

# **Information Technology: The Foundation for 21st Century Research**

---

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.

Science of the 21st Century will require an adequate information infrastructure. Those with access will participate in the transformation of 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

---

- Biotechnology and information technology will be the “magic” technologies of the 21st Century.

# Topics

---

- Biotechnology and information technology will be the “magic” technologies of the 21st Century.
- Moore’s Law constantly transforms IT (and everything else).

# Topics

---

- Biotechnology and information technology will be the “magic” technologies of the 21st Century.
- Moore’s Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology.

# Topics

---

- Biotechnology and information technology will be the “magic” technologies of the 21st Century.
- Moore’s Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology.
- 21st-Century biology will be based on bioinformatics.

# Topics

---

- Biotechnology and information technology will be the “magic” technologies of the 21st Century.
- Moore’s Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology.
- 21st-Century biology will be based on bioinformatics.
- **Bioinformatics is emerging as an independent discipline.**

# Topics

---

- Biotechnology and information technology will be the “magic” technologies of the 21st Century.
- Moore’s Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology.
- 21st-Century biology will be based on bioinformatics.
- Bioinformatics is emerging as an independent discipline.
- **A connected, federated information infrastructure for biology is needed.**



# Topics

---

- Biotechnology and information technology will be the “magic” technologies of the 21st Century.
- Moore’s Law constantly transforms IT (and everything else).
- Information Technology (IT) has a special relationship with biology.
- 21st-Century biology will be based on bioinformatics.
- Bioinformatics is emerging as an independent discipline.
- A connected, federated information infrastructure for biology is needed.
- Current support for public bio-information infrastructure seems inadequate.

# Introduction

---

## *Magical Technology*

# 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.

# 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.

Information technology so advanced that access to information is immediate and universal.

# Moore's Law

---

*Transforms InfoTech  
(and everything else)*

# Moore's Law: *The Statement*

---

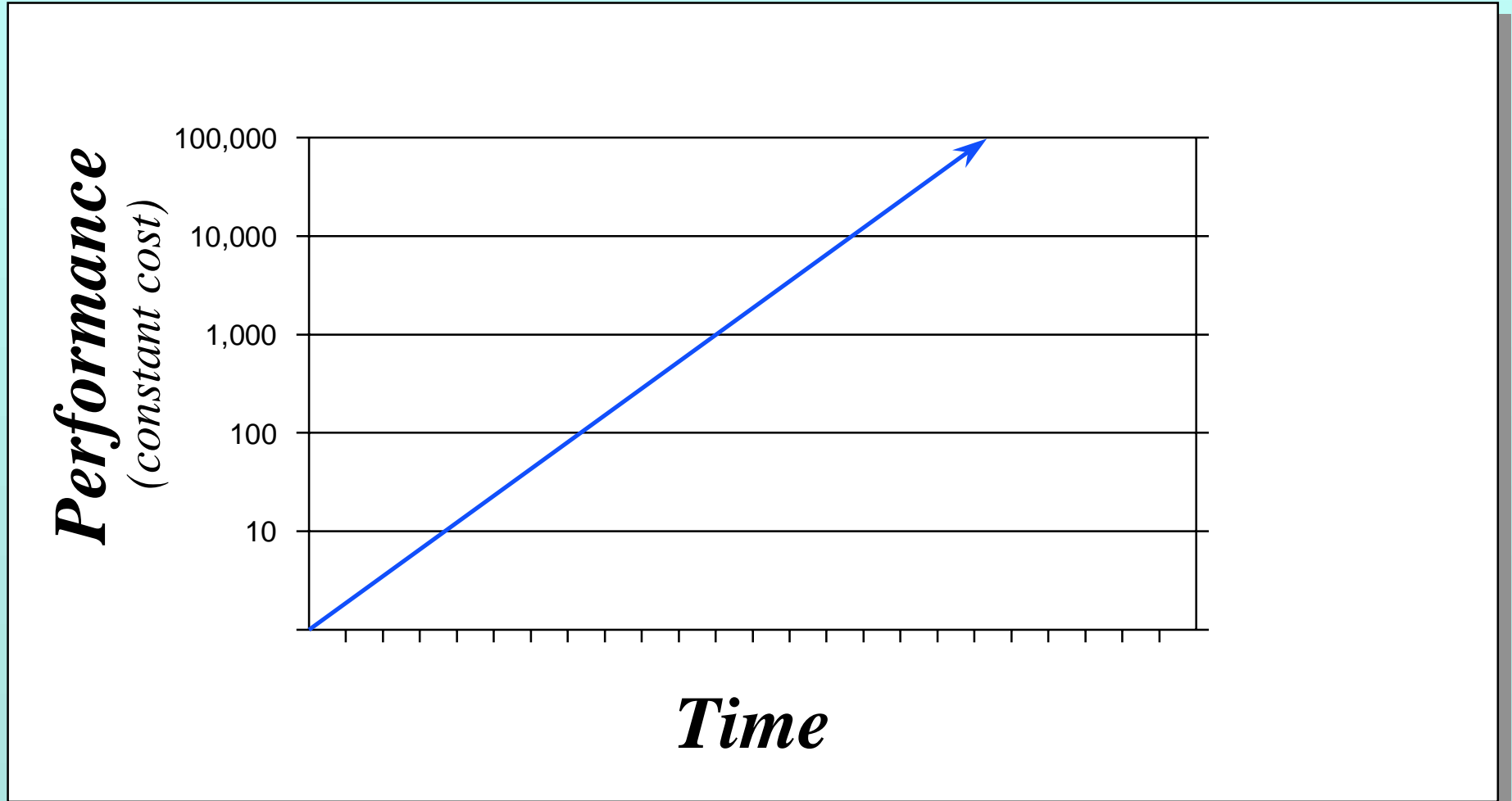
Every eighteen months, the number of transistors that can be placed on a chip doubles.

Gordon Moore, co-founder of Intel...

