

Economic Impact of Bioinformatics

<http://www.esp.org/rjr/atcc.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

Biotechnology, the "magic technology" of the 21st Century, depends on genomic research and genomics in turn depends upon information technology. Computers are the instruments that allow us to "see" genomes, just as electron microscopes allow us to see viruses.

When information technology becomes critical for any activity, the economics of that field changes substantially. In the last 25 years, field after field of human endeavor has been transformed when the relentless effects of Moore's Law, delivering exponentially better computers at exponentially lower prices, eventually provide sufficient computational power at affordable prices.

Bioinformatics is the application of information technology to the problems of biology. With \$2500 desktop PCs now delivering more raw computing power than the first Cray, bioinformatics is rapidly becoming the critical technology for 21st Century biology.

DNA is legitimately seen as a biological mass-storage device, making bioinformatics a sine qua non for effective genomic research. Others areas of biological investigation are also information rich -- for example, an exhaustive tabulation of the Earth's biodiversity would involve a cross-index of the millions of known species against the approximately 500,000,000,000,000 square meters of the Earth's surface.

Access to bioinformatics support and logistics skills are emerging as rate limiting steps for much biomedical research. As 21st Century biology becomes increasing dependent upon bioinformatics and logistics, competitive dynamics will change, possibly leading to the movement of substantial areas of biological research from the public to the private sector.

Topics

- Biotechnology will be the magic technology of the 21st Century.

Topics

- Biotechnology will be the magic technology of the 21st Century.
- Moore's Law constantly transforms IT (and everything else).

Topics

- Biotechnology will be the magic technology of the 21st Century.
- Moore's Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology, especially genetics and genomics.

Topics

- Biotechnology will be the magic technology of the 21st Century.
- 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

- Biotechnology will be the magic technology of the 21st Century.
- 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.
- Currently, support for public bio-information infrastructure seems inadequate.

Topics

- Biotechnology will be the magic technology of the 21st Century.
- 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.
- Currently, support for public bio-information infrastructure seems inadequate.
- In the future, much current public research may move into the private sector.

Biotechnology

21st Century Magic

Magic

To a person from 1897, much current technology would seem like magic.

Magic

To a person from 1897, much current technology would seem like magic.

What technology of 2097 would seem magical to a person from 1997?

Magic

To a person from 1897, much current technology would seem like magic.

What technology of 2097 would seem magical to a person from 1997?

Candidate: Biotechnology so advanced that the distinction between living and non-living is blurred.

Moore's Law

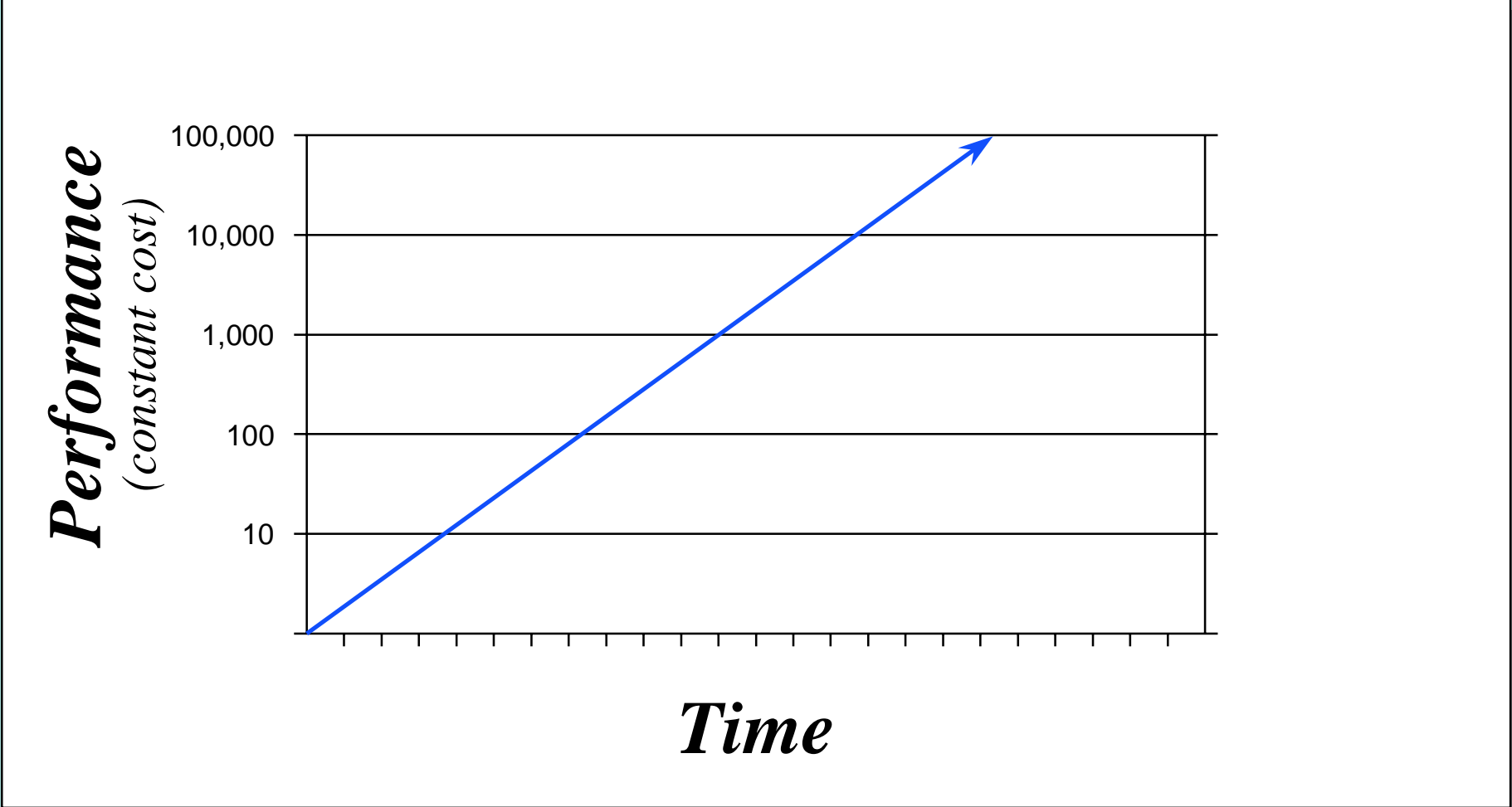
*Transforms InfoTech
(and everything else)*

Moore's Law: *The Statement*

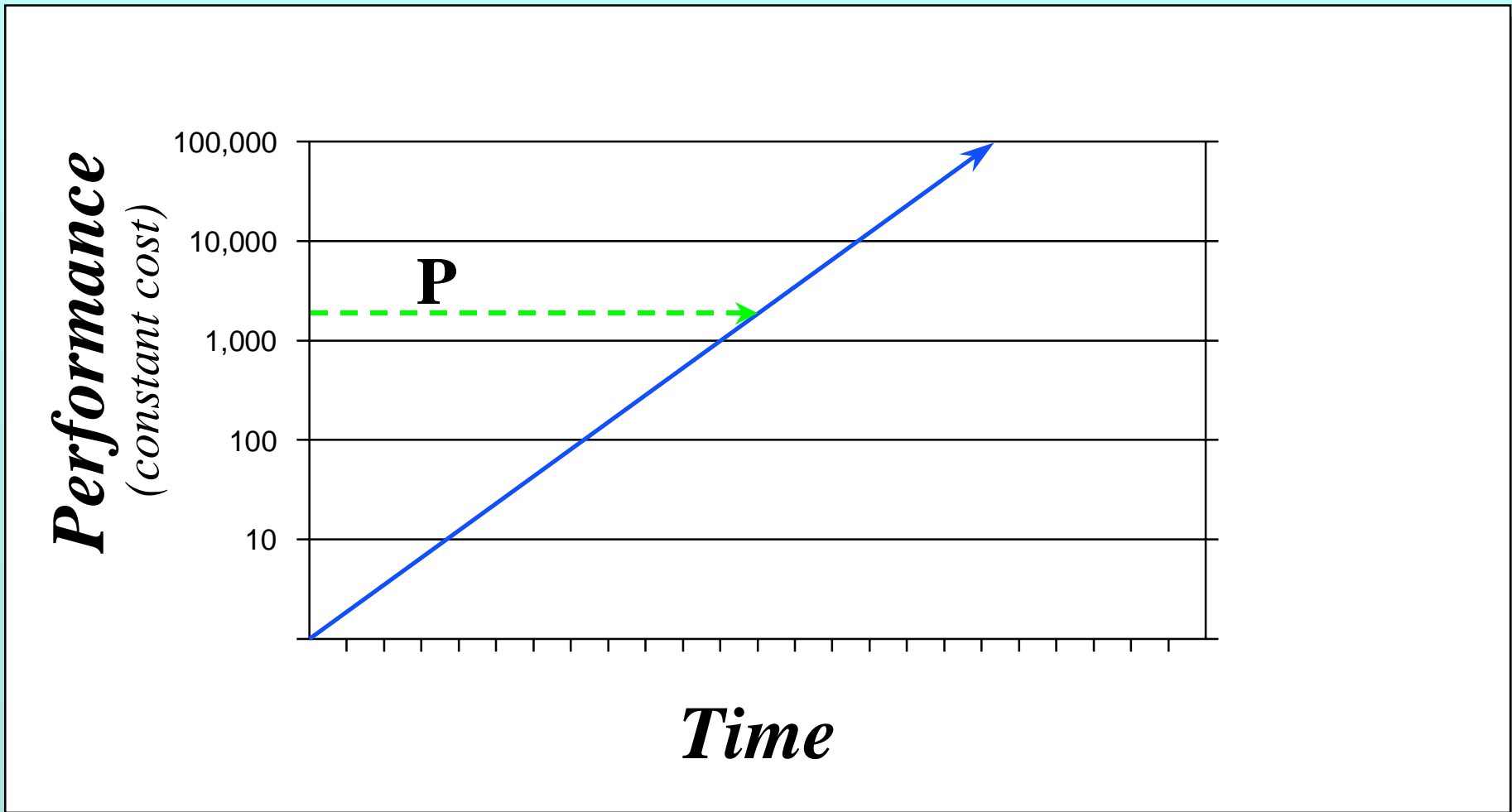
Every eighteen months, the number of transistors that can be placed on a chip doubles.

Gordon Moore, co-founder of Intel...

