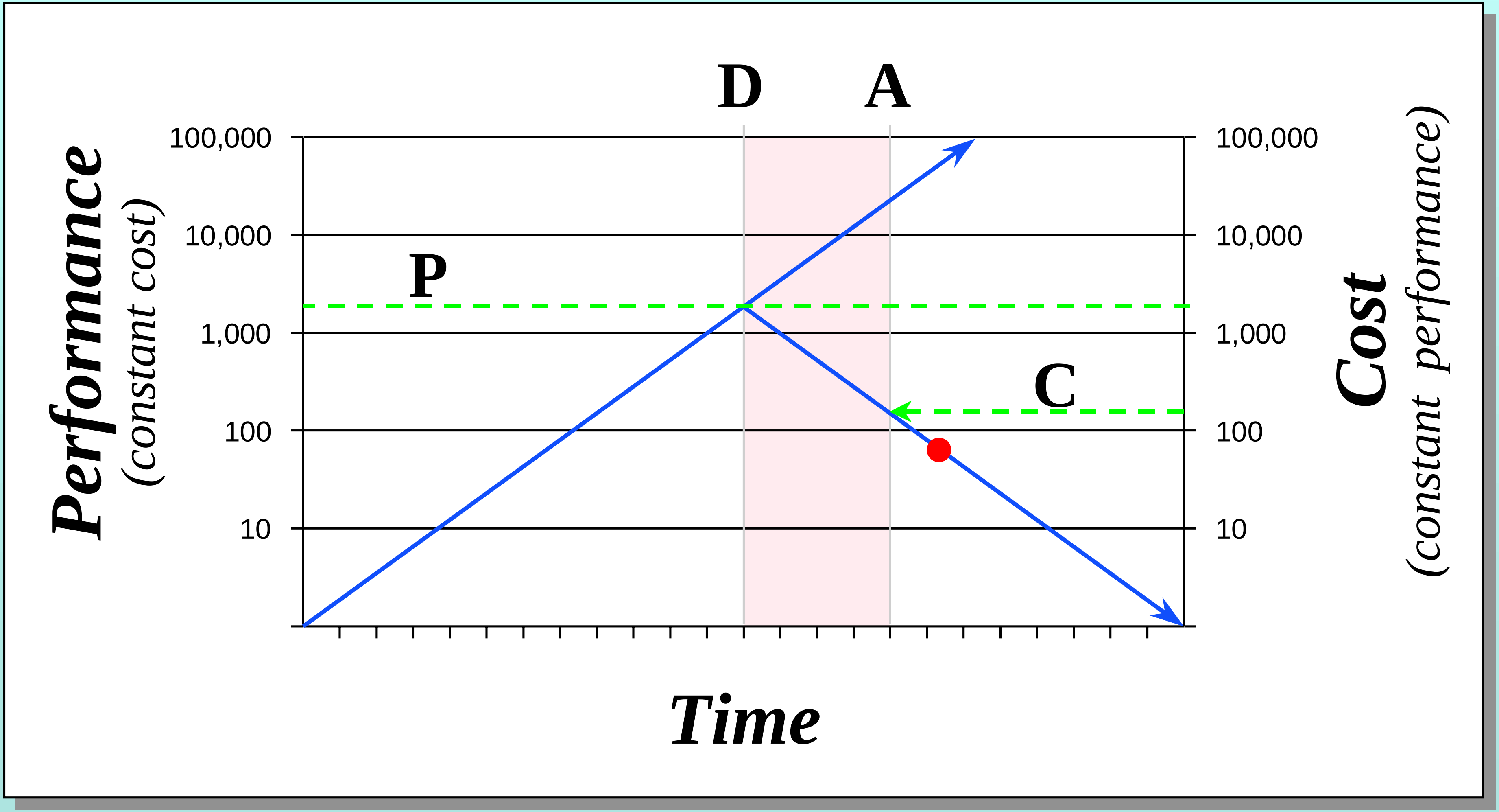
Moore's Law: *The Effect*



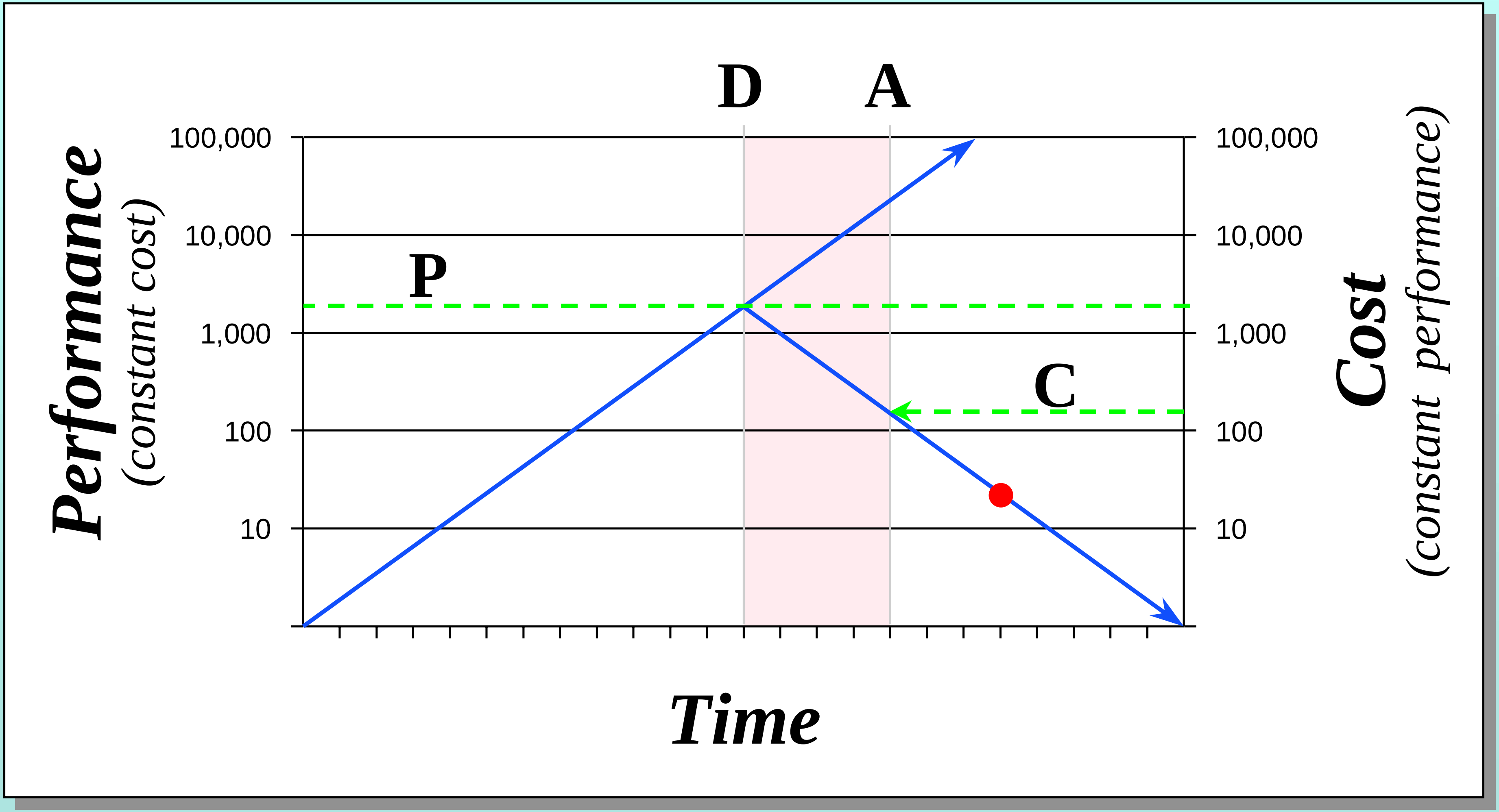
Moore's Law: *The Effect*



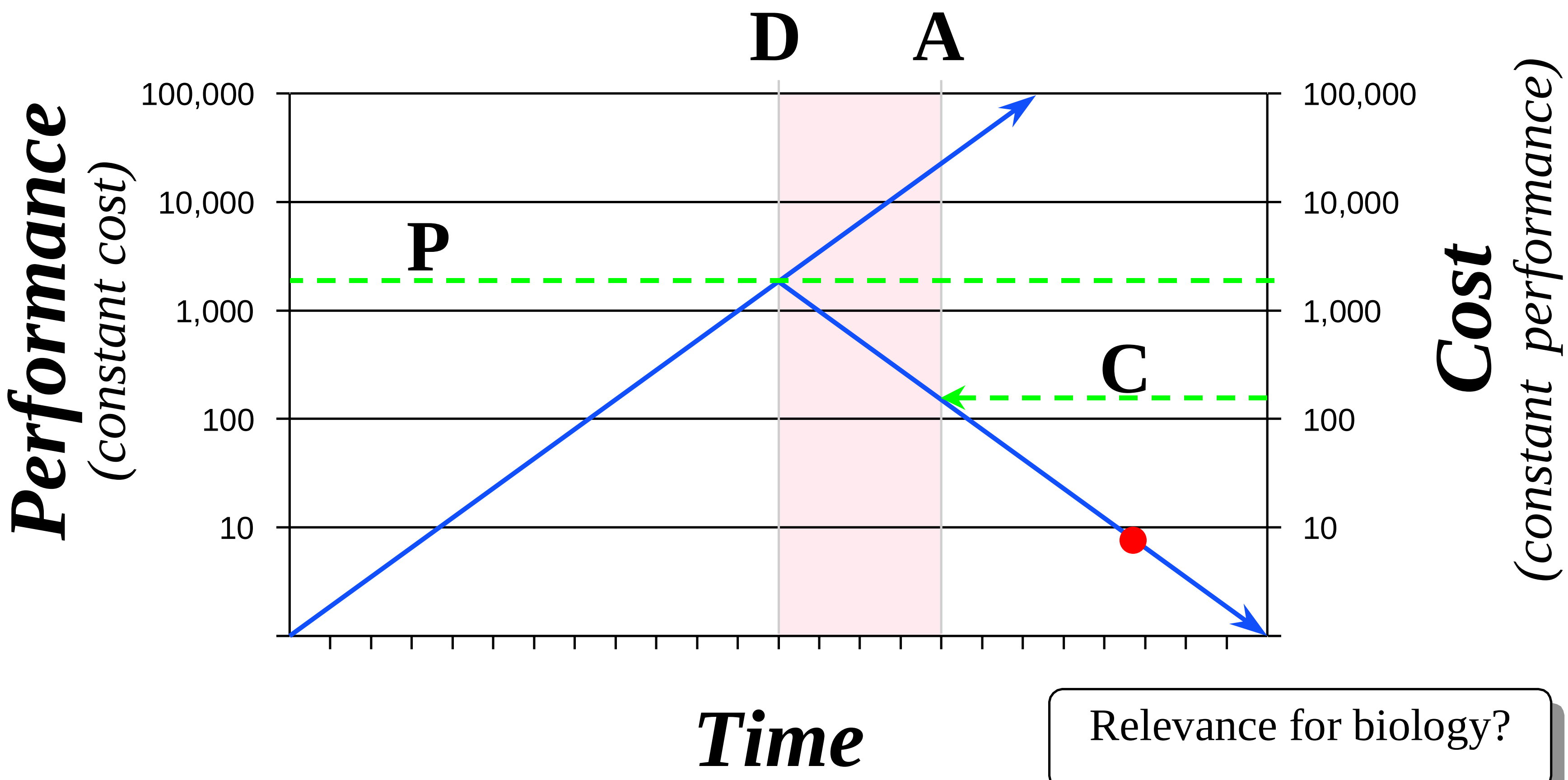
Moore's Law: *The Effect*



Moore's Law: *The Effect*

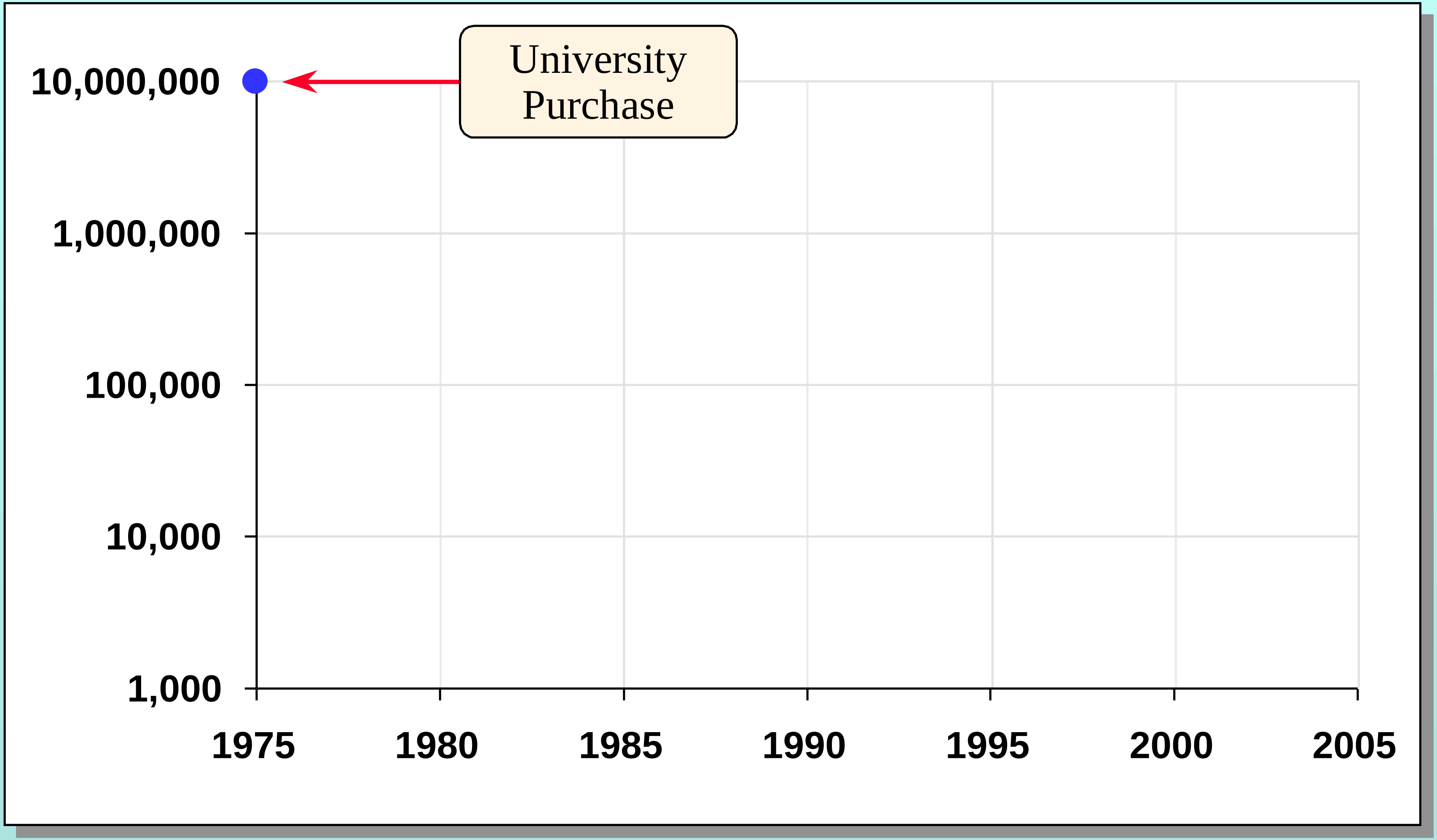


Moore's Law: *The Effect*

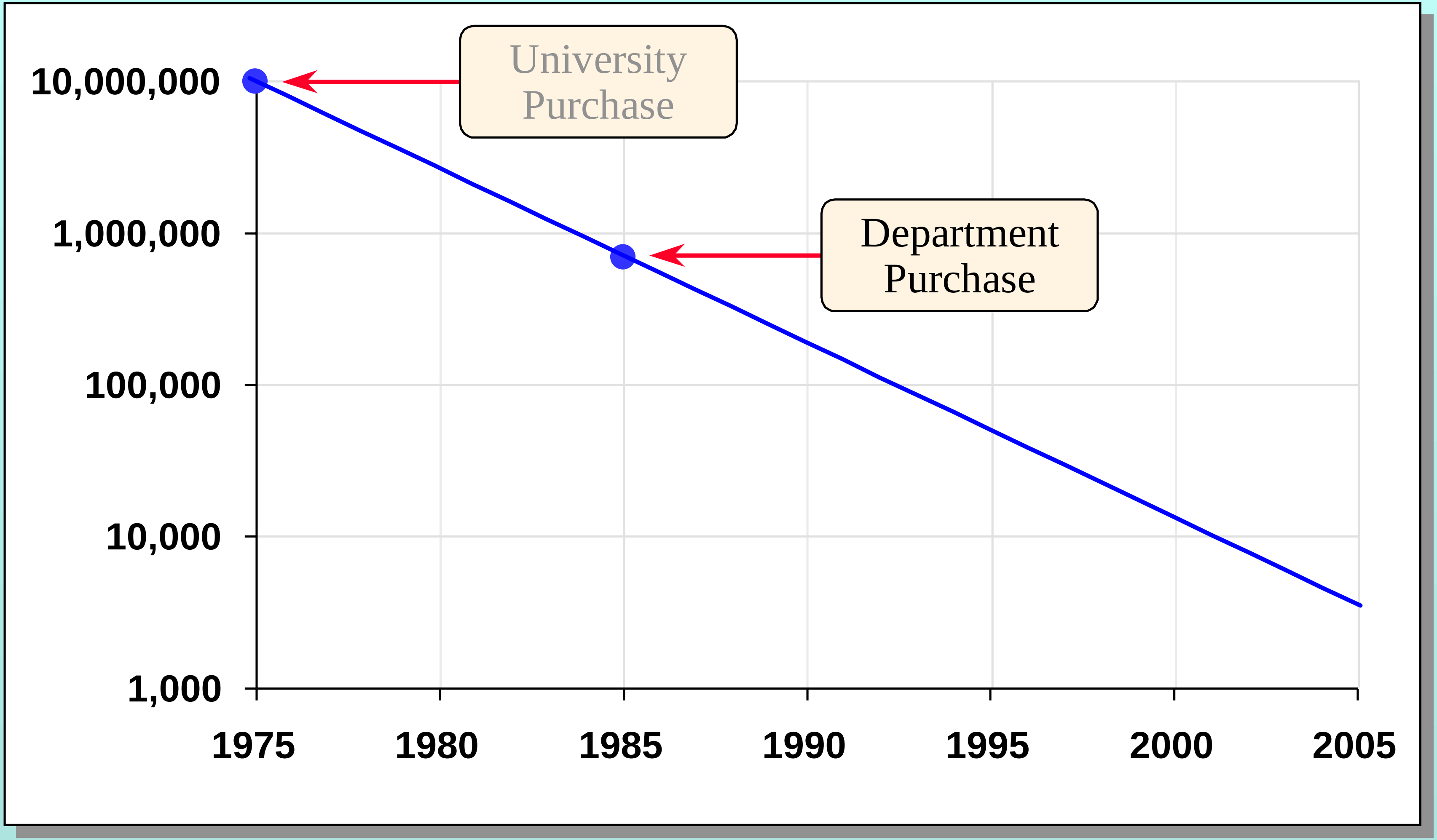


Relevance for biology?

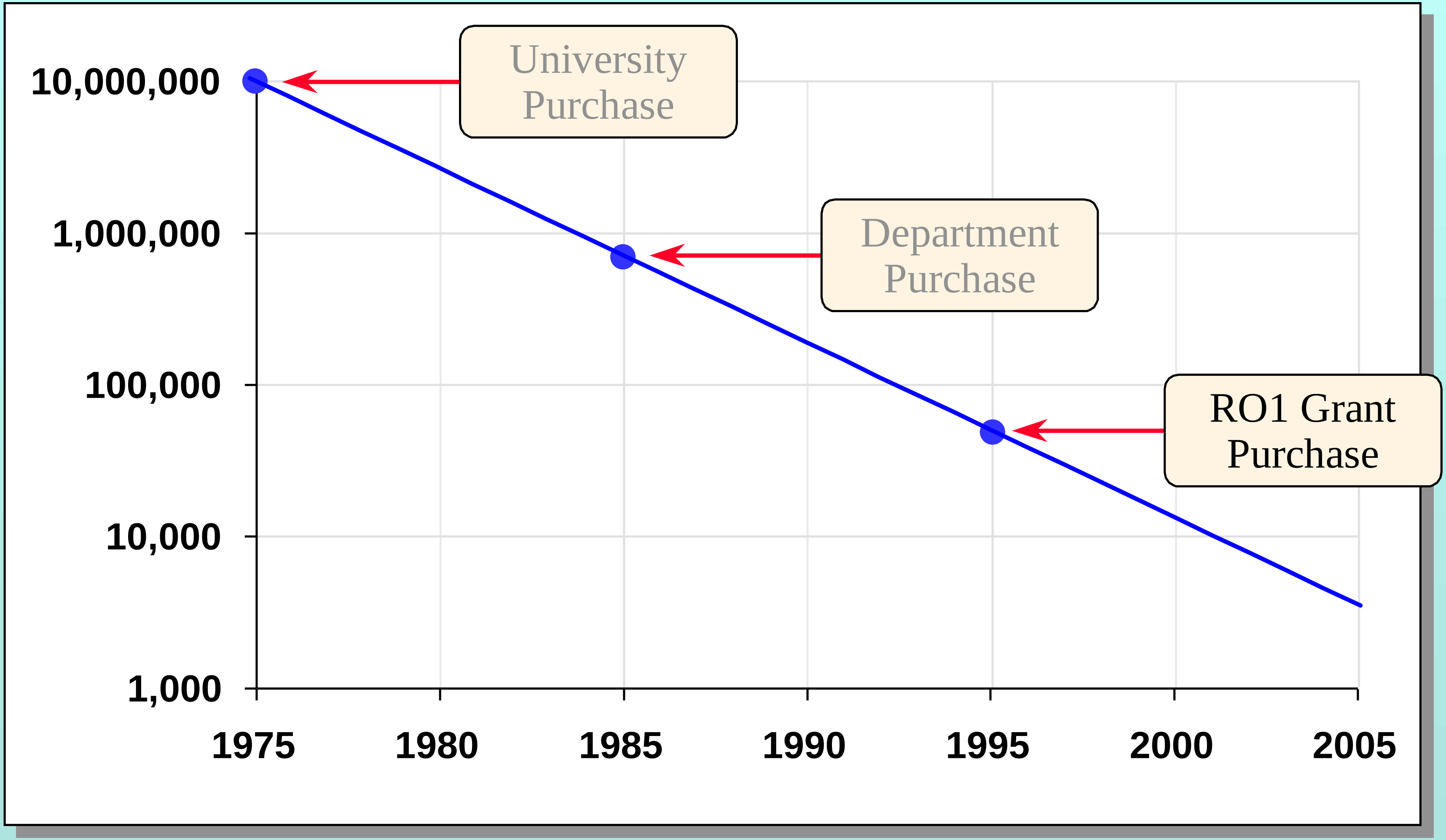
Cost (constant performance)



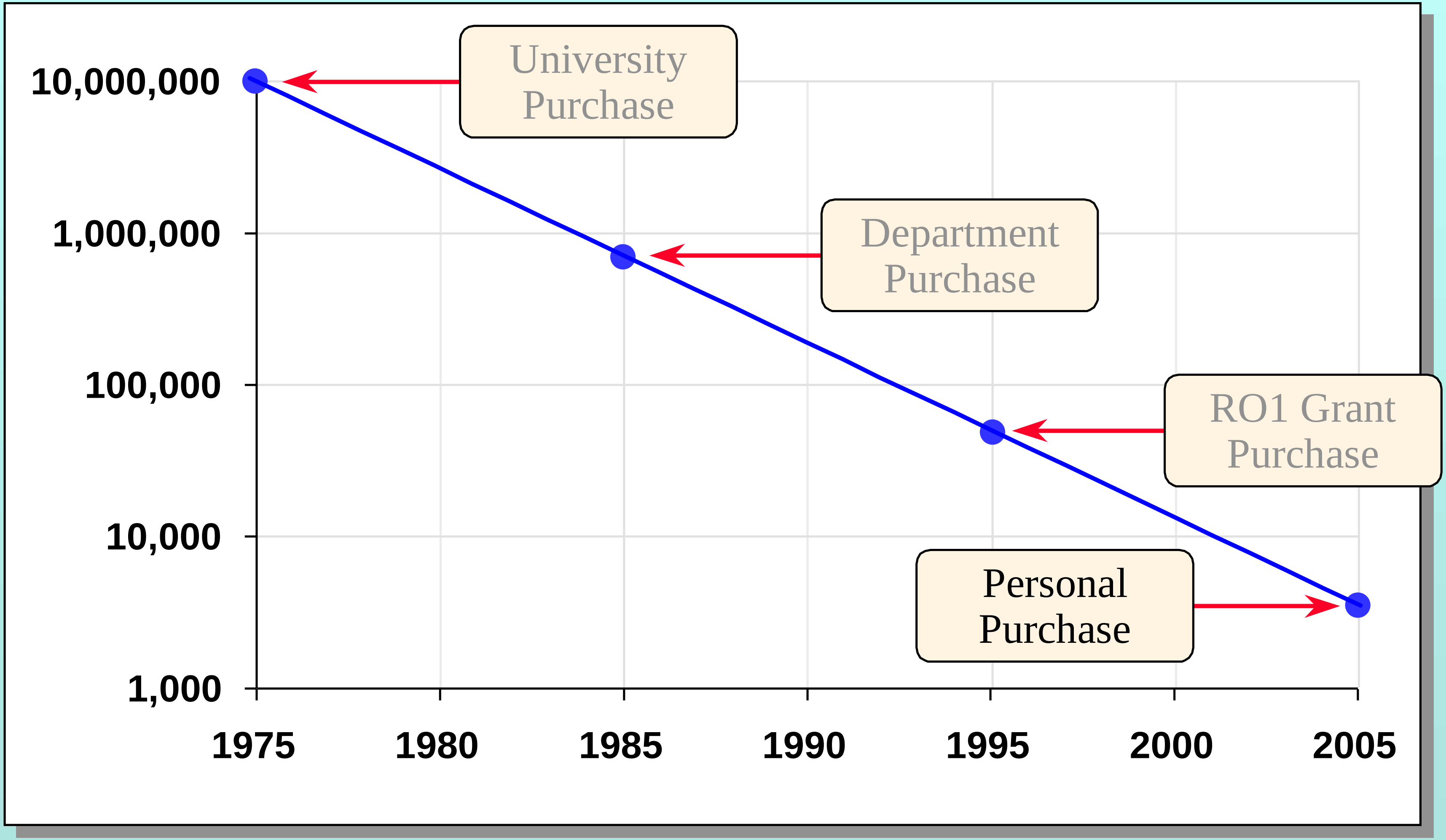
Cost (constant performance)



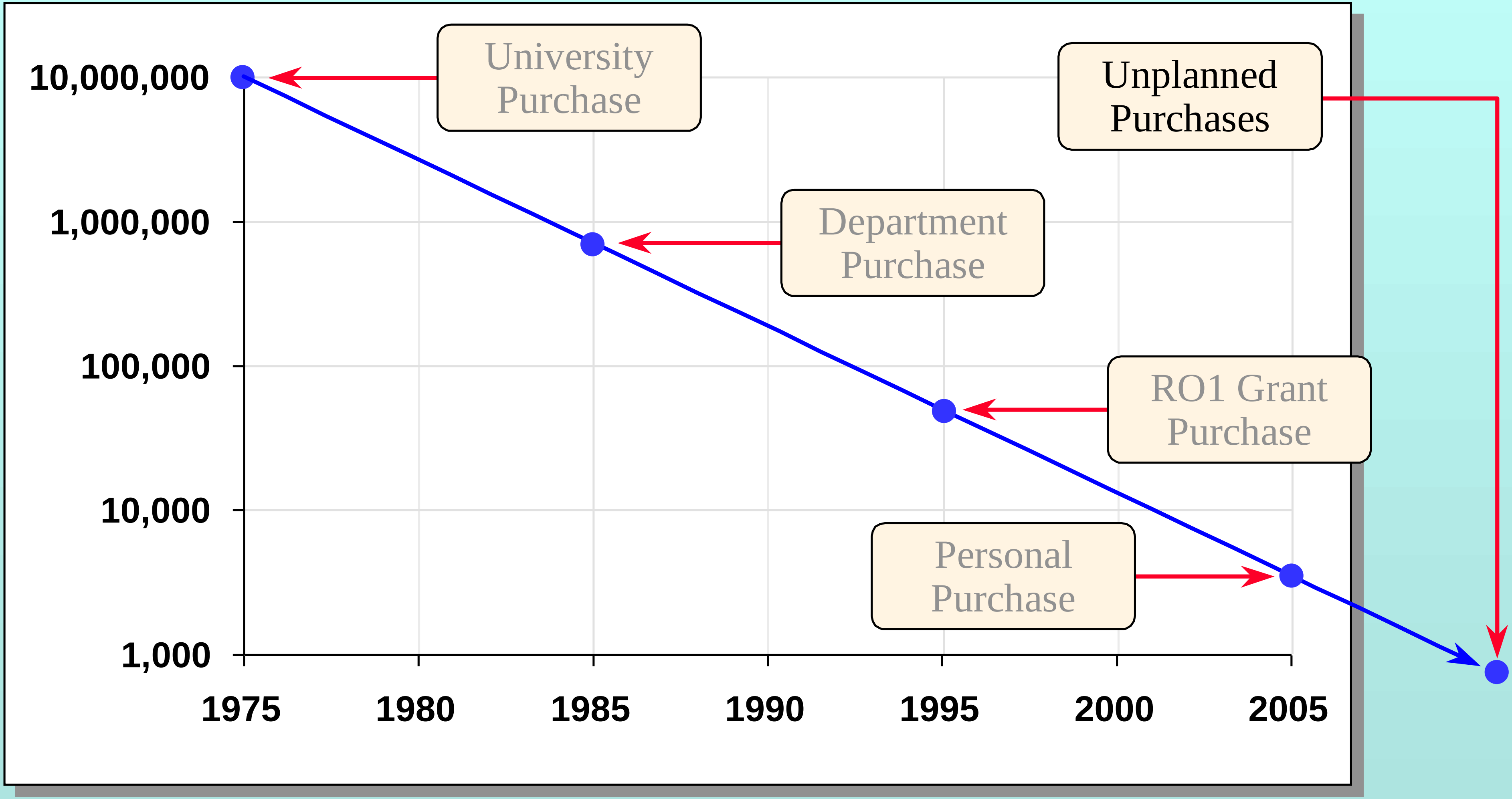
Cost (constant performance)



Cost (constant performance)



Cost (constant performance)



IT-Biology Synergism

IT is Special

Information Technology:

- *affects the performance **and** the management of tasks*

IT is Special

Information Technology:

- *affects the performance **and** the management of tasks*
- *allows the manipulation of huge amounts of highly complex data*

IT is Special

Information Technology:

- *affects the performance **and** the management of tasks*
- *allows the manipulation of huge amounts of highly complex data*
- *is incredibly plastic*

  (programming and poetry are both exercises in pure thought)

IT is Special

Information Technology:

- *affects the performance **and** the management of tasks*
- *allows the manipulation of huge amounts of highly complex data*
- *is incredibly plastic*
  (programming and poetry are both exercises in pure thought)
- *improves exponentially* *(Moore's Law)*

Biology is Special

Life is Characterized by:

- *individuality*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*
- *high (digital) information content*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*
- *high (digital) information content*

No law of large numbers...