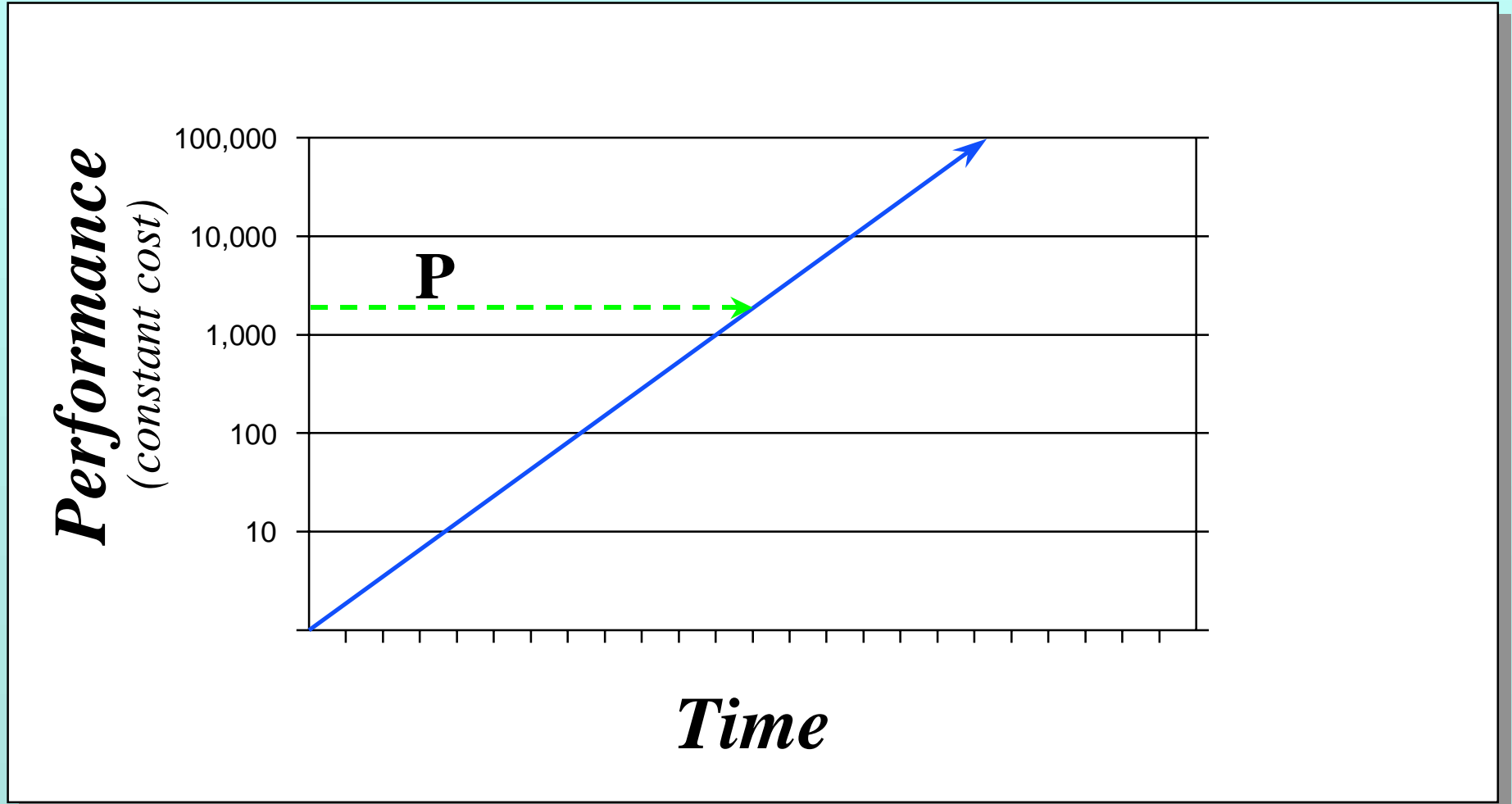


# Moore's Law: *The Effect*

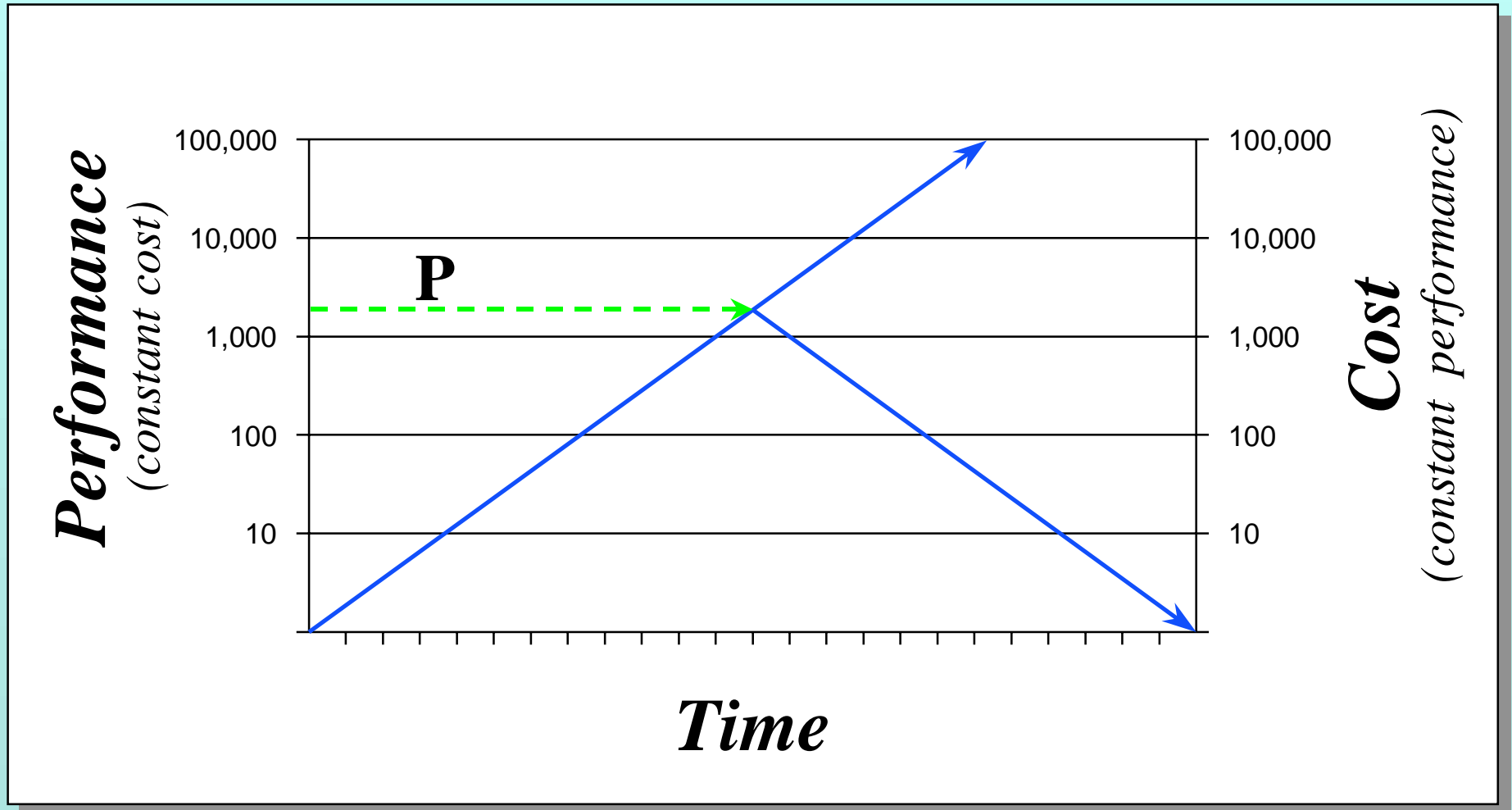
---



# 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

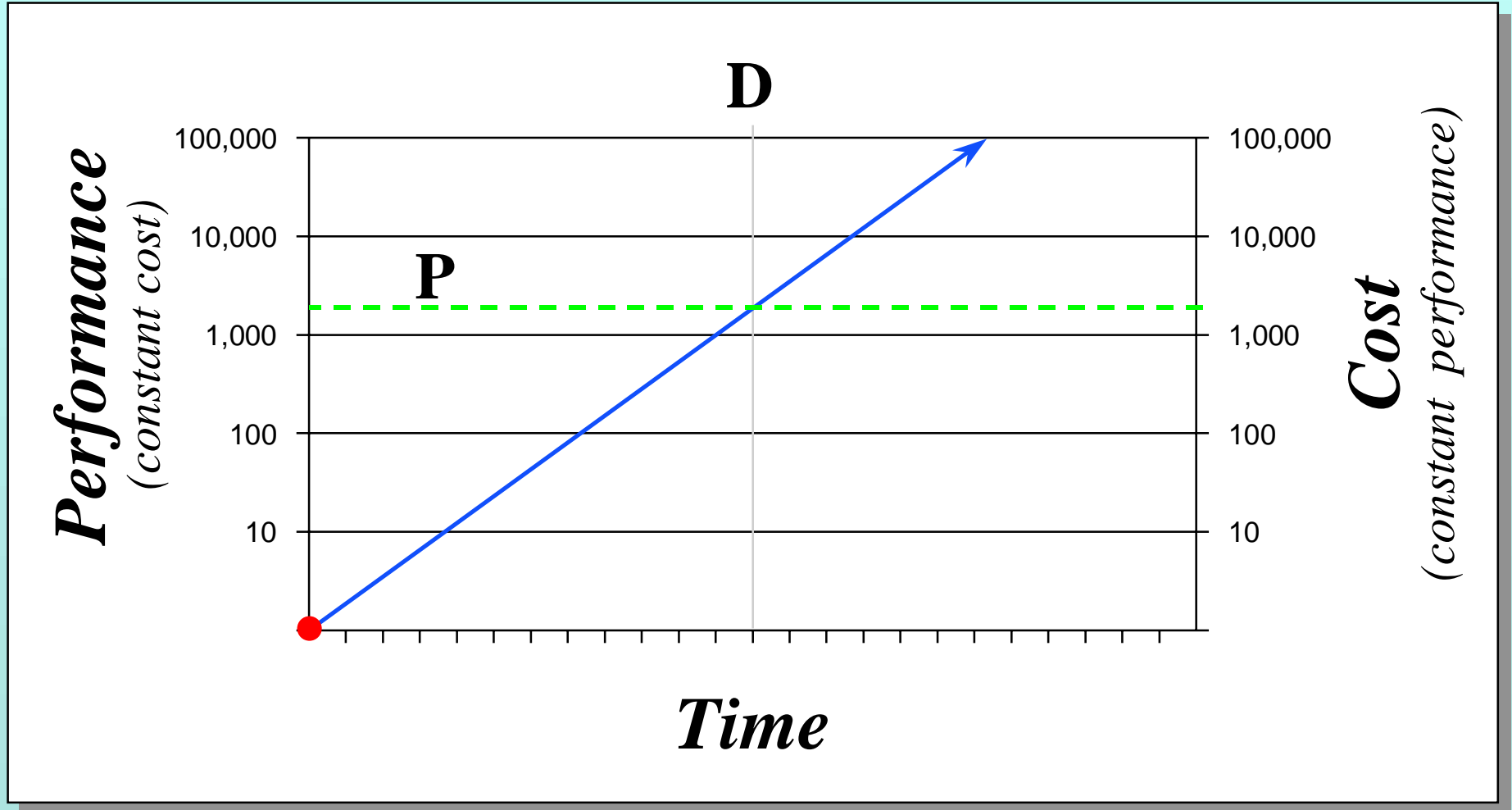
# Moore's Law: *The Effect*

---

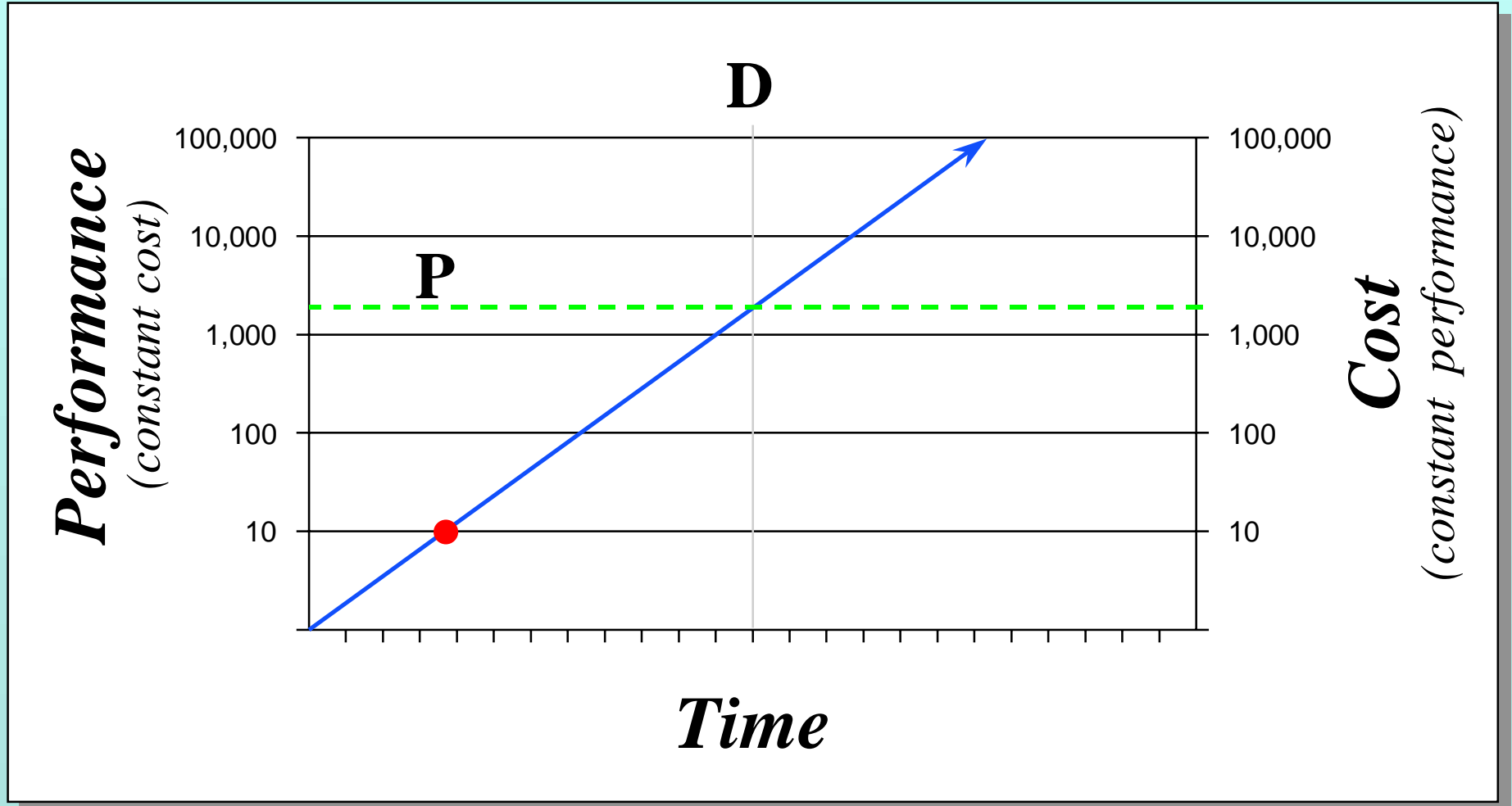
## Three Phases of Novel IT Applications