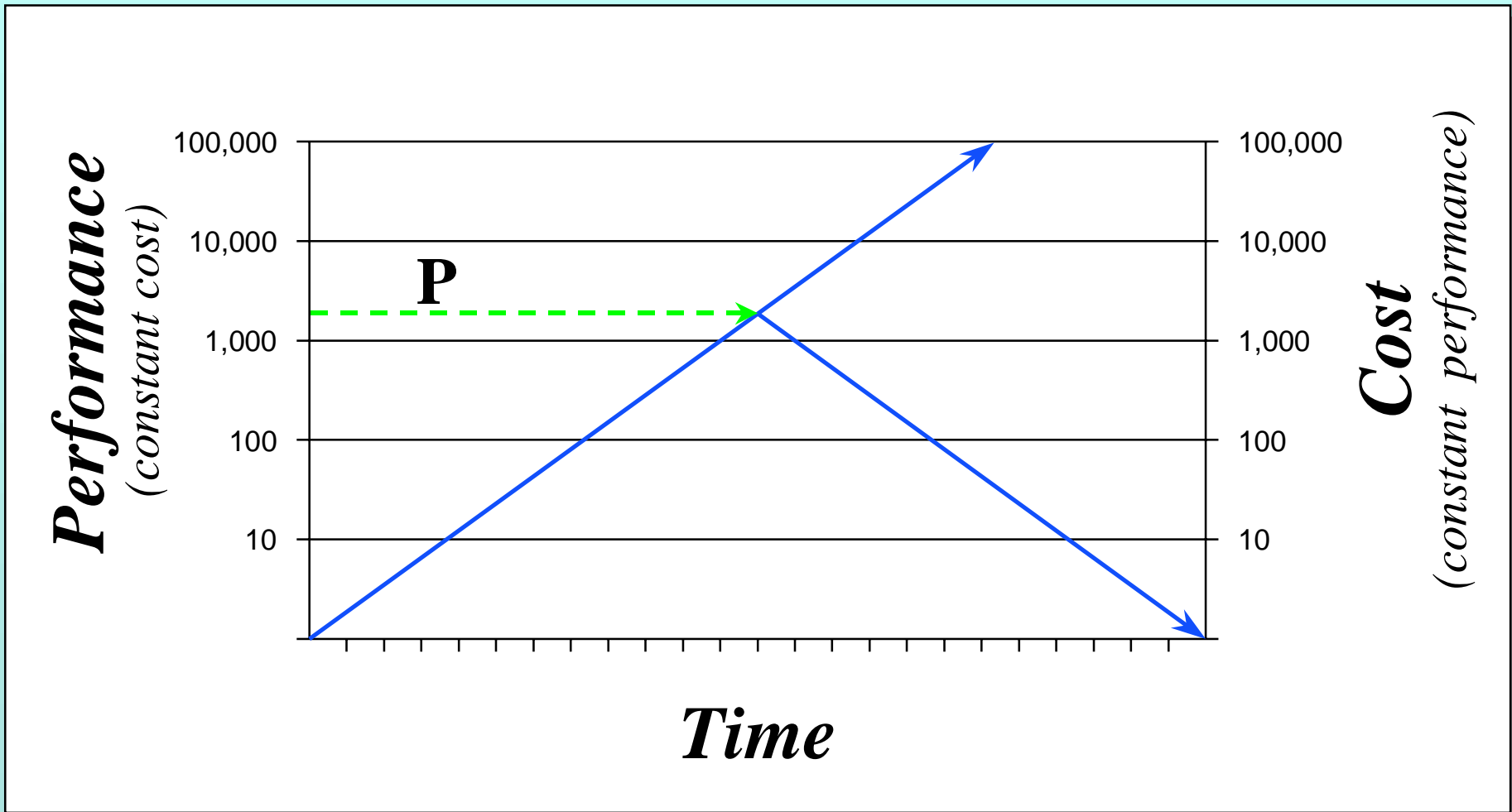
Moore's Law: *The Effect*



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

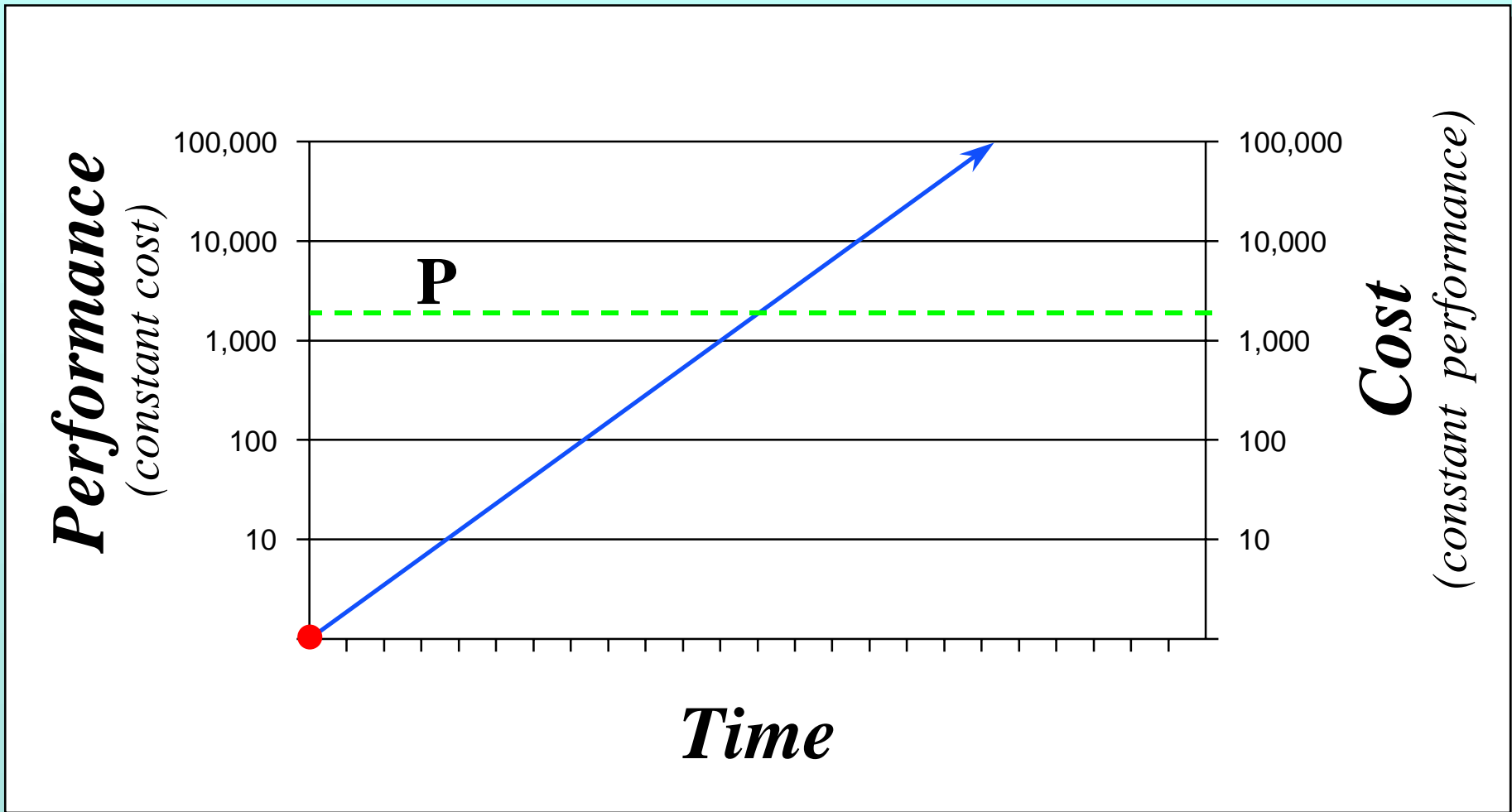
- It's Impossible
- It's Impractical

Moore's Law: *The Effect*

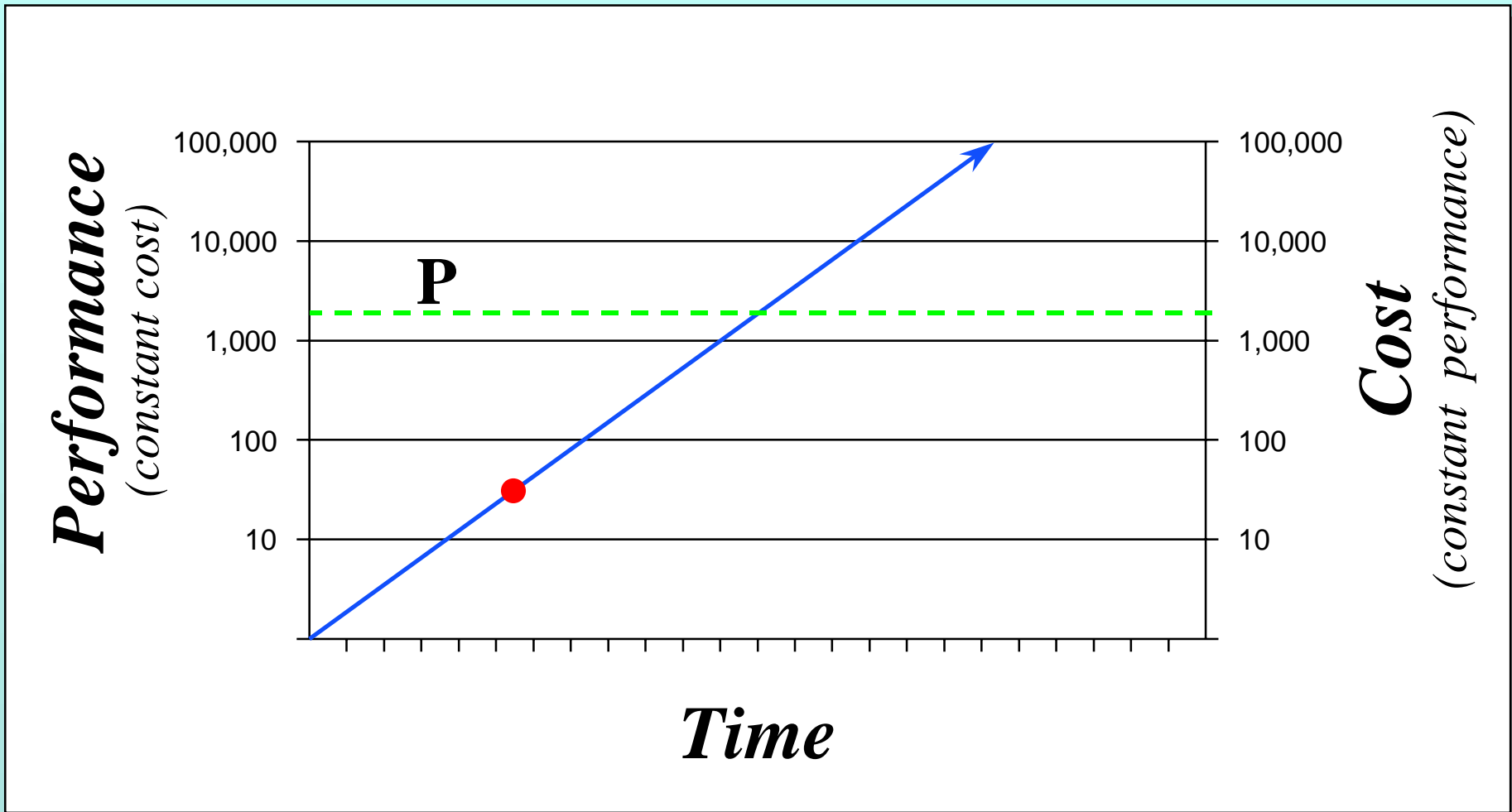
Three Phases of Novel IT Applications

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

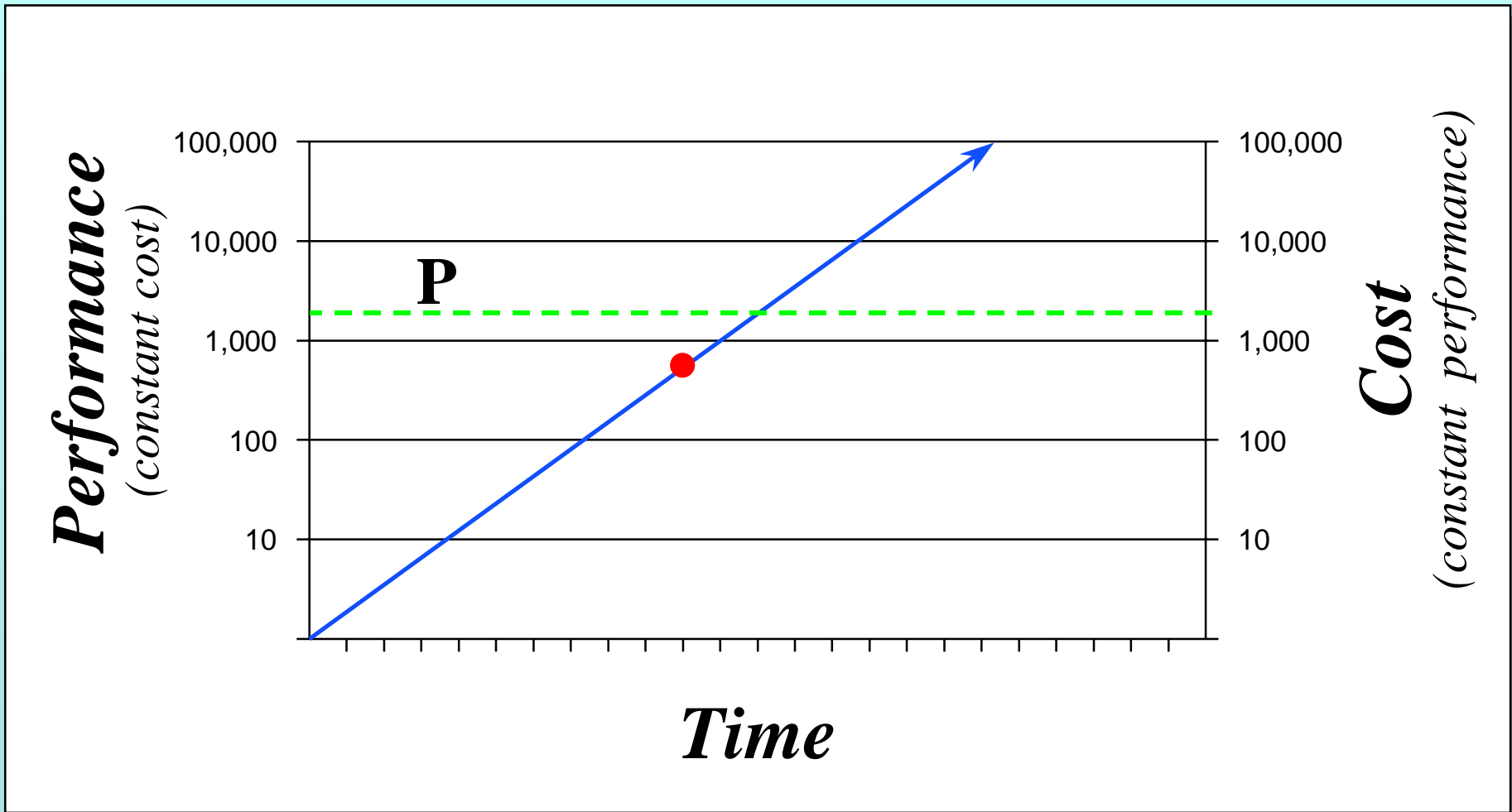
Moore's Law: *The Effect*



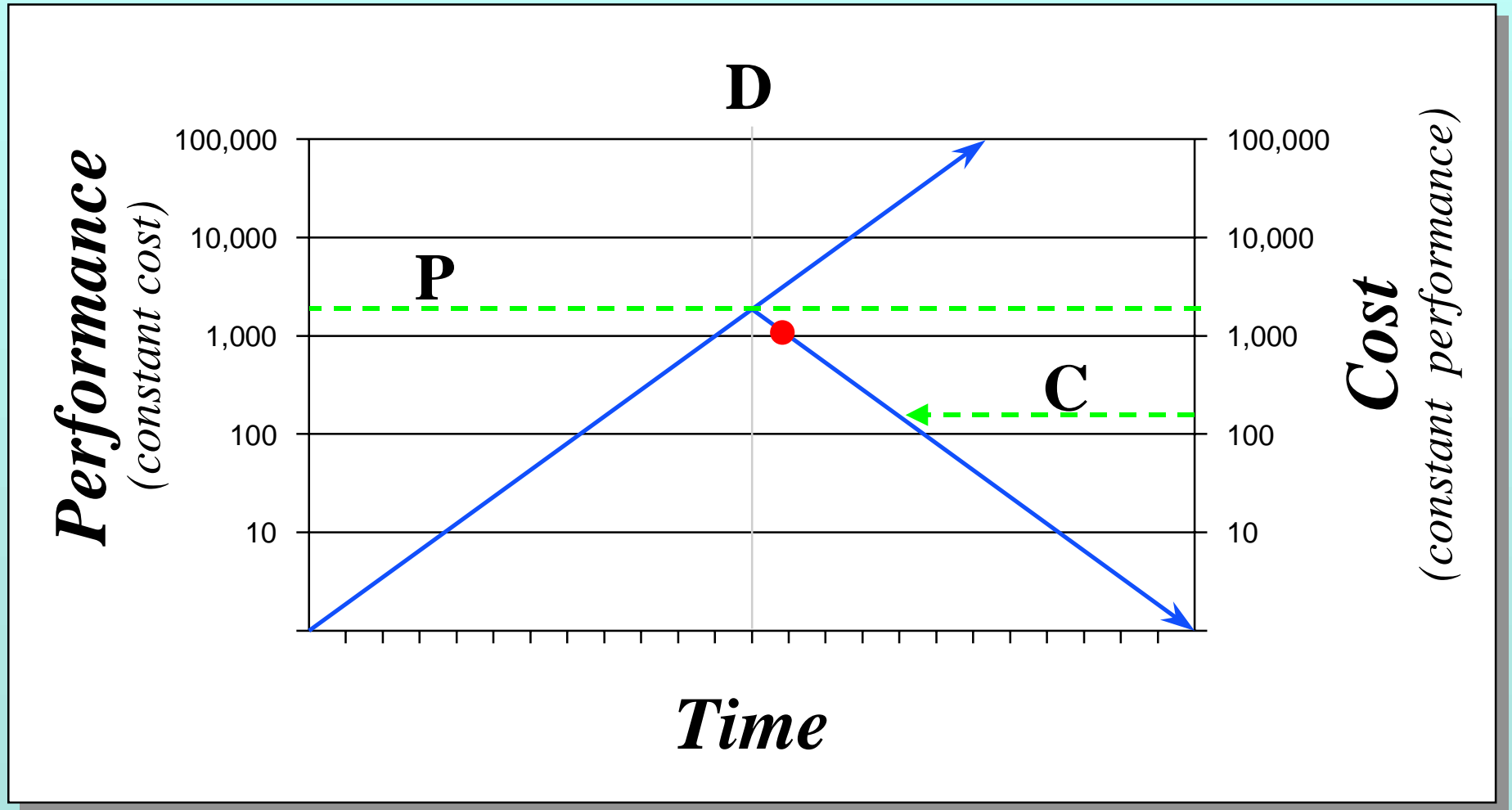
Moore's Law: *The Effect*



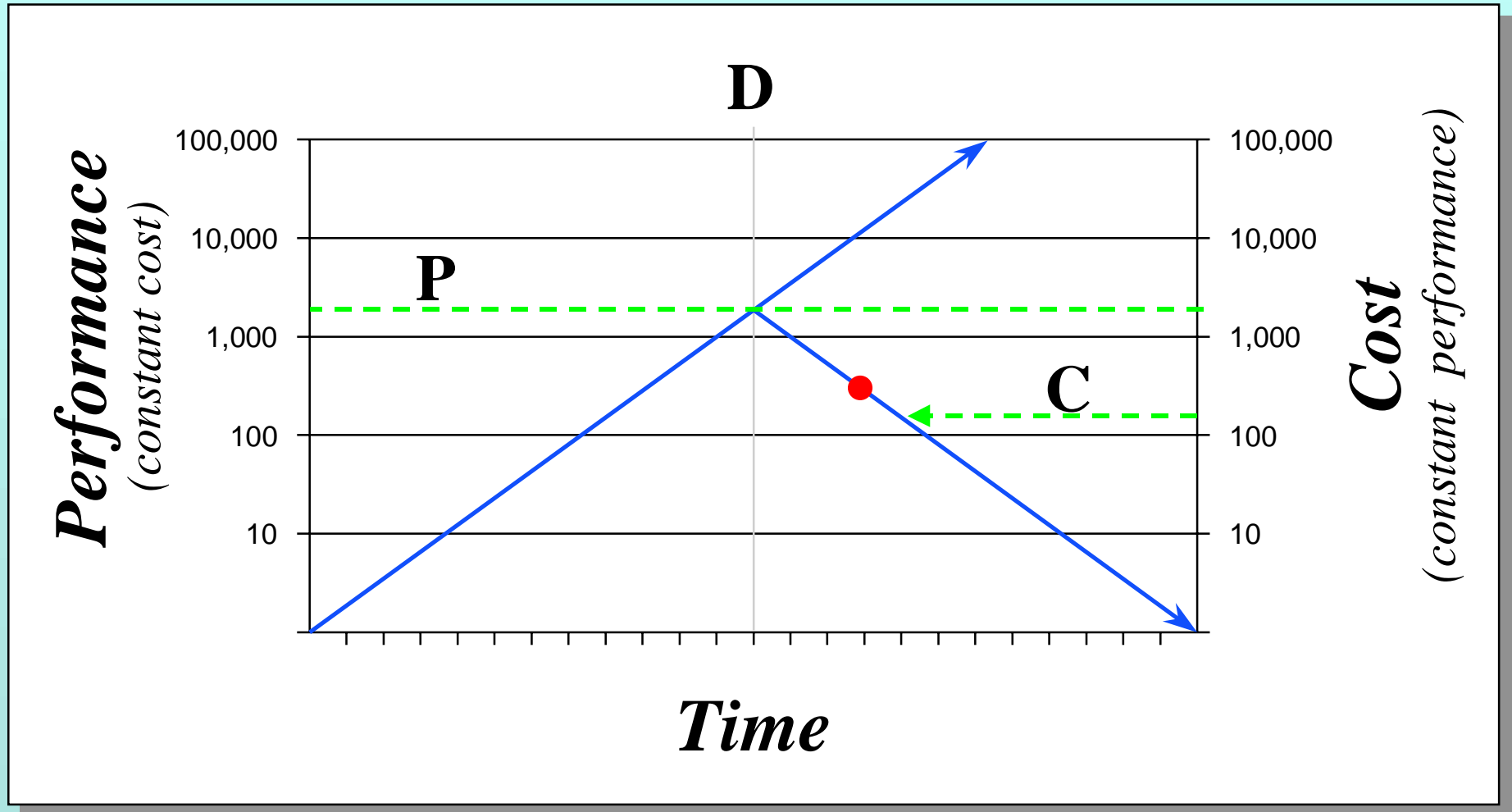
