

CS 598sm

Probabilistic &
Approximate
Computing

<http://misailo.web.engr.illinois.edu/courses/cs598>

Probability

“the chance that something will happen”

Approximation

“an amount or figure that is almost correct and is not intended to be exact”

Uncertainty

“something that is doubtful or unknown“

Probability quantitatively represents uncertainty
(captures the degree of belief)

Approximation efficiently copes with uncertainty
(ignores it or tractably computes with it)

“All models are wrong,
but some are useful!”

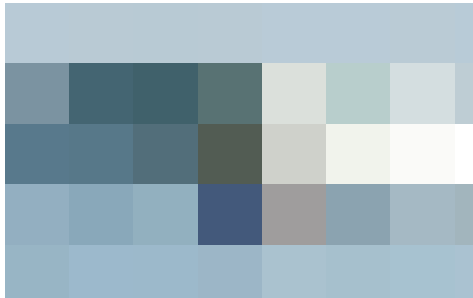
George E. P. Box



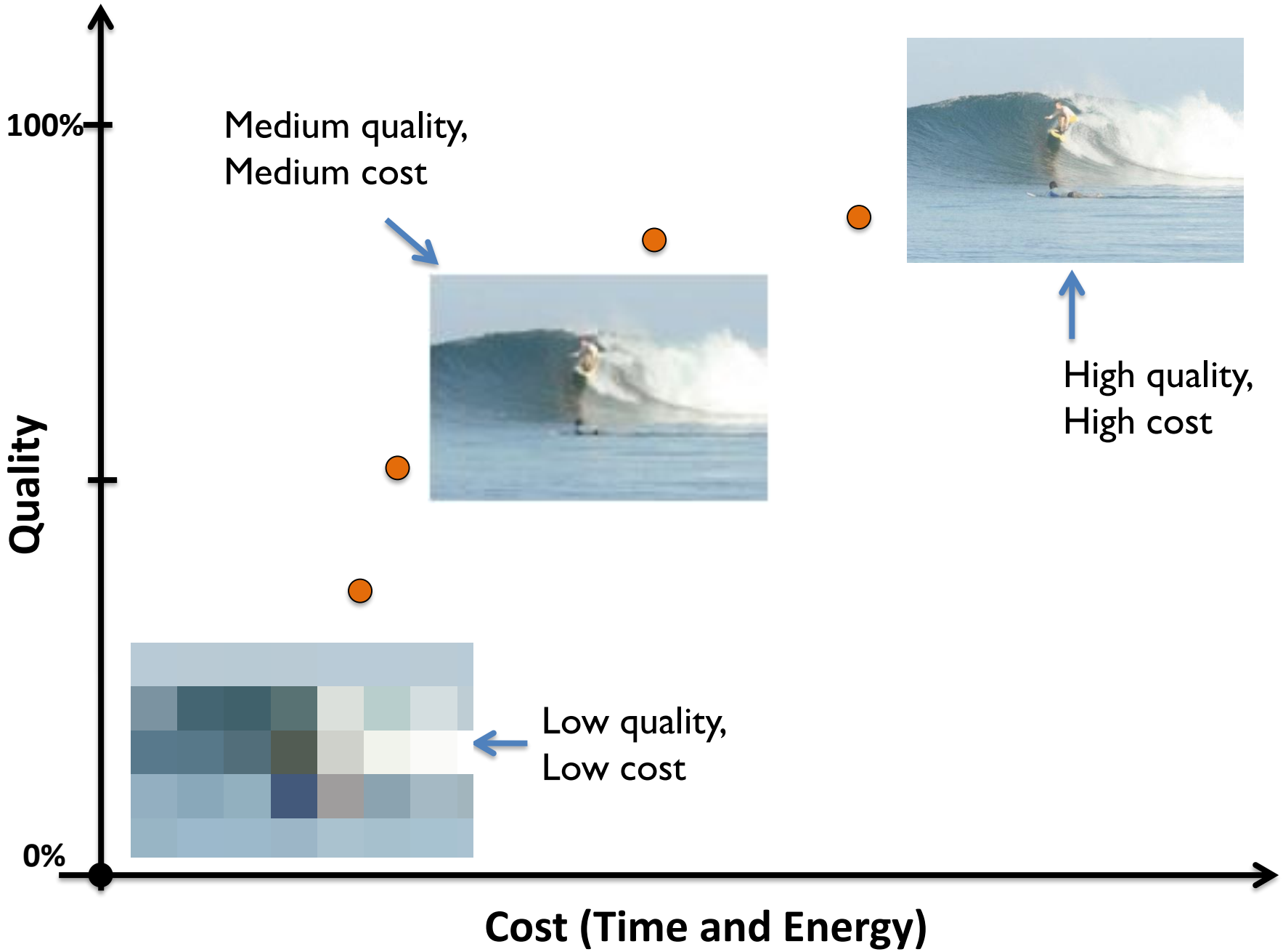
Medium quality,
Medium cost

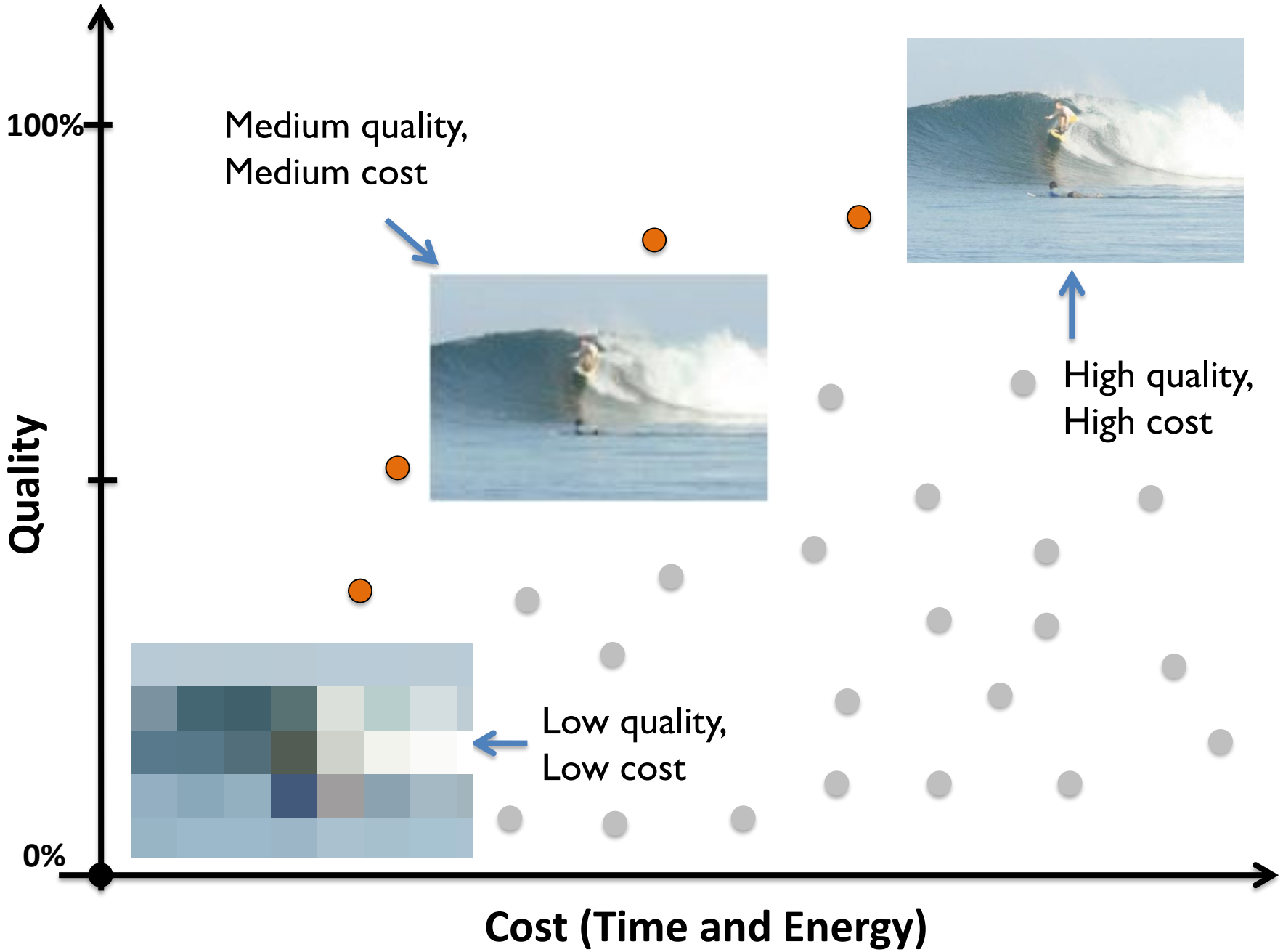


High quality,
High cost



Low quality,
Low cost



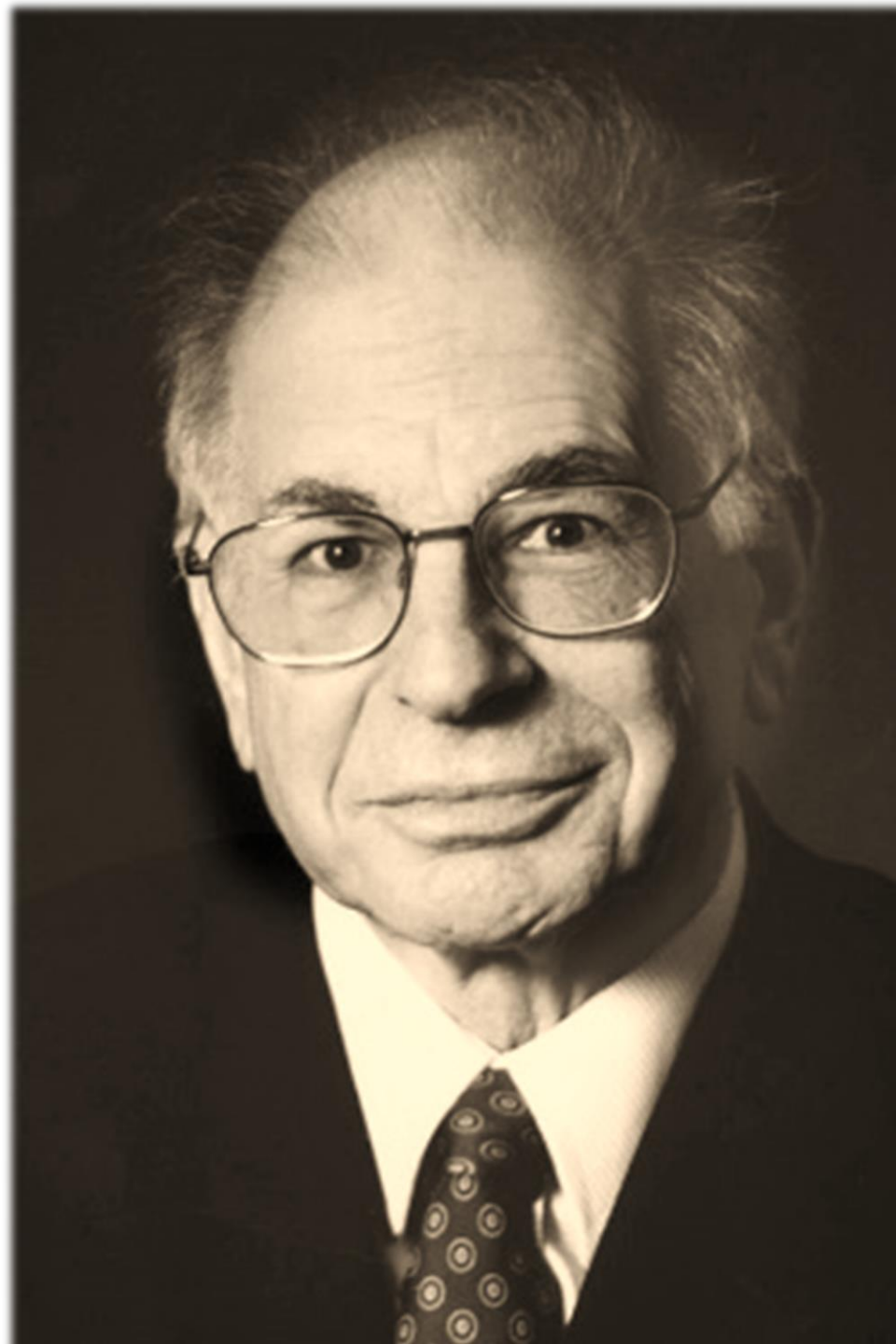


Two ways the brain
forms thoughts:

System 1: fast,
instinctive, and
emotional

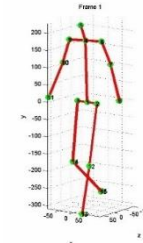
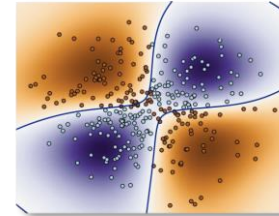
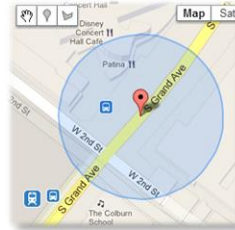
System 2: slow,
conscious, and
logical

Daniel Kahneman



Classical Computing Stack

Applications



Program
System

Translates the program code to *identical* machine code

Perfection!

aware

Executes the machine code *in unambiguous manner*

Observation (1965):

“The number of gates on a chip doubles every 24 months”

Gordon Moore



Dennard Scaling (1974)*:

Voltage and current should be proportional to the linear dimensions of a transistor.

