

Title: Sustaining the Exponential Growth of Embedded Digital Signal Processing Capability¹

Authors:

Dr. Gary A. Shaw (*first author/corresponding author/presenting author*)
MIT Lincoln Laboratory
244 Wood Street
Lexington, Massachusetts 02420-9185
Voice: 781-981-7994
Fax: 781-981-4608
E-mail: shaw@ll.mit.edu

Dr. Mark A. Richards
Georgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, GA 30332-0250
Voice: 404-894-2714
Fax: 404-894-0560
E-mail: mark.richards@ece.gatech.edu

Candidate Sessions (drawn from HPEC abstract submission web site):

- Algorithm Mapping to High Performance Architectures
- Future Program Office Needs for Embedded Computing Technologies
- Case Study Examples of High Performance Embedded Computing

Author's Comments:

This proposed paper is less of a technical exposition than a retrospective and exhortation regarding the importance of sustained efforts in algorithm and software research and development. The paper compares and contrasts the contributions of hardware and algorithm improvements to the rate of improvement in high performance computing and signal processing. We conjecture that as IC shrinkage and attendant performance improvements begin to slow, the exponential rate of improvement we have become accustomed to for embedded applications will be sustainable only through a faster pace of improvement in algorithms and software.

¹ *This work was sponsored by the Department of the Air Force under Air Force Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.*

Report Documentation Page

Form Approved
OMB No. 0704-0188

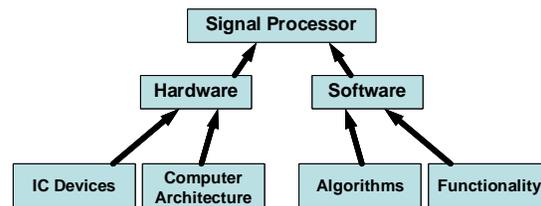
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 01 FEB 2005	2. REPORT TYPE N/A	3. DATES COVERED -			
4. TITLE AND SUBTITLE Sustaining the Exponential Growth of Embedded Digital Signal Processing Capability		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MIT Lincoln Laboratory 244 Wood Street Lexington, Massachusetts 02420-9185; Georgia Institute of Technology School of Electrical and Computer Engineering Atlanta, GA 30332-0250		8. PERFORMING ORGANIZATION REPORT NUMBER			
		10. SPONSOR/MONITOR'S ACRONYM(S)			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
		12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited			
13. SUPPLEMENTARY NOTES See also ADM001742, HPEC-7 Volume 1, Proceedings of the Eighth Annual High Performance Embedded Computing (HPEC) Workshops, 28-30 September 2004. , The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 28	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

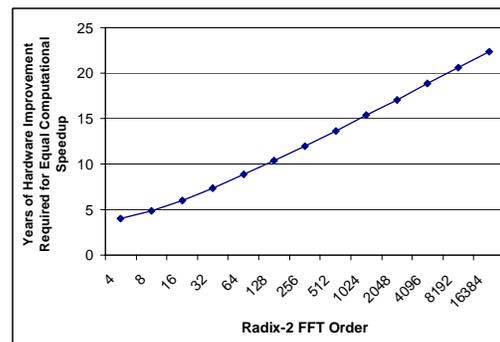
Sustaining the Exponential Growth of Embedded Digital Signal Processing Capability

Gary A. Shaw and Mark A. Richards

The exponential improvement in IC device density and throughput observed over the past 40+ years, first articulated by Gordon Moore in 1965 and canonized as Moore's Law, is well known and has had an immense impact on our society. Obvious, but less appreciated, is the fact that the computational capability of a digital signal processor, and more generally any high performance computing system, is the sum of its hardware capability and its software capability. By "hardware" we mean the physical implementation, which includes both individual ICs and the system architecture. "Software" is the computational procedure, which includes both the mathematical functionality and the particular algorithm by which it is implemented. For example, the discrete Fourier transform (DFT) is a particular mathematical function, while the fast Fourier transform (FFT) is a particular sequence of computations that implements the DFT efficiently.