Moore's Law: *The Effect*



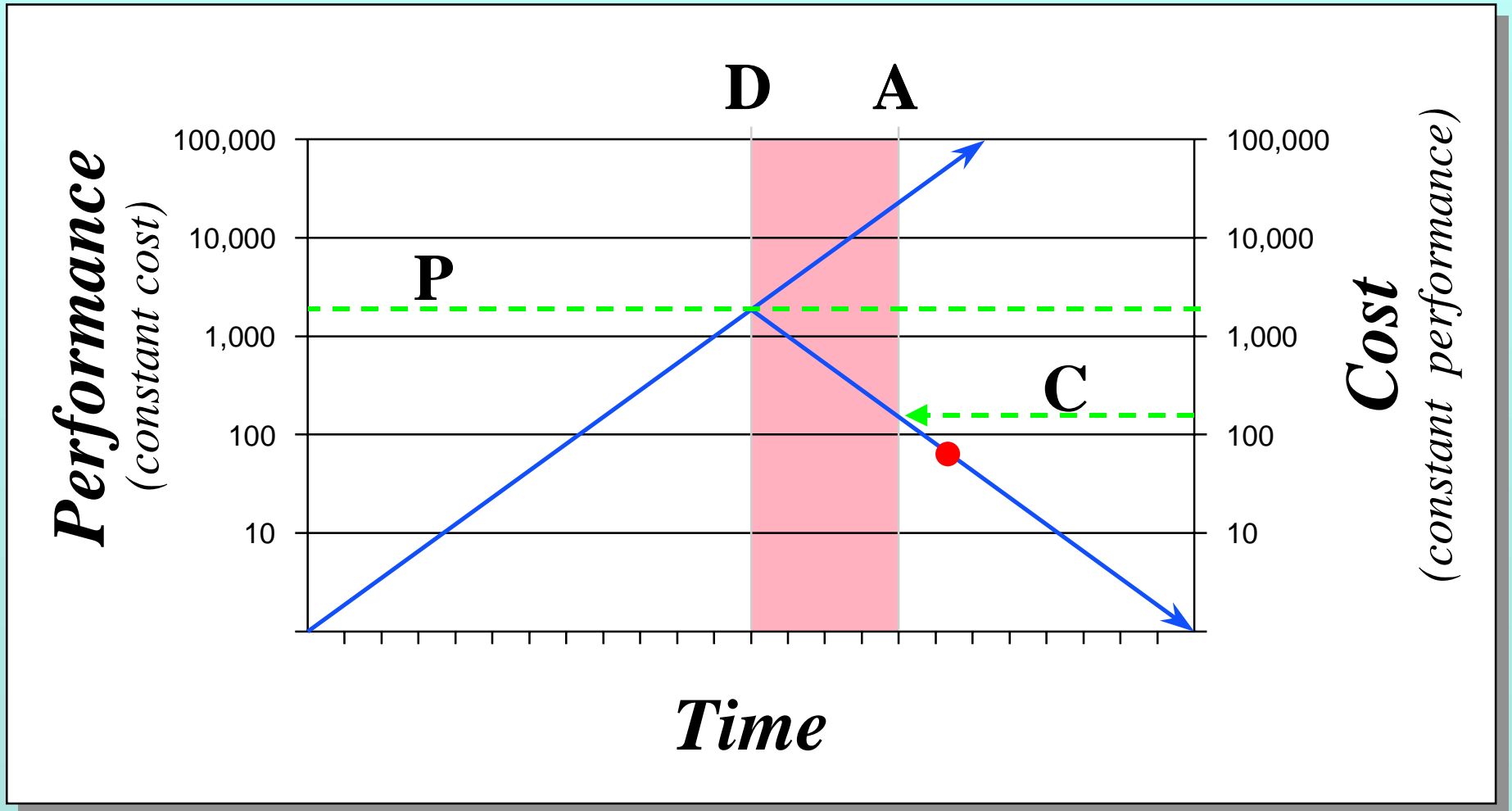
Moore's Law: *The Effect*



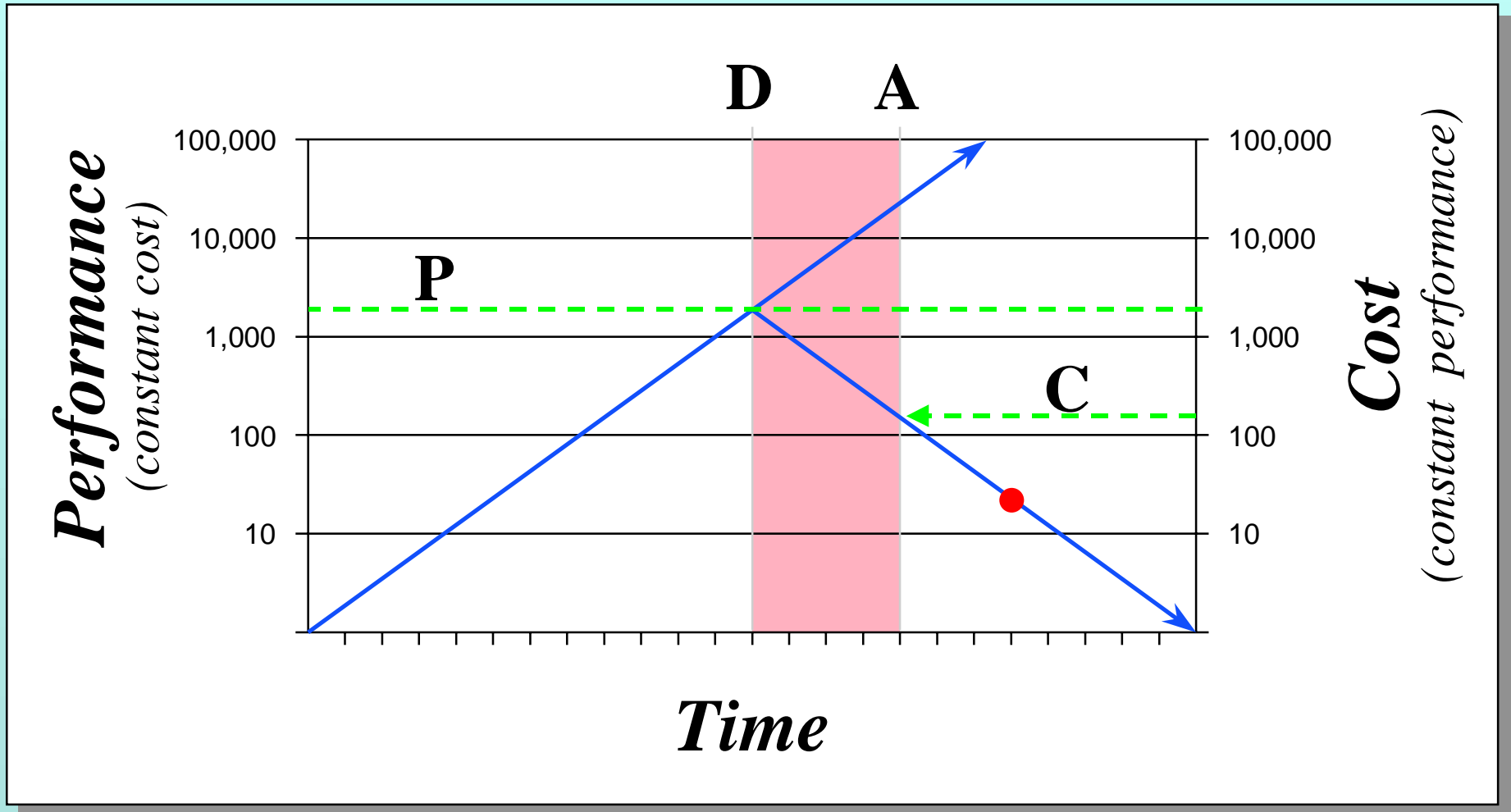
Moore's Law: *The Effect*



Moore's Law: *The Effect*



Moore's Law: *The Effect*



Moore's Law: *The Effect*

- *The first to understand and deploy new IT capabilities often seize great competitive advantage.*

Moore's Law: *The Effect*

- *The first to understand and deploy new IT capabilities often seize great competitive advantage.*
- *As improved uses of technology are developed, “business” processes change.*