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

# Moore's Law: *The Effect*

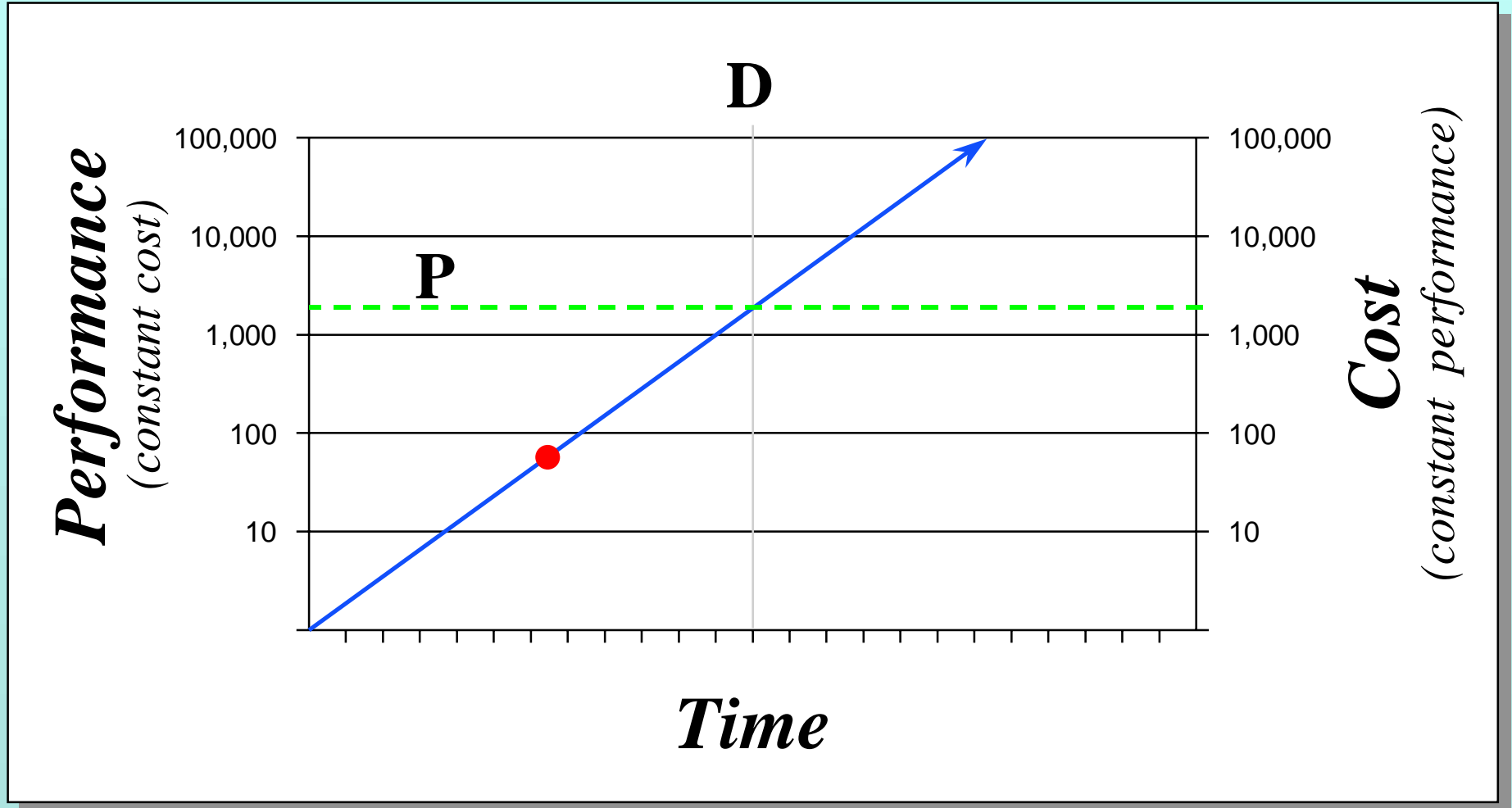


# Moore's Law: *The Effect*

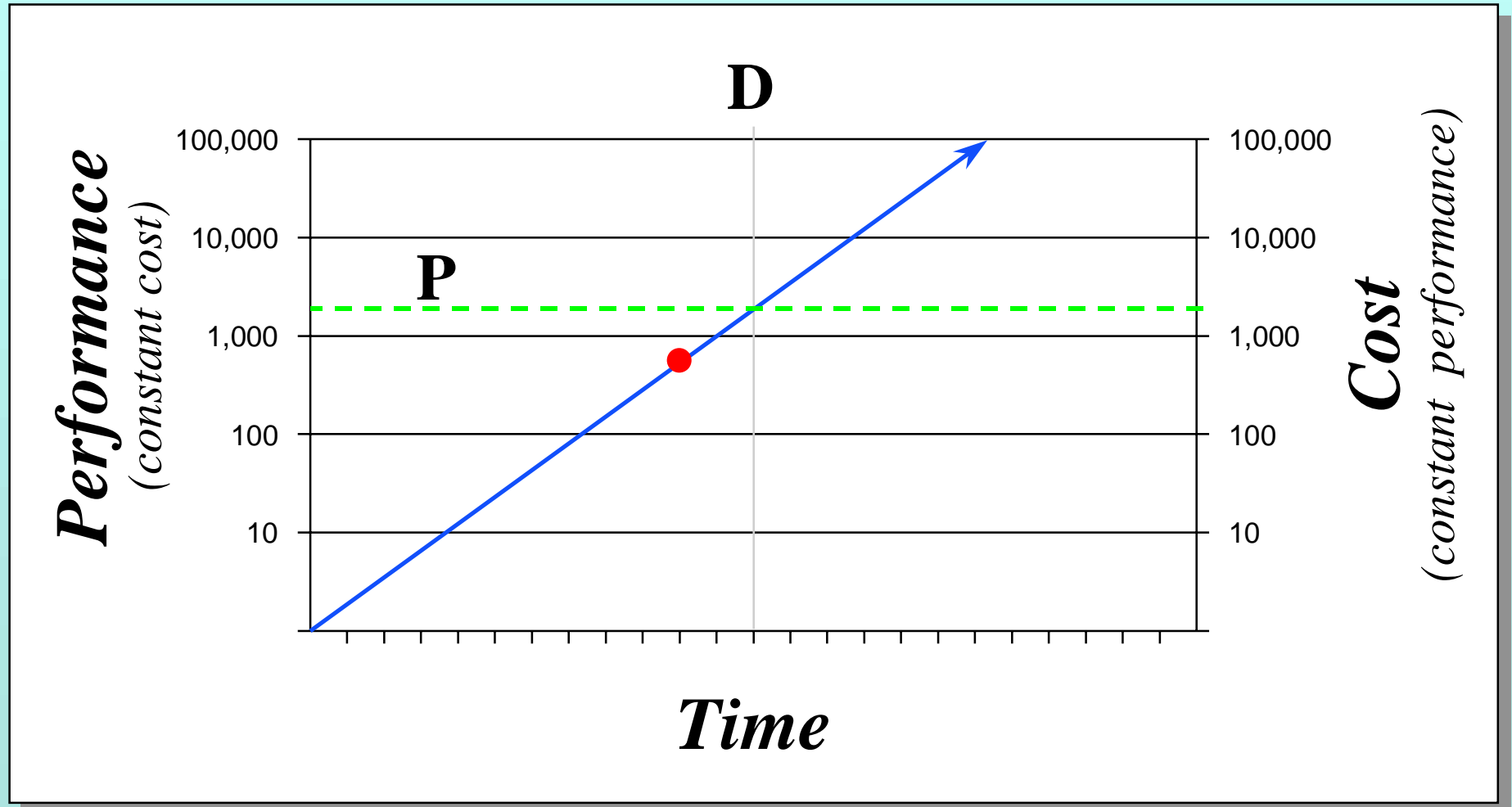




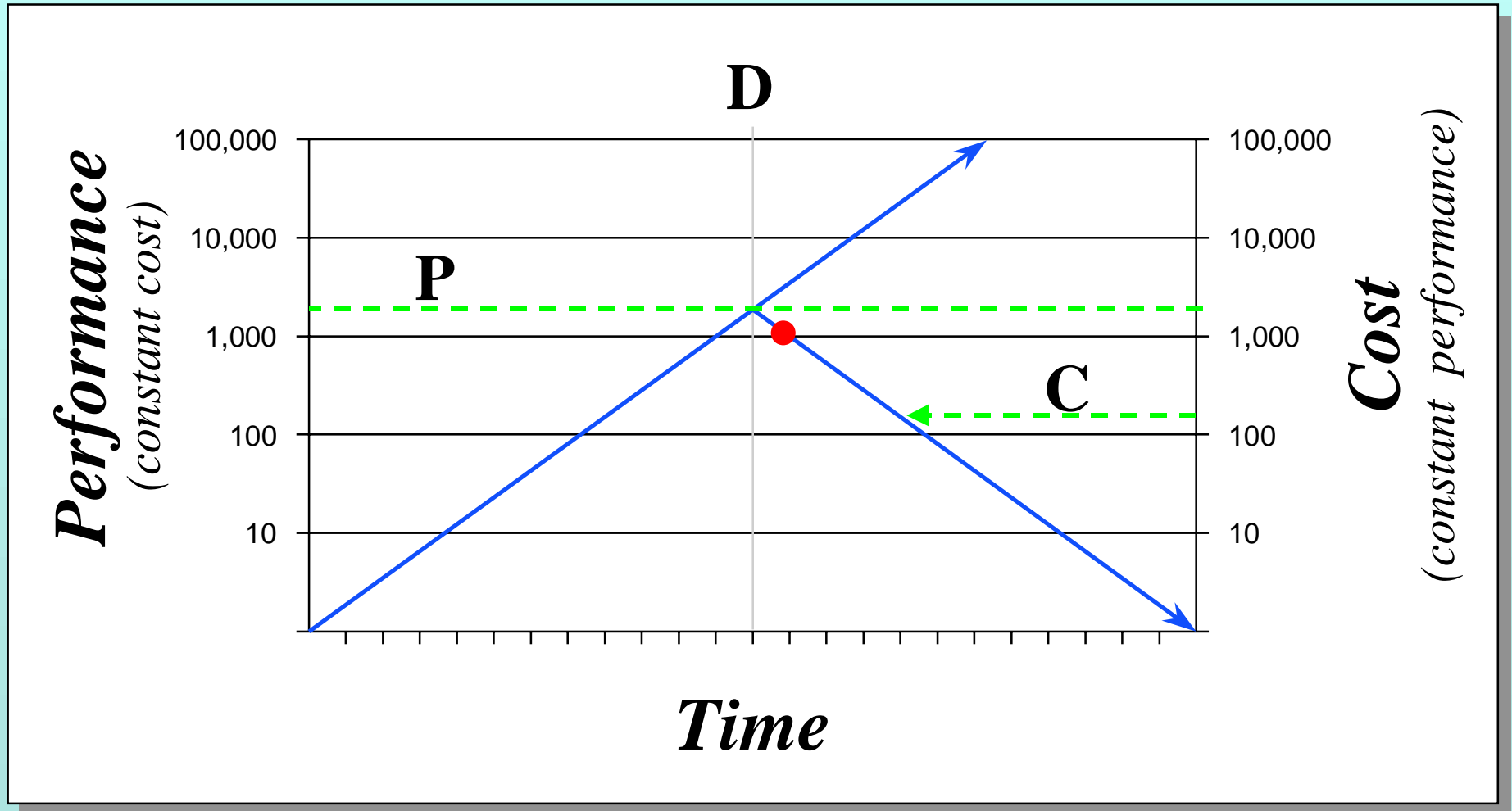
# Moore's Law: *The Effect*



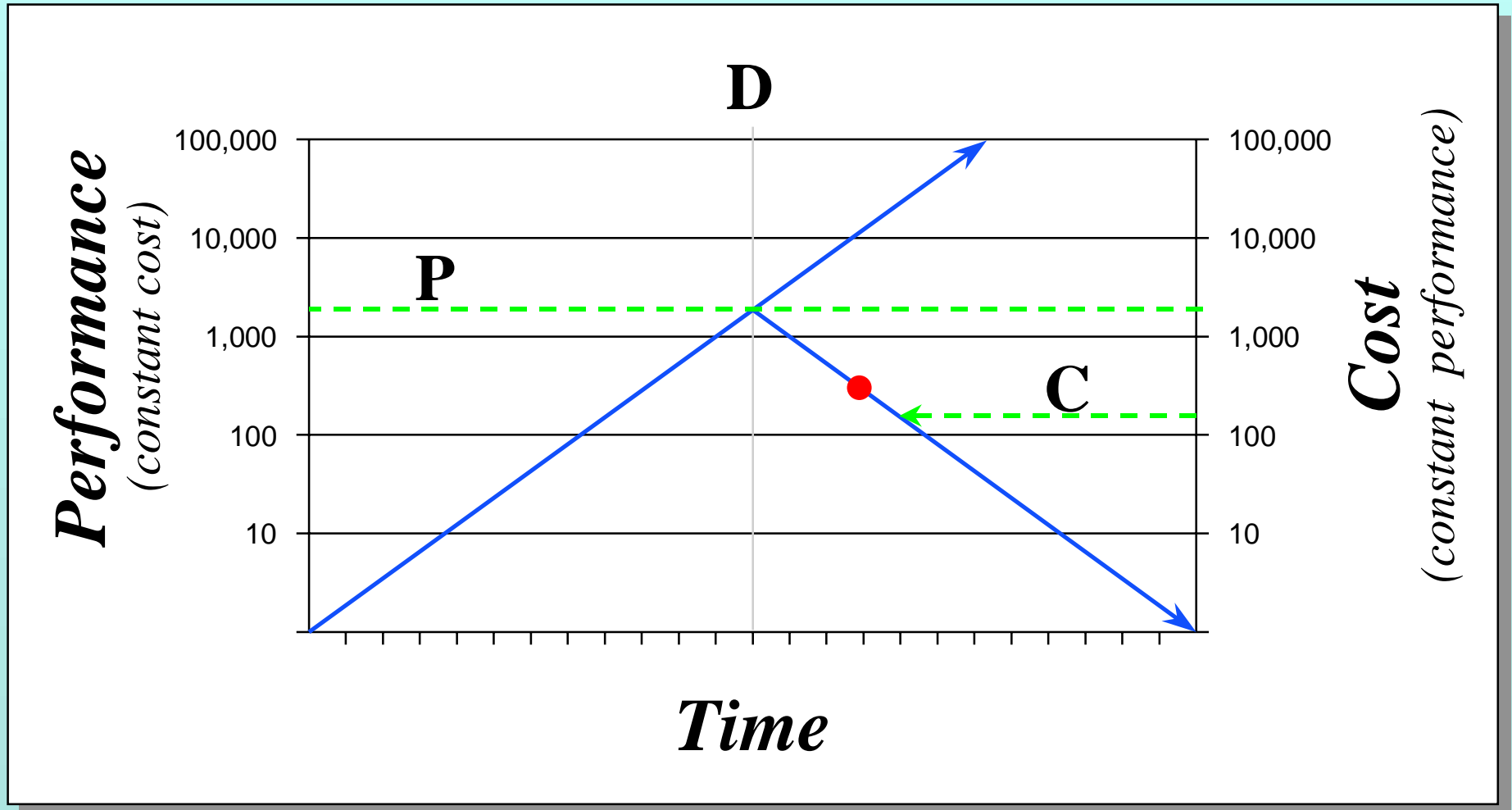
# Moore's Law: *The Effect*



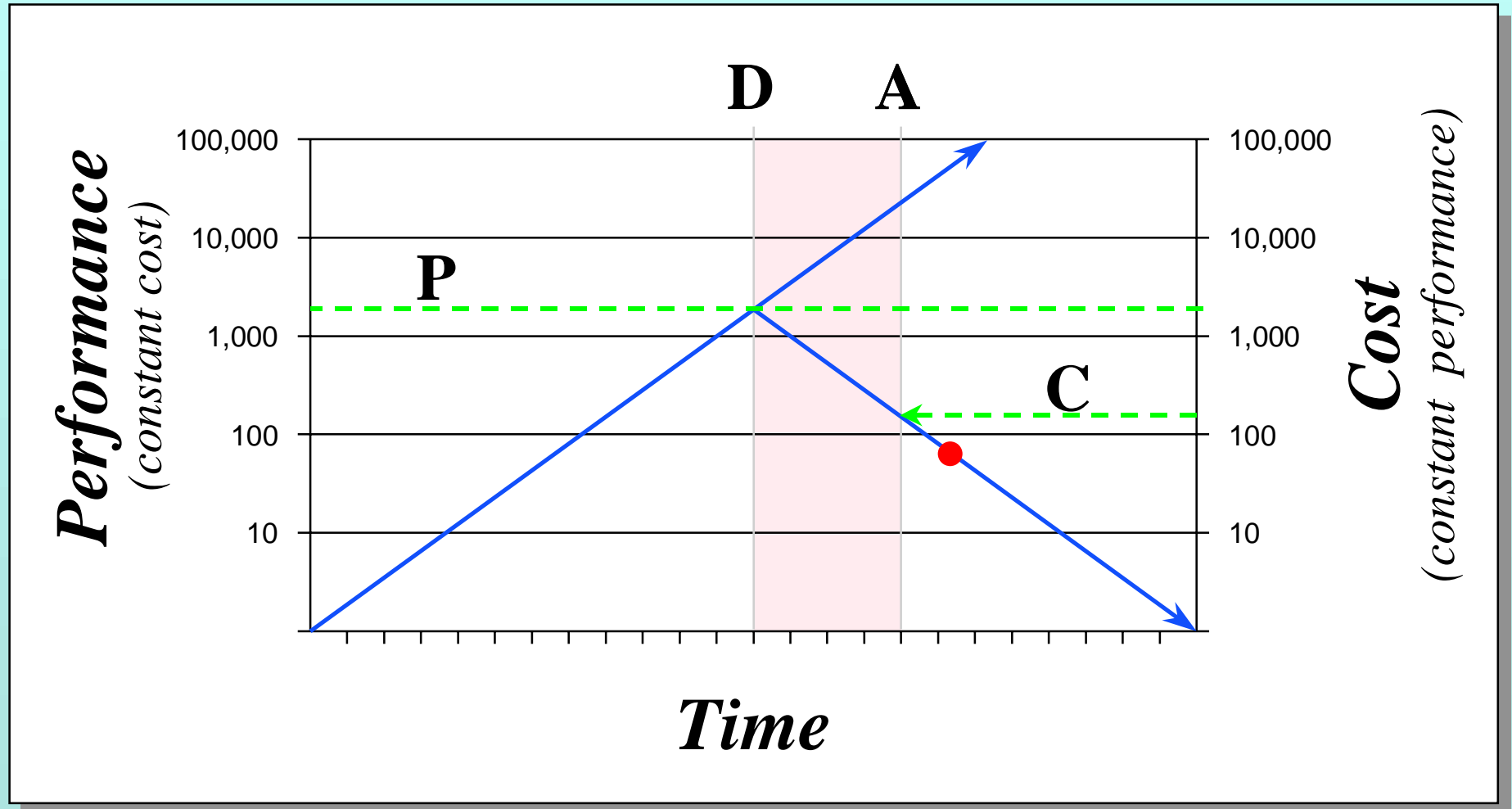
# Moore's Law: *The Effect*



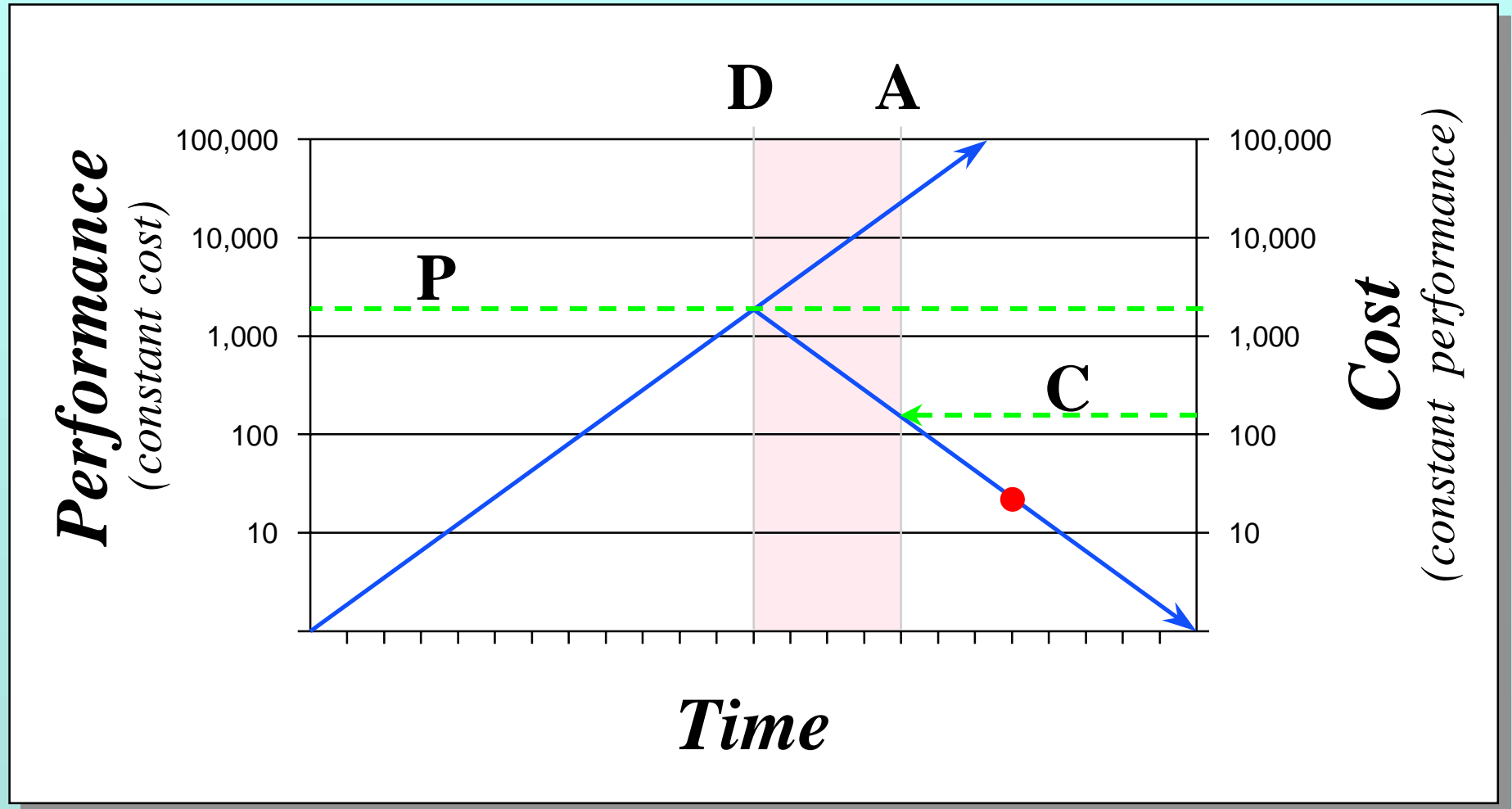
# Moore's Law: *The Effect*



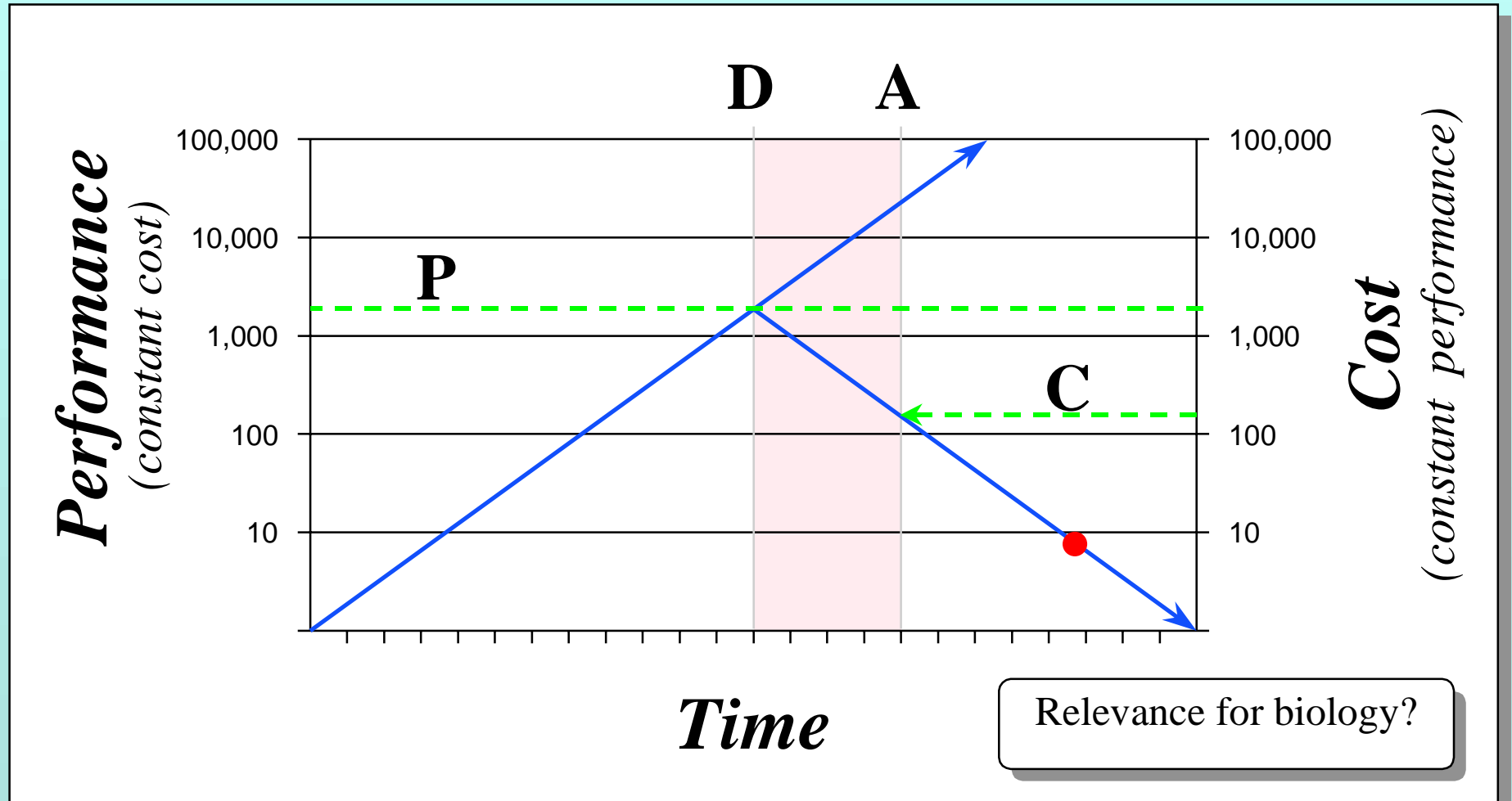
# Moore's Law: *The Effect*



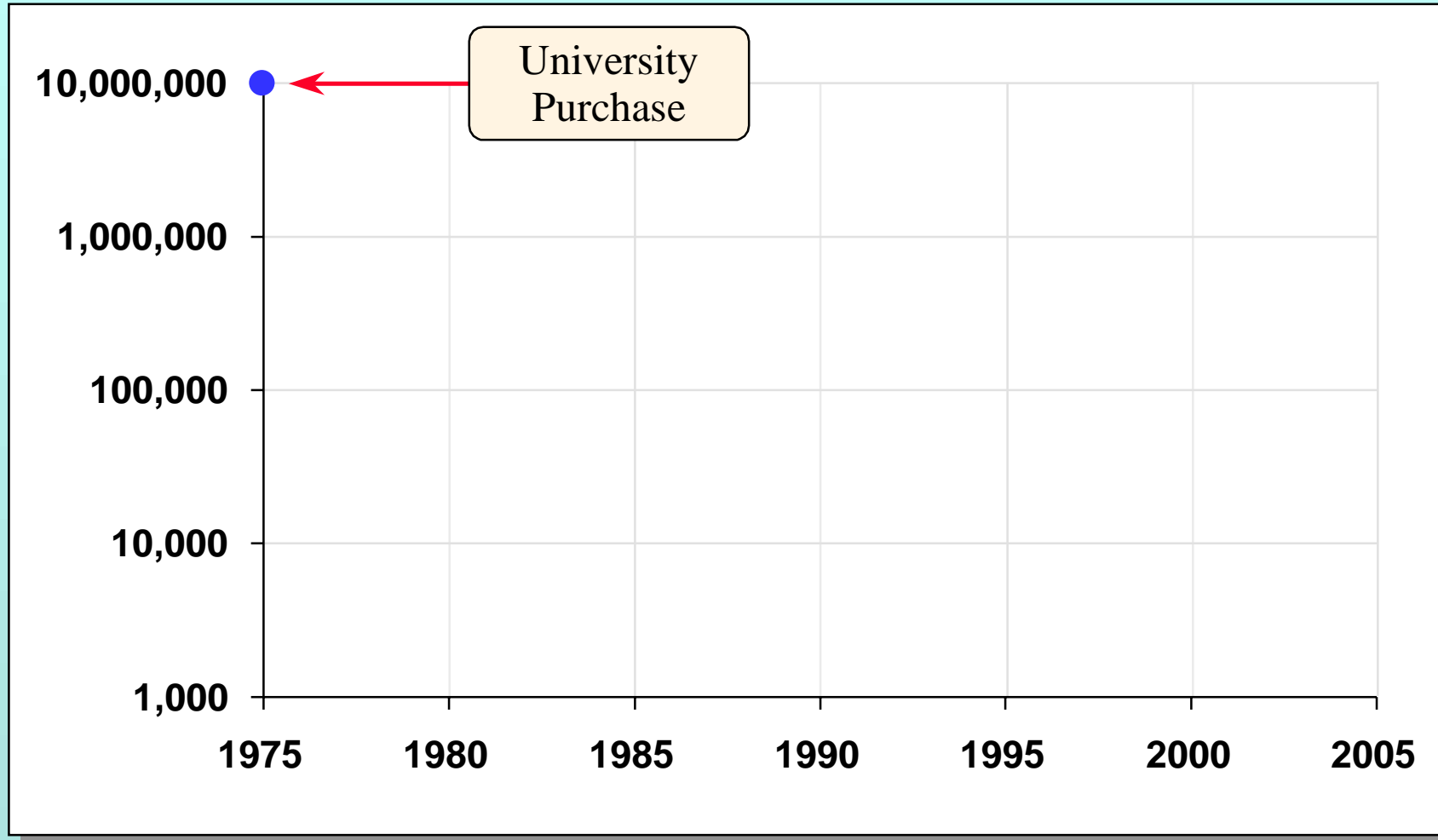