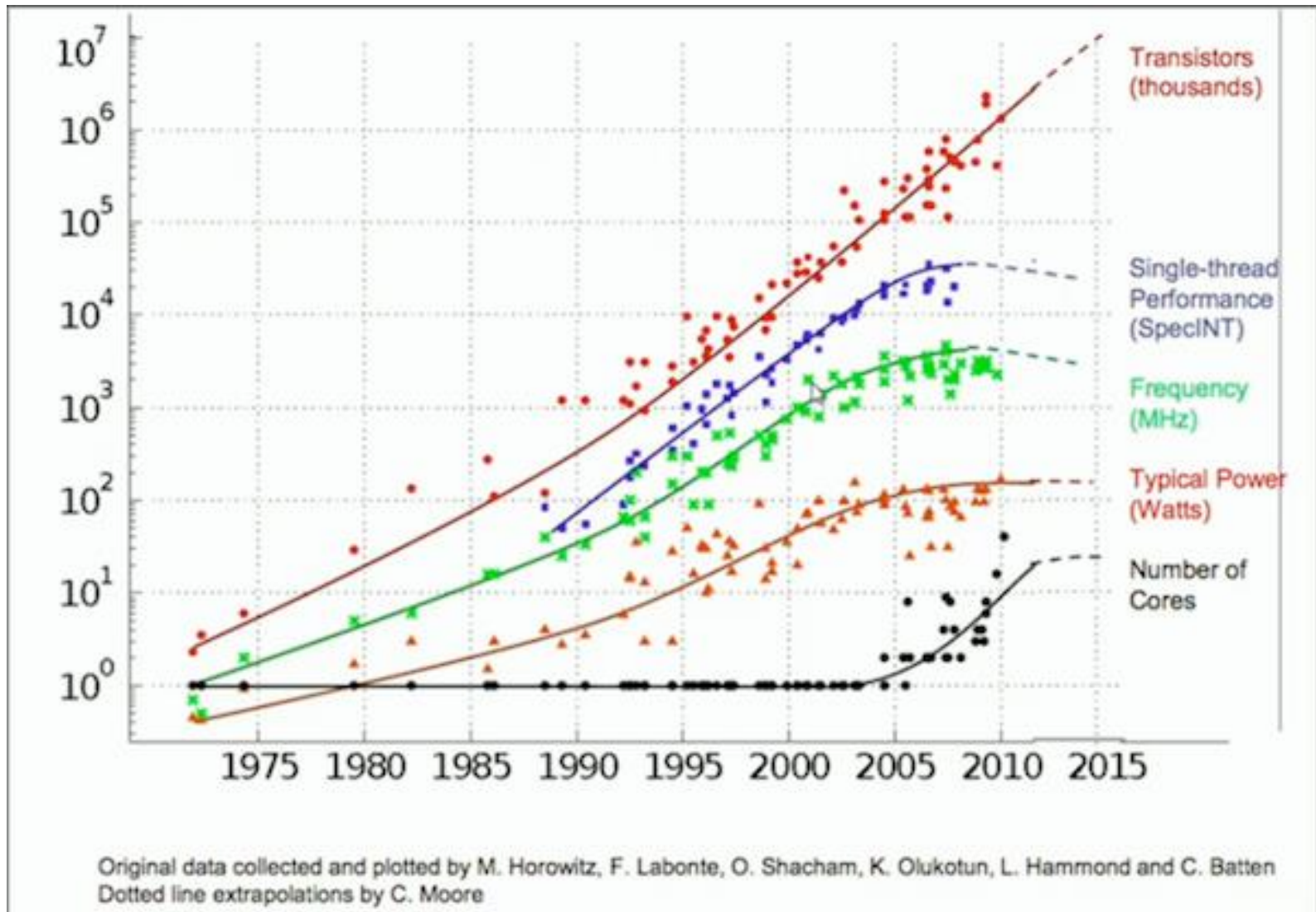
Thus, as transistors shrink, so do necessary voltage and current.