Moore's Law: *The Effect*

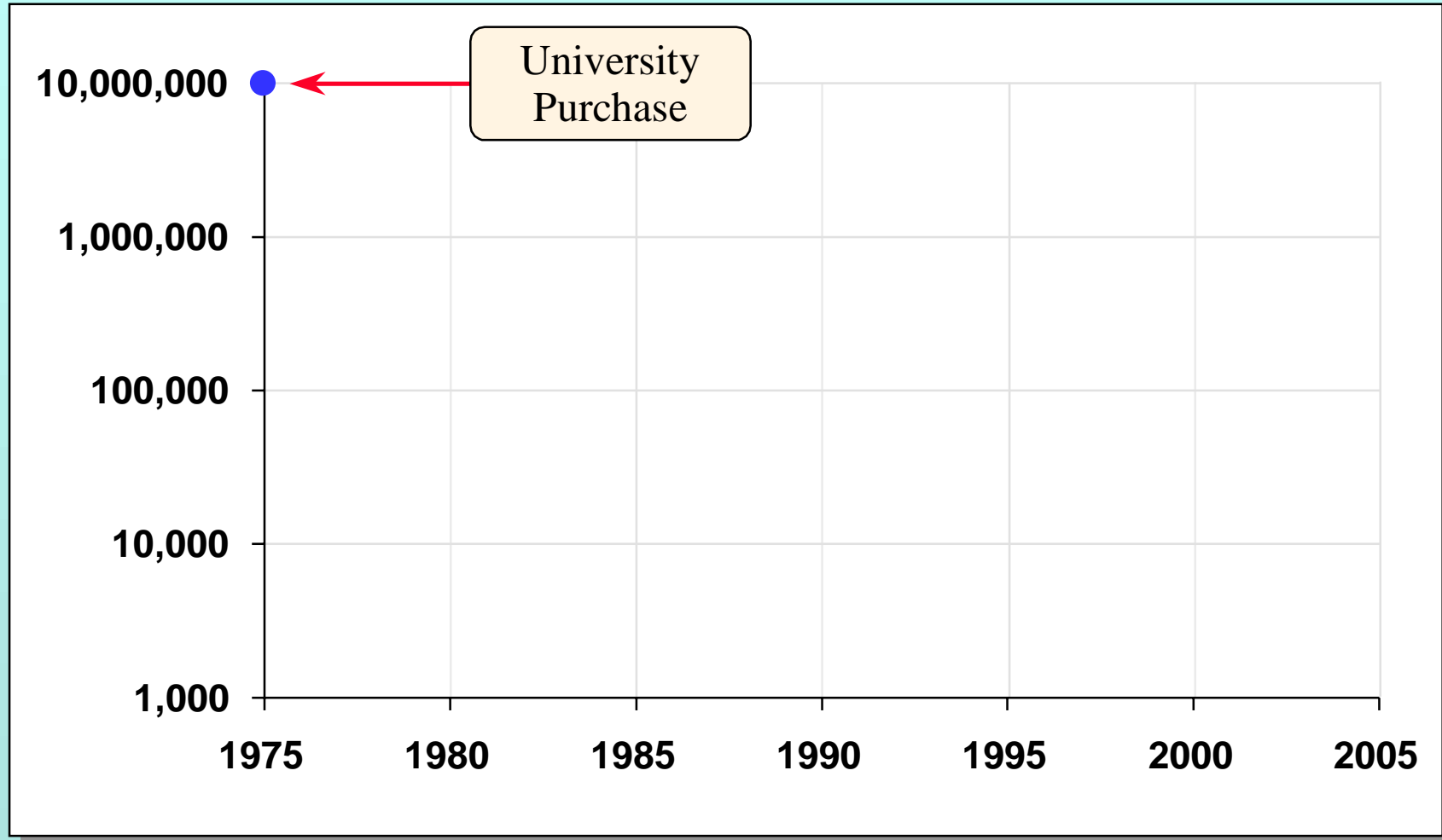
- *The first to understand and deploy new IT capabilities often seize great competitive advantage.*
- *As improved uses of technology are developed, “business” processes change.*
- *Ultimately, access to appropriate IT becomes essential for simple existence.*

Moore's Law: *The Effect*

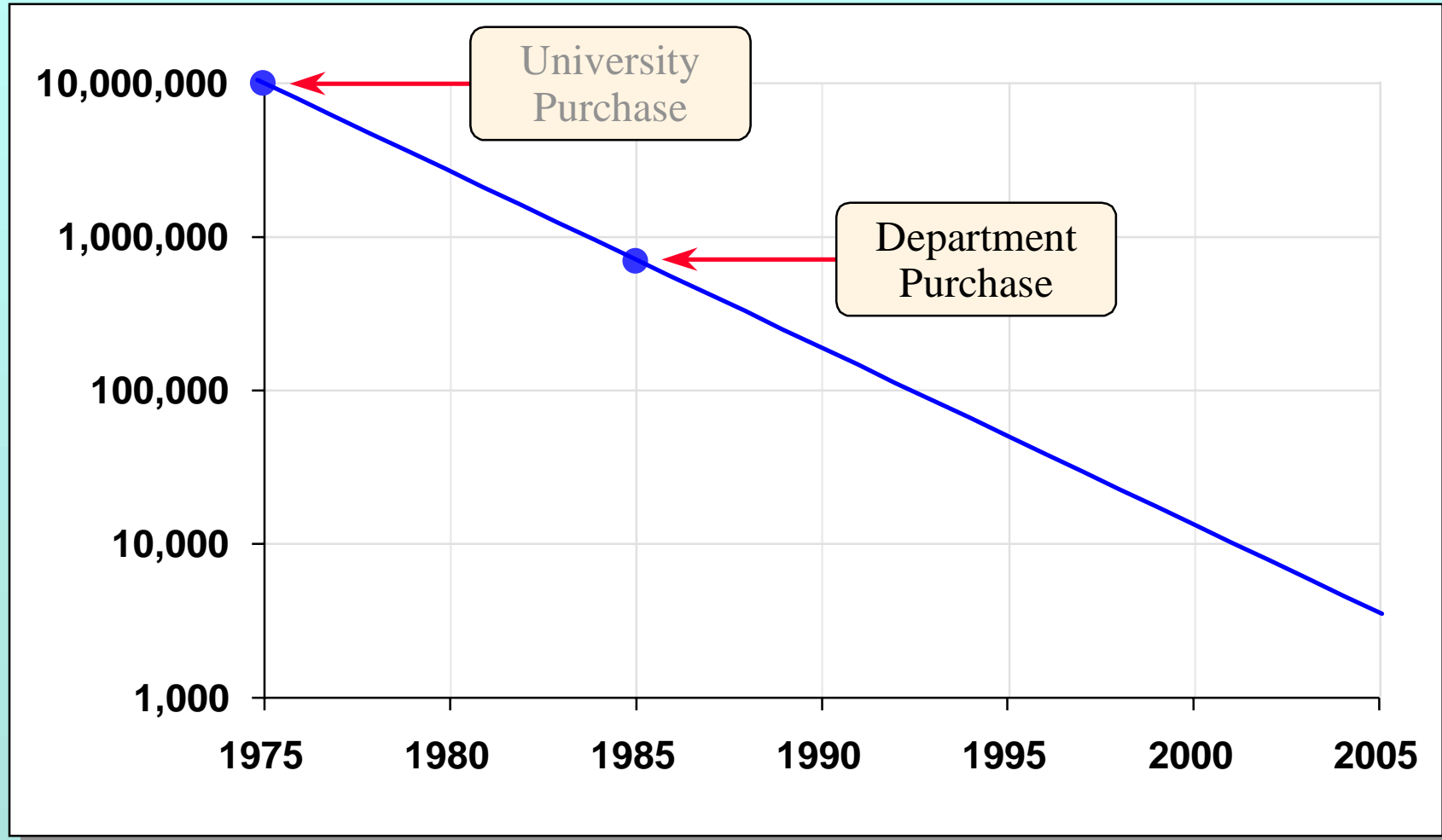
- *The first to understand and deploy new IT capabilities often seize great competitive advantage.*
- *As improved uses of technology are developed, “business” processes change.*
- *Ultimately, access to appropriate IT becomes essential for simple existence.*

Relevance for biology?

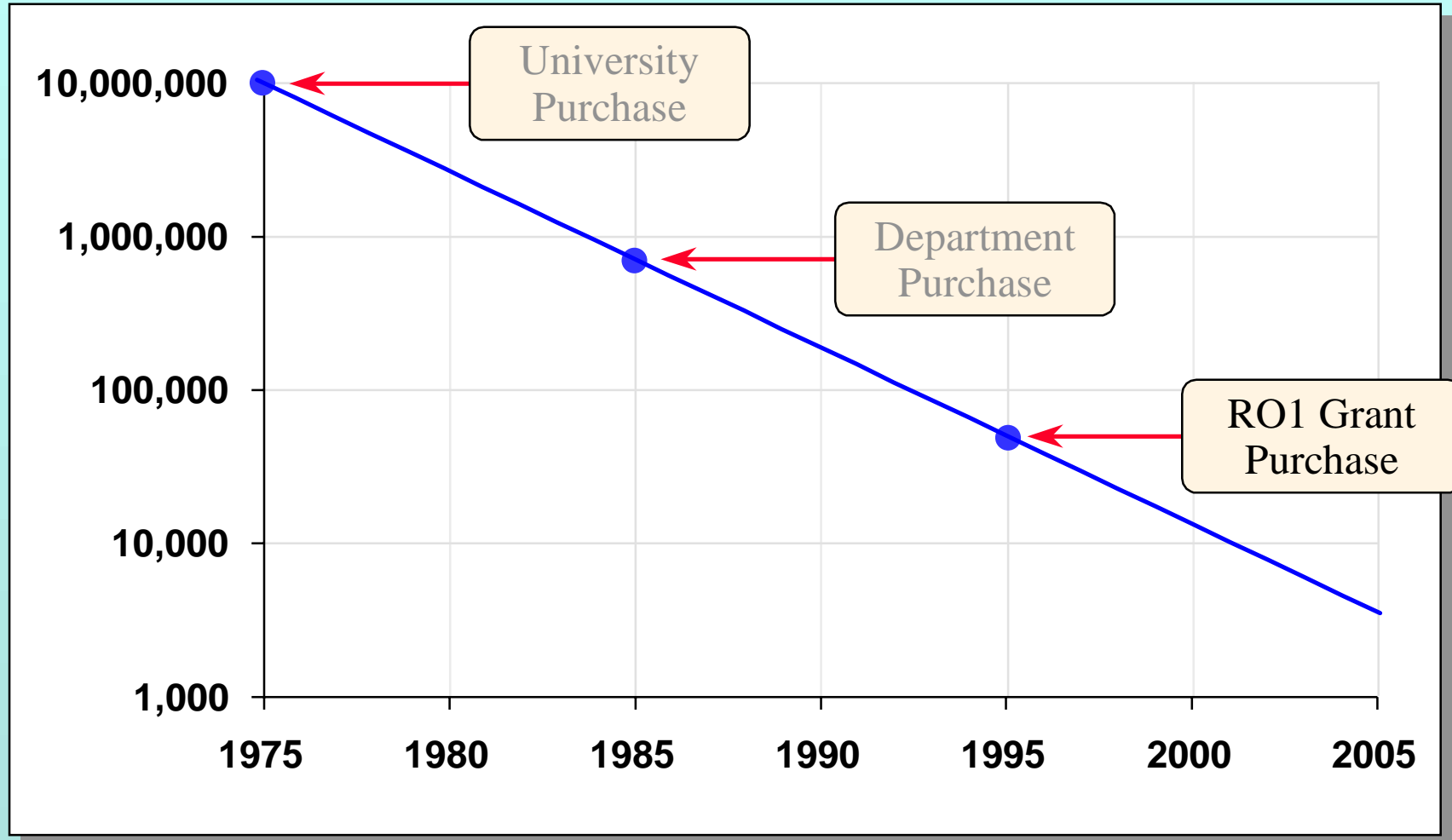
Cost (constant performance)