Observed over long periods of time, software improvements, particularly fast algorithms, have contributed exponential rate improvements that equal or exceed the improvements accruing from faster hardware. The best known example in the DSP community is, of course, the FFT. Depending upon the problem order, the computational speedup afforded by the basic radix-2 FFT is equivalent to anywhere from one to 20+ years of device improvement. The FFT is



not unique in terms of its impact on the acceleration of HPEC applications. In an example of long-term evolution of scientific computing algorithms, Bentley² documents improvements in the solution of 3-D elliptic partial differential equations. He shows that from 1945 to 1985, operation counts for problems computed on an $N \times N \times N$ grid were reduced through a succession of algorithmic improvements by a typical factor of $N^4/60$. For a broadly representative problem size of $N = 64$, the improvement is a factor of about 3×10^5 , or just under 1.4 orders of magnitude per decade. This rate of improvement is on par with that observed for IC devices. Other examples of long-term, as well as sudden, algorithmic improvements can be found in applications ranging from dial-up modems to computer chess. The improvements associated with these application areas will be illustrated in detail in the presentation.

¹ This work was sponsored by the Department of the Air Force under Air Force Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

² Jon Bentley, "Programming Pearls," *Comm. ACM*, v. 27(11), pp. 1087-1092, Nov. 1984.

While evidence exists for equal rate improvements from algorithms and hardware, there are nonetheless significant differences in the manner in which hardware and software improvements are manifested. Hardware throughput performance increases exponentially and *predictably* as a function of time. As long as Moore's Law remains viable, we can count on the fact that a computationally complex algorithm not currently realizable in real time will eventually become realizable, and we can even predict approximately when it will be realizable! In contrast, while reduced complexity algorithms are discovered unpredictably in time, they increase execution speed exponentially and predictably *as a function of problem dimension or order*. Discovery of a fast algorithm thus acts in effect like the discovery of a "worm-hole" in time evolution of an application. Thus, rather than waiting for the time evolution of Moore's Law hardware improvements, the discovery of a fast algorithm creates an instantaneous leap ahead in performance, with the "time-compression" benefit of the worm-hole increasing in proportion to problem order.

In comparison to the historically predictable payoff in performance associated with investments in IC development, the unpredictability of new algorithm discovery makes investing in signal processing capability somewhat more risky. However, the higher risk of algorithm investment is more than offset by its most fundamental payoff: the development of entirely new capabilities resulting from the application of new concepts and mathematics. Improvements in hardware speed enable application performance improvements by supporting more of the existing functionality within a given amount of time, space, power, or other resource. In contrast, breakthroughs in functionality add entirely new tools to the signal processing toolbox, and may change the entire nature of the signal processing problem and implementation complexity.

With regard to new algorithm concepts and capabilities, the scope of research opportunities is actually expanding. Many of the traditional examples of algorithm breakthroughs we have cited were new computational procedures that substantially reduced operation counts for a specific function; the solutions of PDEs, sorting, and the FFT are all examples. New mathematical techniques provide new opportunities for similar improvements. Algorithms that achieve speedups by clever matching of mathematical problem structure to computer architecture represent a very different avenue of attack. More fundamentally, emerging research focusing on knowledge-based and cognitive systems opens up to scrutiny entirely new types of both functionality and computational complexity.

The investment needed to maintain, and even expand, an active and robust embedded processing research community is a fraction of the investment needed to keep semiconductors on the Moore's Law growth curve, primarily because algorithm research costs are dominated by salaries, which rise at a much slower rate than the cost of new fabrication facilities. Therefore, relatively modest increases in algorithm and software research are sufficient to maintain the total progress in signal processing performance. In particular, as we move into a future where the certainty of Moore's Law hardware improvements is in question, and the investment gap between algorithm research and improved semiconductor fabrication facilities continues to widen, increasing our investment in algorithm research appears to be an attractive and proven means to sustaining exponential growth in embedded application performance.