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*
- *high (digital) information content*

No law of large numbers, since every living thing is genuinely unique.

IT-Biology Synergism

- *Physics needs calculus, the method for manipulating information about statistically large numbers of vanishingly small, independent, equivalent things.*

IT-Biology Synergism

- *Physics needs calculus, the method for manipulating information about statistically large numbers of vanishingly small, independent, equivalent things.*
- *Biology needs information technology, the method for manipulating information about large numbers of dependent, historically contingent, individual things.*

Biology is Special

For it is in relation to the statistical point of view that the structure of the vital parts of living organisms differs so entirely from that of any piece of matter that we physicists and chemists have ever handled in our laboratories or mentally at our writing desks.

Erwin Schrödinger. 1944. *What is Life*.

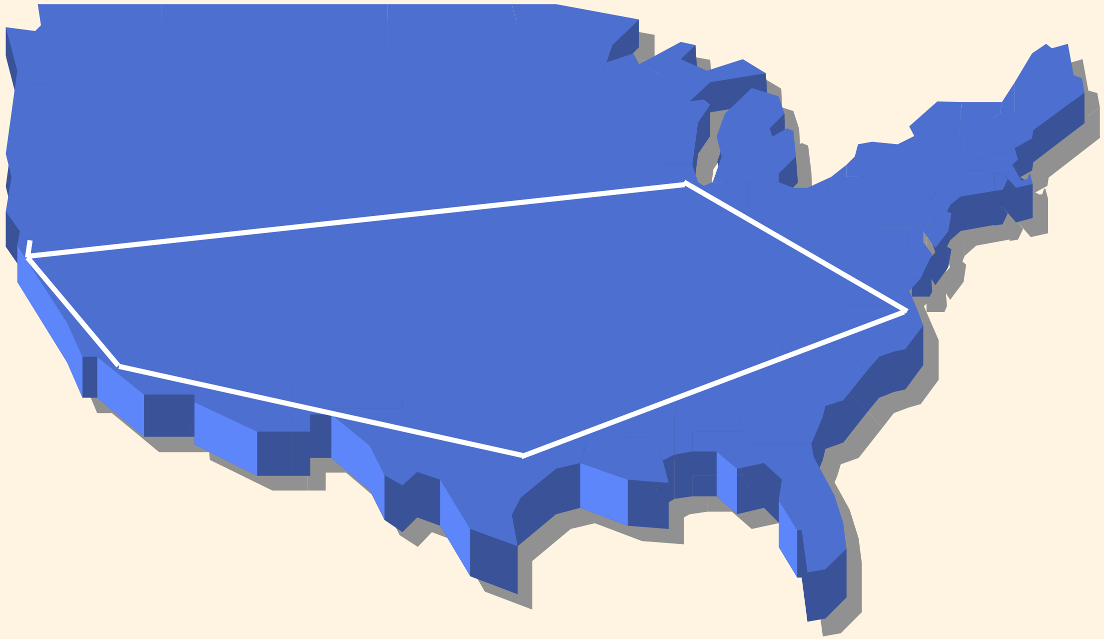
Genetics as Code

[The] chromosomes ... contain in some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state. ... [By] code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, could tell from their structure whether [an egg carrying them] would develop, under suitable conditions, into a black cock or into a speckled hen, into a fly or a maize plant, a rhodo-dendron, a beetle, a mouse, or a woman.

Erwin Schrödinger. 1944. *What is Life*.

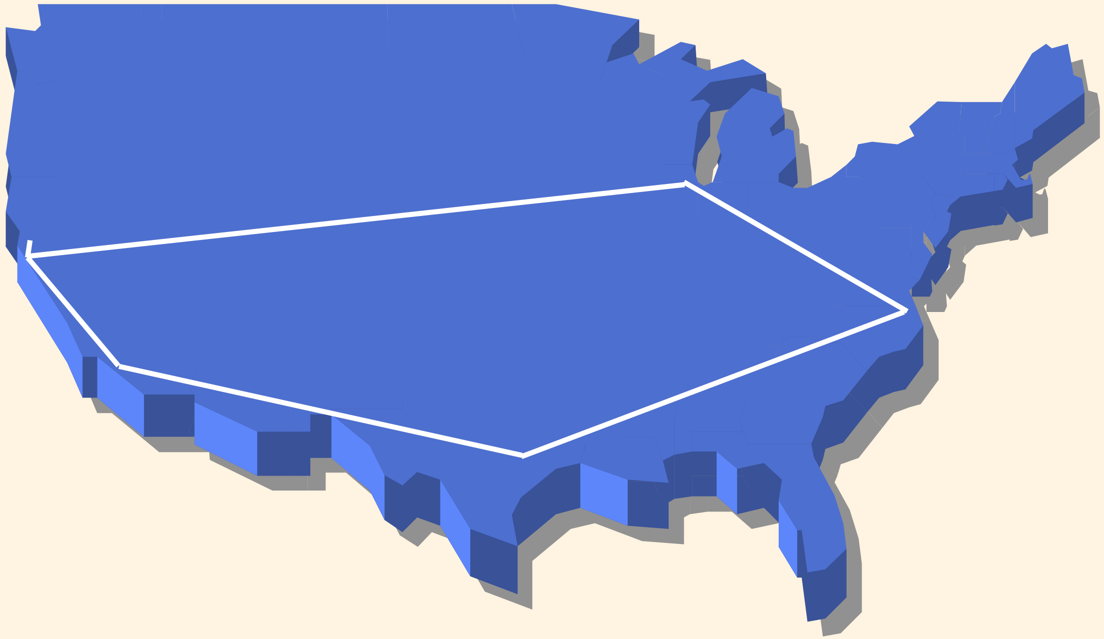
One Human Sequence

We now know that Schrödinger's mysterious human "code-script" consists of 3.3 billion base pairs of DNA.



One Human Sequence

We now know that Schrödinger's mysterious human "code-script" consists of 3.3 billion base pairs of DNA.



Typed in 10-pitch font, one human sequence would stretch for more than 5,000 miles. Digitally formatted, it could be stored on one CD-ROM. Biologically encoded, it fits easily within a single cell.

Bio-digital Information

DNA is a highly efficient digital storage device:

- There is more mass-storage capacity in the DNA of a side of beef than in all the hard drives of all the world's computers.

Bio-digital Information

DNA is a highly efficient digital storage device:

- There is more mass-storage capacity in the DNA of a side of beef than in all the hard drives of all the world's computers.
- Storing all of the (redundant) information in all of the world's DNA on computer hard disks would require that the entire surface of the Earth be covered to a depth of three miles in Conner 1.0 gB drives.

Genomics:
An Example

Human Genome Project - Goals

– construction of a high-resolution genetic map of the human genome;

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

- construction of a high-resolution genetic map of the human genome;
- production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

- construction of a high-resolution genetic map of the human genome;
- production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;
- determination of the complete sequence of human DNA and of the DNA of selected model organisms;

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

- construction of a high-resolution genetic map of the human genome;
- production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;
- determination of the complete sequence of human DNA and of the DNA of selected model organisms;
- development of capabilities for collecting, storing, distributing, and analyzing the data produced;

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

– construction of a high-resolution genetic map of the human genome;

– production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;

– determination of the complete sequence of human DNA and of the DNA of selected model organisms;

– development of capabilities for collecting, storing, distributing, and analyzing the data produced;

– creation of appropriate technologies necessary to achieve these objectives.

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Infrastructure and the HGP

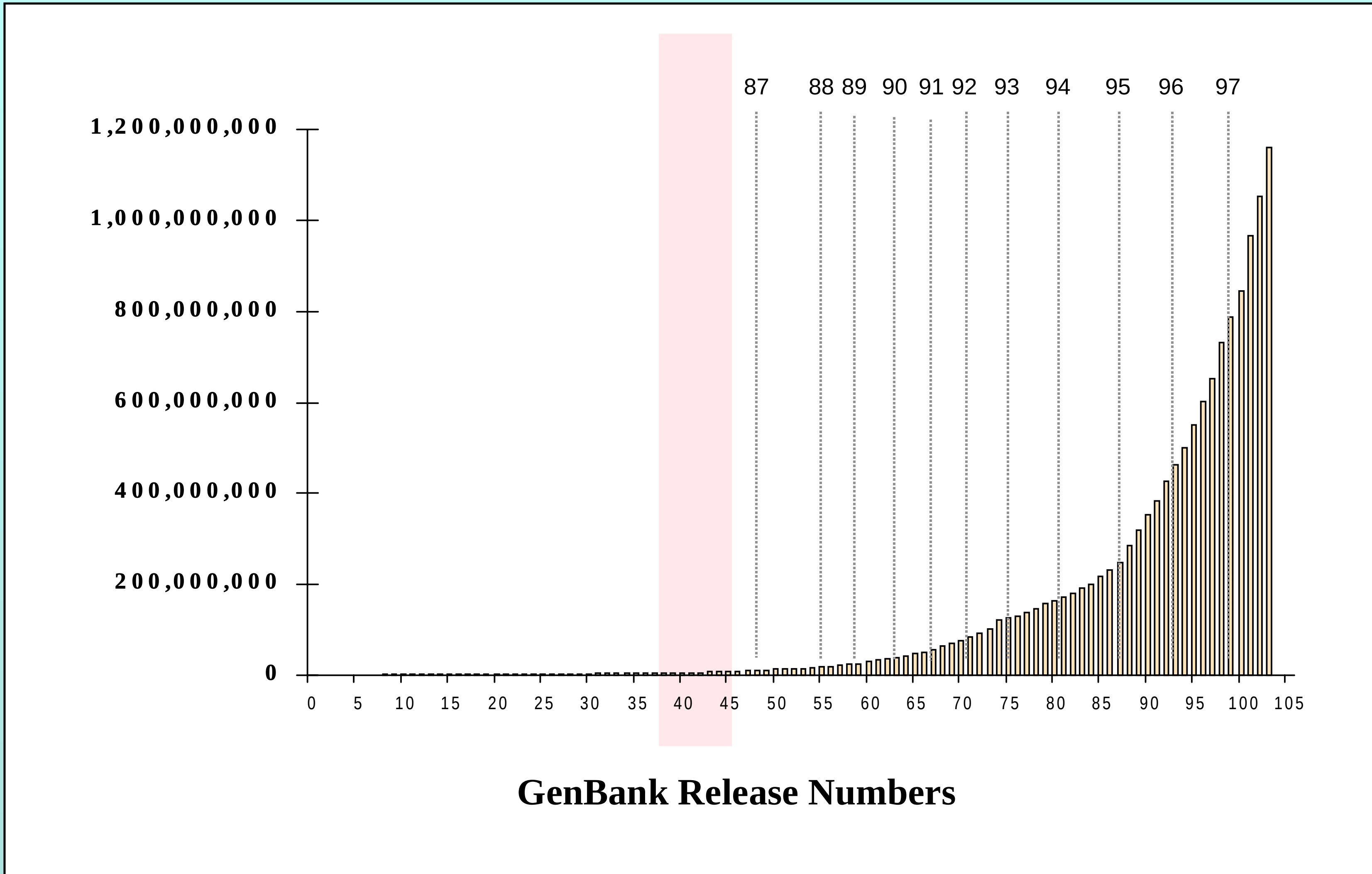
Progress towards all of the [Genome Project] goals will require the establishment of well-funded centralized facilities, including a stock center for the cloned DNA fragments generated in the mapping and sequencing effort and a data center for the computer-based collection and distribution of large amounts of DNA sequence information.