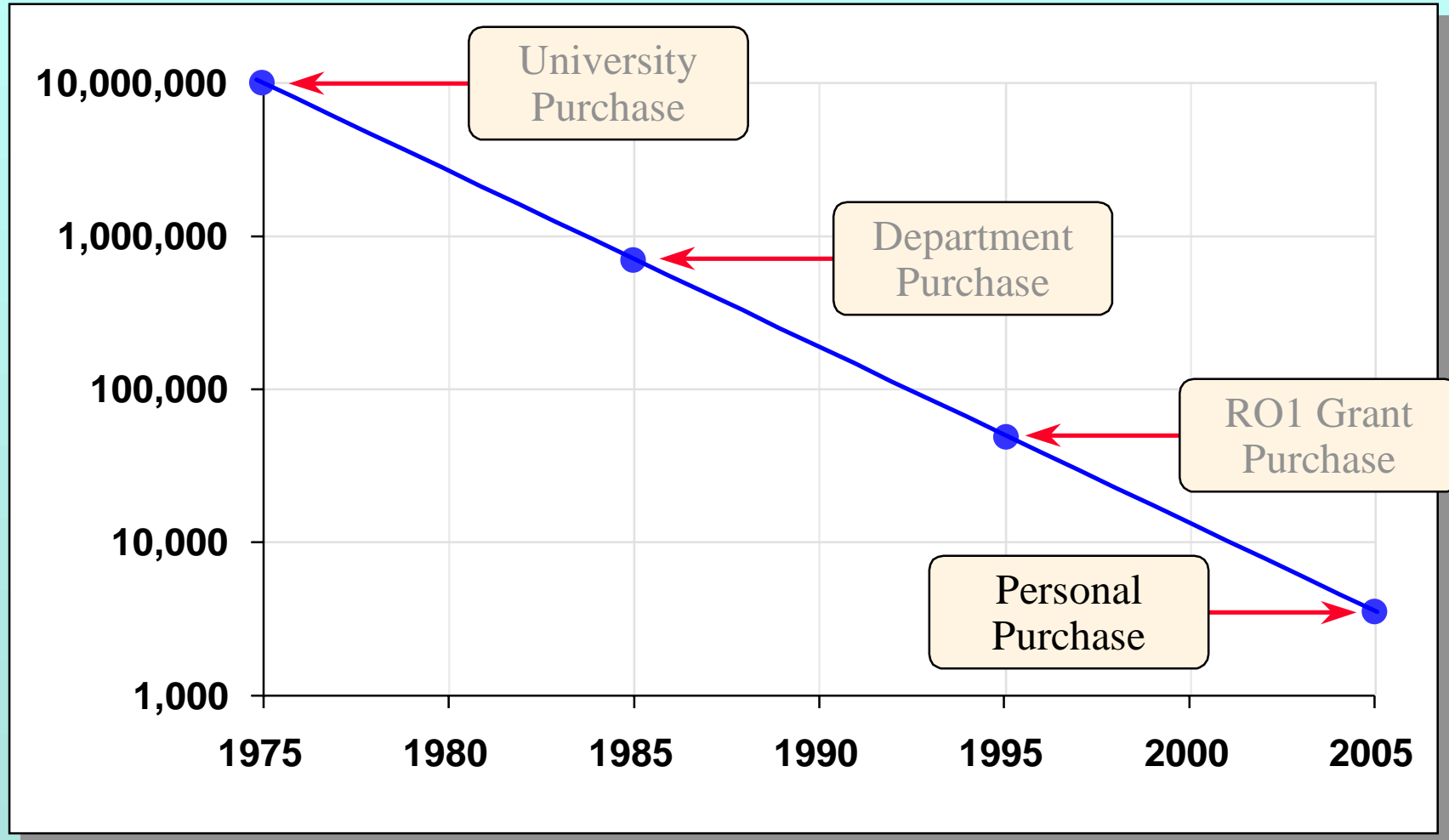
Cost (constant performance)



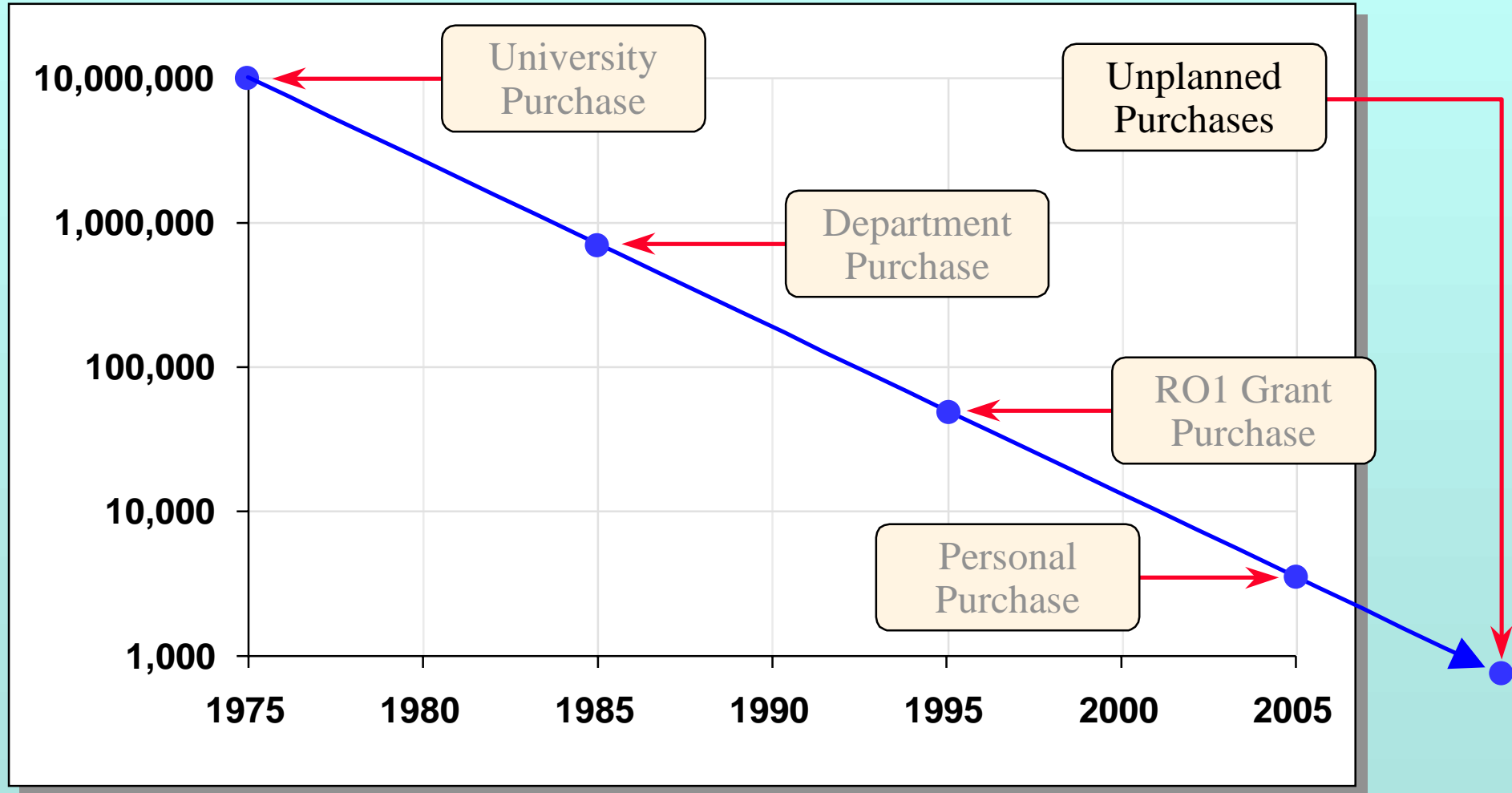
Cost (constant performance)



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*.

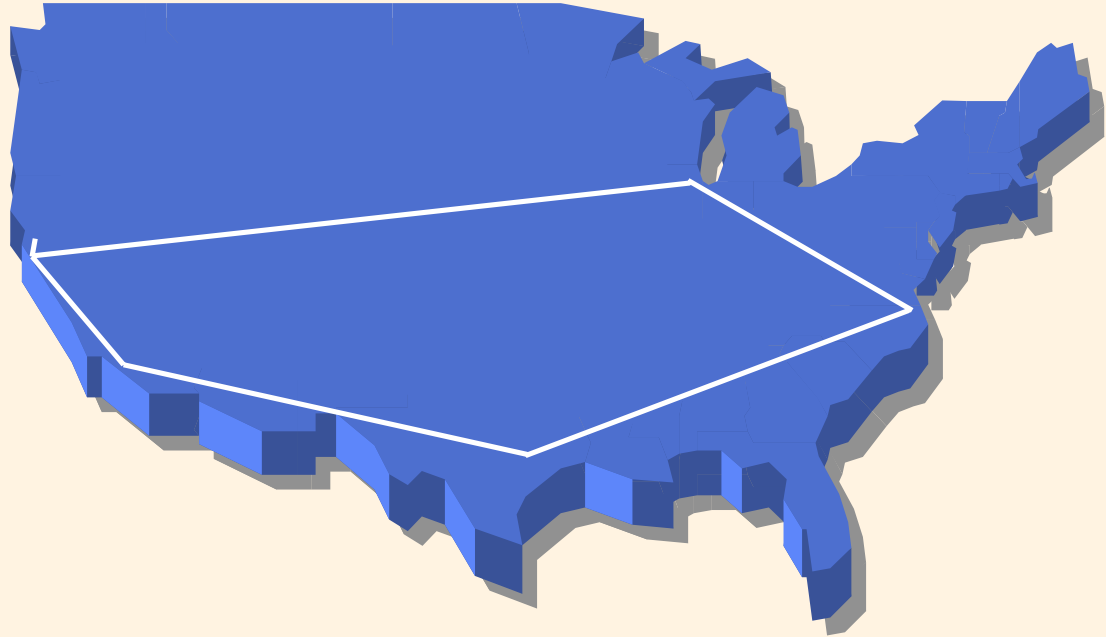
Genetics as Code

[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*.

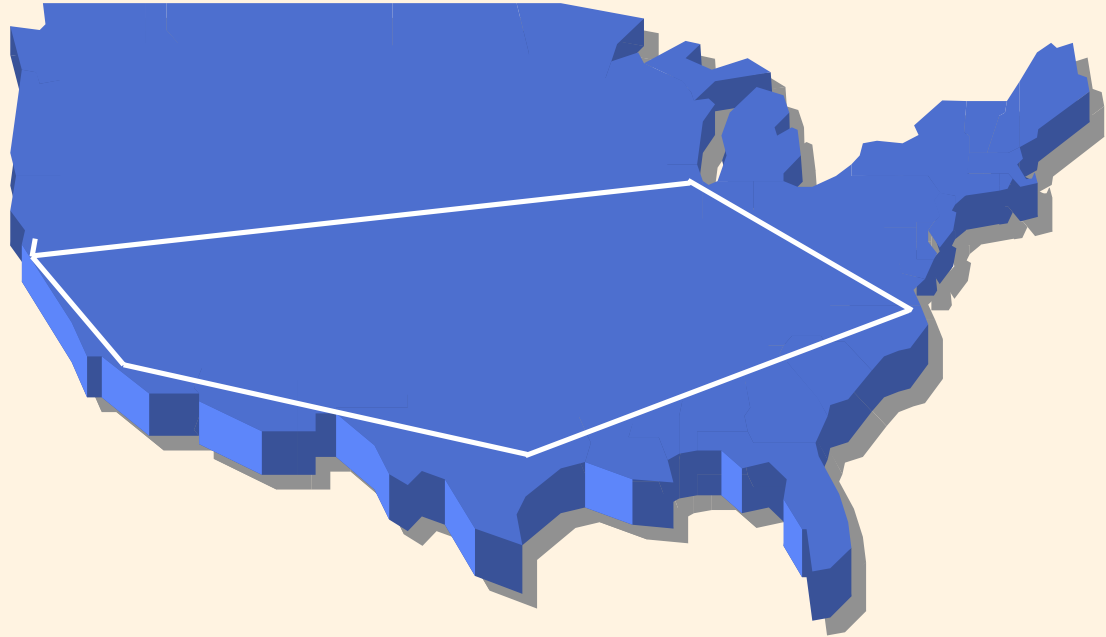
One Human Sequence

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

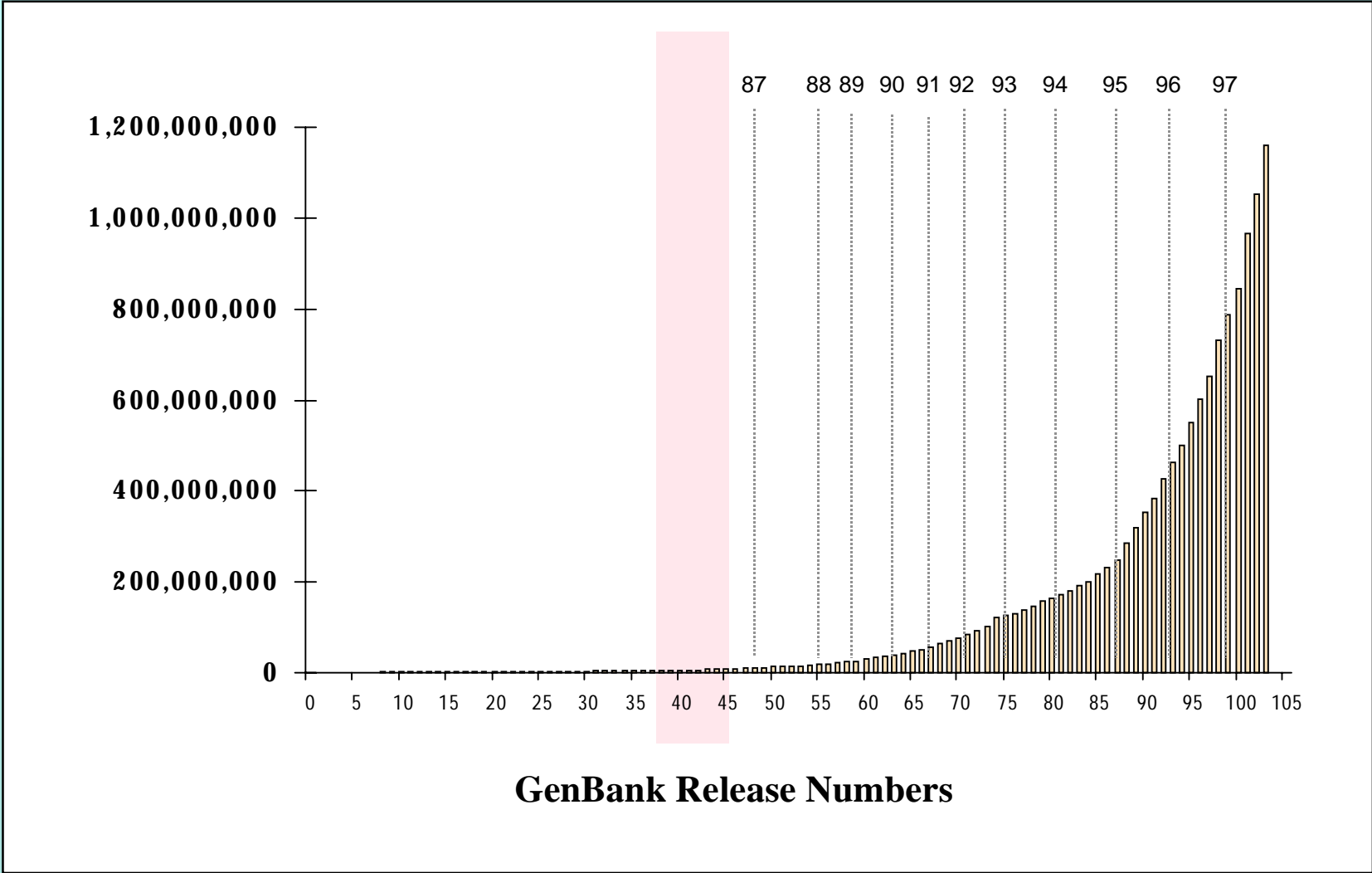
Progress towards all of the [Genome Project] goals will require the establishment of well-funded centralized facilities, including a stock center for the cloned DNA fragments generated in the mapping and sequencing effort and a data center for the computer-based collection and distribution of large amounts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