Sustaining the Exponential Growth of Embedded DSP Capability[†]

High Performance Embedded Computing Workshop

28 September 2004

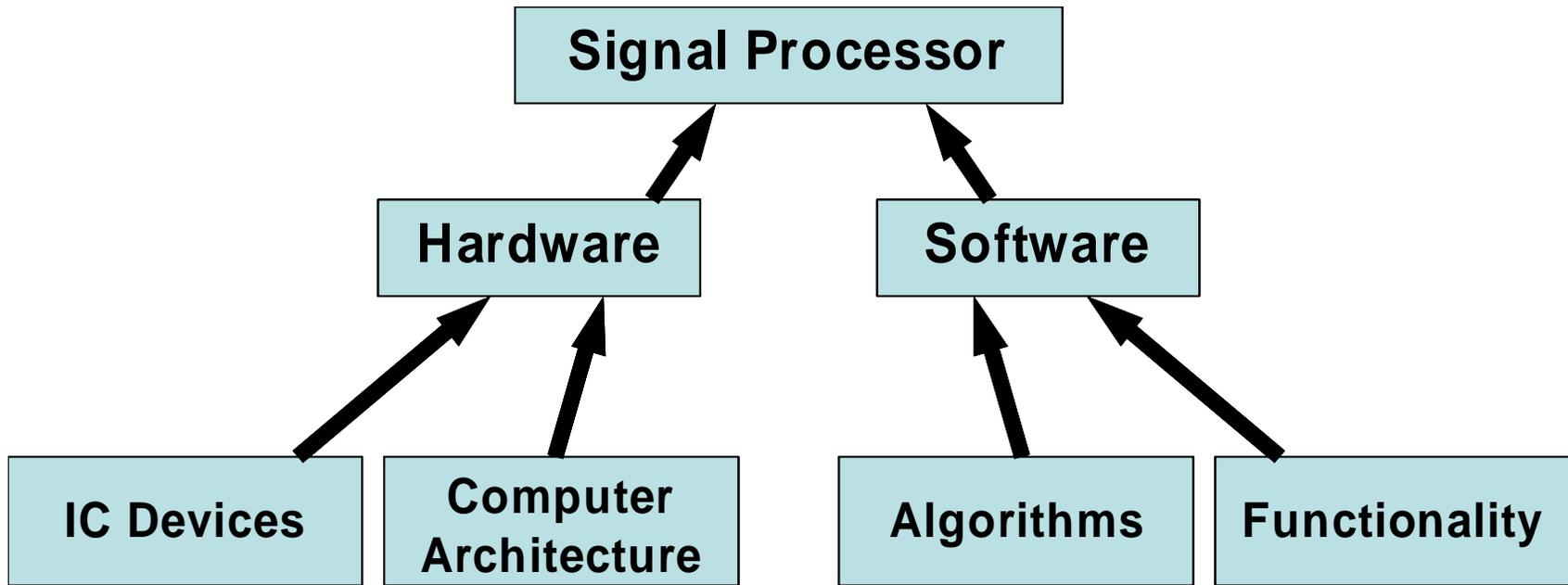
**Dr. Gary Shaw
MIT Lincoln Laboratory**

**Dr. Mark Richards
Georgia Institute of Technology**

[†] This work was sponsored by the Department of the Air Force under Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.



Elements Contributing to Embedded Processor Performance





Outline

- **Historical perspective – fulfillment of Moore's Law**
- **Impediments to continued IC density growth**
- **Algorithms – the softer side of exponential growth**
- **Implications regarding sustaining exponential growth**
- **Summary and Conclusions**