National Research Council. 1988. *Mapping and Sequencing the Human Genome*. Washington, DC: National Academy Press. p. 3

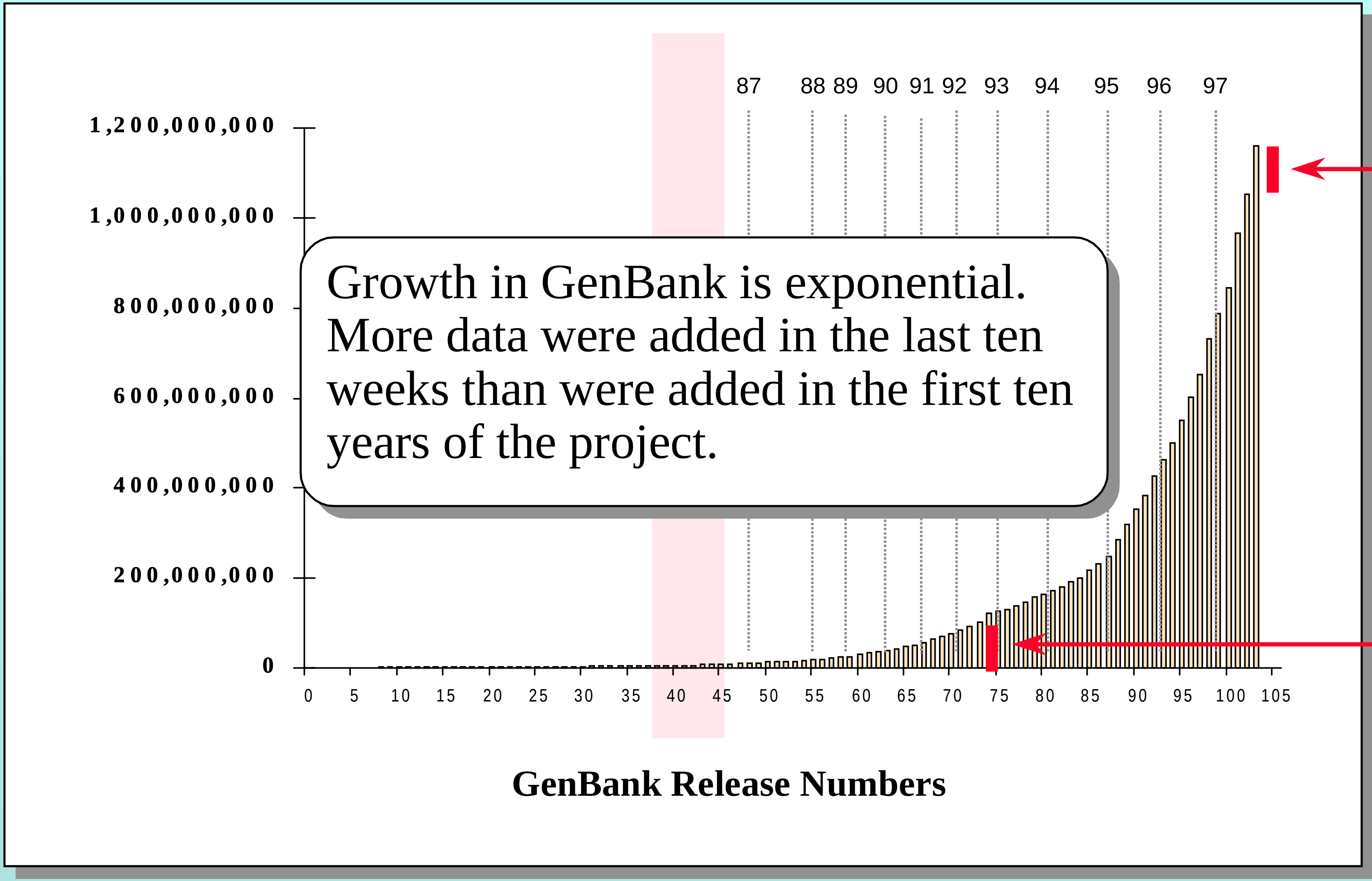
GenBank Totals *(Release 103)*

DIVISION	Entries	Per Cent	Base Pairs	Per Cent
Phage Sequences (PHG)	1,313	0.074%	2,138,810	0.184%
Viral Sequences (VRL)	45,355	2.568%	44,484,848	3.834%
Bacteria (BCT)	38,023	2.153%	88,576,641	7.634%
Plant, Fungal, and Algal Sequences (PLN)	44,553	2.523%	92,259,434	7.951%
Invertebrate Sequences (INV)	29,657	1.679%	105,703,550	9.110%
Rodent Sequences (ROD)	36,967	2.093%	45,437,309	3.916%
Primate Sequences (PRI1–2)	75,587	4.280%	134,944,314	11.630%
Other Mammals (MAM)	12,744	0.722%	12,358,310	1.065%
Other Vertebrate Sequences (VRT)	17,713	1.003%	17,040,159	1.469%
High-Throughput Genome Sequences (HTG)	1,120	0.063%	72,064,395	6.211%
Genome Survey Sequences (GSS)	42,628	2.414%	22,783,326	1.964%
Structural RNA Sequences (RNA)	4,802	0.272%	2,487,397	0.214%
Sequence Tagged Sites Sequences (STS)	52,824	2.991%	18,161,532	1.565%
Patent Sequences (PAT)	87,767	4.970%	27,593,724	2.378%
Synthetic Sequences (SYN)	2,577	0.146%	5,698,945	0.491%
Unannotated Sequences (UNA)	2,480	0.140%	1,933,676	0.167%
EST1-17	**1,269,737**	**71.905%**	**466,634,317**	**40.217%**
TOTALS	**1,765,847**	**100.000%**	**1,160,300,687**	**100.000%**

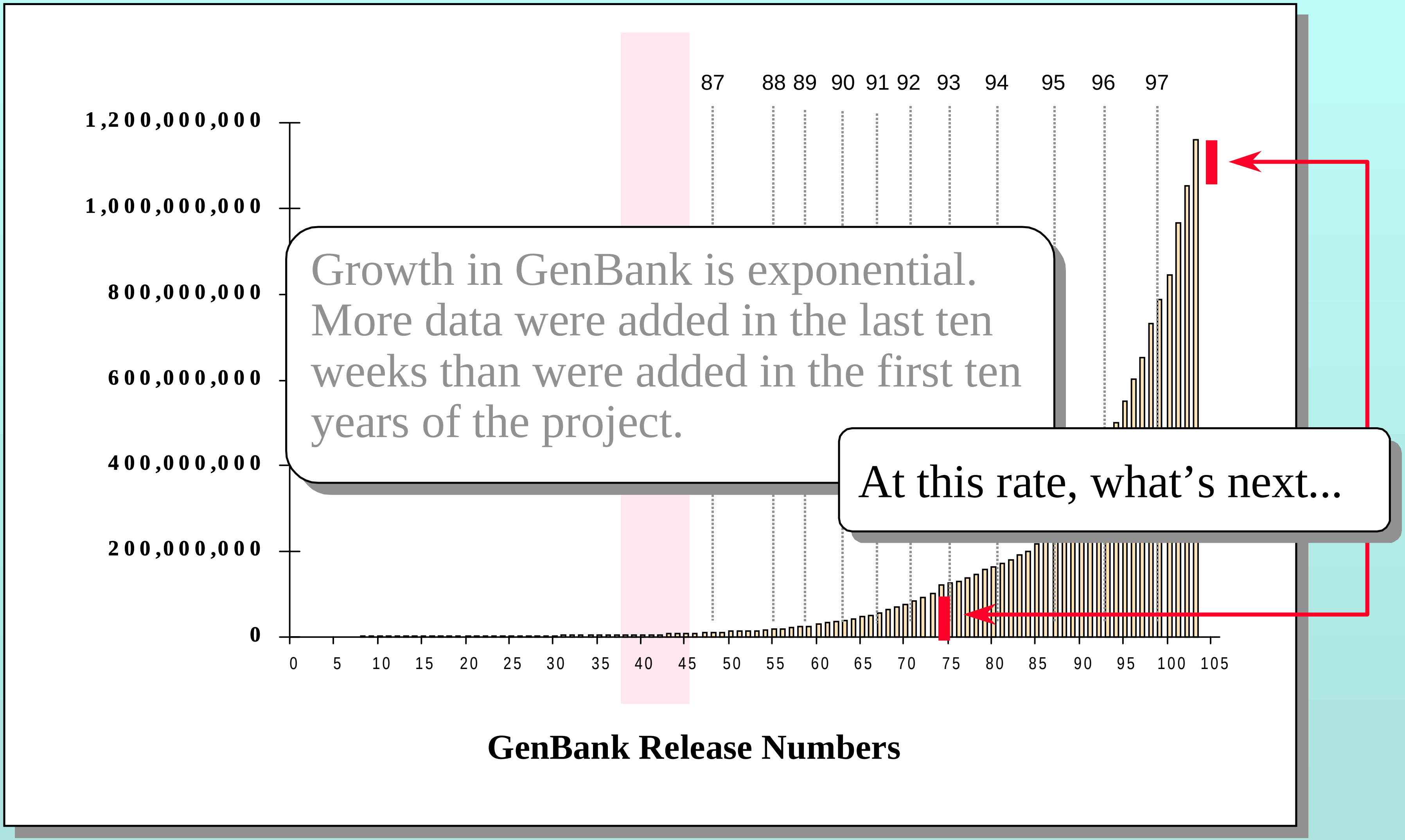
Base Pairs in GenBank



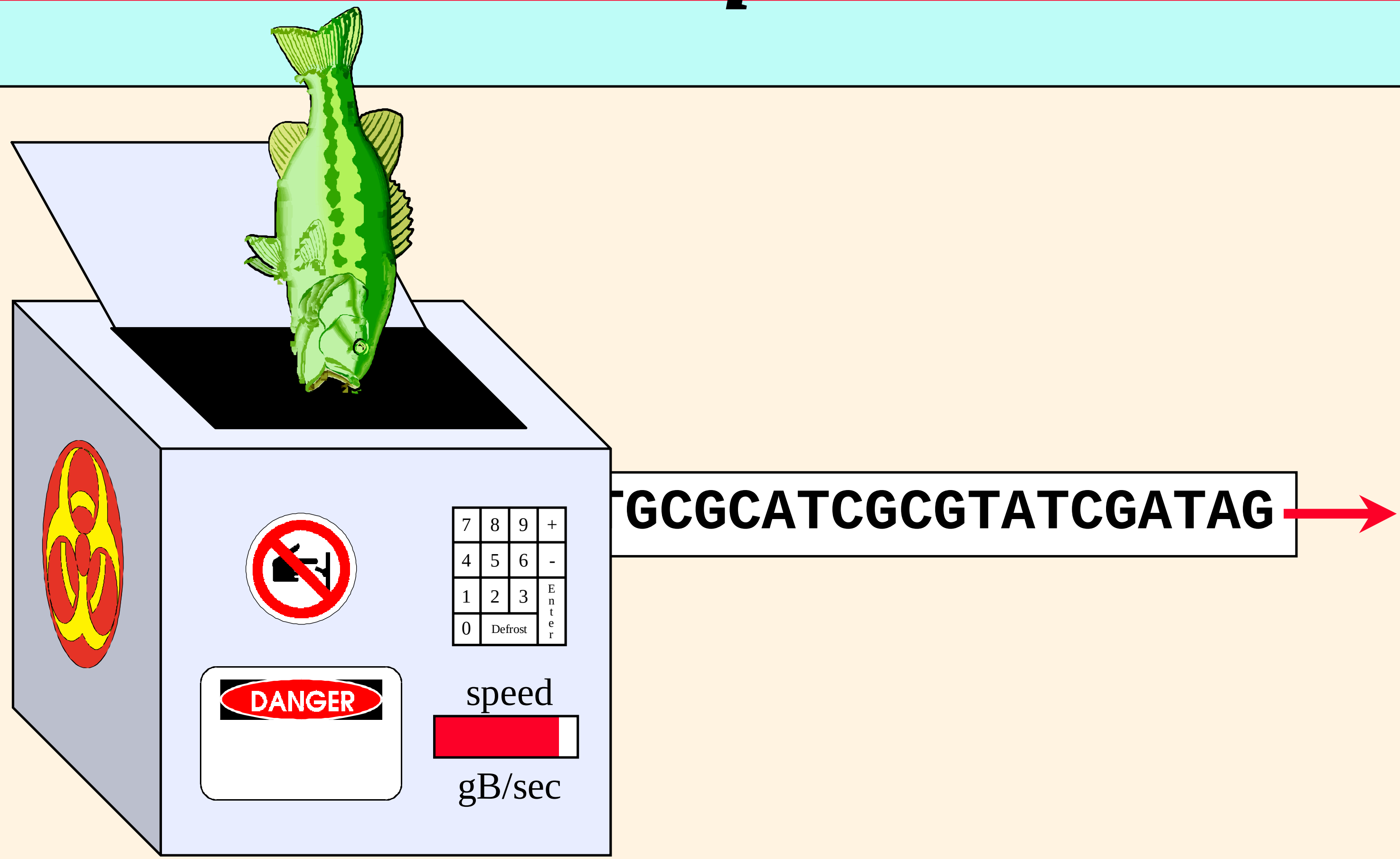
Base Pairs in GenBank



Base Pairs in GenBank



ABI *Bass-o-Matic Sequencer*



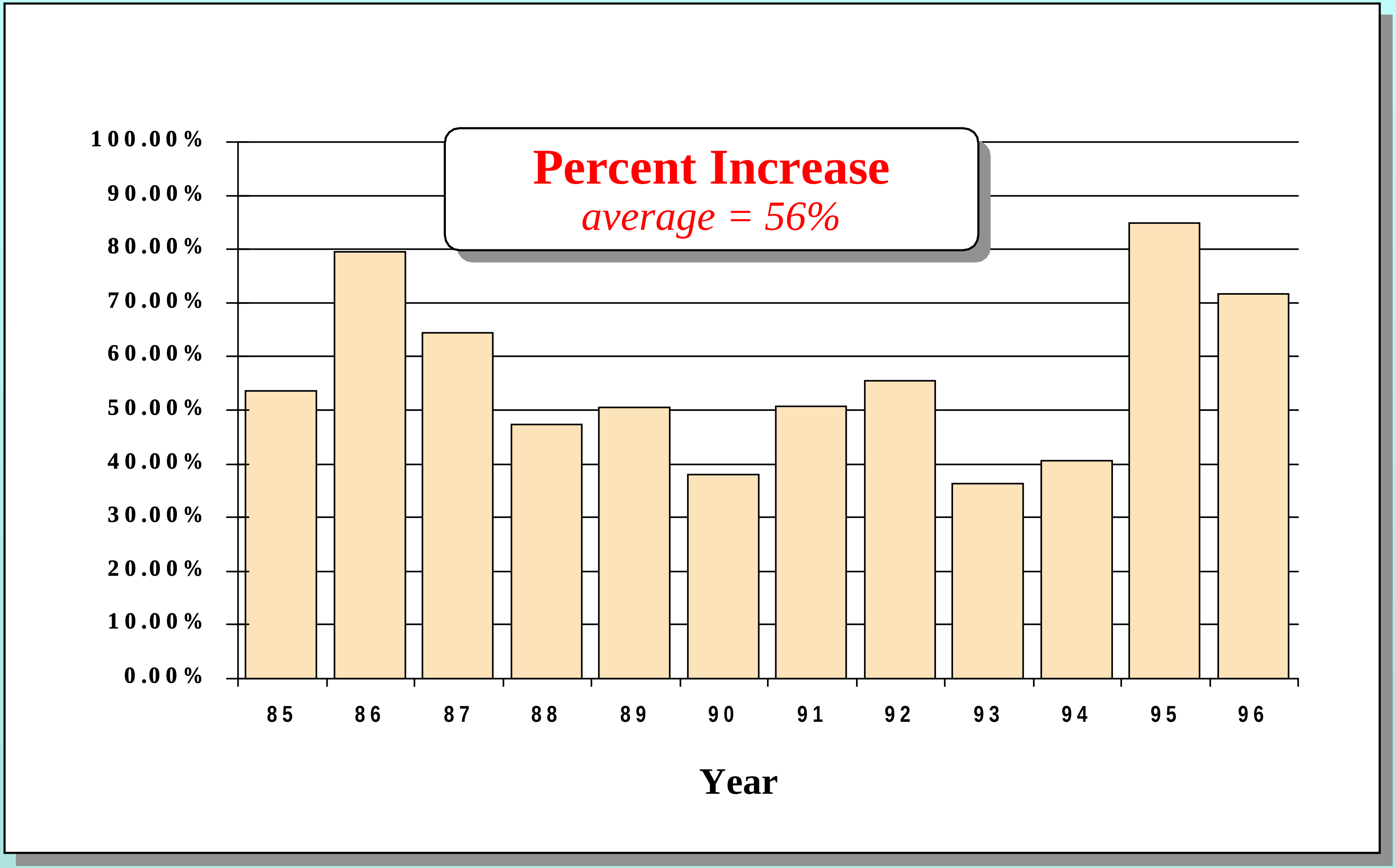
In with the sample, out with the sequence...