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

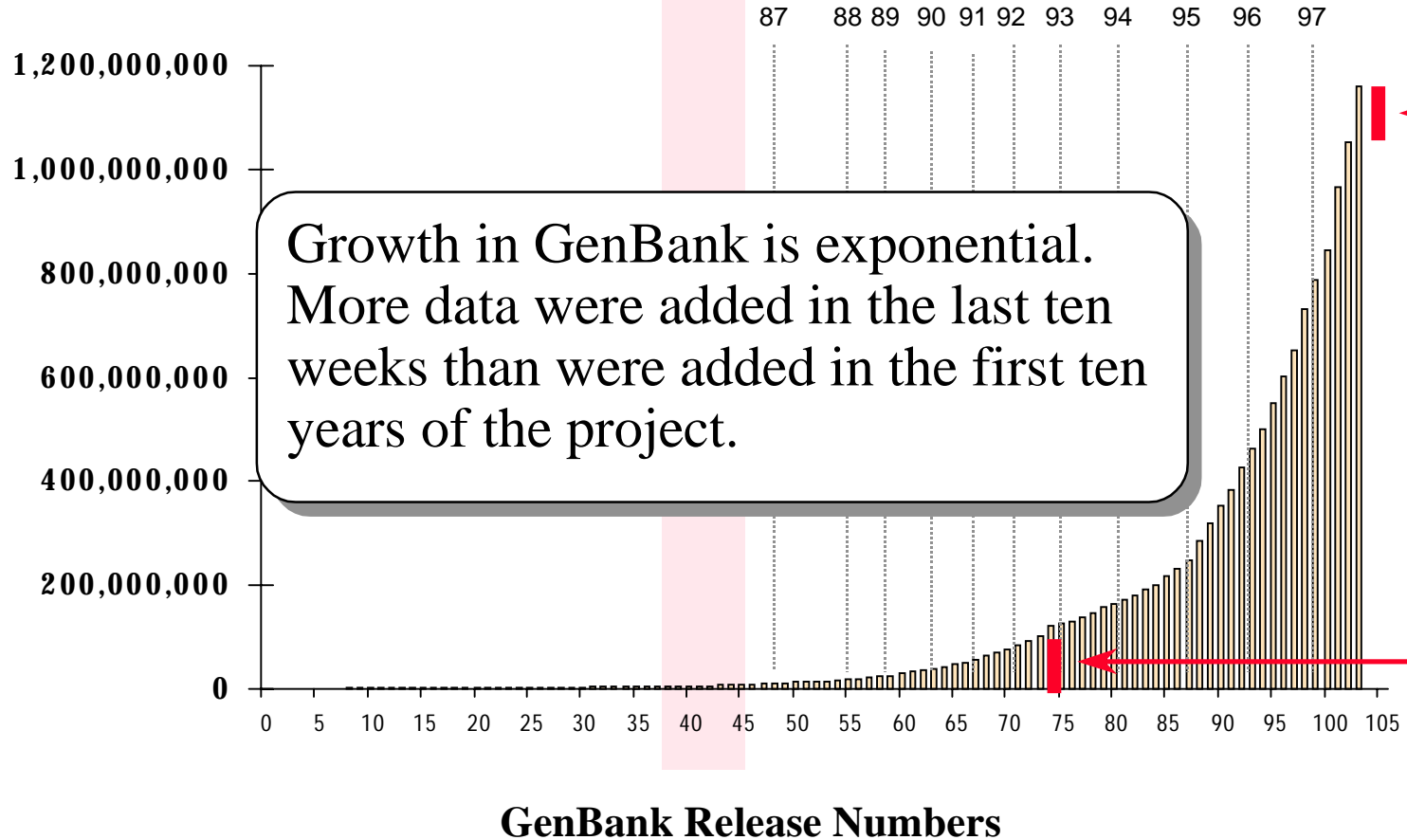
GenBank Totals *(Release 103)*

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%

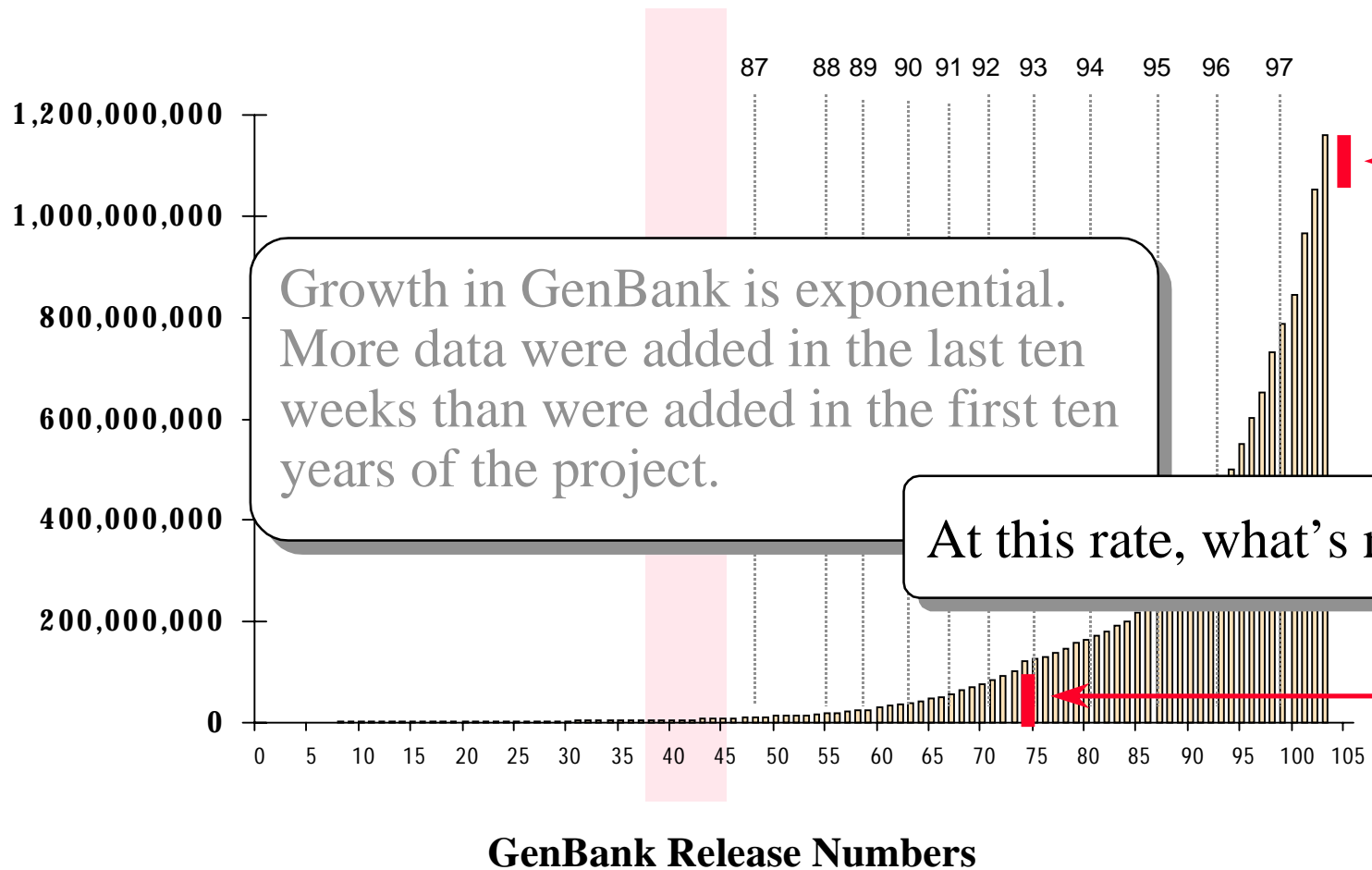
Base Pairs in GenBank



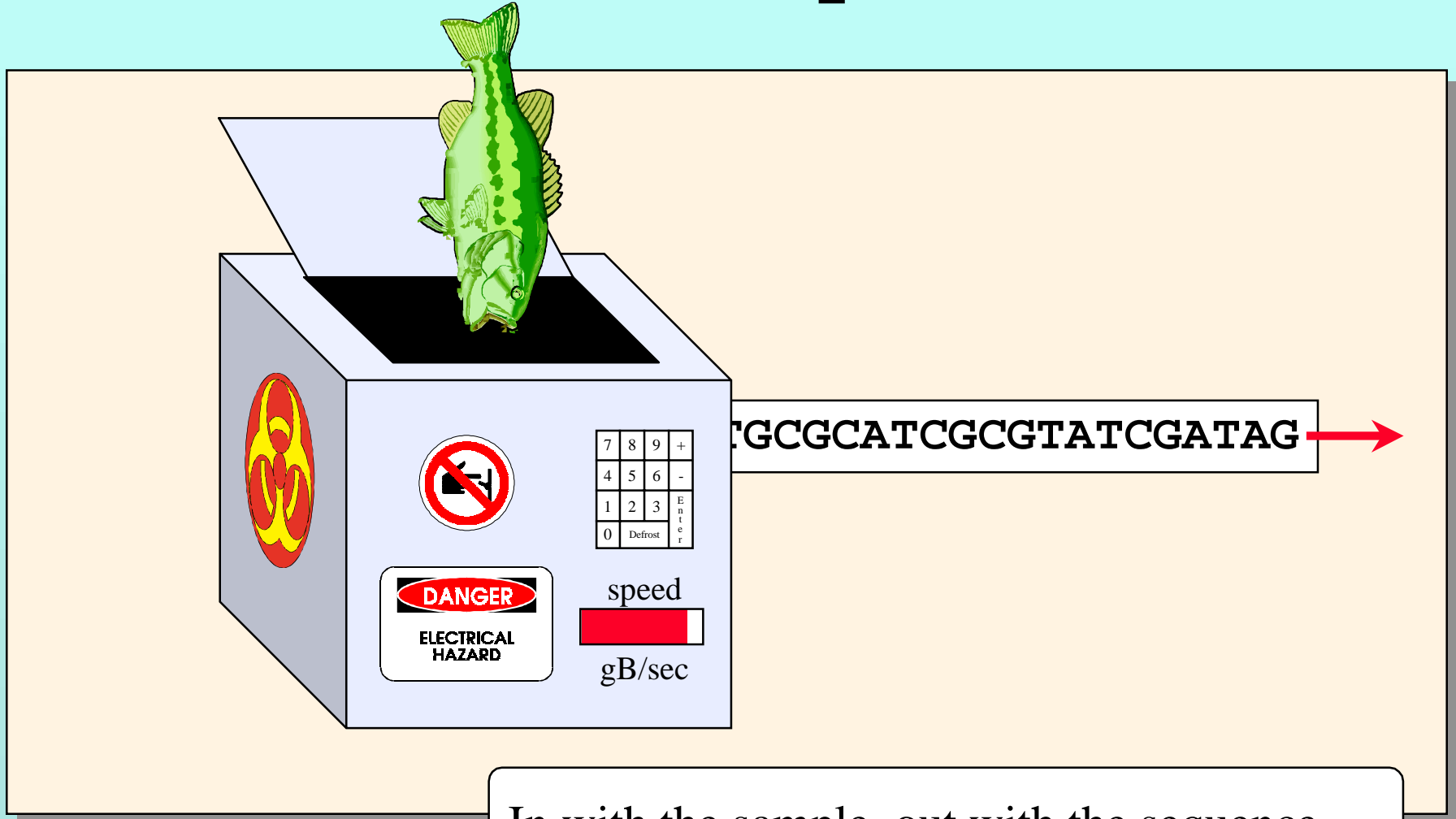
Base Pairs in GenBank



Base Pairs in GenBank



ABI Bass-o-Matic Sequencer



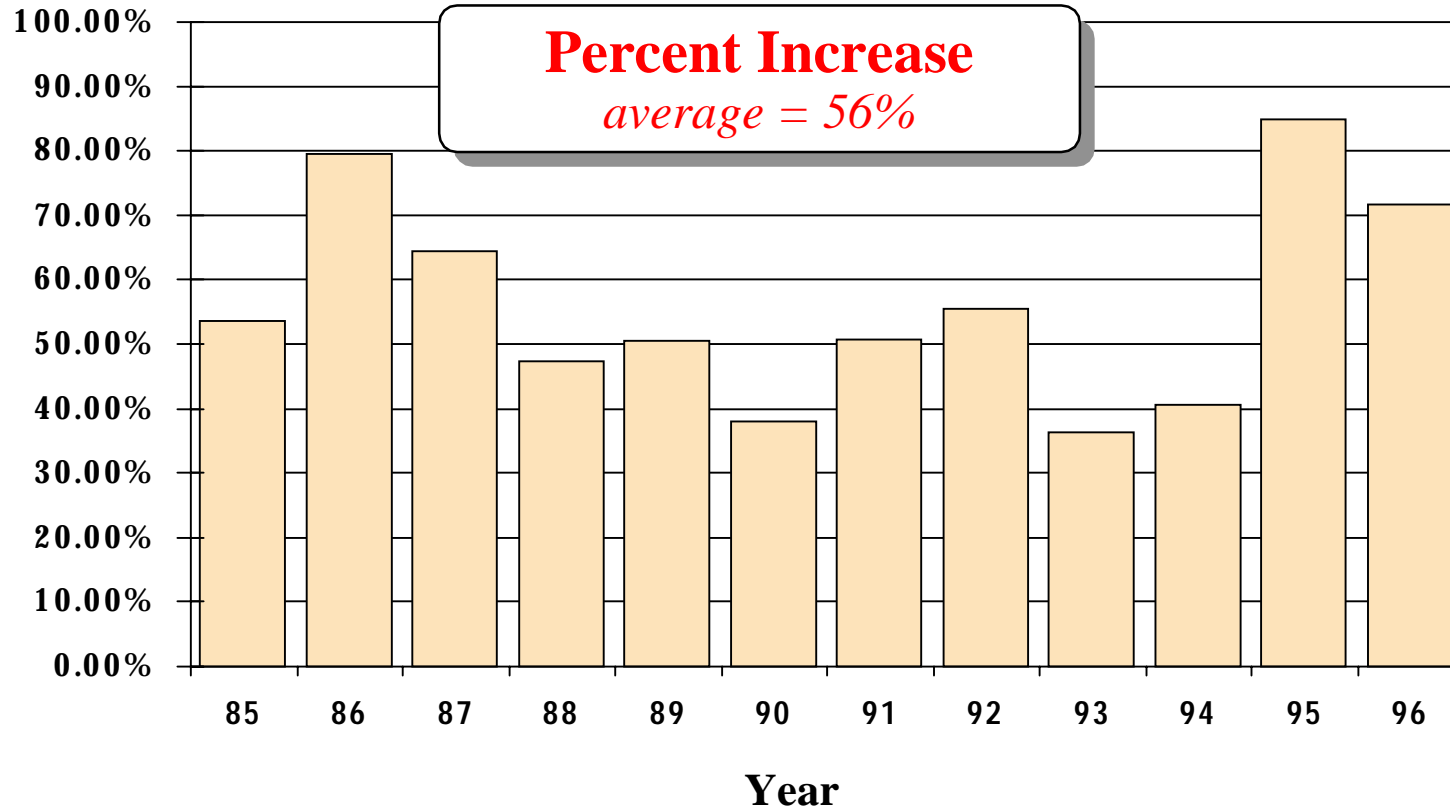
In with the sample, out with the sequence...