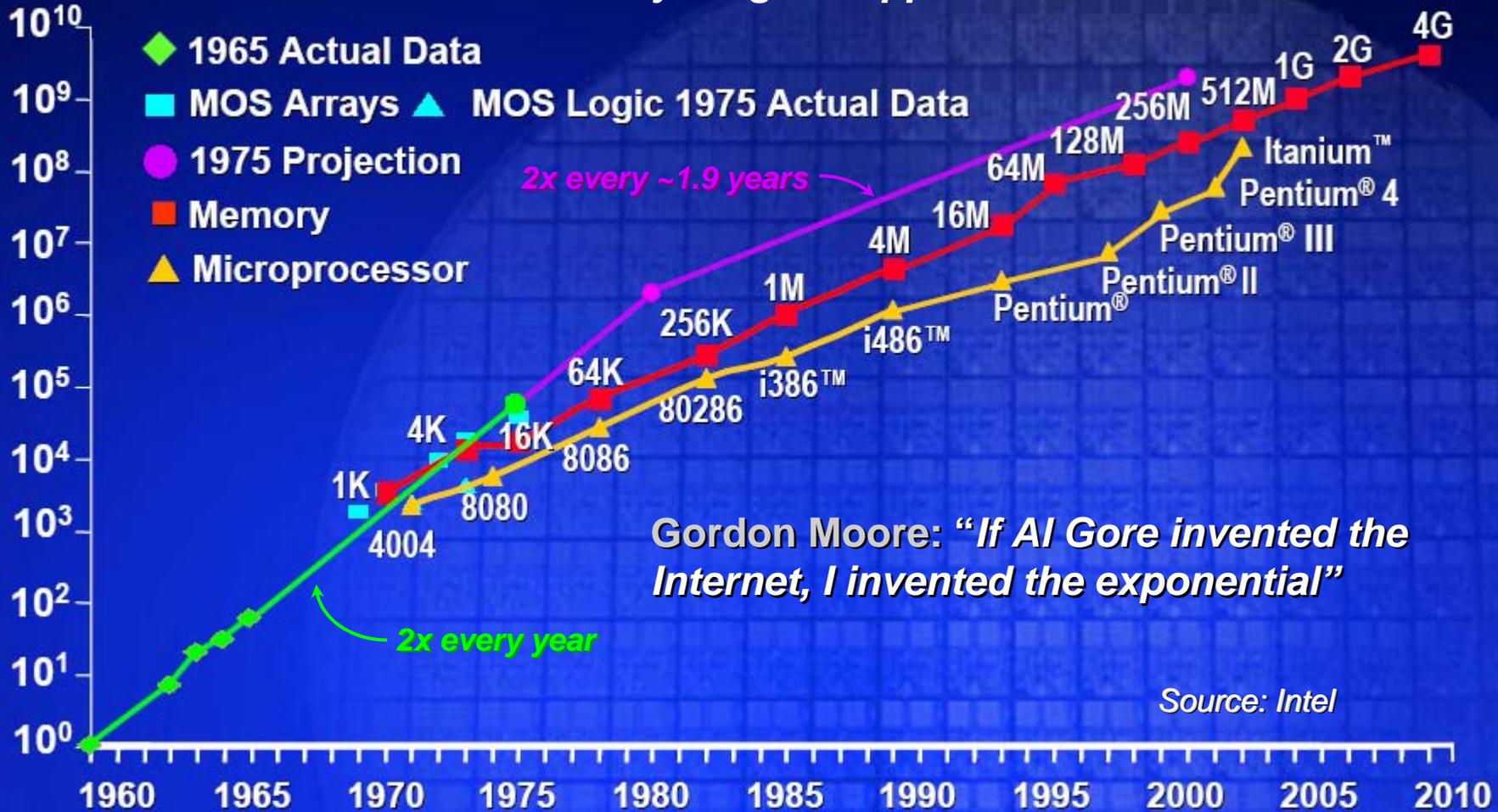


Moore's Law: Prediction and Realization

Transistors
Per Die

John von Neumann: "Truth is much too complicated to allow anything but approximations."