What’s Really Next

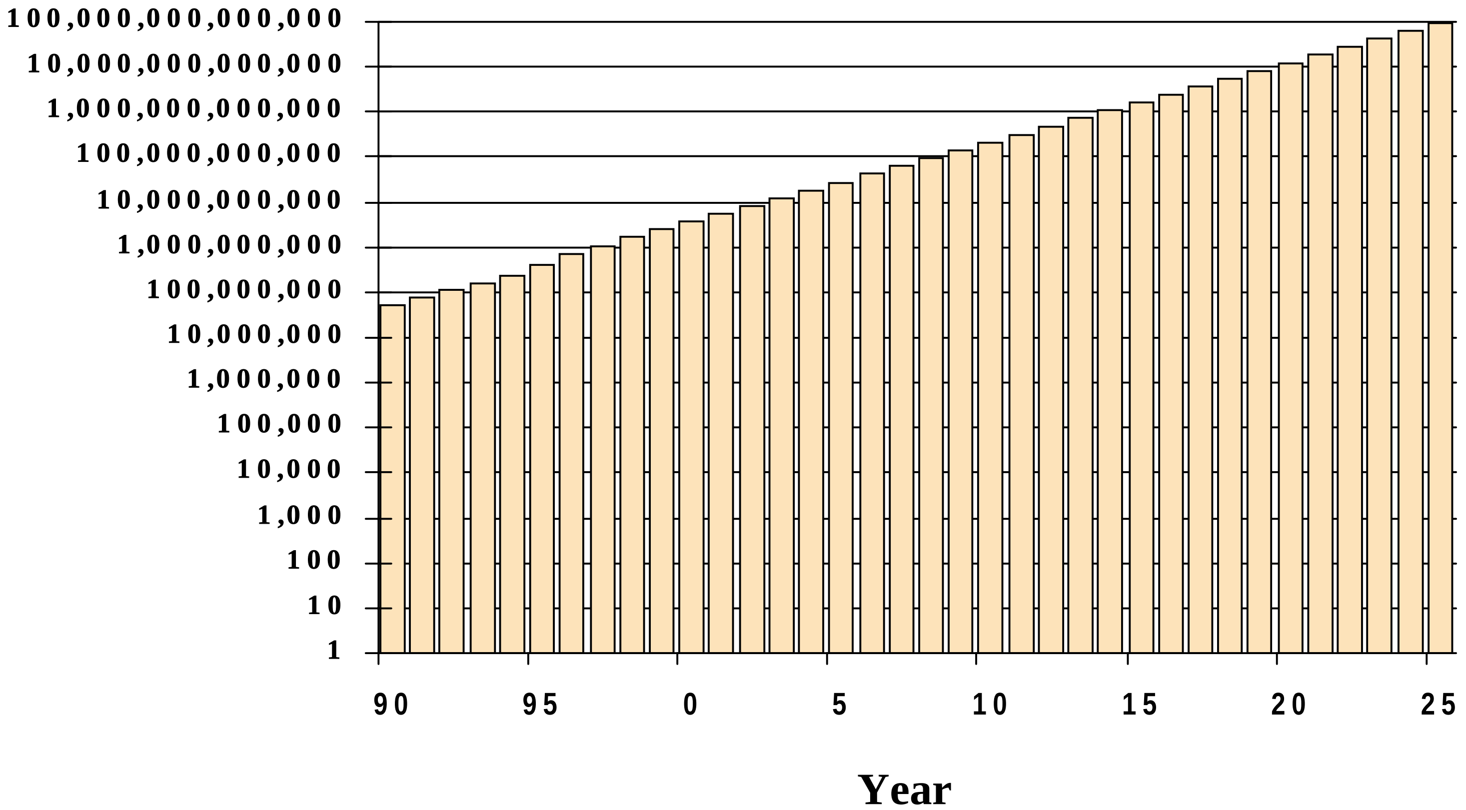
The post-genome era in biological research will take for granted ready access to huge amounts of genomic data.

The challenge will be ***understanding*** those data and using the understanding to solve real-world problems...

Base Pairs in GenBank *(Percent Increase)*



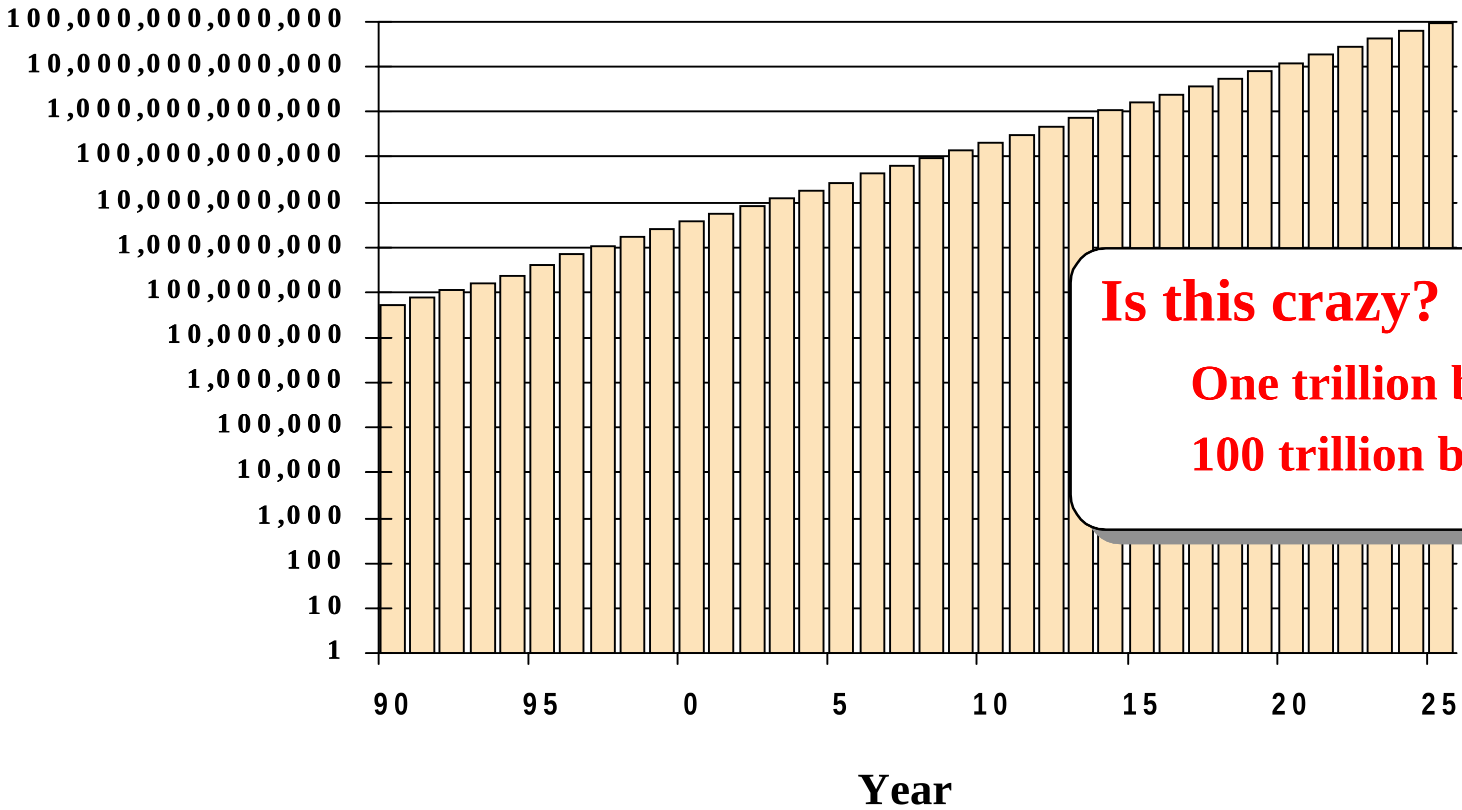
Projected Base Pairs



Assumed annual growth rate: 50%

(less than current rate)

Projected Base Pairs



Is this crazy?

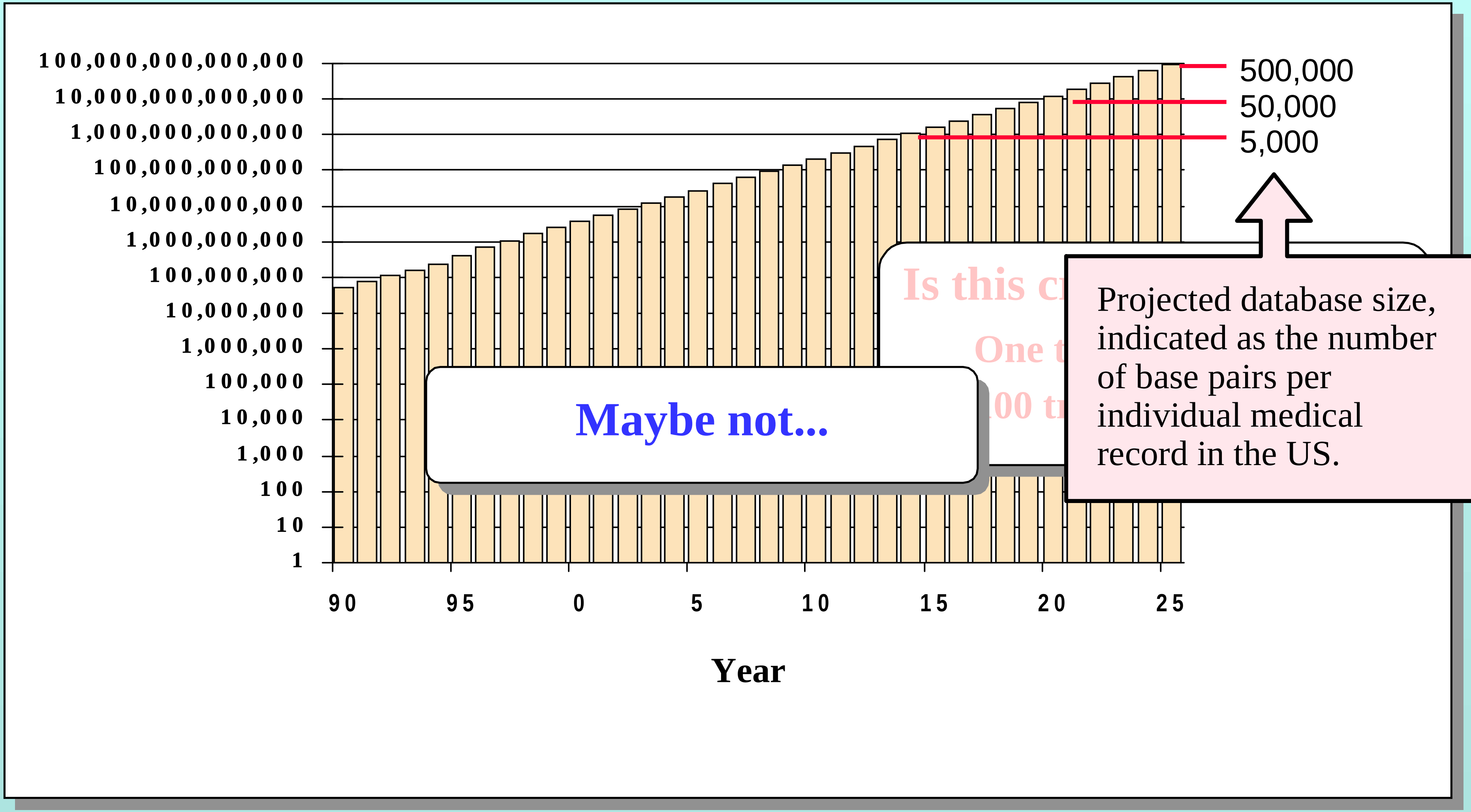
One trillion bp by 2015

100 trillion by 2025

Assumed annual growth rate: 50%

(less than current rate)

Projected Base Pairs



21st Century Biology

Post-Genome Era

The Post-Genome Era

Post-genome research involves:

- applying genomic tools and knowledge to more general problems
- asking new questions, tractable only to genomic or post-genomic analysis
- moving beyond the structural genomics of the human genome project and into the functional genomics of the post-genome era

The Post-Genome Era

Suggested definition:

- functional genomics = biology

The Post-Genome Era

An early analysis:

Walter Gilbert. 1991. Towards a paradigm shift in biology. *Nature*, 349:99.

Paradigm Shift in Biology

To use [the] flood of knowledge, which will pour across the computer networks of the world, biologists not only must become computer literate, but also change their approach to the problem of understanding life.

Walter Gilbert. 1991. Towards a paradigm shift in biology. *Nature*, 349:99.

Paradigm Shift in Biology

The new paradigm, now emerging, is that all the 'genes' will be known (in the sense of being resident in databases available electronically), and that the starting point of a biological investigation will be theoretical. An individual scientist will begin with a theoretical conjecture, only then turning to experiment to follow or test that hypothesis.

Walter Gilbert. 1991. Towards a paradigm shift in biology. *Nature*, 349:99.

Paradigm Shift in Biology

Case of Microbiology

< 5,000	known and described bacteria
5,000,000	base pairs per genome
25,000,000,000	TOTAL base pairs

If a full, annotated sequence were available for all known bacteria, the practice of microbiology would match Gilbert's prediction.