# Moore's Law: *The Effect*



# Moore's Law: *The Effect*

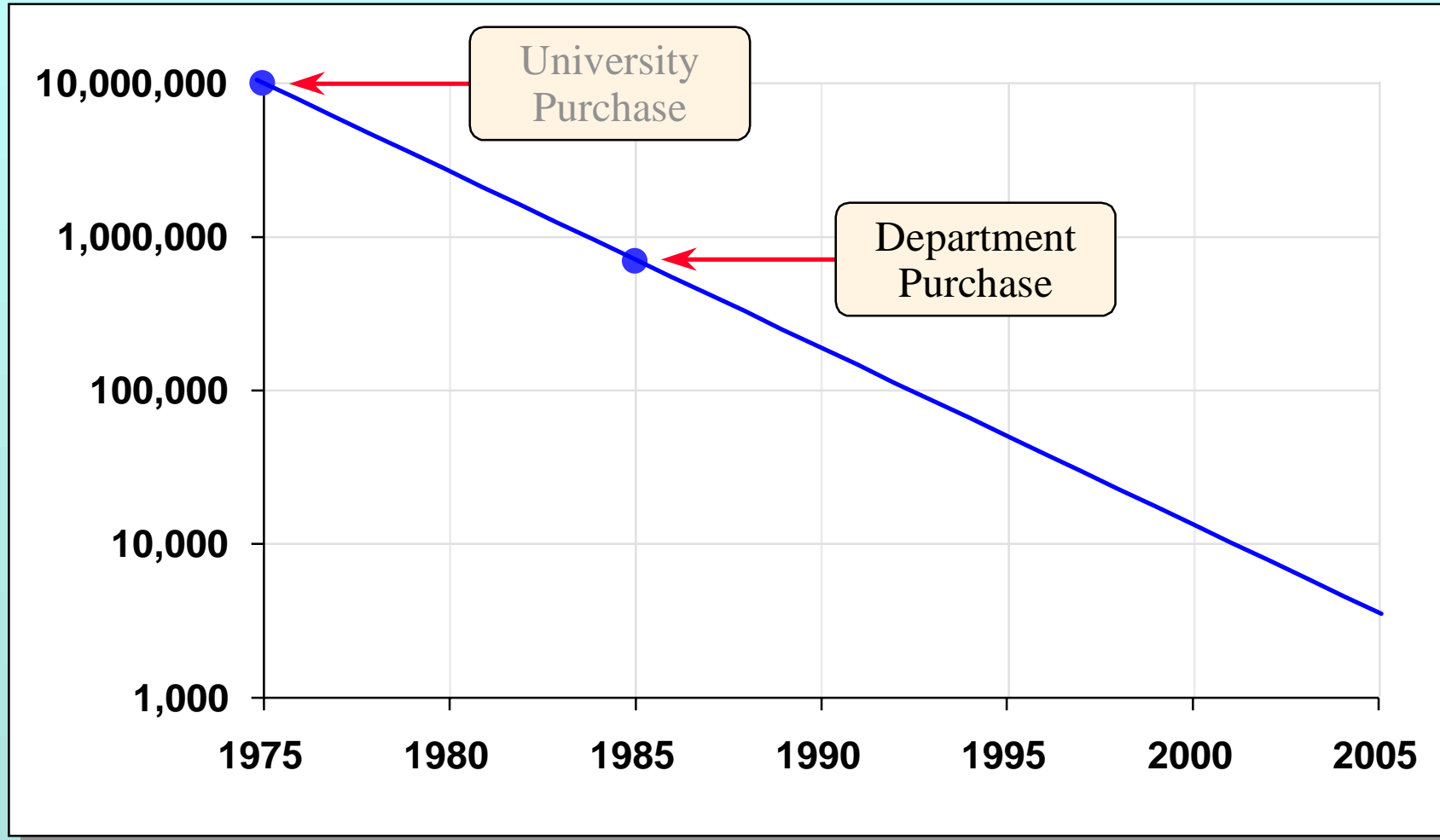


# Cost (constant performance)

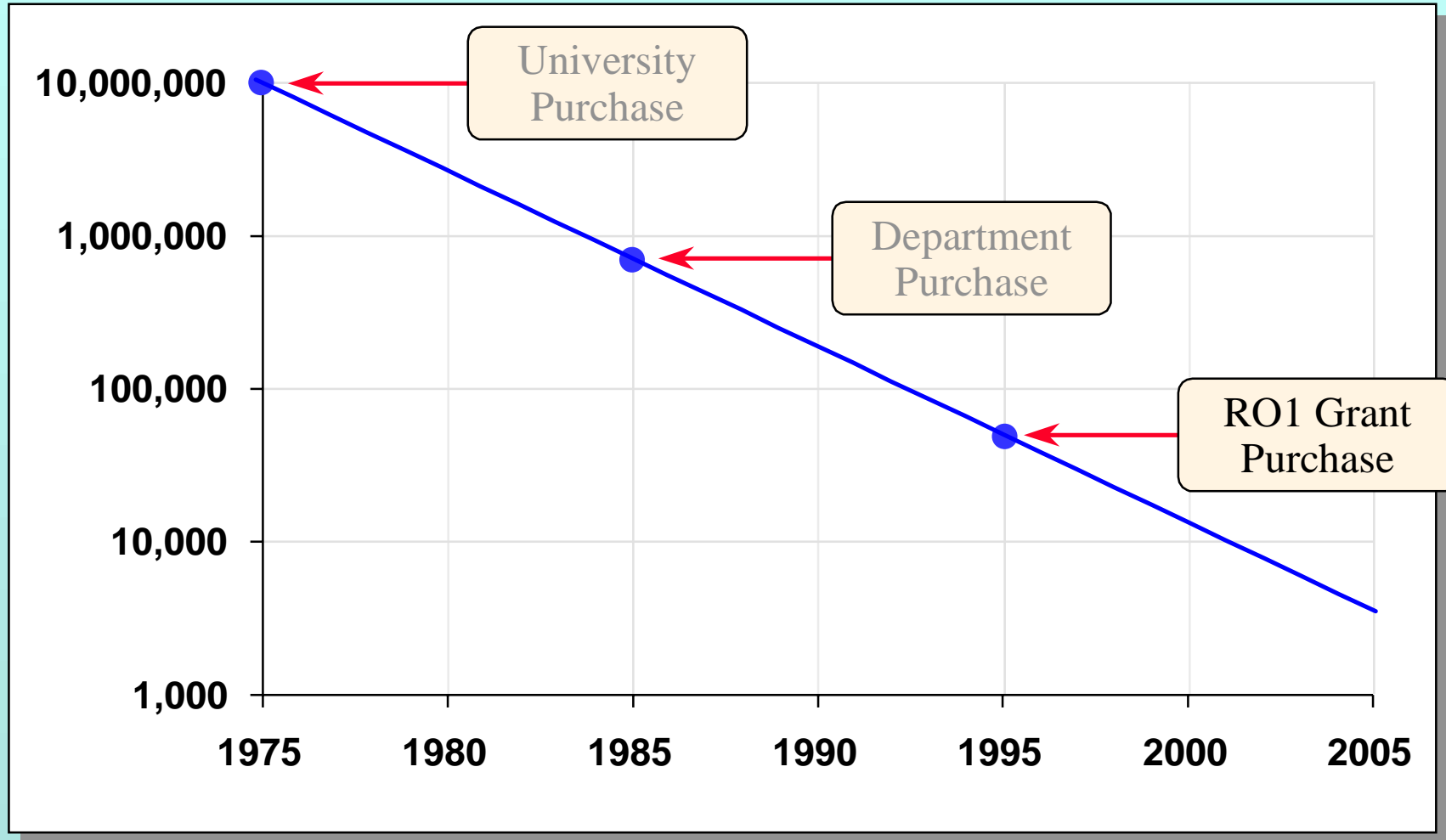




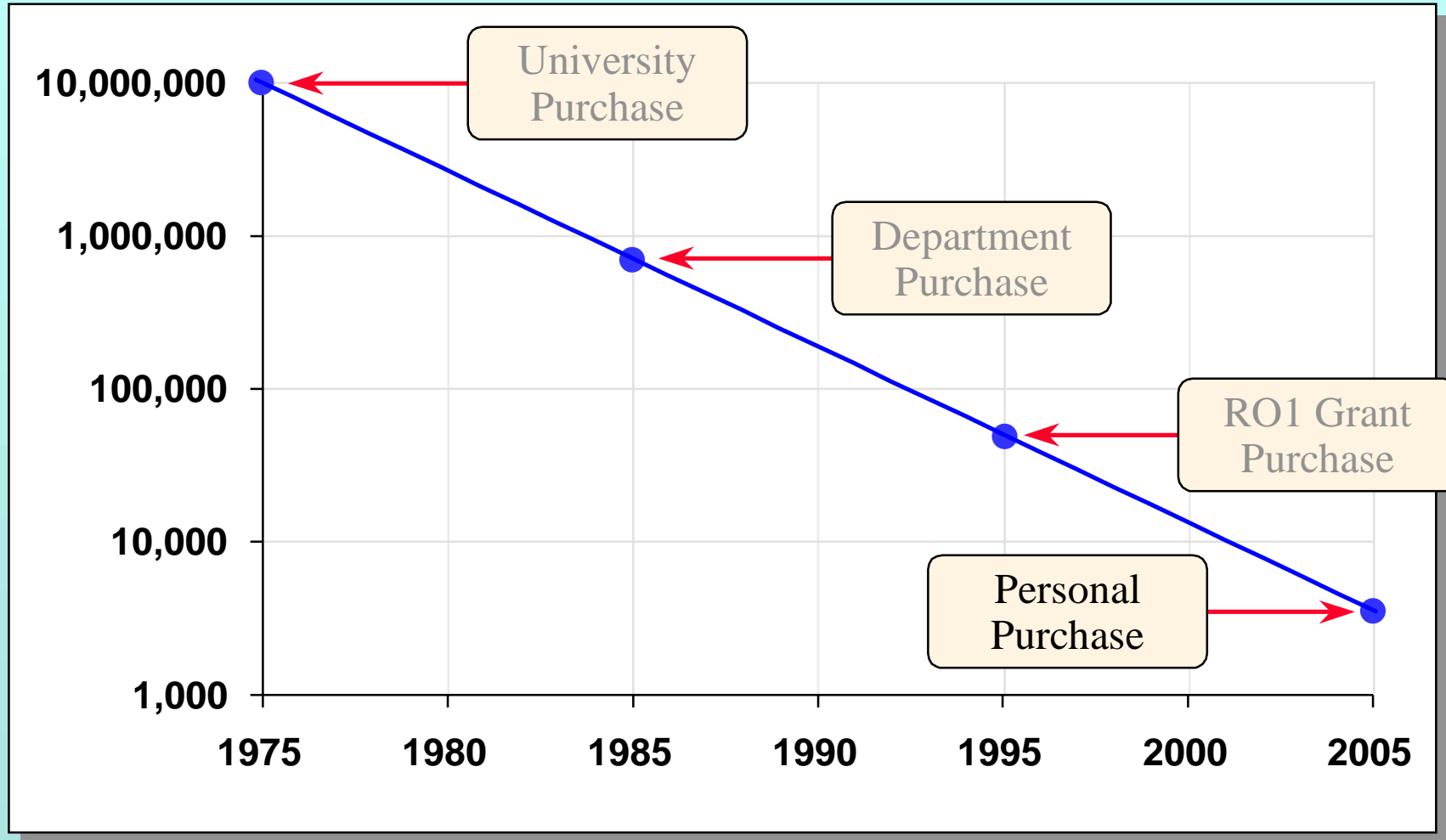
# Cost (constant performance)



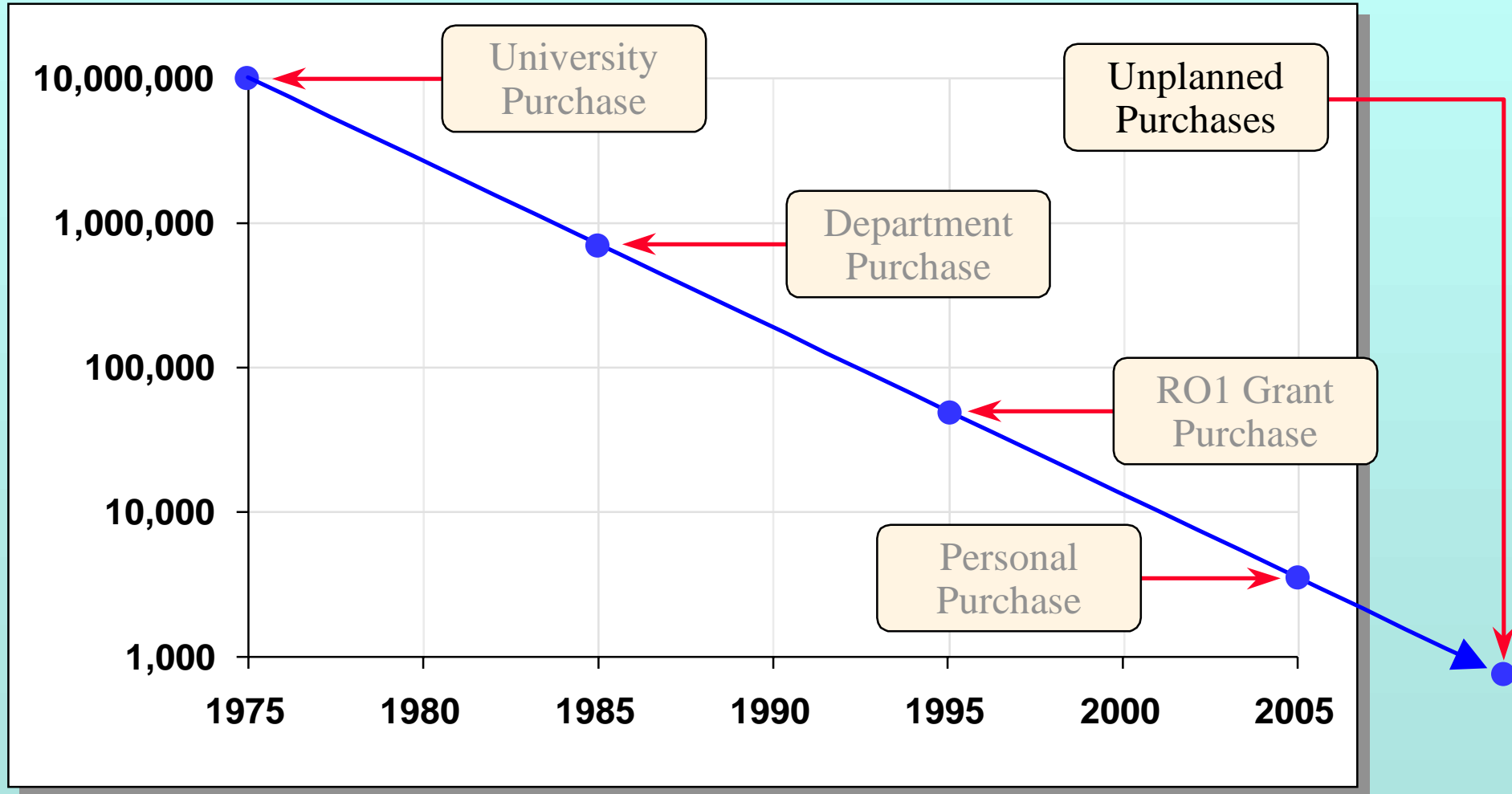
# Cost (constant performance)



# 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

---

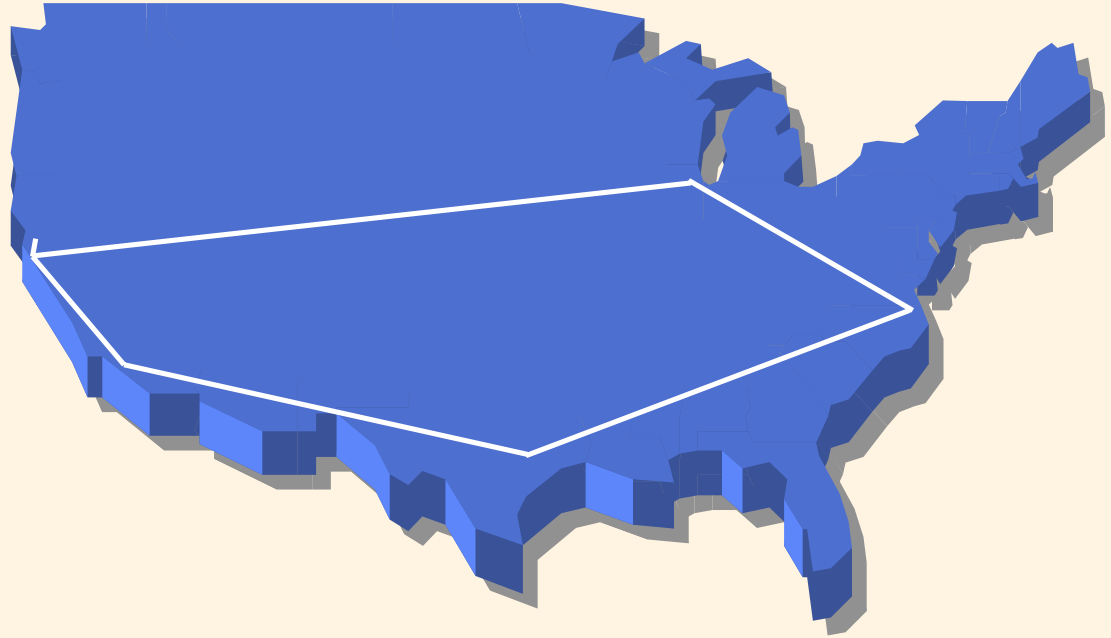
[The] chromosomes ... contain in some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state. ... [By] code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, 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