Power is proportional to the area of the transistor (while the transistor is still reliable)

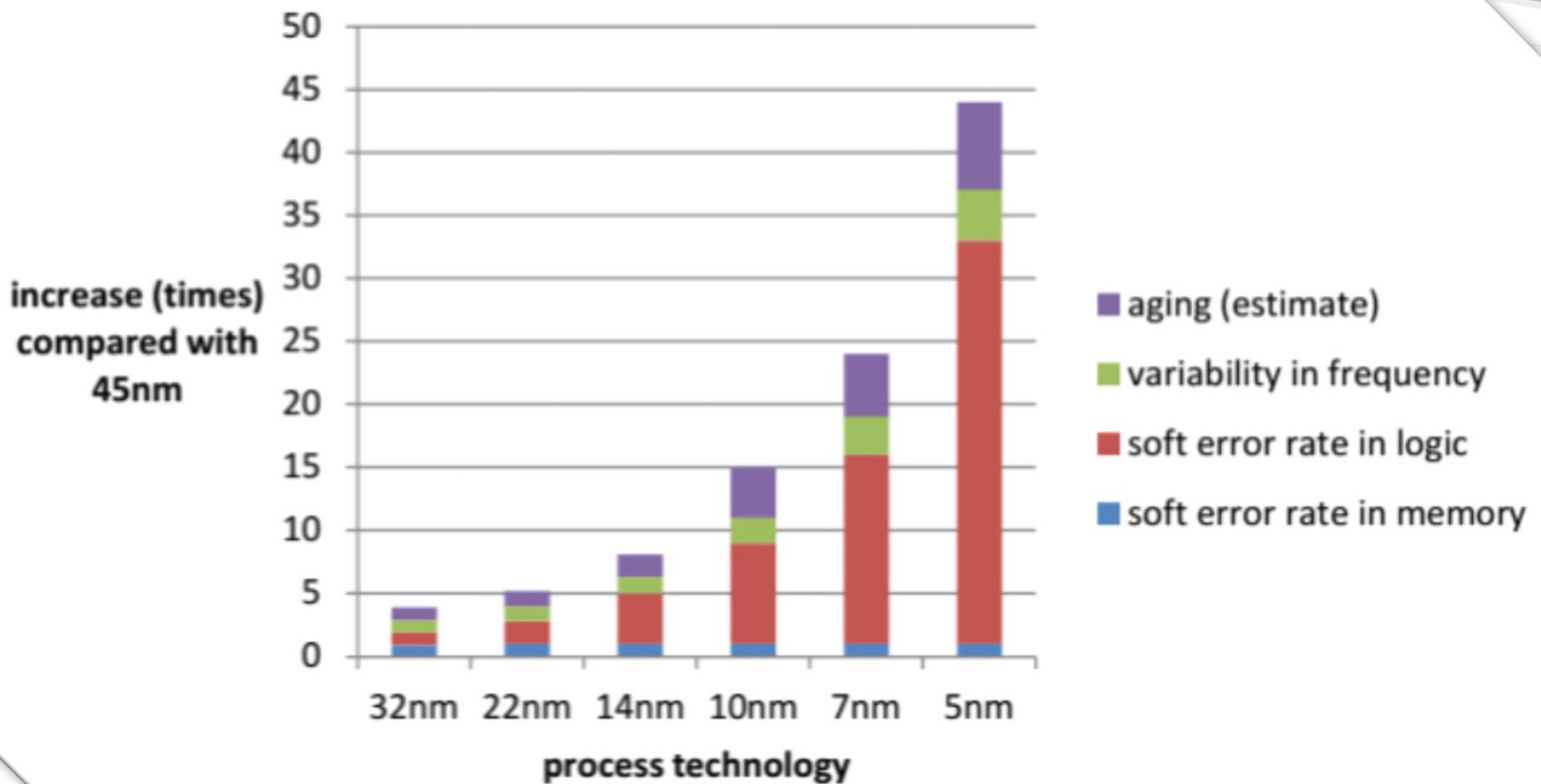
Robert Dennard



End of Dennard Scaling (2005-now)