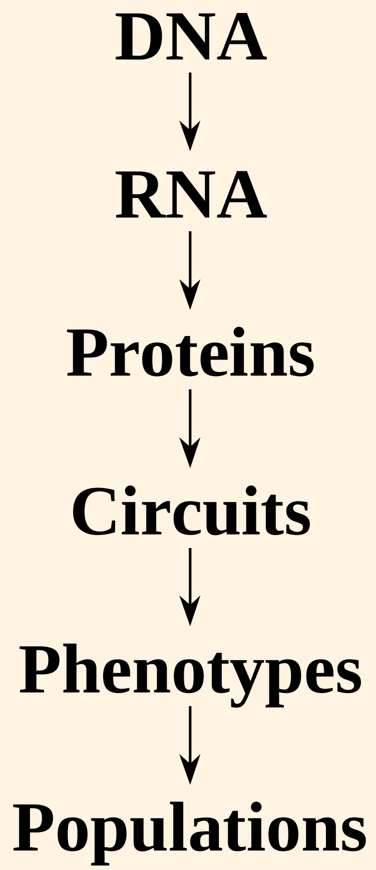
21st Century Biology

The Science

Fundamental Dogma

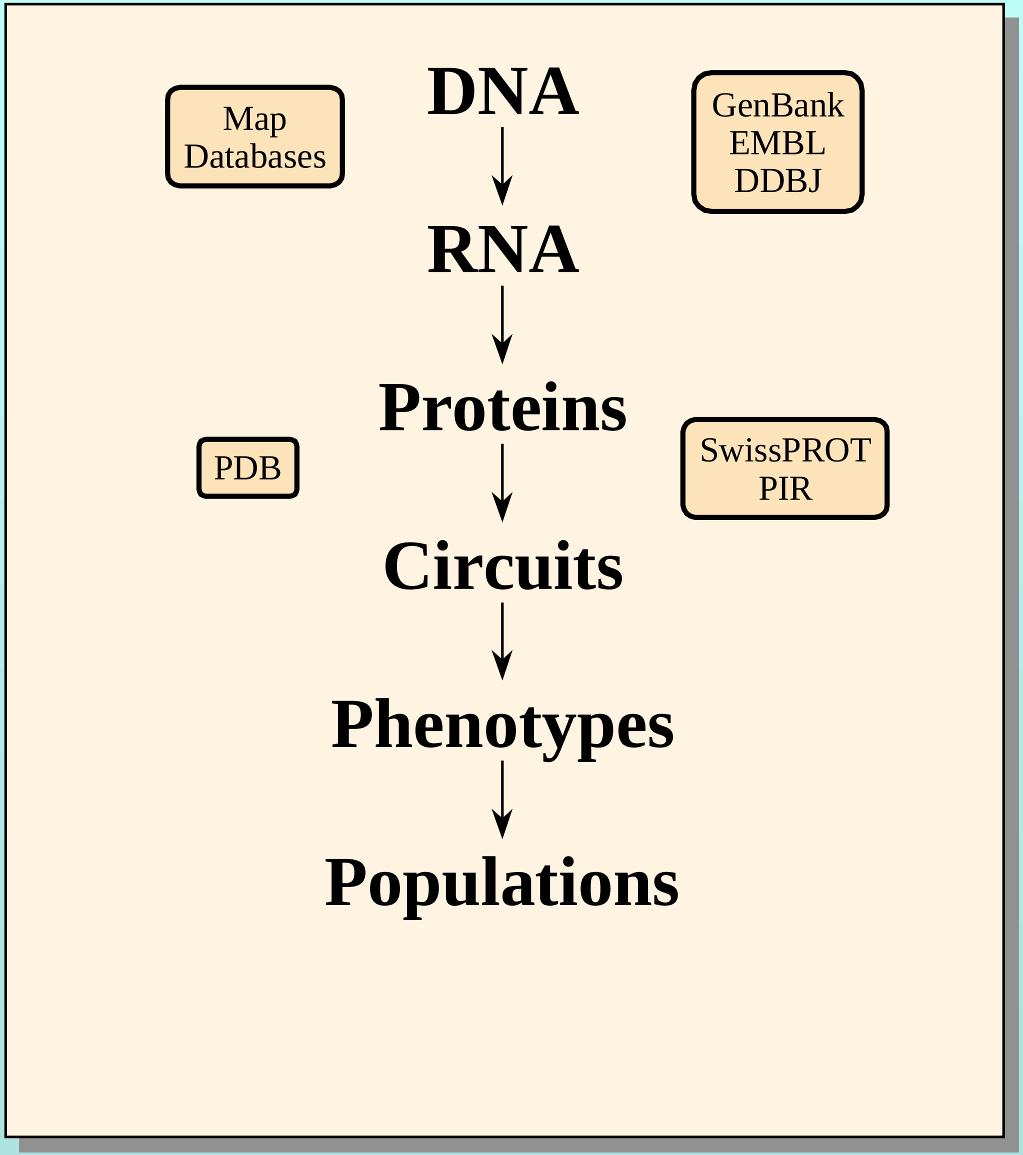
The fundamental dogma of molecular biology is that genes act to create phenotypes through a flow of information from DNA to RNA to proteins, to interactions among proteins (regulatory circuits and metabolic pathways), and ultimately to phenotypes.

Collections of individual phenotypes, of course, constitute a population.



Fundamental Dogma

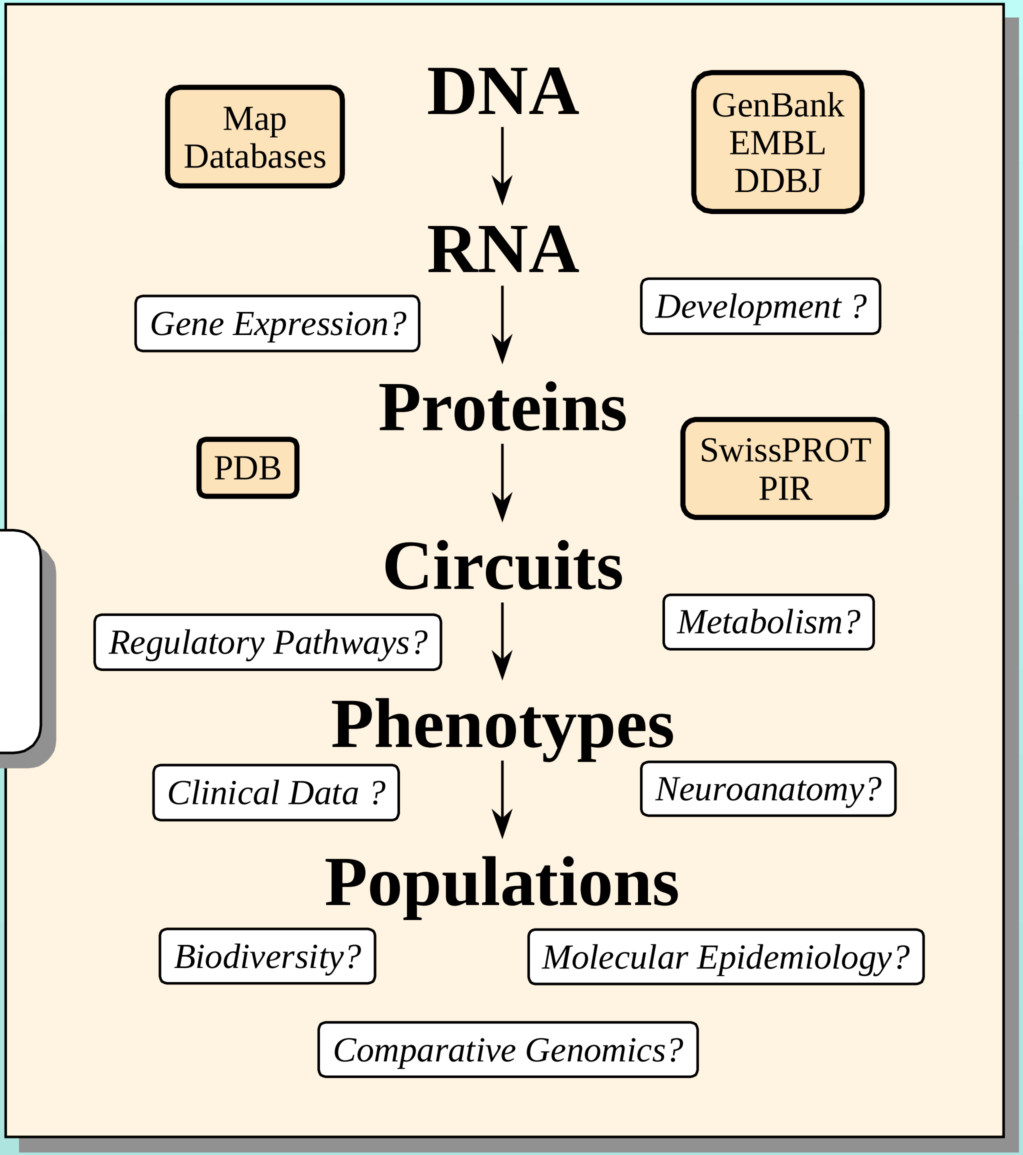
Although a few databases already exist to distribute molecular information,



Fundamental Dogma

Although a few databases already exist to distribute molecular information,

the post-genomic era will need many more to collect, manage, and publish the coming flood of new findings.



21st Century Biology

The People

Human Resources Issues

- Reduction in need for non-IT staff

Human Resources Issues

- Reduction in need for non-IT staff
- Increase in need for IT staff, especially "information engineers"

Human Resources Issues

- Reduction in need for non-IT staff
- Increase in need for IT staff, especially "information engineers"

In modern biology, a general trend is to convert expert work into staff work and finally into computation. New expertise is required to design, carry out, and interpret continuing work.

Human Resources Issues

Elbert Branscomb: “You must recognize that some day you may need as many computer scientists as biologists in your labs.”

Human Resources Issues

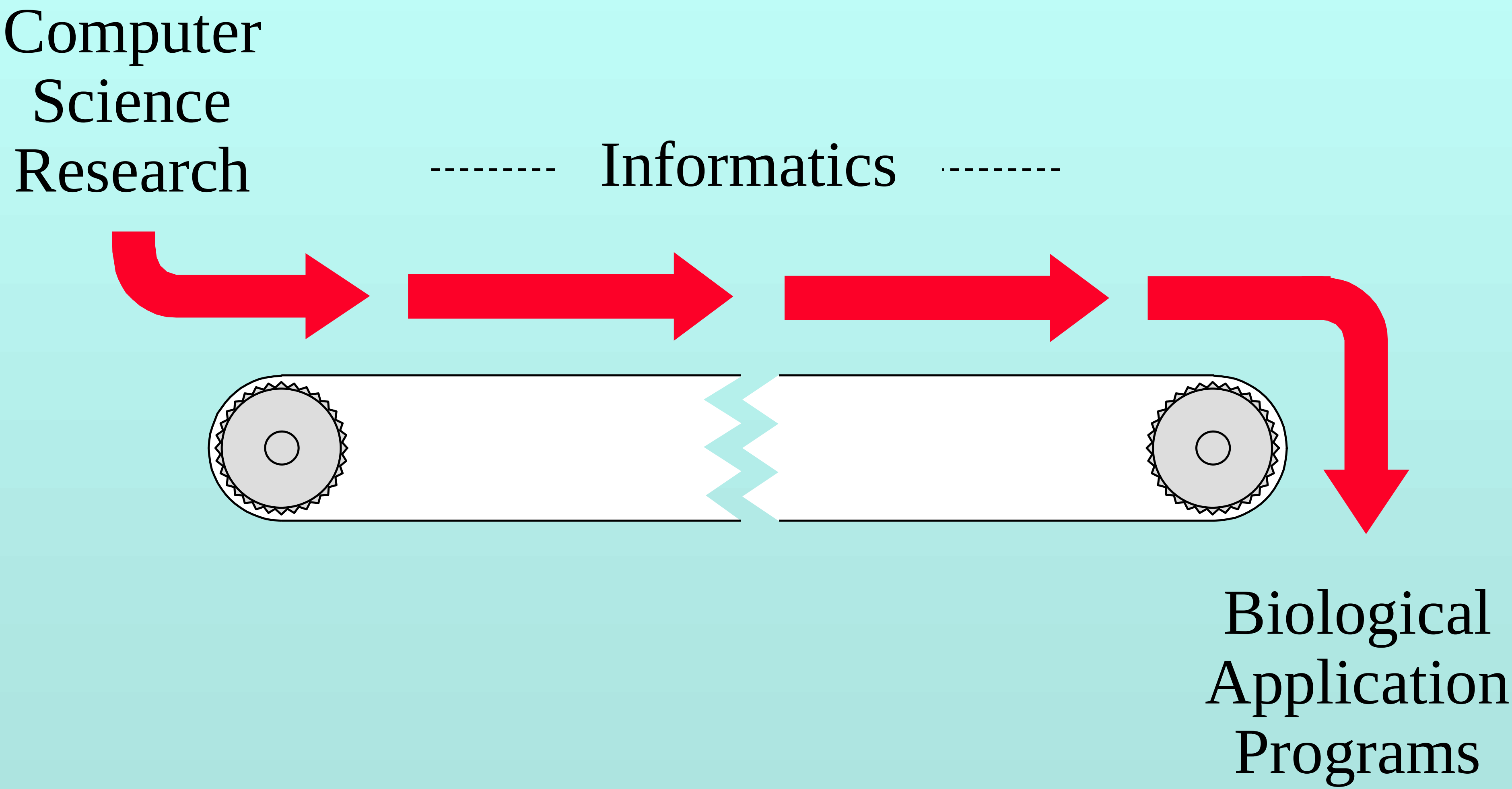
Elbert Branscomb: "You must recognize that some day you may need as many computer scientists as biologists in your labs."

Craig Venter: "At TIGR, we already have twice as many computer scientists on our staff."

Exchange at DOE workshop on high-throughput sequencing.

New Discipline of Informatics

What is Informatics?