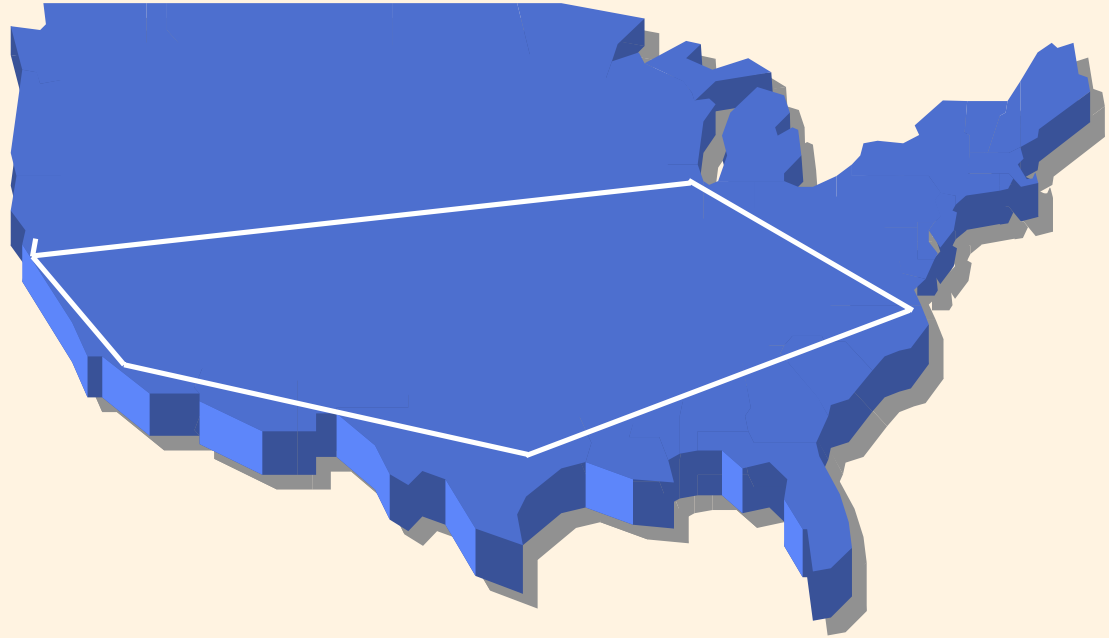
---

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

---

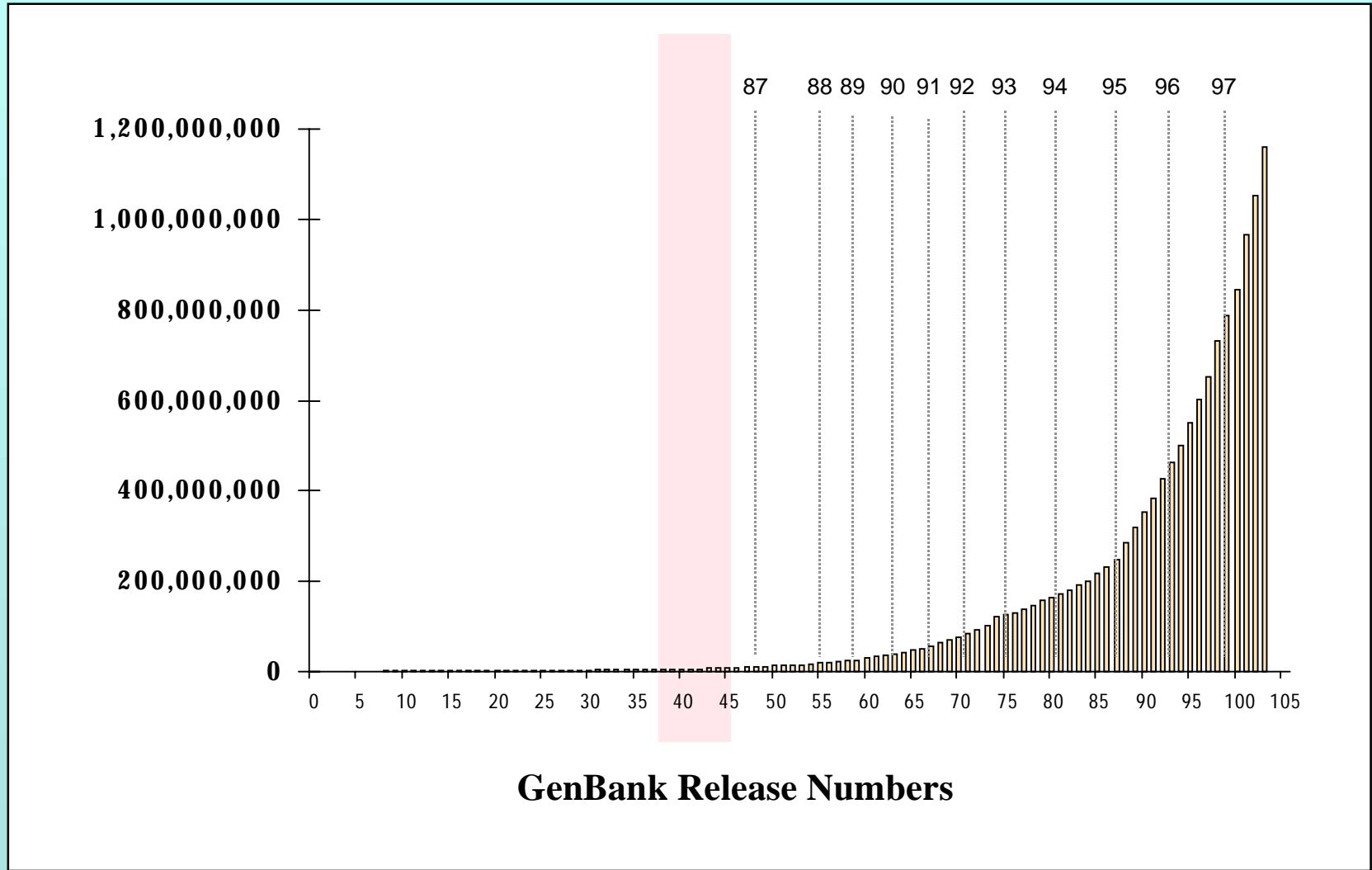
Progress towards all of the [Genome Project] goals will require the establishment of well-funded centralized facilities, including a stock center for the cloned DNA fragments generated in the mapping and sequencing effort and a data center for the computer-based collection and distribution of large amounts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

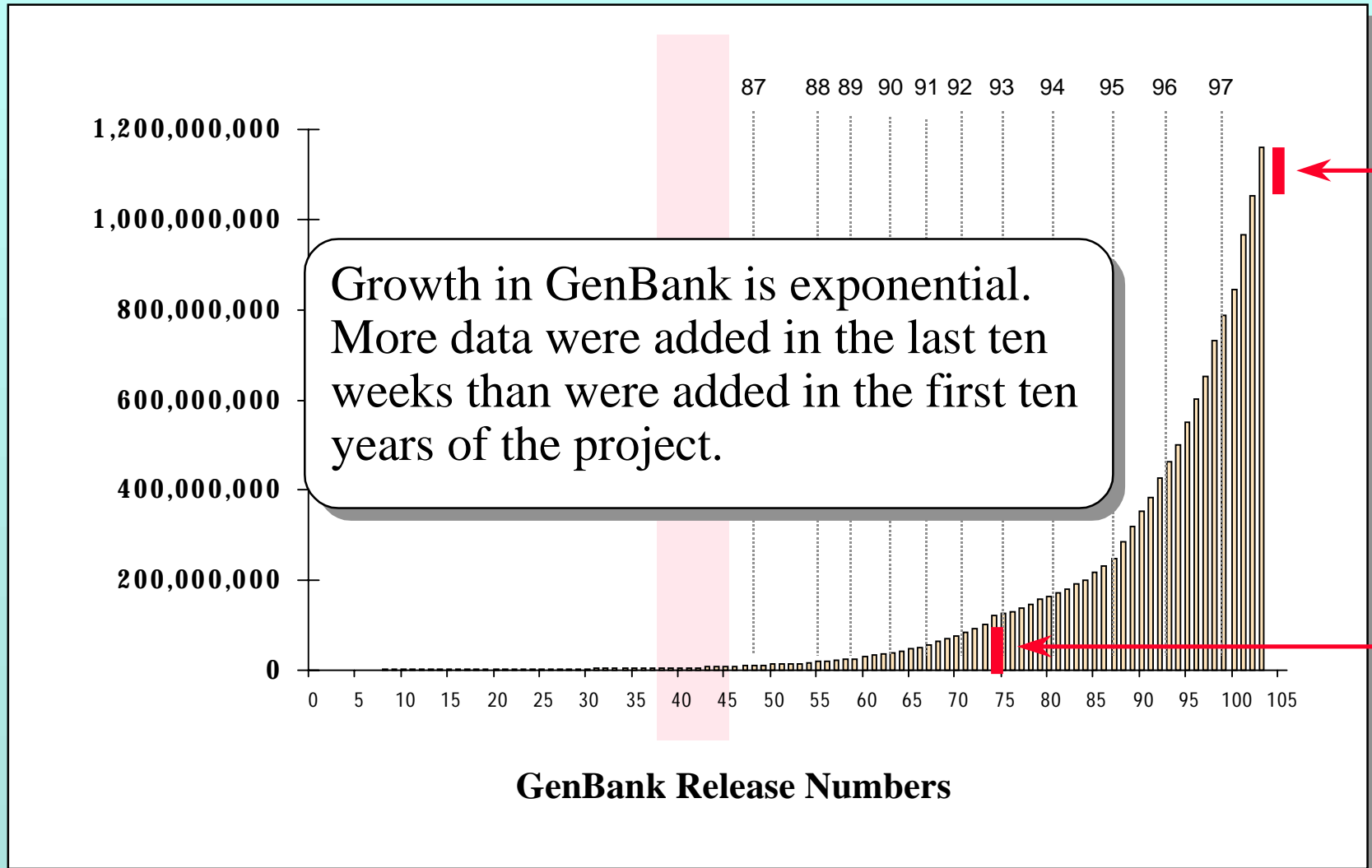
DIVISION	Entries	Per Cent	Base Pairs	Per Cent
Phage Sequences (PHG)	1,313	0.074%	2,138,810	0.184%
Viral Sequences (VRL)	45,355	2.568%	44,484,848	3.834%
Bacteria (BCT)	38,023	2.153%	88,576,641	7.634%
<b>Plant, Fungal, and Algal Sequences (PLN)</b>	<b>44,553</b>	<b>2.523%</b>	<b>92,259,434</b>	<b>7.951%</b>
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%
<b>High-Throughput Genome Sequences (HTG)</b>	<b>1,120</b>	<b>0.063%</b>	<b>72,064,395</b>	<b>6.211%</b>
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%
<b>EST1-17</b>	<b>1,269,737</b>	<b>71.905%</b>	<b>466,634,317</b>	<b>40.217%</b>
<b>TOTALS</b>	<b>1,765,847</b>	<b>100.000%</b>	<b>1,160,300,687</b>	<b>100.000%</b>