What's Really Next

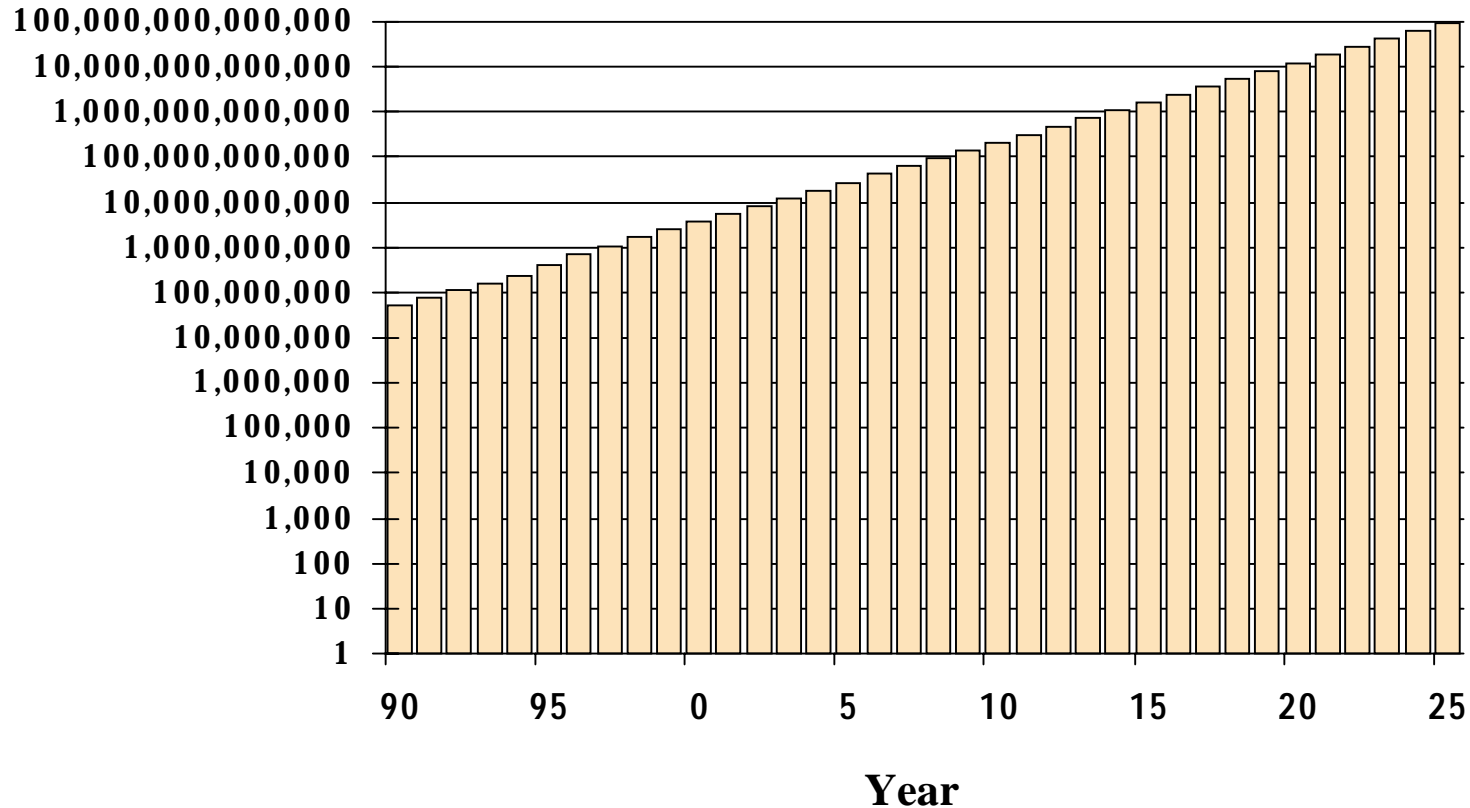
The post-genome era in biological research will take for granted ready access to huge amounts of genomic data.

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

Base Pairs in GenBank (*Percent Increase*)

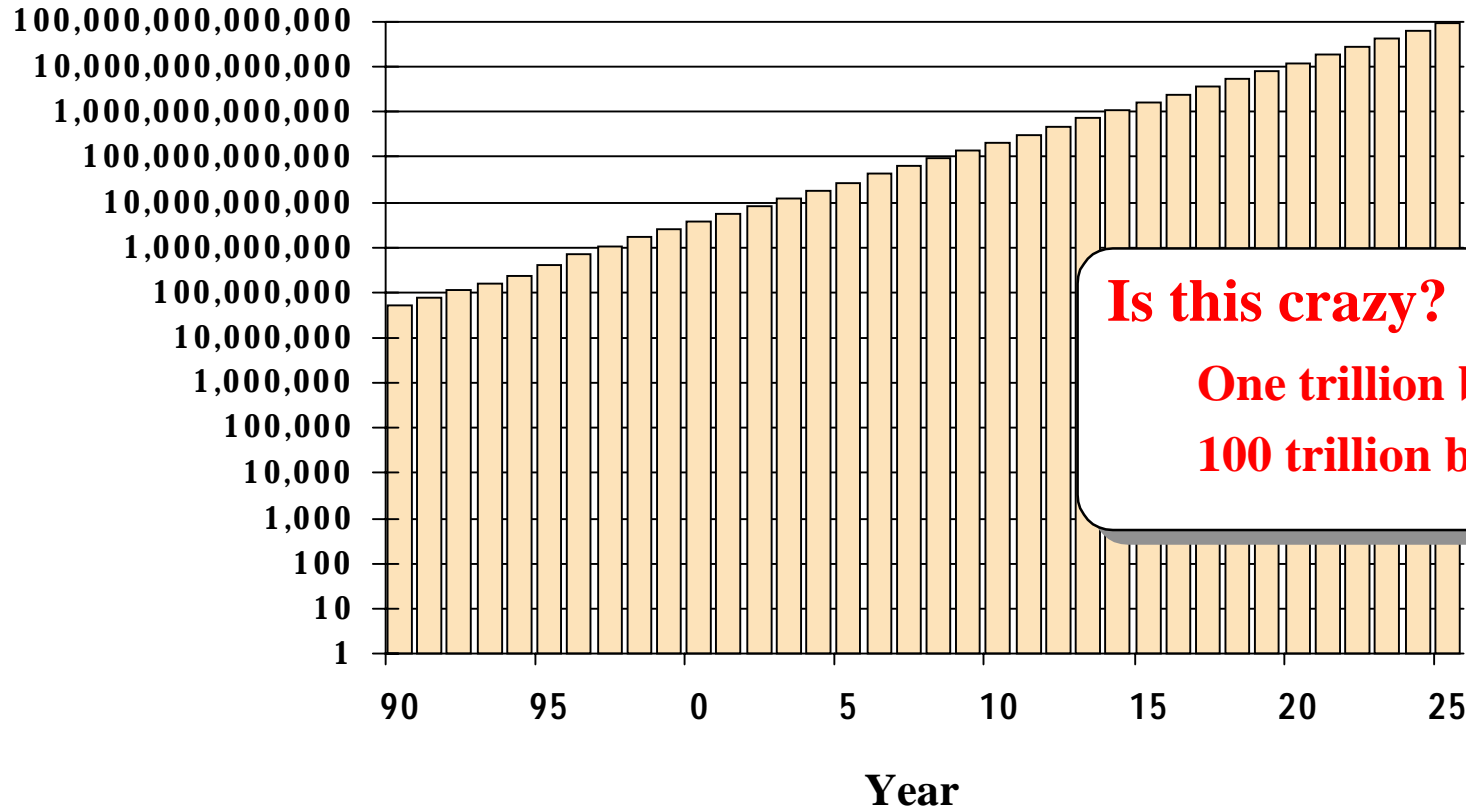


Projected Base Pairs



Assumed annual growth rate: 50%
(less than current rate)

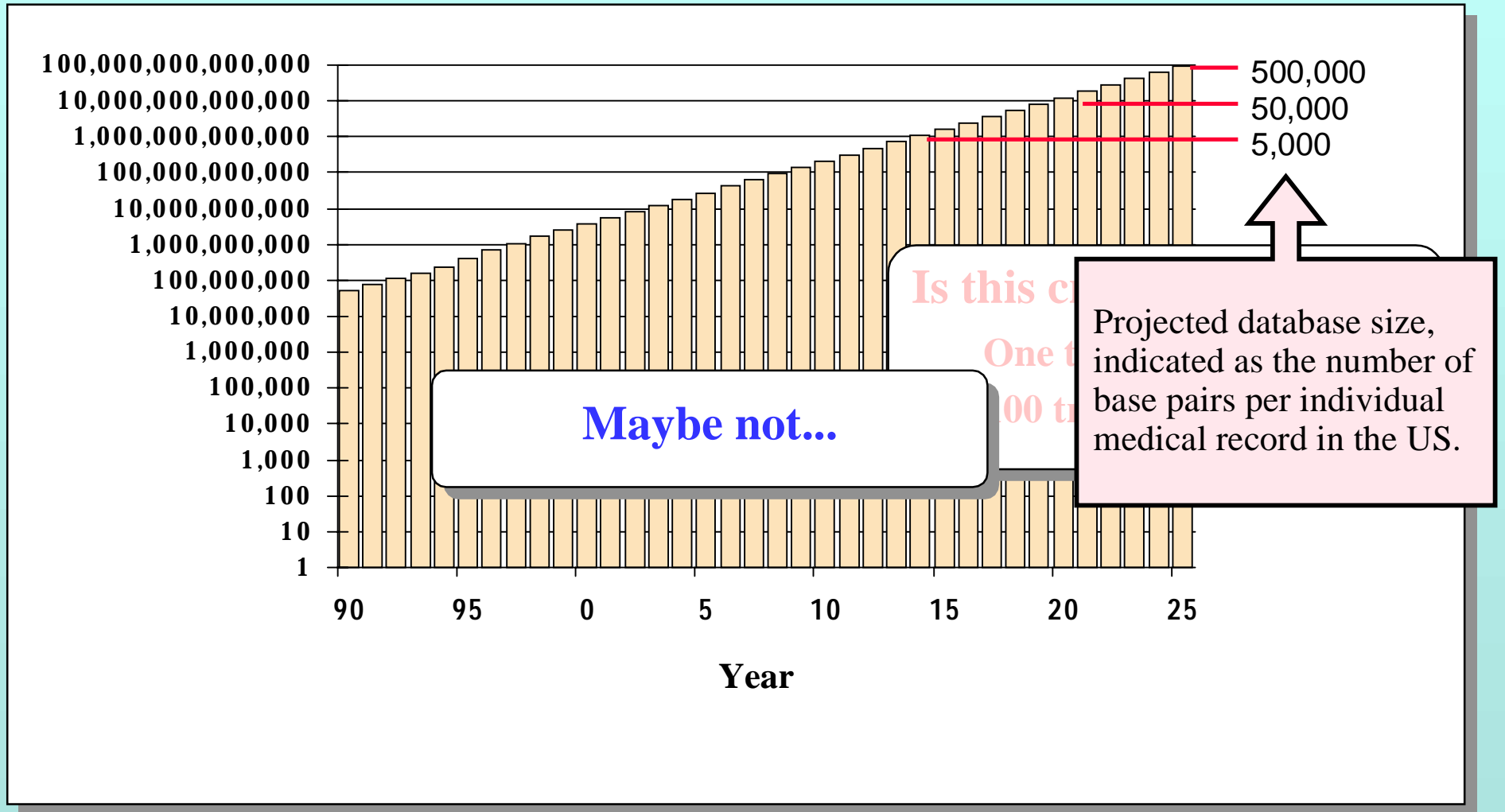
Projected Base Pairs



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.