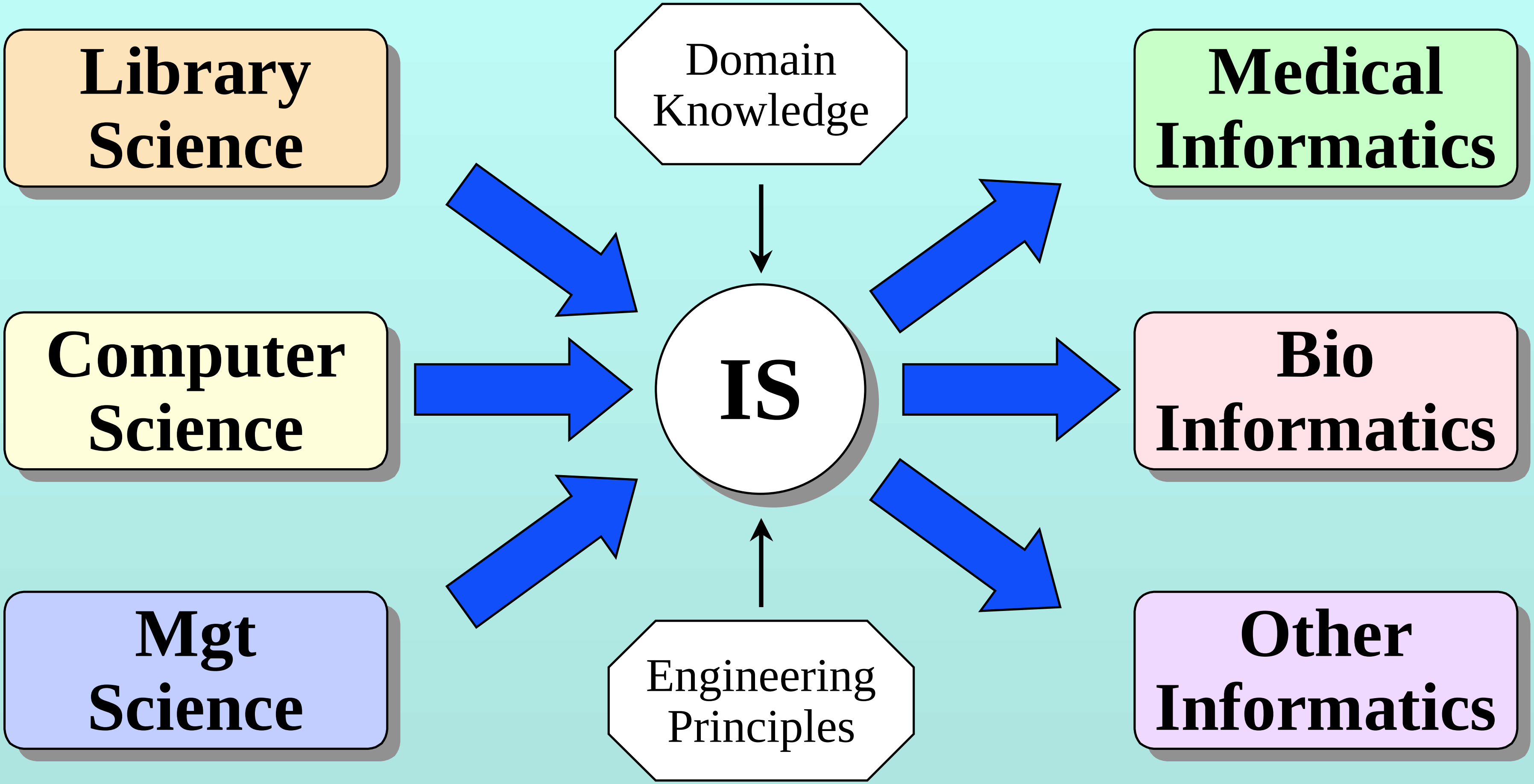
What is Informatics?

Informatics combines expertise from:

- *domain science (e.g., biology)*
- *computer science*
- *library science*
- *management science*

All tempered with an engineering mindset...

What is Informatics?



Engineering Mindset

Engineering is often defined as the use of scientific knowledge and principles for practical purposes. While the original usage restricted the word to the building of roads, bridges, and objects of military use, today's usage is more general and includes chemical, electronic, and even mathematical engineering.

Parnas, David Lorge. 1990. *Computer*, 23(1):17-22.

Engineering Mindset

Engineering is often defined as the use of scientific knowledge and principles for practical purposes. While the original usage restricted the word to the building of roads, bridges, and objects of military use, today's usage is more general and includes chemical, electronic, and even mathematical engineering.

Parnas, David Lorge. 1990. *Computer*, 23(1):17-22.

... or even information engineering.

Engineering Mindset

> Engineering education ... stresses finding good, as contrasted with workable, designs. Where a scientist may be happy with a device that validates his theory, an engineer is taught to make sure that the device is efficient, reliable, safe, easy to use, and robust.

Parnas, David Lorge. 1990. *Computer*, 23(1):17-22.

Engineering Mindset

> Engineering education ... stresses finding good, as contrasted with workable, designs. Where a scientist may be happy with a device that validates his theory, an engineer is taught to make sure that the device is efficient, reliable, safe, easy to use, and robust.
>
> Parnas, David Lorge. 1990. *Computer*, 23(1):17-22.

The assembly of working, robust systems, on time and on budget, is the key requirement for a federated information infrastructure for biology.

Funding for Bio-Information Infrastructure

Call for Change

Among the many new tools that are or will be needed (for 21st-century biology), some of those having the highest priority are:

- bioinformatics
- computational biology
- functional imaging tools using biosensors and biomarkers
- transformation and transient expression technologies
- nanotechnologies

Impact of Emerging Technologies on the Biological Sciences: Report of a Workshop. NSF-supported workshop, held 26-27 June 1995, Washington, DC.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.
- Current levels of support for public bio-information infrastructure are too low.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.
- Current levels of support for public bio-information infrastructure are too low.
- Reallocation of federal funding is difficult, and subject to political pressures.

The Problem

- IT moves at "Internet Speed" and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.
- Current levels of support for public bio-information infrastructure are too low.
- Reallocation of federal funding is difficult, and subject to political pressures.
- Federal-funding decision processes are ponderously slow and inefficient.

Federal Funding of Bio-Databases

The challenges:

Federal Funding of Bio-Databases

The challenges:

- providing adequate funding levels

Federal Funding of Bio-Databases

The challenges:

- providing adequate funding levels
- making timely, efficient decisions

IT Budgets

A Reality Check

Rhetorical Question

Which is likely to be more complex:

- identifying, documenting, and tracking the whereabouts of **all parcels** in transit in the US at one time

Rhetorical Question

Which is likely to be more complex:

- identifying, documenting, and tracking the whereabouts of **all parcels** in transit in the US at one time
- identifying, documenting, and analyzing the structure and function of **all individual genes in all economically significant organisms**; then analyzing **all significant gene-gene and gene-environment interactions** in those organisms and their environments

Business Factoids

United Parcel Service:

- uses two redundant 3 Terabyte (yes, 3000 GB) databases to track all packages in transit.
- has 4,000 full-time employees dedicated to IT
- spends one billion dollars per year on IT
- has an income of 1.1 billion dollars, against revenues of 22.4 billion dollars

Business Comparisons

Company	Revenues	IT Budget	Pct
Chase-Manhattan	16,431,000,000	1,800,000,000	10.95 %
AMR Corporation	17,753,000,000	1,368,000,000	7.71 %
Nation’s Bank	17,509,000,000	1,130,000,000	6.45 %
Sprint	14,235,000,000	873,000,000	6.13 %
IBM	75,947,000,000	4,400,000,000	5.79 %
MCI	18,500,000,000	1,000,000,000	5.41 %
Microsoft	11,360,000,000	510,000,000	4.49 %
United Parcel	22,400,000,000	1,000,000,000	4.46 %
Bristol-Myers Squibb	15,065,000,000	440,000,000	2.92 %
Pfizer	11,306,000,000	300,000,000	2.65 %
Pacific Gas & Electric	10,000,000,000	250,000,000	2.50 %
Wal-Mart	104,859,000,000	550,000,000	0.52 %
K-Mart	31,437,000,000	130,000,000	0.41 %

Federal Funding of Biomedical-IT

Appropriate funding level:

- approx. 5-10% of research funding
- *i.e.*, 1 - 2 **billion** dollars per year

Federal Funding of Biomedical-IT

Appropriate funding level:

- approx. 5-10% of research funding
- *i.e.*, 1 - 2 **billion** dollars per year

Source of estimate:

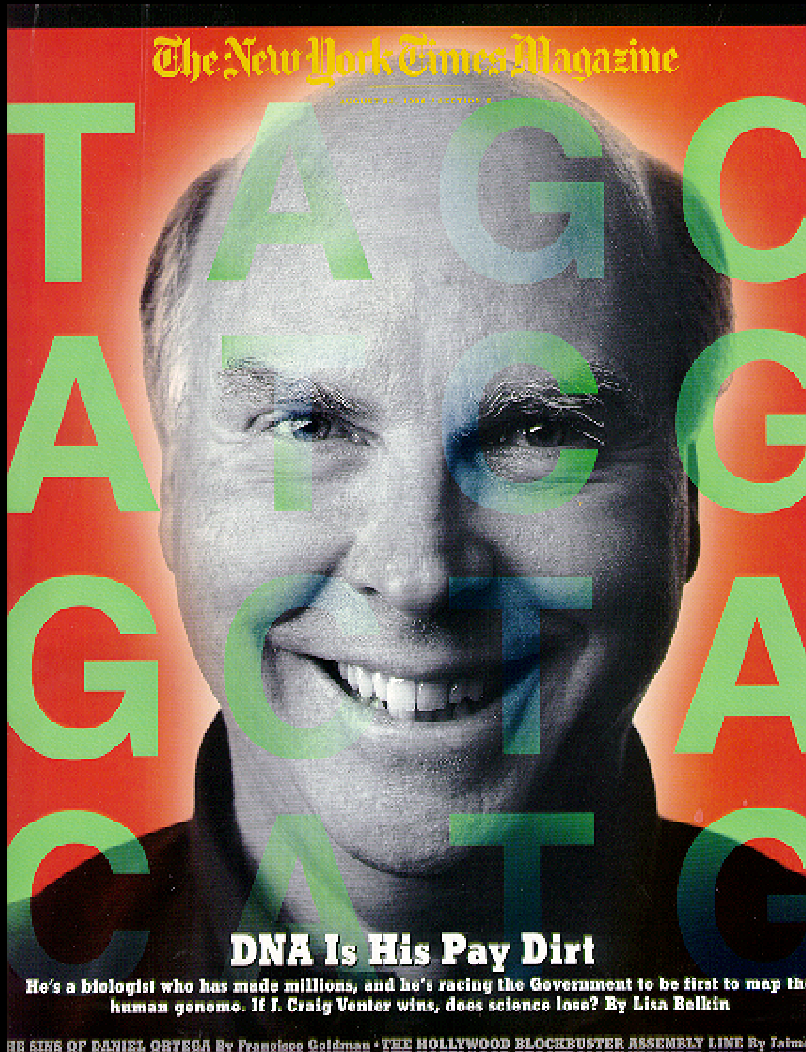
- Experience of IT-transformed industries.

- Current support for IT-rich biological research.

Private Sector

Information Technology
&
Genomic Research

Who is this man?



Who is this man?

And why should you care?

Craig Venter

Accomplishments - cDNA:

Early in the genome project, Venter argued that cDNA data held much of the economically significant content of genomic information.

Craig Venter

Accomplishments - cDNA:

Early in the genome project, Venter argued that cDNA data held much of the economically significant content of genomic information.

Although most genome researchers scoffed, Venter persevered and ultimately the importance of cDNA data was recognized, both by the scientific community (more than 70% of the sequences in GenBank are now cDNA sequences) and by the financial community (Venter received millions in private sector funding to create TIGR - The Institute for Genomic Research).

Craig Venter

Accomplishments - prokaryotic sequencing:

At TIGR, Venter recognized that sophisticated application of computer technology should allow whole-genome shotgun sequencing, thereby eliminating the time-consuming step of creating a physical map prior to sequencing.

Craig Venter

Accomplishments - prokaryotic sequencing:

At TIGR, Venter recognized that sophisticated application of computer technology should allow whole-genome shotgun sequencing, thereby eliminating the time-consuming step of creating a physical map prior to sequencing.

Although an NIH review panel rejected his idea as impossible, Venter persevered and soon produced the first whole-organism genomic sequence -- that of *Haemophilus influenzae*. Now TIGR has produced more whole-genome microbial sequences than any other organization.

Craig Venter

Audacious Goal- eukaryotic sequencing:

In collaboration with Perkin-Elmer, Venter has now formed a new organization, Celera, with the goal of becoming the world's leading source of whole-genome sequence data from eukaryotes.

Craig Venter

Audacious Goal- eukaryotic sequencing:

In collaboration with Perkin-Elmer, Venter has now formed a new organization, Celera, with the goal of becoming the world's leading source of whole-genome sequence data from eukaryotes.

Although the genome research community is once again discounting Venter's likelihood of success, his track record is two for two when it comes to applying a clear vision for the application of advanced technology to genomic research.

Craig Venter

Audacious Goal- eukaryotic sequencing:

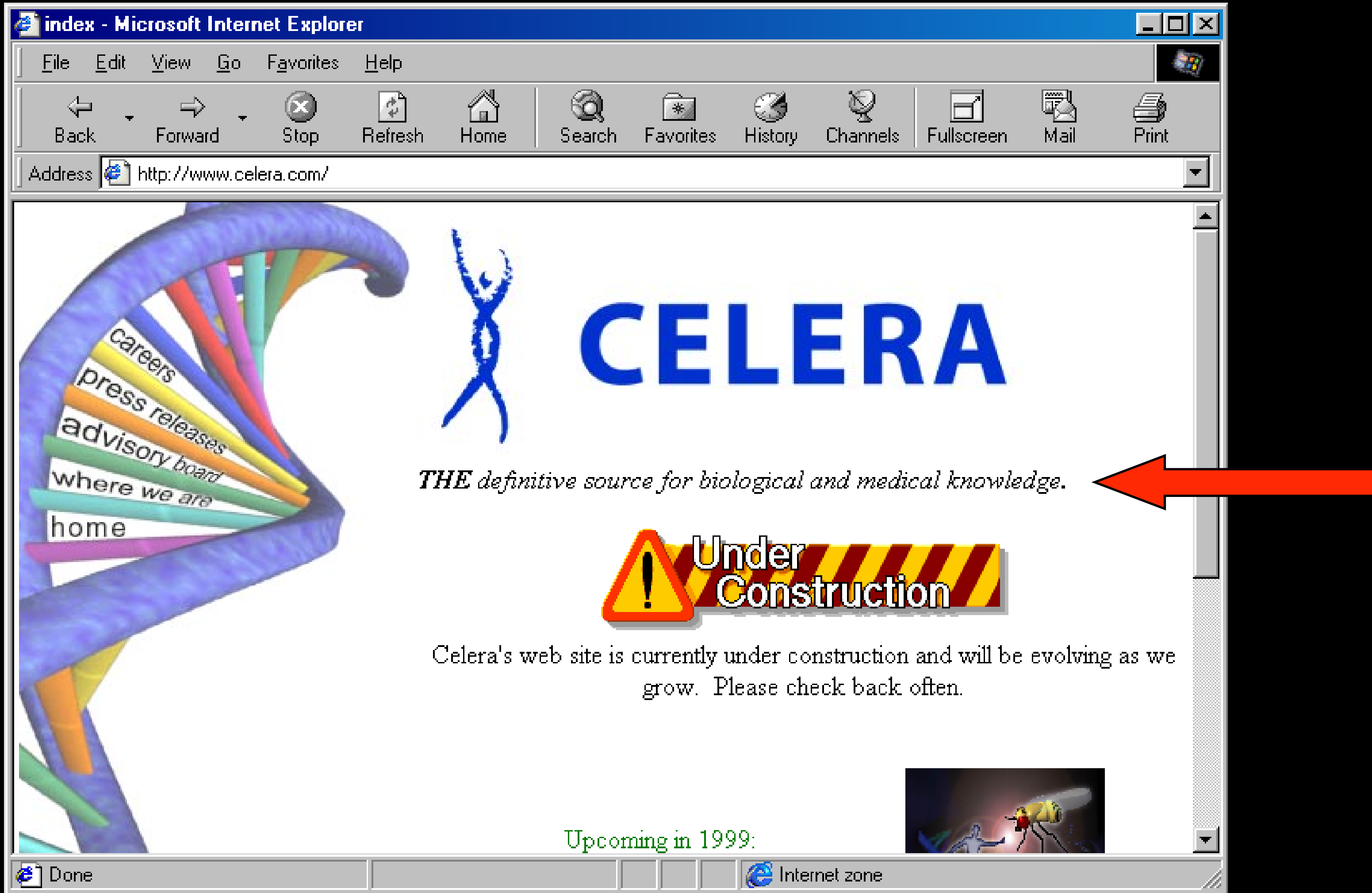
Of course, Venter may fail this time. But, if he doesn't, what is the likely implication of the success of Celera for, say, biomedical research?