# Base Pairs in GenBank

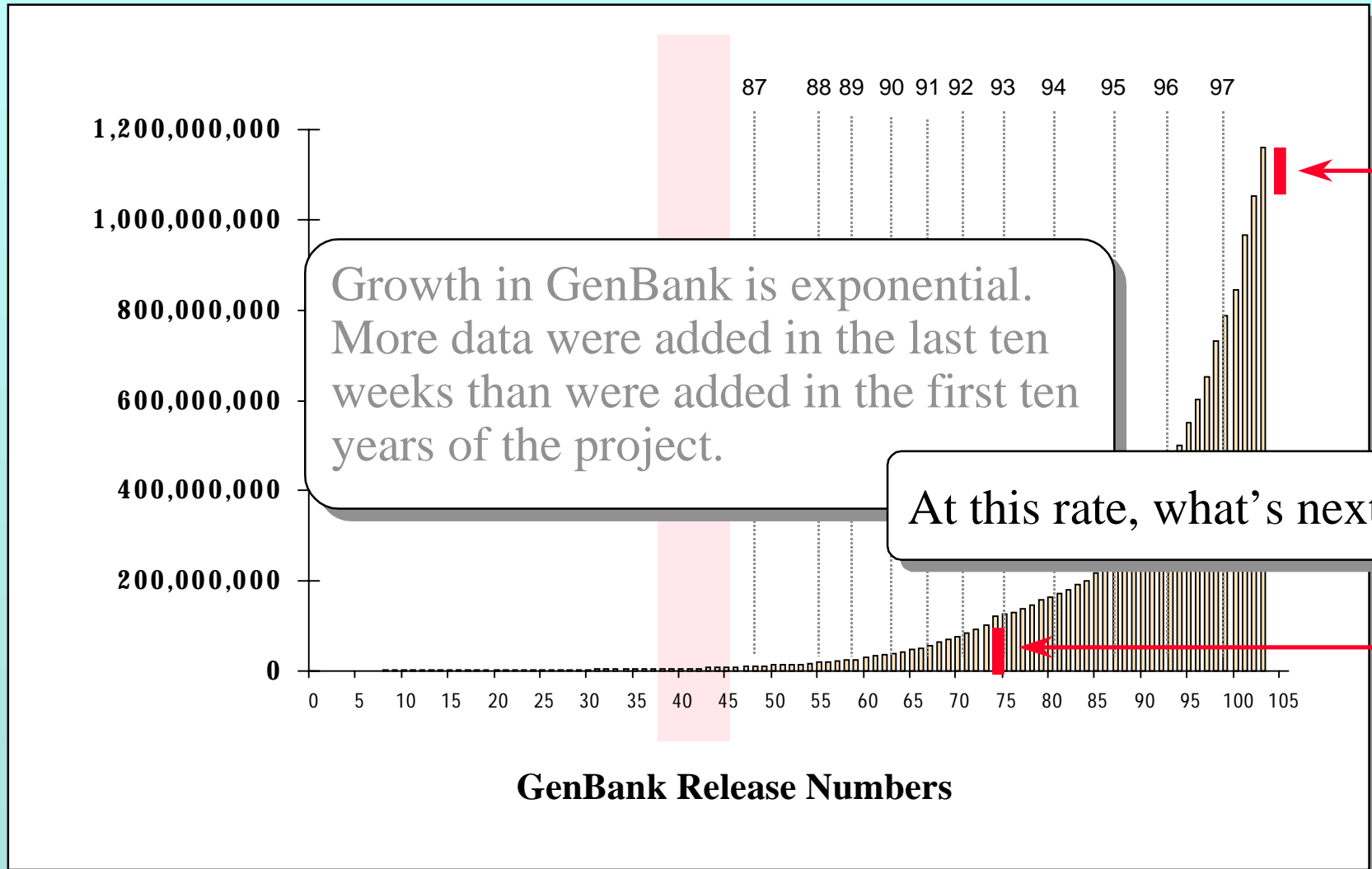




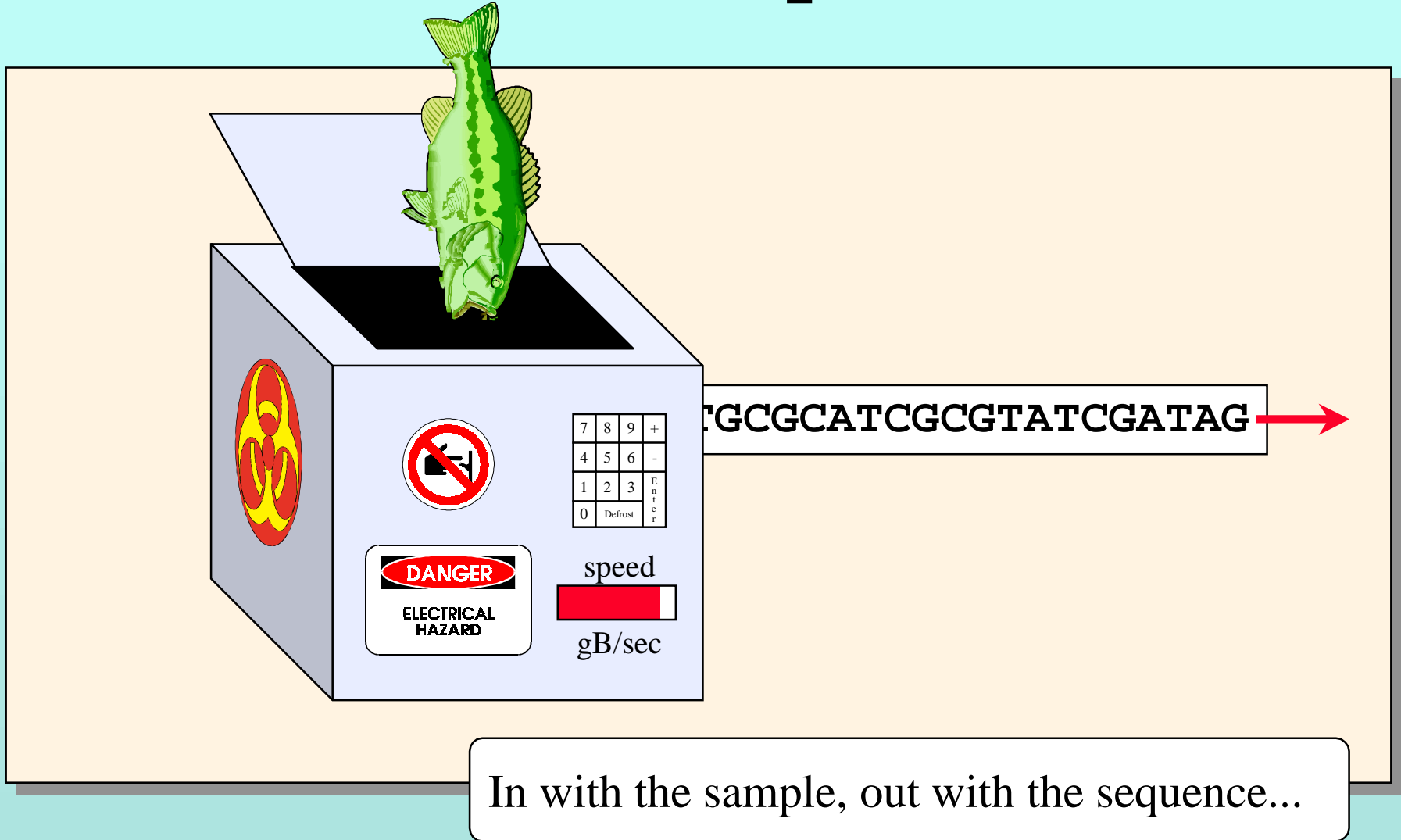
# Base Pairs in GenBank



# Base Pairs in GenBank



# *ABI Bass-o-Matic Sequencer*



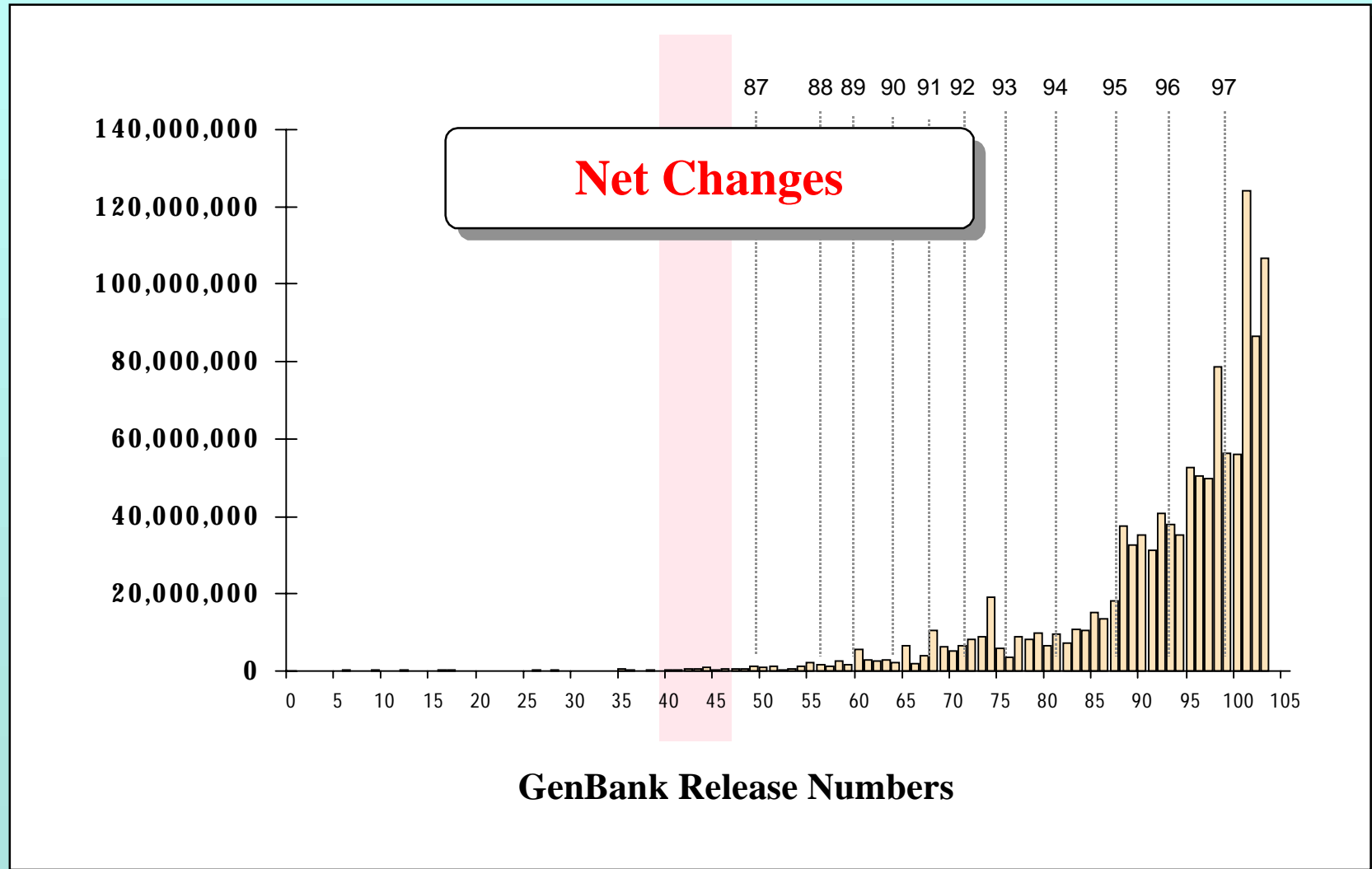
# What's Really Next

---

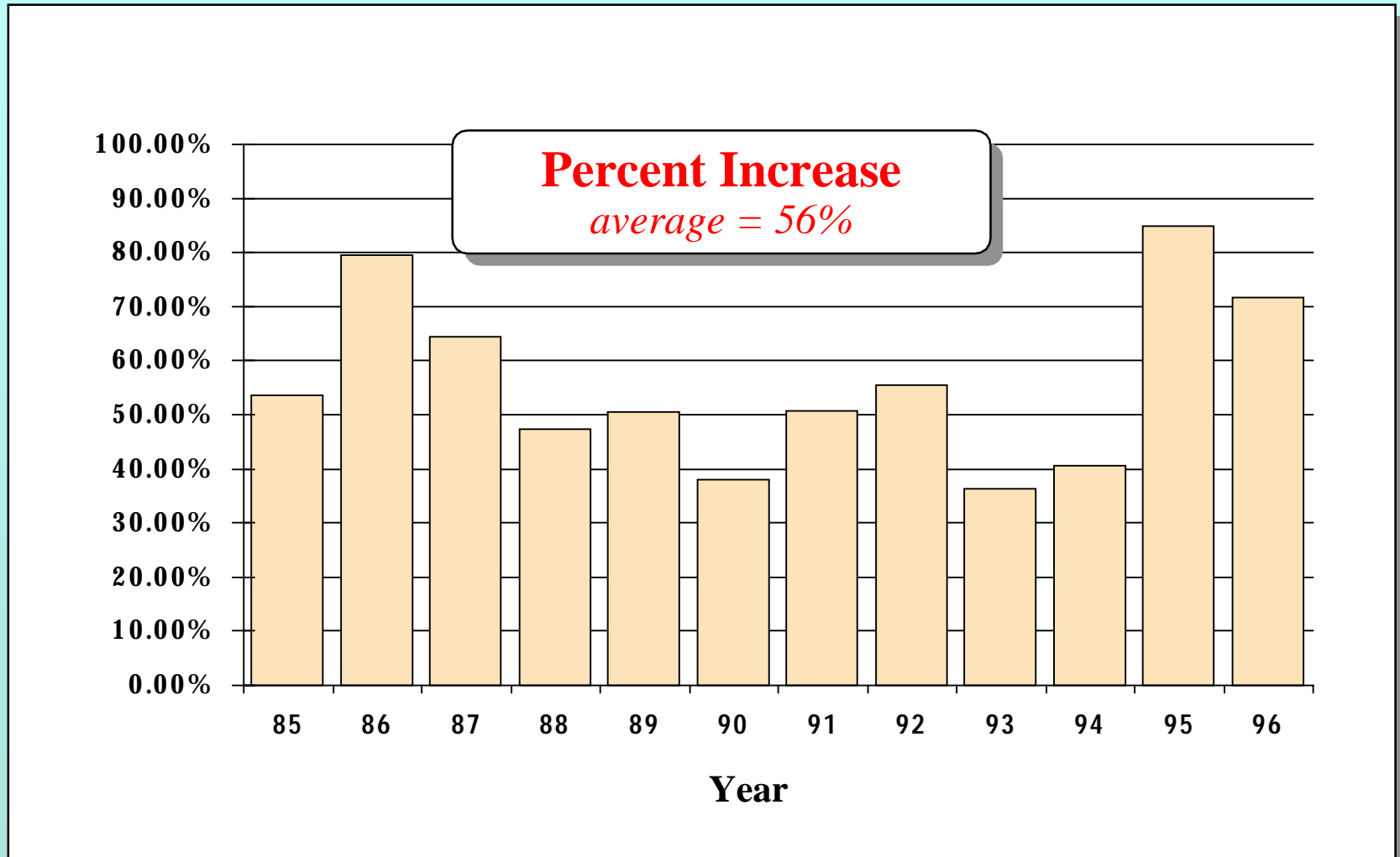
The post-genome era in biological research will take for granted ready access to huge amounts of genomic data.