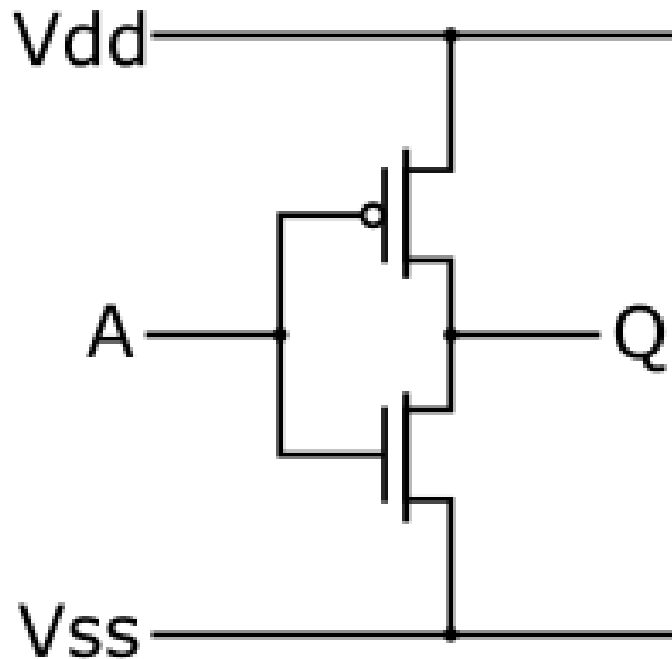
How Much Can We Shrink?



Source: Inter-Agency Workshop on HPC Resilience at Extreme Scale, Feb 2012.

CMOS Transistors

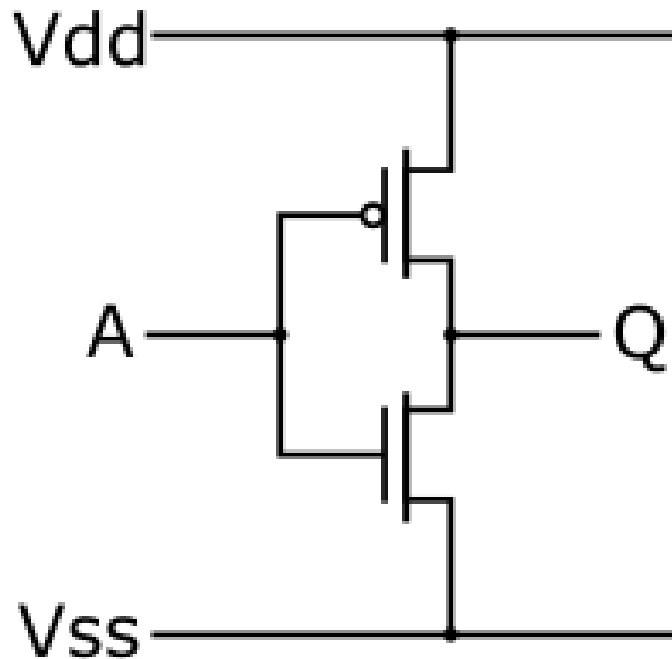
Simple Inverter



A	Q
0	1
1	0

Probabilistic CMOS Transistors

Simple Inverter



A	Q	
0	1	$P > 0.99$
0	0	$P < 0.01$
1	0	$P > 0.99$
1	1	$P < 0.01$

Breaking Digital Abstraction

Approximate and Unreliable Hardware

- **Process Variation and Aging**

Tiwari et al., ISCA '07; Mohapatra et al., ISLPED '09;
Rahimi et al., DAC '13; Karpuzcu et al., DNS '12;
Namaki-Shoushtari et al., CODES+ISSS '13...

- **Timing Errors & Soft Faults**

Ernst et al., MICRO '03; Sarangi et al., MICRO '08;
Kruijff et al., ISCA '10; Leem et al. DATE '10;
Sampson et al., PLDI '11; Ku He et al., DATE '11;
Esmaeilzadeh et al., ASPLOS '12 ...

- **Inexact Circuits & Storage**

Palem et al., SSDM '04; Narayanan et al., DATE '10; Chippa et al., DAC '10; Liu et al., ASPLOS '11;
Esmaeilzadeh et al., ASPLOS '12; Esmaeilzadeh et al., MICRO '12; Sampson et al., MICRO '13;
Venkataramani et al., MICRO '13; Venkataramani et al., ISLPED '14; Kozhikkottu et al., ISLPED '14;
Miao et al., ICCAD '14; St Amant et al., ISCA '14; Düben et al., Phil. Trans. R. Soc. '14 ...



© MIT News

John Von Neumann, 1952

PROBABILISTIC LOGICS AND THE SYNTHESIS OF RELIABLE ORGANISMS FROM UNRELIABLE COMPONENTS

J. von Neumann

1. INTRODUCTION

The paper that follows is based on notes taken by Dr. R. S. Pierce on five lectures given by the author at the California Institute of Technology in January 1952. They have been revised by the author but they reflect, apart from minor changes, the lectures as they were delivered.