Craig Venter

Audacious Goal- eukaryotic sequencing:

Of course, Venter may fail this time. But, if he doesn't, what is the likely implication of the success of Celera for, say, biomedical research?

Let's take a look at Celera's official, publicly announced goals...


index - Microsoft Internet Explorer
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorites History Channels Fullscreen Mail Print
Address http://www.celera.com/
careers
press releases
advisory board
where we are
home
CELERA
THE definitive source for biological and medical knowledge.
Under Construction
Celera's web site is currently under construction and will be evolving as we grow. Please check back often.
Upcoming in 1999:
Done
Internet zone

Celera

We want this new company to be **the** definitive source of genomic and related medical information, which scientists can use to better understand human biology and to deliver improved healthcare options.

Tony White, chief executive officer of Perkin-Elmer,
Teleconference, May 11, 1998

Celera

We want this new company to be **the** definitive source of genomic and related medical information, which scientists can use to better understand human biology and to deliver improved healthcare options. We believe that this company combines compelling technology with unique sequencing strategies, resulting in a genomics sequencing facility with an expected capacity ...

Tony White, chief executive officer of Perkin-Elmer, Teleconference, May 11, 1998

Celera

We want this new company to be **the** definitive source of genomic and related medical information, which scientists can use to better understand human biology and to deliver improved healthcare options. We believe that this company combines compelling technology with unique sequencing strategies, resulting in a genomics sequencing facility with an expected capacity greater than that of the current combined world output.

Tony White, chief executive officer of Perkin-Elmer, Teleconference, May 11, 1998

Celera

We are not a philanthropic organisation, we have a revenue model for this. We are sure people will want to buy the information.

Tony White, chief executive officer of Perkin-Elmer, quoted in *The Guardian*, 13 May 1998

Private Sector

Logistics Mastery
&
Biotechnology

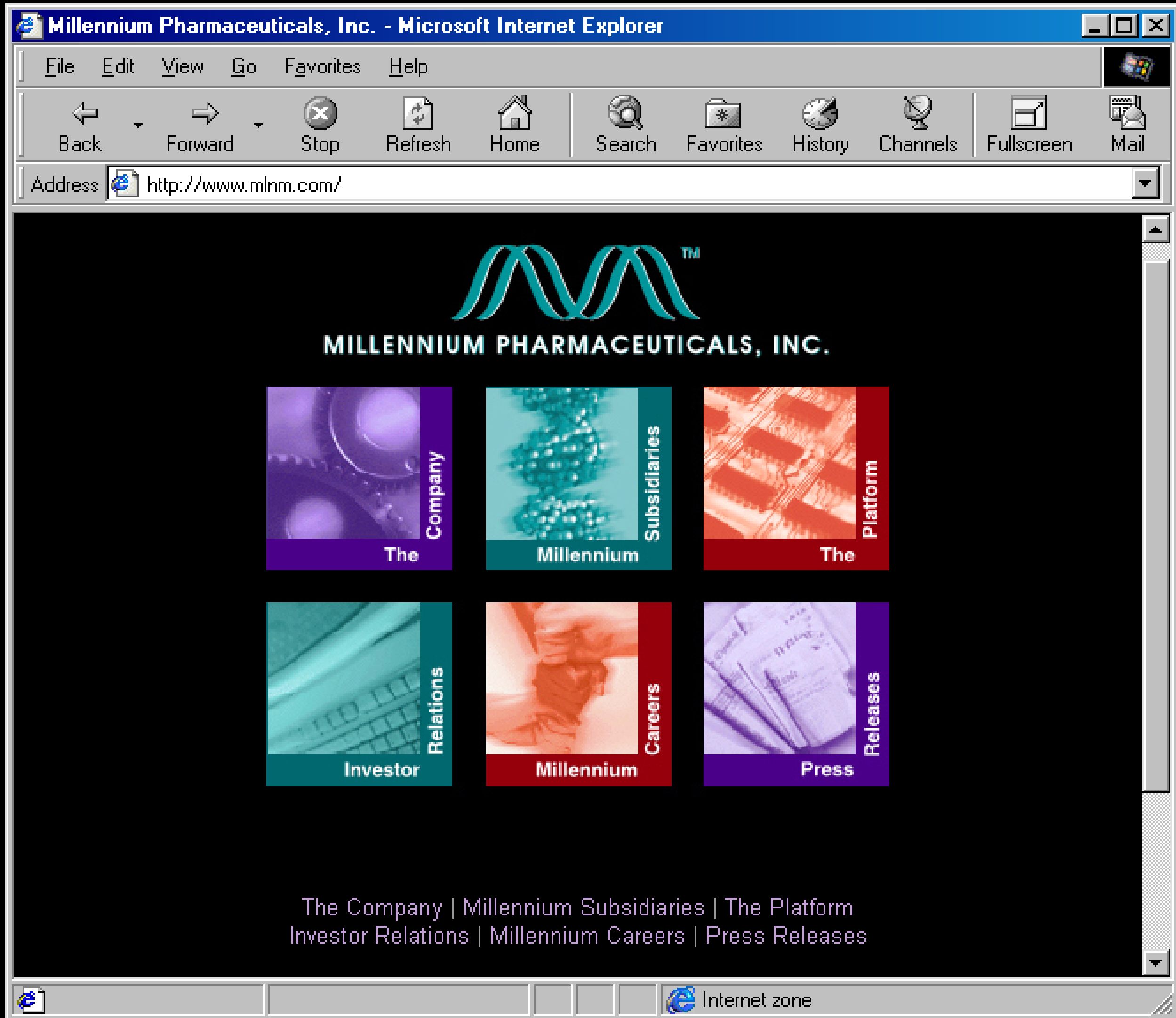
Logistics Mastery

What role does mastery of logistics and process control play in successful biotechnology ventures?

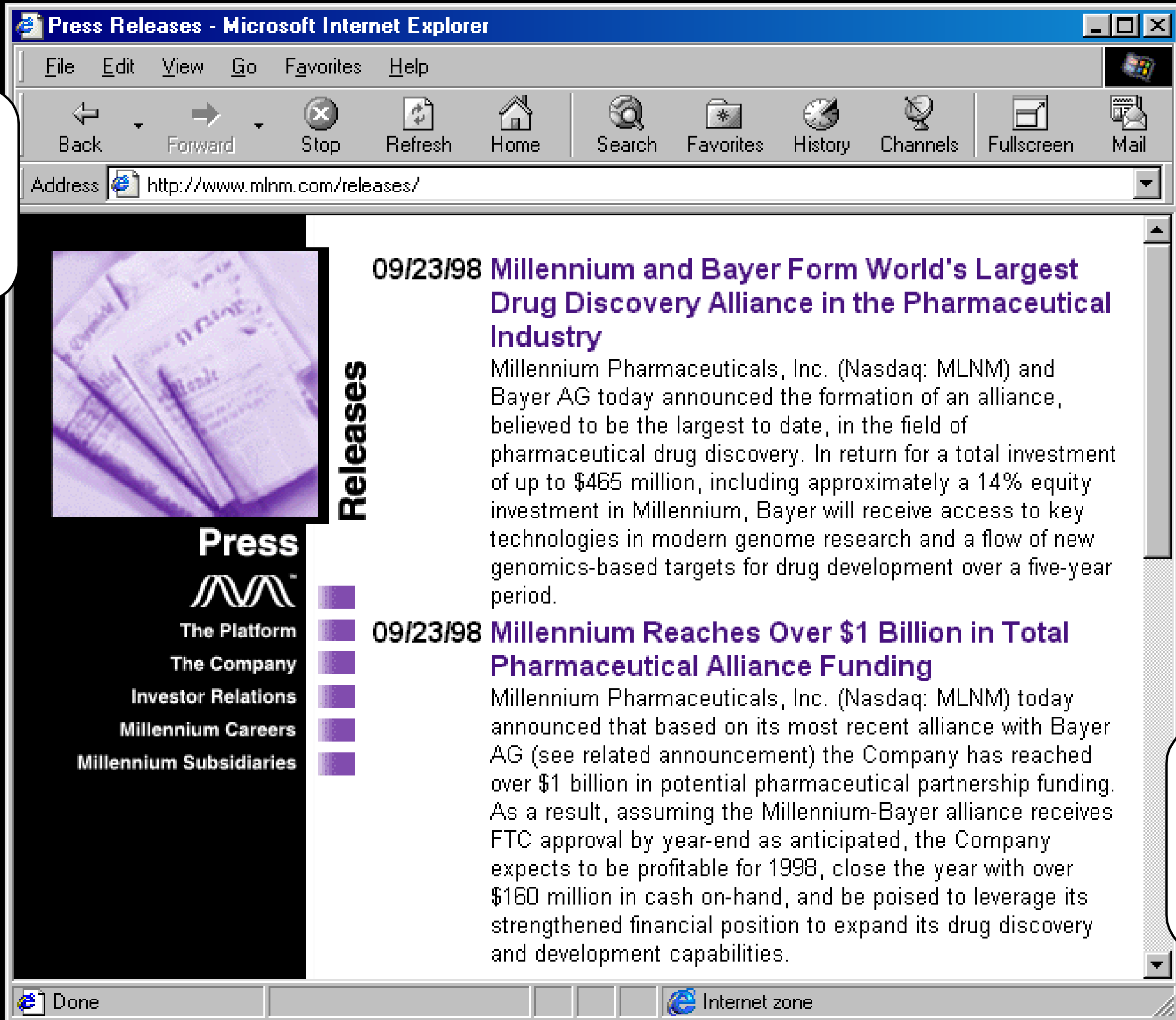
Logistics Mastery

What role does mastery of logistics and process control play in successful biotechnology ventures?

Consider *Millennium Pharmaceuticals* ...


Millennium Pharmaceuticals, Inc. - Microsoft Internet Explorer
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorites History Channels Fullscreen Mail
Address http://www.mlnm.com/
MILLENNIUM PHARMACEUTICALS, INC.
The Company
Millennium Subsidiaries
The Platform
Investor Relations
Millennium Careers
Press Releases
The Company | Millennium Subsidiaries | The Platform
Investor Relations | Millennium Careers | Press Releases
Internet zone

Millennium recently signed one of the most lucrative deals in the biotechnology industry.



Press Releases - Microsoft Internet Explorer

File Edit View Go Favorites Help

Back Forward Stop Refresh Home Search Favorites History Channels Fullscreen Mail

Address http://www.mlnm.com/releases/

Press Releases

The Platform
The Company
Investor Relations
Millennium Careers
Millennium Subsidiaries

09/23/98 Millennium and Bayer Form World's Largest Drug Discovery Alliance in the Pharmaceutical Industry

Millennium Pharmaceuticals, Inc. (Nasdaq: MLNM) and Bayer AG today announced the formation of an alliance, believed to be the largest to date, in the field of pharmaceutical drug discovery. In return for a total investment of up to $465 million, including approximately a 14% equity investment in Millennium, Bayer will receive access to key technologies in modern genome research and a flow of new genomics-based targets for drug development over a five-year period.

09/23/98 Millennium Reaches Over $1 Billion in Total Pharmaceutical Alliance Funding

Millennium Pharmaceuticals, Inc. (Nasdaq: MLNM) today announced that based on its most recent alliance with Bayer AG (see related announcement) the Company has reached over $1 billion in potential pharmaceutical partnership funding. As a result, assuming the Millennium-Bayer alliance receives FTC approval by year-end as anticipated, the Company expects to be profitable for 1998, close the year with over $160 million in cash on-hand, and be poised to leverage its strengthened financial position to expand its drug discovery and development capabilities.

Done Internet zone

As documented in *The Economist*, much of Millennium's success traces to its expertise in logistics and process control...

Millenium Pharmaceuticals

Millenium's bugs

Mark Levin is an engineer. That makes him ideally qualified to be a successful biotechnology entrepreneur

The Economist

SEPTEMBER 26TH - OCTOBER 2ND 1998

Millenium Pharmaceuticals

Many other biotech firms, increasingly desperate for cash as investors have shied away from their shares, have queued up to do deals with the pharmaceutical industry at almost any price. Mr Levin, though, has been able to dictate his own terms. Uniquely among biotechnologists, he seems to have mastered the art of having his cake and eating it.

Except that Mr Levin is not really a biotechnologist at all: he is a chemical engineer. He has worked in process control for companies as varied as Miller Brewing and Genentech, a firm of biotech pioneers. Whereas biologists tend to see biotech as the search for a compound, Mr Levin thinks of it as a complex production process. While they concentrate on the bio, he also thinks hard about the technology.

The Economist

SEPTEMBER 26TH - OCTOBER 2ND 1998

Millenium Pharmaceuticals

Mr Levin focuses on trying to make each link in the discovery chain as efficient as possible. He has assembled an impressive array of technologies -- including robotics and information systems as well as molecular biology. He then enhances them and links them together in novel ways to create what the engineer in him likes to call "technology platforms". The idea is that these platforms should help drug searchers to travel rapidly on their long and tortuous journey from gene to treatment. Mr Levin's goal is to boost the productivity of drug discovery by 50%, which would lead to many more new drugs coming on to the market each year.

The Economist

SEPTEMBER 26TH - OCTOBER 2ND 1998

21st Century Research

We have argued that mastery of information technology and logistics may transform future biomedical research.

21st Century Research

We have argued that mastery of information technology and logistics may transform future biomedical research.

The examples of Celera and Millennium Pharmaceuticals suggest that the 21st Century may already have arrived.

21st Century Research

We have argued that mastery of information technology and logistics may transform future biomedical research.

The examples of Celera and Millennium Pharmaceuticals suggest that the 21st Century may already have arrived.

If public sector funding cannot provide public-sector researchers with access to data of the sort that Celera hopes to accumulate, what will be the effects on public sector research?

21st Century Research

We have argued that mastery of information technology and logistics may transform future biomedical research.

The examples of Celera and Millennium Pharmaceuticals suggest that the 21st Century may already have arrived.

If public sector funding cannot provide public-sector researchers with access to data of the sort that Celera hopes to accumulate, what will be the effects on public sector research?

Possibly, many biomedical research fields may disappear entirely into the private sector, when it becomes impossible to access essential resources with public-sector support.

Slides:

http://www.esp.org/rjr/ccaf.pdf