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

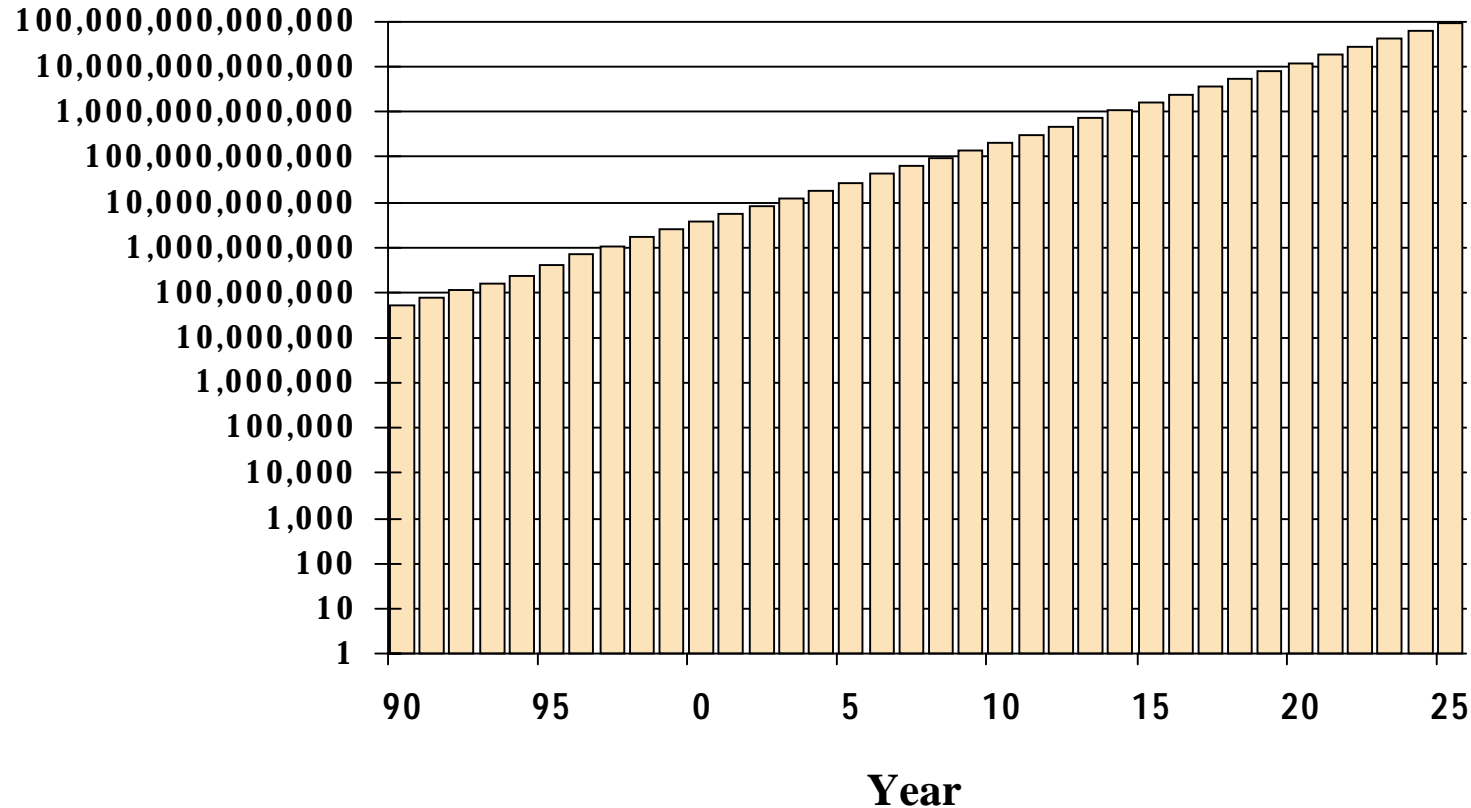
# Base Pairs in GenBank



# Base Pairs in GenBank (*Percent Increase*)

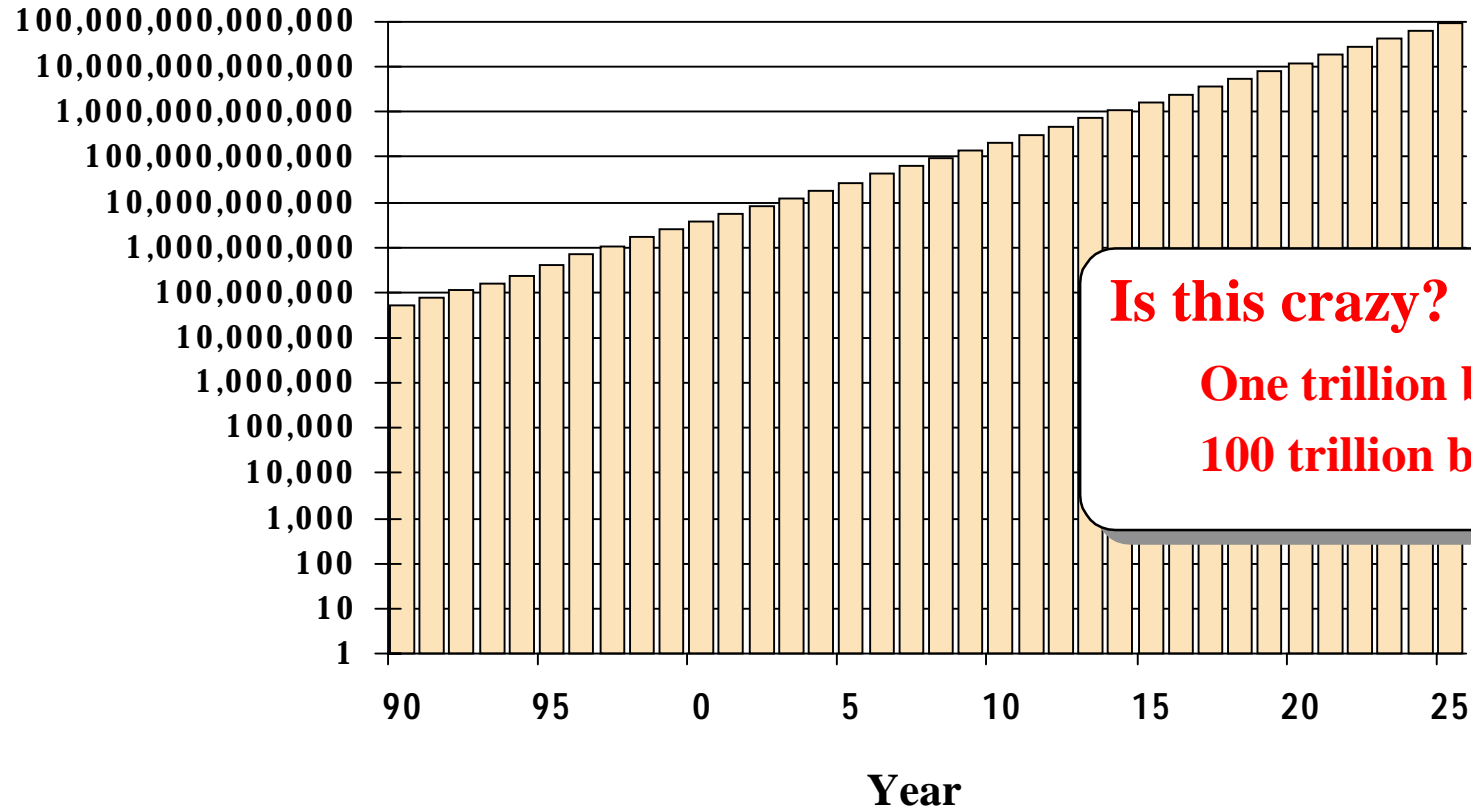


# Projected Base Pairs



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

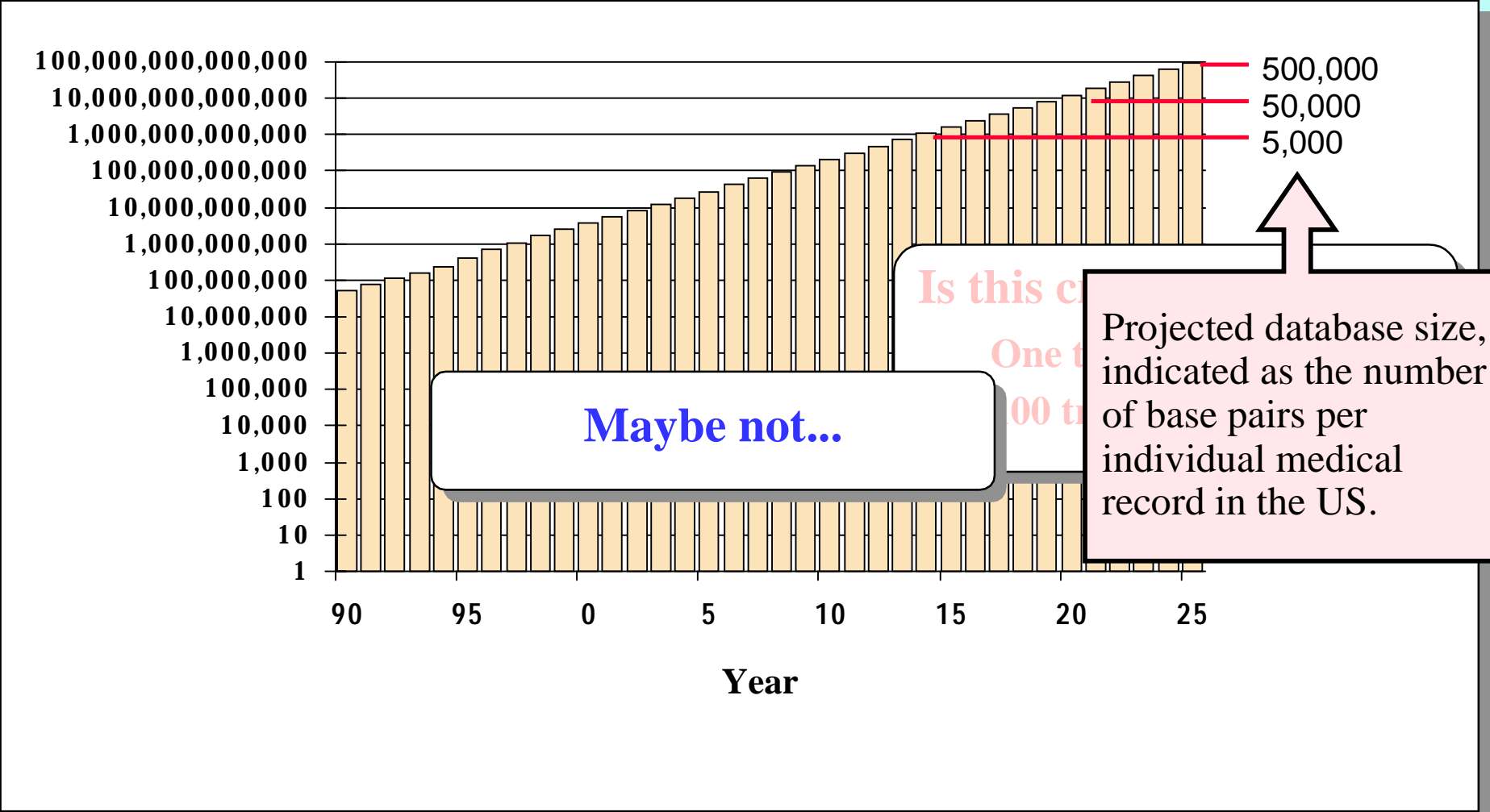
# Projected Base Pairs



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

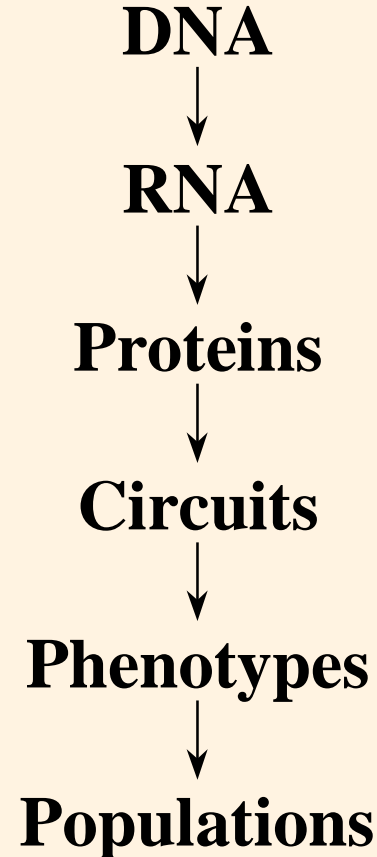
---

*The Science*

# Fundamental Dogma

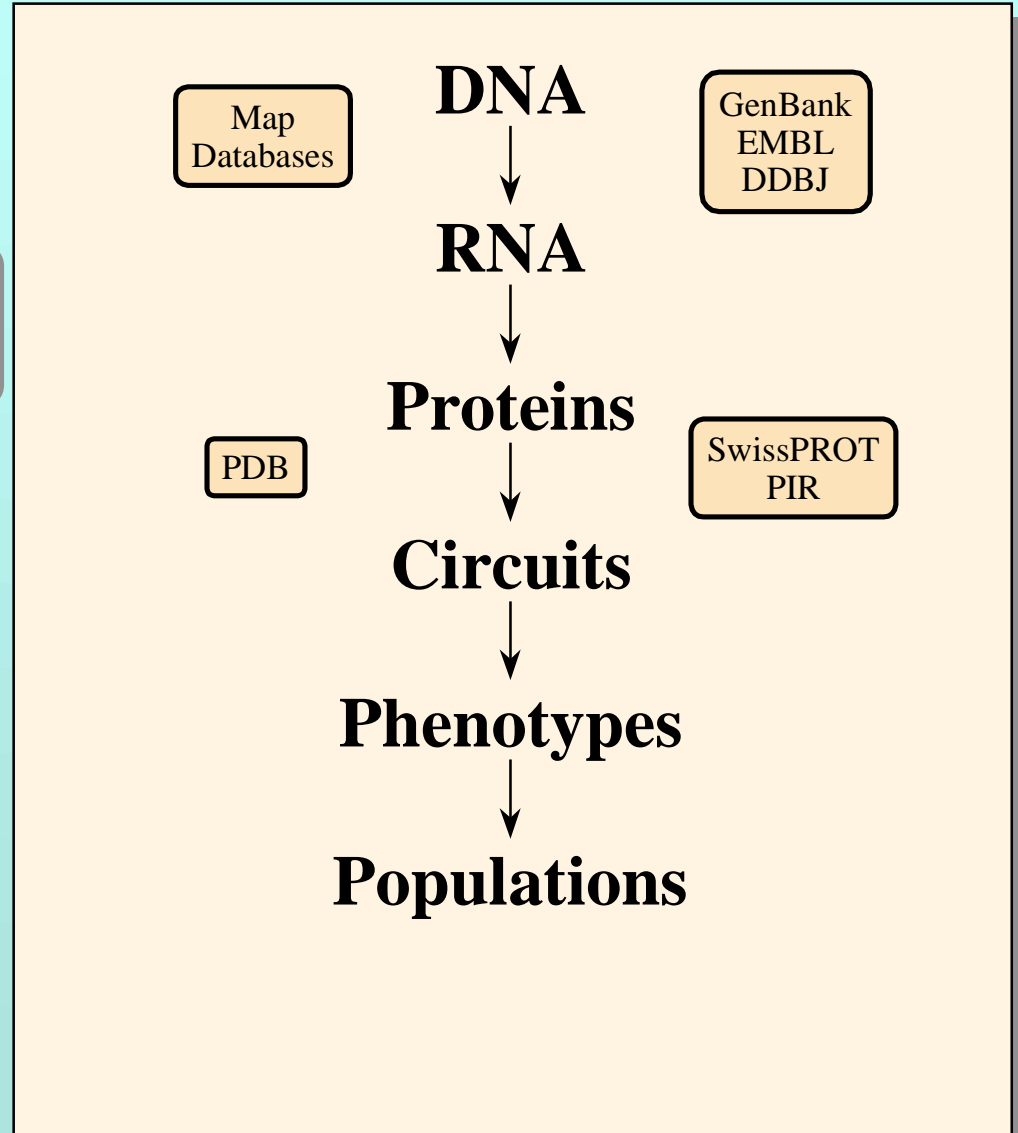
The fundamental dogma of molecular biology is that genes act to create phenotypes through a flow of information from DNA to RNA to proteins, to interactions among proteins (regulatory circuits and metabolic pathways), and ultimately to phenotypes.

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



# Fundamental Dogma

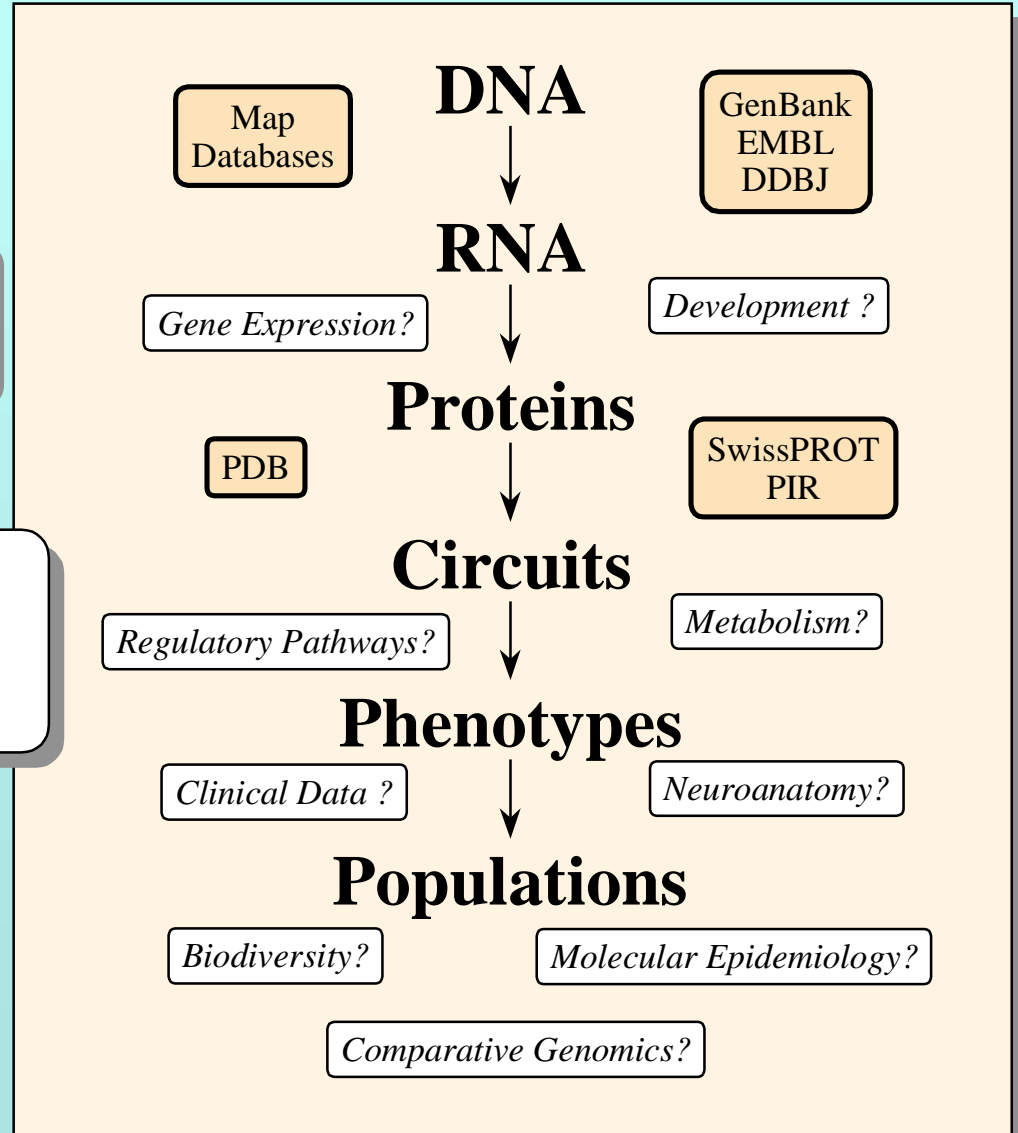
Although a few databases already exist to distribute molecular information,



# 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.”

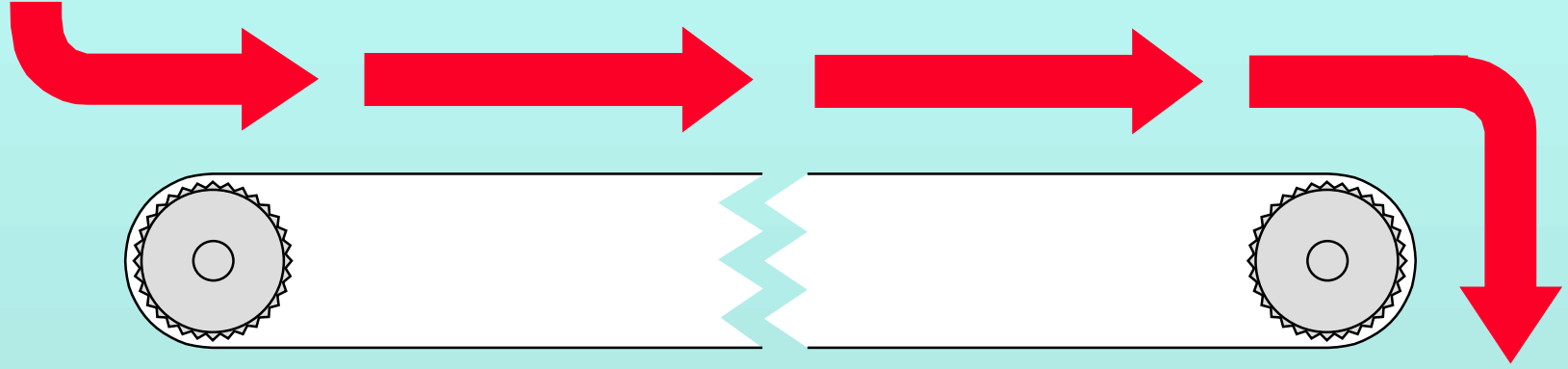
Exchange at DOE workshop on high-throughput sequencing.

# New Discipline of Informatics

# What is Informatics?

Computer  
Science  
Research

----- Informatics -----



Biological  
Application  
Programs

# What is Informatics?

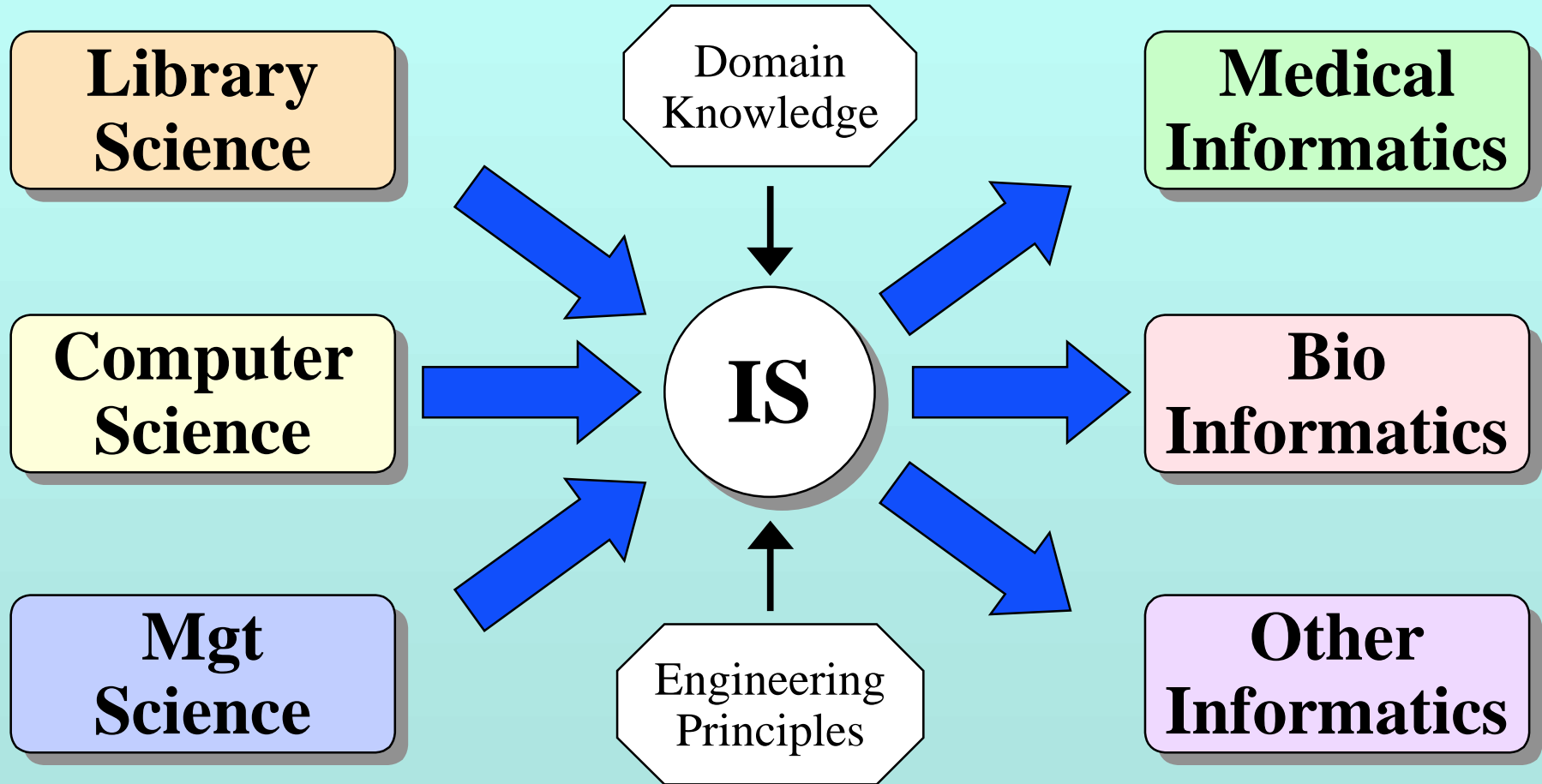
---

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.


# Federated Information Infrastructure

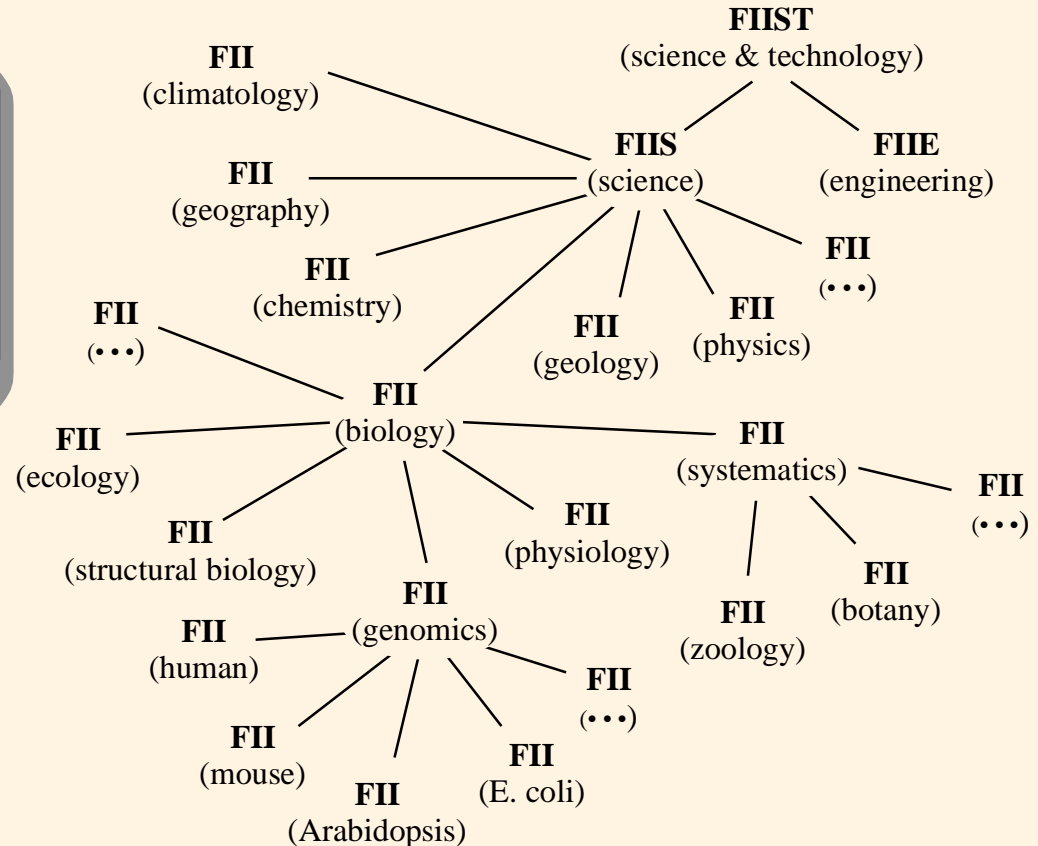
# National Information Infrastructure

	commercial uses		non-commercial uses		
	ETC	other	Edu	Lib	Res
analog					
digital					

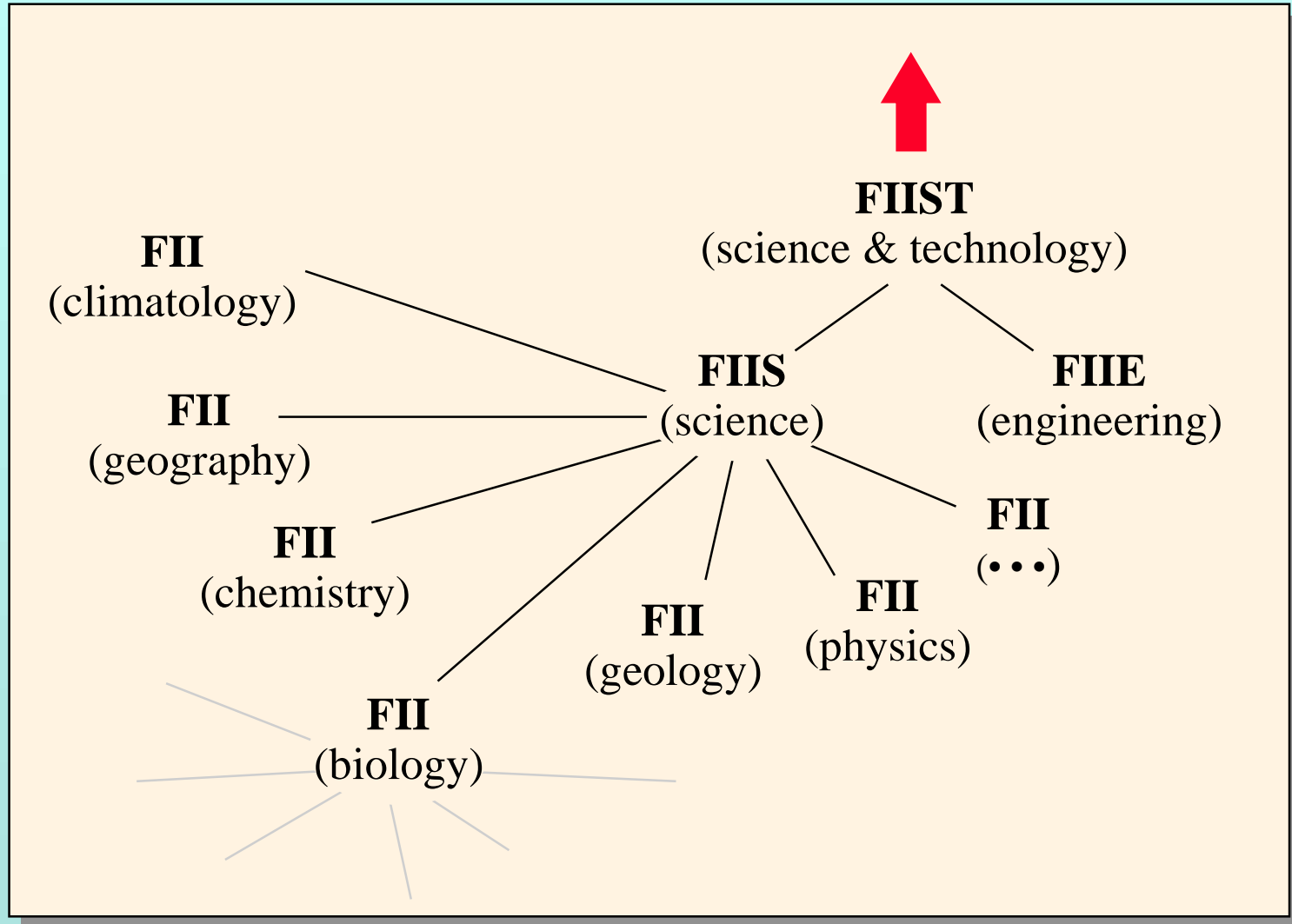
# FIIST & NII

The research component of the NII contains a Federated Information Infrastructure for Science and Technology..