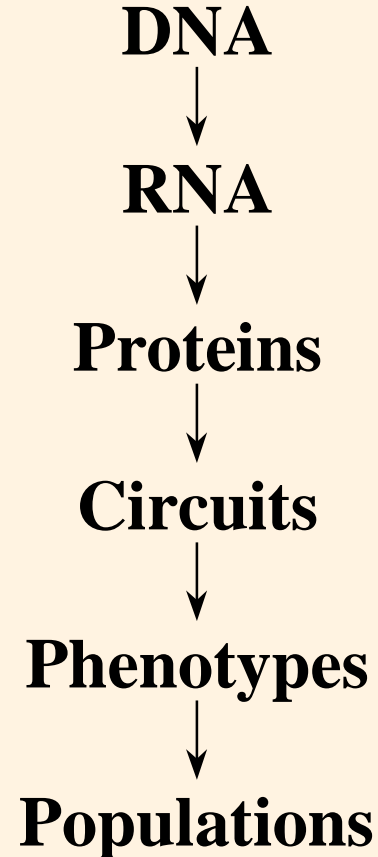
21st Century Biology

The Science

Fundamental Dogma

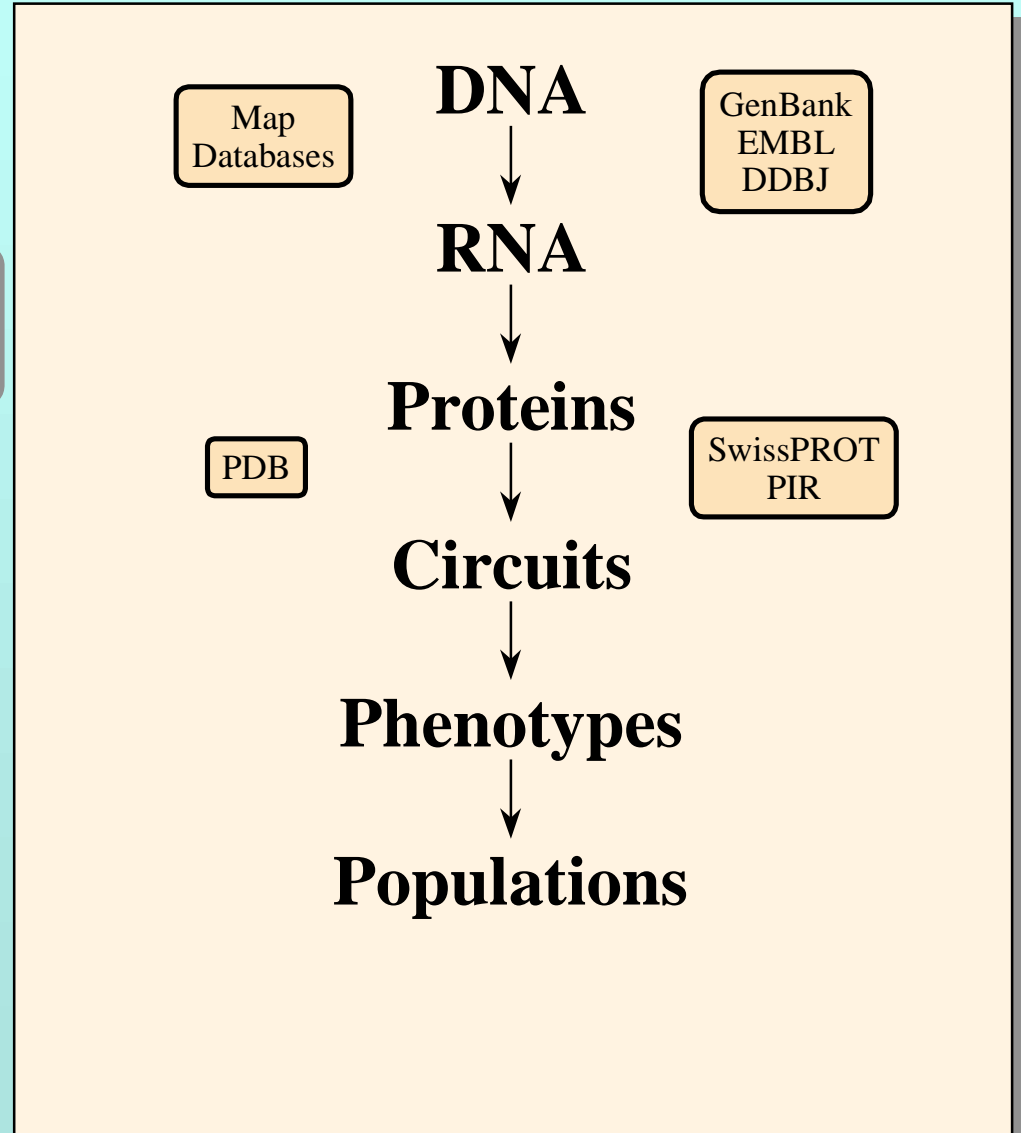
The fundamental dogma of molecular biology is that genes act to create phenotypes through a flow of information from DNA to RNA to proteins, to interactions among proteins (regulatory circuits and metabolic pathways), and ultimately to phenotypes.

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



Fundamental Dogma

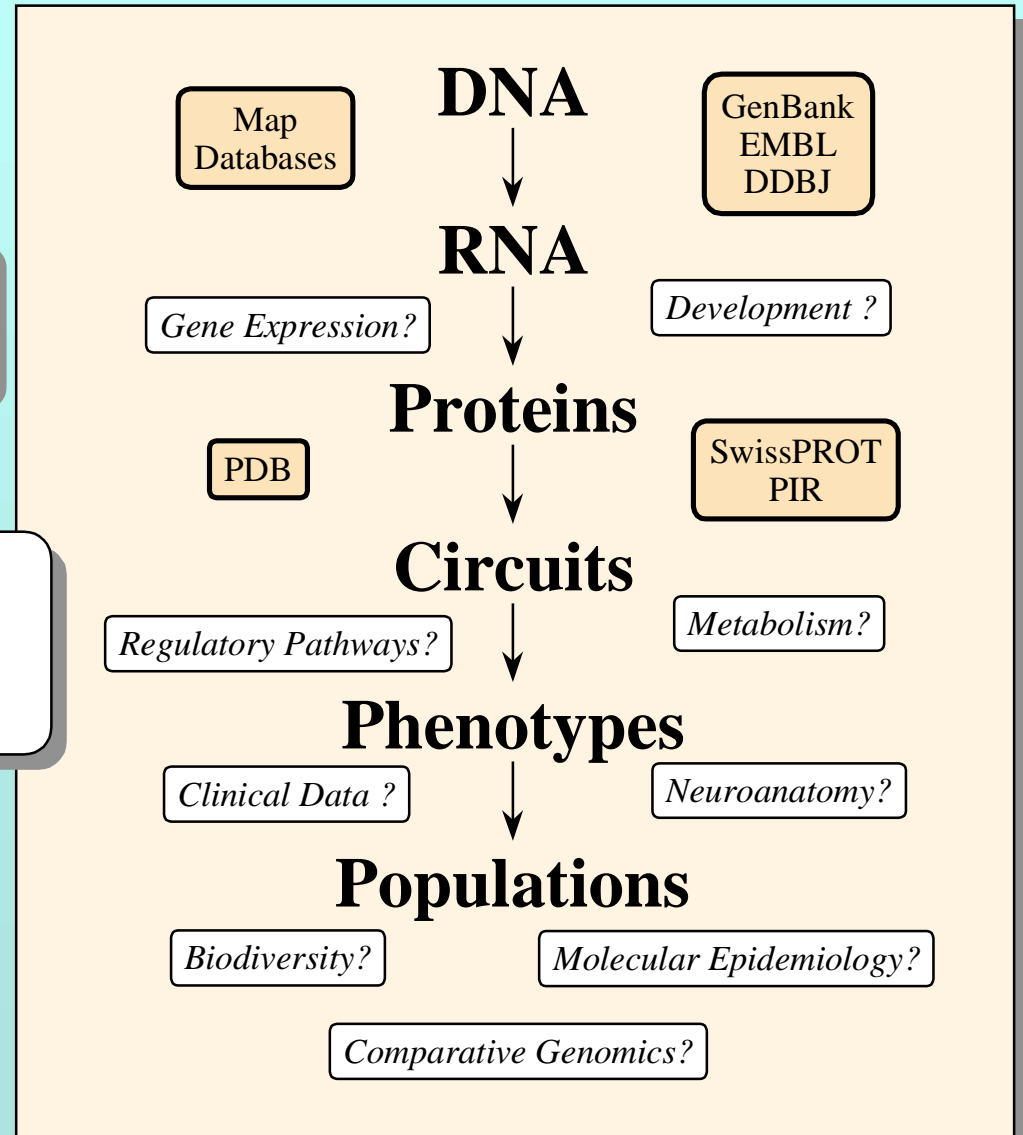
Although a few databases already exist to distribute molecular information,



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.

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

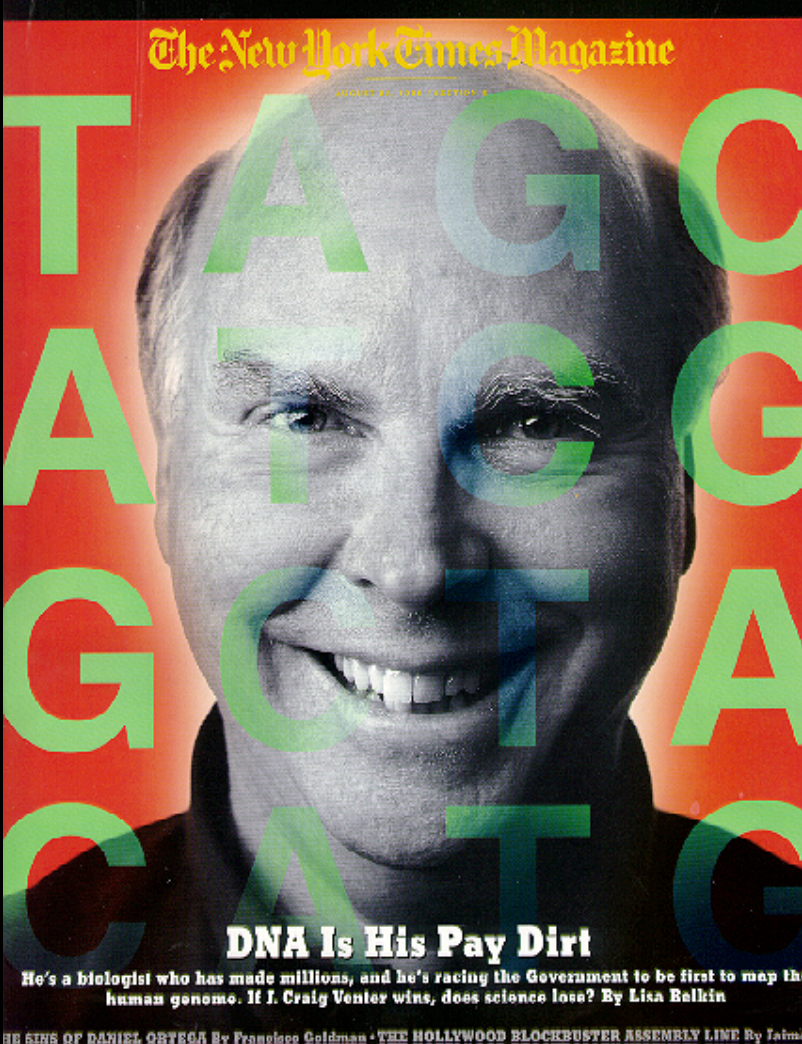
Source of estimate:

- Experience of IT-transformed industries.
- Current support for IT-rich biological research.

Private Sector

Future Reality

Who is this man?



Who is this man?




And why should you care?

The Institute for Genomic Research - Microsoft Internet Explorer

File Edit View Go Favorites Help

Back Forward Stop Refresh Home Search Favorites History Channels Fullscreen Mail Print

Address <http://www.tigr.org/>



About TIGR

TIGR Database

Software and Lab Methods

FTP Site

What's New

Conferences, Education & Training

Career Opportunities

Other Links

Welcome to
The Institute for
Genomic Research

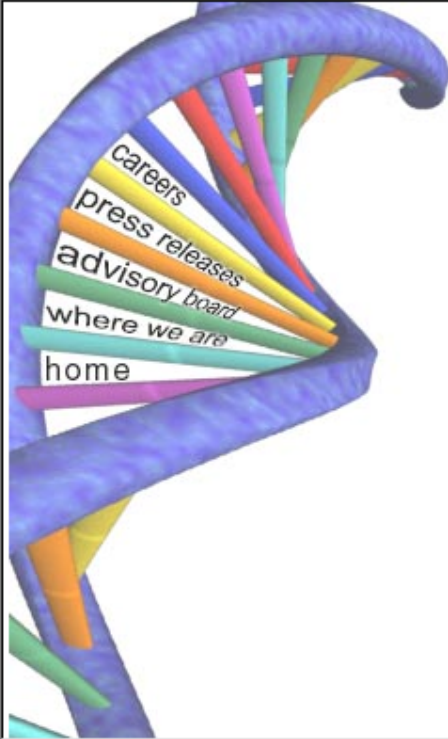
about/ Internet zone


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/





Press Releases

- Search
- PE Home
- Company Information
- Analytical Instruments
- PE Applied Biosystems
- Press Releases
- Investor Information
- Career Opportunities
- Financing
- Today's Stock Price
- Request Literature

Perkin-Elmer, Dr. J. Craig Venter, And TIGR Announce Formation Of New Genomics Company

Plan to Sequence Human Genome Within Three Years

NORWALK, CT and ROCKVILLE, MD, May 9, 1998 -- The Perkin-Elmer Corporation (NYSE:PKN), Dr. J. Craig Venter, and The Institute for Genomic Research (TIGR) announced today that they have signed letters of intent to form a new joint venture, Celera Genomics, to sequence the human genome.

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

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.mlhm.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

Millenium Pharmaceuticals

Millenium's bugs

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



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.

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.

Slides:

<http://www.esp.org/rjr/atcc.pdf>