The subject-matter, as the title suggests, is the role of error in logics, or in the physical implementation of logics - in automata-synthesis. Error is viewed, therefore, not as an extraneous and misdirected or misdirecting accident, but as an essential part of the process under consideration - its importance in the synthesis of automata being fully comparable to that of the factor which is normally considered, the intended and

Our present treatment of error is unsatisfactory and as such is the author's conviction, voiced over many years, that error should be treated by thermodynamical methods, and be the subject of a thermodynamical theory, as information has been, by the work of L. Szilard and C. E. Shannon [Cf. 5.2]. The present treatment falls far short of achieving this, but it is hoped, it is hoped, some of the building materials, which will have to enter into the final structure.

The author wants to express his thanks to K. A. Brueckner and M. Gell-Mann, then at the University of Illinois, to whose discussions in 1951 he owes some important stimuli on this subject; to Dr. R. S. Pierce at the California Institute of Technology, on whose excellent notes this exposition is based; and to the California Institute of Technology, whose invitation to deliver these lectures combined with the very warm reception by the audience, caused him to write this paper in its present form, and whose cooperation in connection with the present publication is much appreciated.



Programming Language Support

Specifications

$\langle 0.99 * R(x) \rangle f(x) \{ \dots \}$

Rely (OOPLSA'13)

@approx int x = ...

EnerJ (PLDI'11)

x := Gaussian (0,1);

Kozen (JCSS'81)

Verification

assert Pr[Error] < 0.001

assert Expected[Error] = 0

Key Concept

Probability Theory

Approximate Compilers

Loop Perforation (ICSE'10, FSE'11)

```
for (i = 0; i < n; i++) { ... }
```



```
for (i = 0; i < n; i += 2) { ... }
```

Approximate Compilers

Loop Perforation (ICSE'10, FSE'11)

```
for (i = 0; i < n; i++) { ... }
```



```
for (i = 0; i < n; i += 2) { ... }
```

It's not going to work!

Program will produce

incorrect result!



Original



Perforated

Encoded 3x faster



Original



Perforated



Original



Perforated

**Any pixel
difference**



Original



Perforated

**> 1% pixel
difference**



Original



Perforated

**> 5% pixel
difference**

Not a correctness issue

Accuracy issue

Other Accuracy-Aware Transformations

Software-Level Transformations

Call an approximate version of the existing functions

Baek et al., PLDI '10; Hoffmann et al., ASPLOS '11; Ansel et al., CGO '11;
Zhu, Misailovic, Kelner, Rinard POPL '12 ; Samadi et al., ASPLOS '14;

Return saved outputs previously computed for similar inputs

Chaudhuri et al., FSE '11; Samadi et al., ASPLOS '14; Mishra et al., WACAS '14;

Decrease precision of floating-point operations

Rubio-Gonzalez et al., SC '13; Schkufza et al., PLDI '14;

Remove locks or barriers in parallel programs

Rinard ICS '06; Rinard OOPSLA '07; Meng et al., IPDPS'09; Chakradhar et al., DAC '10;
Misailovic et al., MIT-TR '10, TECS PEC '13; Renganarayana et al., RACES '12;
Rinard, HotPar '13; Campanoni et al., CGO '15

Goal of the Course

Embark on a journey to
rethink computing in the world
where absolute correctness is elusive

- Learn from recent papers
- Discuss new research ideas
- **Do a fun project**

CS 598 SM

COURSE LOGISTICS

Schedule

Twice a week – Tuesdays and Thursdays 11am-12:15pm

We first do several lectures: tutorial style introductions to

- Approximate computing
- Probabilistic programming
- Covers key ideas and classical results

In the majority of the course, we will discuss recent papers

- Typically, discussion focus is on one paper at a time
- One student presents the paper
- Everyone participates in the discussion

Course Format

Research-oriented Course:

- Discussing latest research
- Reading from primary literature (papers)
- Focus on finding new ideas and building new systems, not lecturing and grading

Research project is the main outcome of the course

- Be able to publish your work at a conference
- It is **hard! Unpredictable + requires a lot of time and effort**

Prerequisites

Basic Probability (e.g., CS 36 I)

Compilers and/or PL course (e.g., CS 42 I)

Basic computer architecture (e.g., CS 233)

(or a commitment to learn as you go)

Real Prerequisites

Experience doing research

(If you don't know what you're getting into,
talk to me after the class)

Grading

Miniquizes	10%
Reviews & Discussion	15%
Paper Presentation	20%
Project	50%
Homework	5% + 10% XC

Miniquizzes

Tests background knowledge (like the one today)

- Concept from the paper, math, intuition
- 5 minutes at the beginning of the class
- We will discuss solution immediately afterwards

Each miniquiz is **worth 0.5%** (up to 10%).