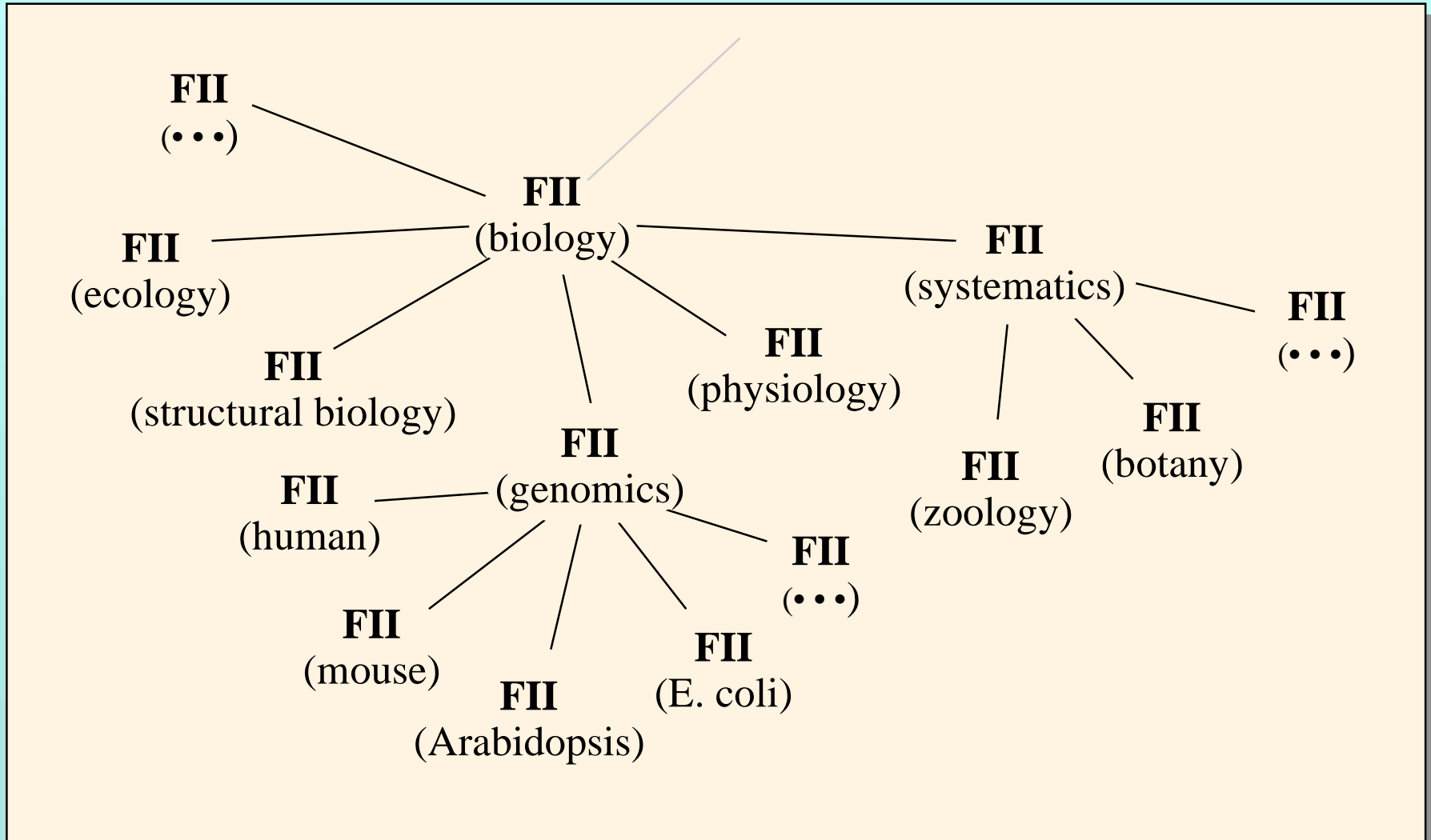
	commercial uses		non-commercial uses		
	ETC	other	Edu	Lib	Res
analog					
digital					



# FIIST



# FII



# Public Funding of Databases

---

## Stand-alone Criteria:

- Is there a need?
- Will this meet the need?
- Can they do it?
- Is it worth it?



# Public Funding of Databases

---

## Global Criteria:

- Does it adhere to standards?
- Will it interoperate?
- Is there commitment to federation?
- Is it worth it?

# Information Resources and the GII

---

## Guiding Principles:

- Global value explosion
- Componentry
- Anonymous interoperability
- Technical scalability
- Social scalability
- Value additivity

# 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.

# Basics

---

*Business 101*

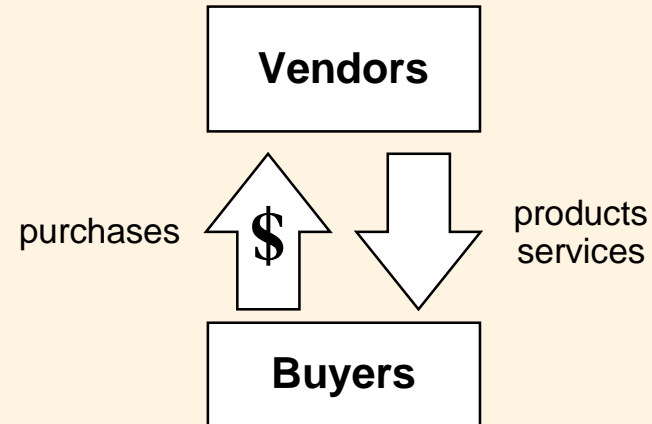
# Market Forces

In a simple market economy, vendors try to anticipate the needs of buyers and offer products and services to meet those needs.

Real users decide whether or not to buy a product or service, depending upon whether or not it meets a real need at a reasonable price.

## Business 101 Insight:

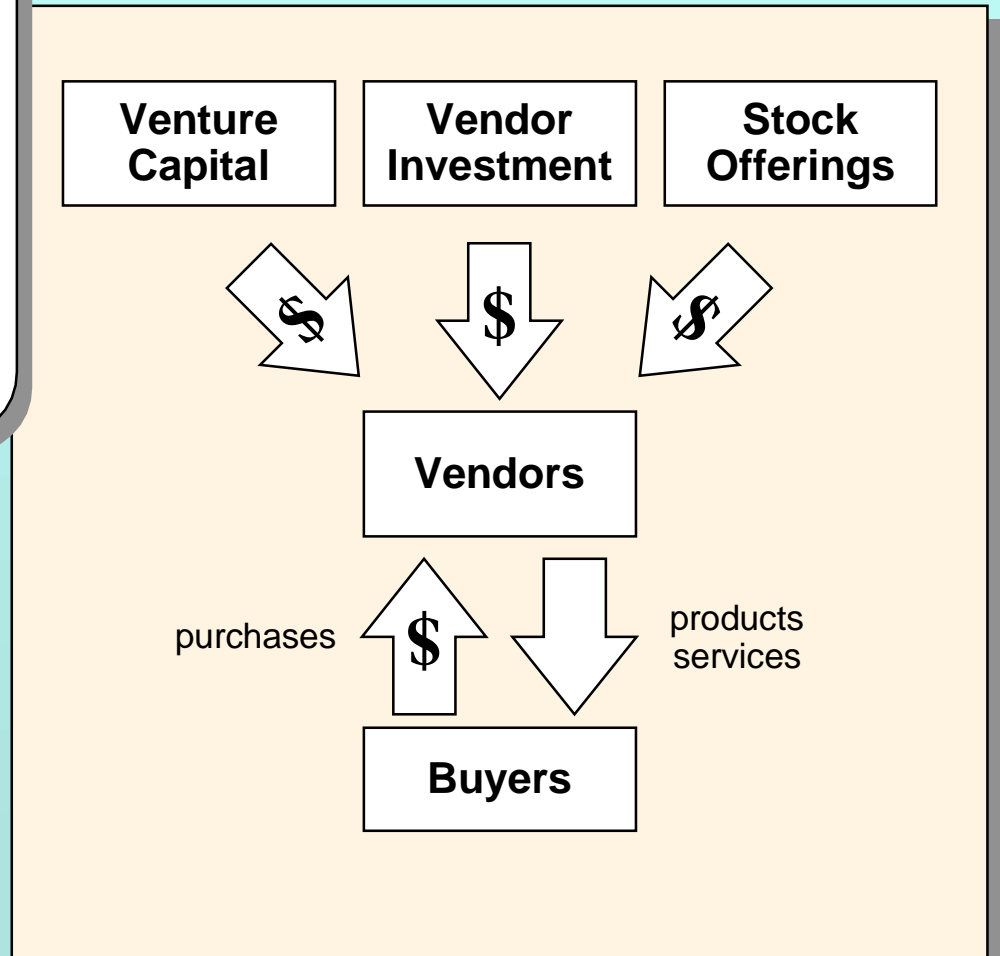
Successful vendors target a niche and excel at meeting the needs of that niche.



# Market Forces

Funding to initiate the development of products and services come from investors, not from buyers.

Investors decide whether or not to provide start-up funding based upon the estimated ability of the vendor to create products and services that will meet real needs at competitive prices.

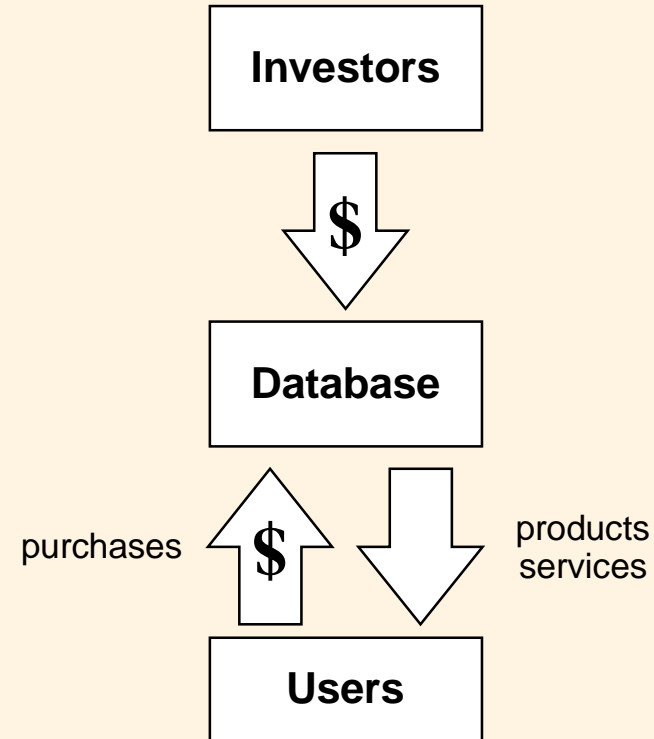


# Federal Funding

If biological databases were driven by market forces, individual users would choose what services they need and individual database providers would choose what services to make available.

Investors would provide start-up money on the likelihood of successful products and services being developed.

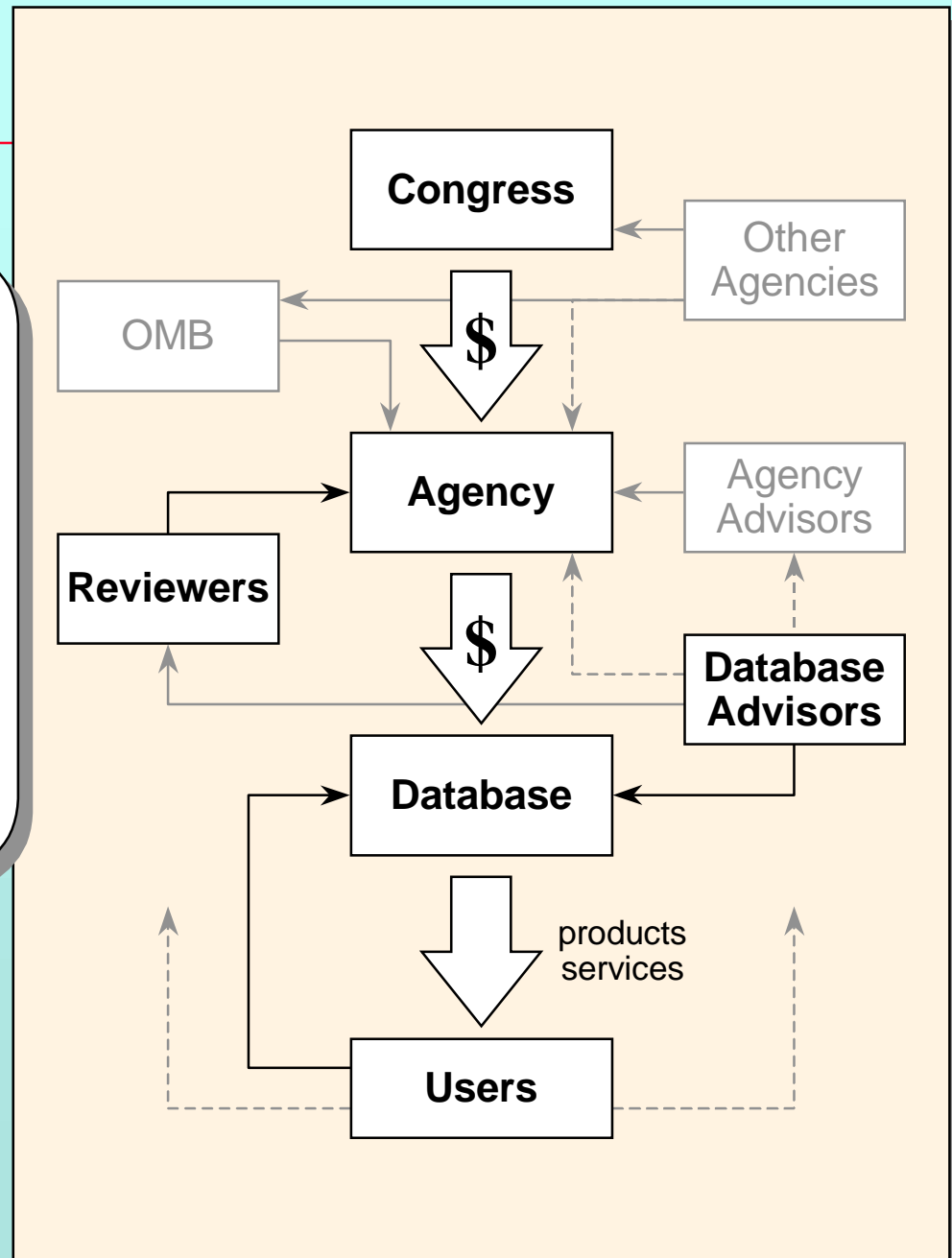
Ultimate success would depend on meeting the needs of real users. Decisions could be made rapidly, in response to changing needs and emerging opportunities.



# Federal Funding

Instead, funding decisions for grant-supported biological databases can follow a ponderously slow course, with almost no opportunity for real-time input from real users.

Even with the best of intentions at all levels, this process is slow, inefficient, risk-averse, and non-responsive to the real and changing needs of users.





# Federal Funding of Bio-Databases

---

## Possible solutions:

- increase the direct support of federal service organizations providing information infrastructure (*e.g.*, NCBI).
- reduce support for investigator-initiated, grant-funded public database projects.
- create market forces, initially through subsidization, later simply through direct support for affected science (*e.g.*, NSFnet into internet).

# Federal Funding of Bio-Databases

---

## Creating market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.

# Federal Funding of Bio-Databases

---

## Creating market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.

# Federal Funding of Bio-Databases

---

## Creating market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.
- provide guaranteed supplementary funding, redeemable only for access to bio-databases.

# Federal Funding of Bio-Databases

---

## Creating market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.
- provide guaranteed supplementary funding, redeemable only for access to bio-databases.
- data stamps

# Federal Funding of Bio-Databases

---

## Creating market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.
- provide guaranteed supplementary funding, redeemable only for access to bio-databases.
- data stamps, AKA *food (for-thought) stamps* ?!

# Food (for thought) Stamps

---

## Funding Agencies could:

- provide a 10% supplement to **every** research grant in the form of “stamps” redeemable only at database providers.
- allow the “stamps” to be transferable among scientists, so that a market for them could emerge.
- provide funding only after the stamps have been redeemed at a database provider.

# Food (for thought) Stamps

---

## Problems:

- how to estimate the amount of FFT stamps that would actually be redeemed (and thus the required budget set-aside).
- how to identify “approved” database providers.
- how to initiate the FFT system.
- etc etc



# Food (for thought) Stamps

---

## Alternatives (if no solution emerges):

- increasingly inefficient research activities (abject failure will occur when it becomes simpler to repeat research than to obtain prior results).
- loss of access to bio-databases for public-sector research.
- movement of majority of “important” biological research into the private sector.
- loss of American pre-eminence (if other countries solve the problems first).

# Slides:

---

**<http://www.esp.org/rjr/self.pdf>**