- Self-graded, the main purpose is to bring everyone to the same page before we start the discussions
- In total 25 quizzes; can miss 5 without penalty

Papers

For the majority of the class, we will **jointly read and discuss** recent research results

We focus on one paper in discussions

- but for most part we put it in context with at least another related piece of research

Make sure you can make it to the class on the day you're presenting the paper!

Selecting Papers

Submit at least 5 candidate papers you'd like to present

- List of papers is on the website (use the week and number)
- Split in the **primary read** – the one we will focus the most in the discussion and **secondary read** – the one we will use to expand our understanding of the topic
- If you'd like a paper outside of the list, email me and make a case

Submission deadline is **September 5**

- Link: <http://misailo.web.engr.illinois.edu/courses/598sm/>
- Will get back with the assignments by the class after

Reviews and Discussions

For each paper, write a review of up to 600 words:

- Summarize the **primary** paper:
state main contributions in 2-3 sentences (use your own words!)
- Discuss pros and cons:
give a honest critique of the approach (main part)
- Two questions:
about the paper, the general topic, its impact, or extensions (key!)
- Summarize the **secondary** paper after reading only its introduction (and maybe example) sections
state main contributions in 2-3 sentences
- Relate the two papers
What is similar and what is different (extrapolate what the 2nd paper does from the first two sections)

Reviews and Discussions

Send reviews before the lecture

- By midnight the day before
- Submission forms: next to the paper on the course site

Purpose of reviews is to prepare for the in-class discussion

- Light feedback on the reviews, lead where discussion goes

Participate in the discussion during the class

- Purpose: practice how to be loud
(at the conferences, board meetings, home...)

Presentation

Week before: Meet with the instructor

- **Mandatory!** (typically during the office hour, e.g. 30 mins)
- Discuss outline and questions so far (ok if still rough!)

30 minute slot per presentation:

- Explain motivation for the work
- Clearly present the technical solution and results
- Use your own example (not the one from the paper)
- Outline limitations / improvements
- Focus on concepts, leave out nonessential details
- Discuss the impact on the related/follow-up work

Grading Presentations

Presentation quality:

- How well did you understand the work?
- How well did you present it (clarity and grace)?
- How well did you answer the questions?

We will take into account the **paper difficulty**

Project

Teams of two (if the number of students is even)

- Teamwork is a great experience!

Research projects, some ideas:

- New Software and/or hardware approximations
- Dynamic or input-aware approximations
- Optimize approximate inference algorithms
- New program analysis for probabilistic programs
- New probabilistic analysis of approximate programs
- Implement and compare existing approaches
- Survey literature on an emerging topic

Grading Projects

Proposal by **October 3**

- Meet with instructor for a quick discussion

Deliverables:

- Short paper – up to 5 pages ACM 10pt format
- Think of e.g., DATE (<https://www.date-conference.com/>)
- Project overview – 10/15 minutes
- Due **last week of classes (Tuesday)**

Real outcome:

- Prepare (or make a good step toward) a publishable research paper

Grading

Miniquizes

Reviews & Discussion

Paper Presentations

Project

Final Exam

15%

20%

50%

5% +

10% XC

Grading on an absolute scale (no curve!)

CS 598 SM

RESOURCES FOR READING, WRITING AND PRESENTING

Reading Papers

“How to Read a Research Paper”,
by Michael Mitzenmacher

<http://www.eecs.harvard.edu/~michaelm/postscripts/ReadPaper.pdf>

“How to Read an Engineering Research Paper”,
by William Griswold

<http://cseweb.ucsd.edu/~wgg/CSE210/howtoread.html>

Advice compiled by Tao Xie:

<http://taoxie.cs.illinois.edu/advice.htm#review>

Writing Reviews

“The Task of the Referee”, by Allan Smith

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.177.3844>

“Constructive and Positive Reviewing”,
by Mark Hill and Kathryn McKinley

<http://www.cs.utexas.edu/users/mckinley/notes/reviewing.html>

Presenting Research

“How to give strong technical presentations”
by Markus Püschel

<http://users.ece.cmu.edu/~pueschel/teaching/guides/guide-presentations.pdf>

Patrick Winston’s talk @ MIT:

<https://www.youtube.com/playlist?list=PL9F53600IA3C605FC>

Jean Luc Doumont’s talk

<https://www.youtube.com/watch?v=meBXuTIPJQk>

CS 598 SM

QUESTIONS SO FAR?