- ◆ 1965 Actual Data
- MOS Arrays ▲ MOS Logic 1975 Actual Data
- 1975 Projection
- Memory
- ▲ Microprocessor

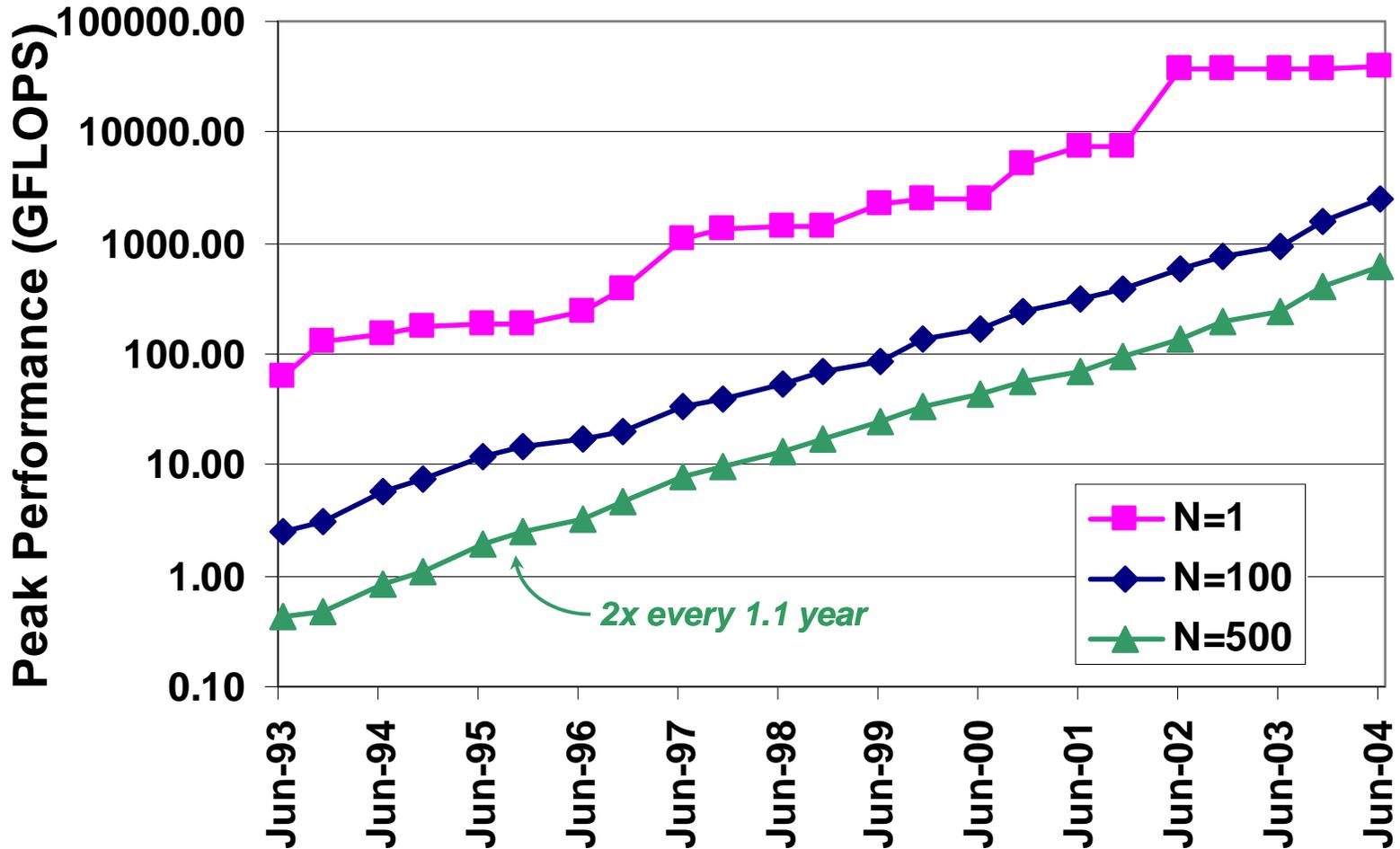


Gordon Moore: "If Al Gore invented the Internet, I invented the exponential"

Source: Intel



Top 500 Computer Growth





Outline

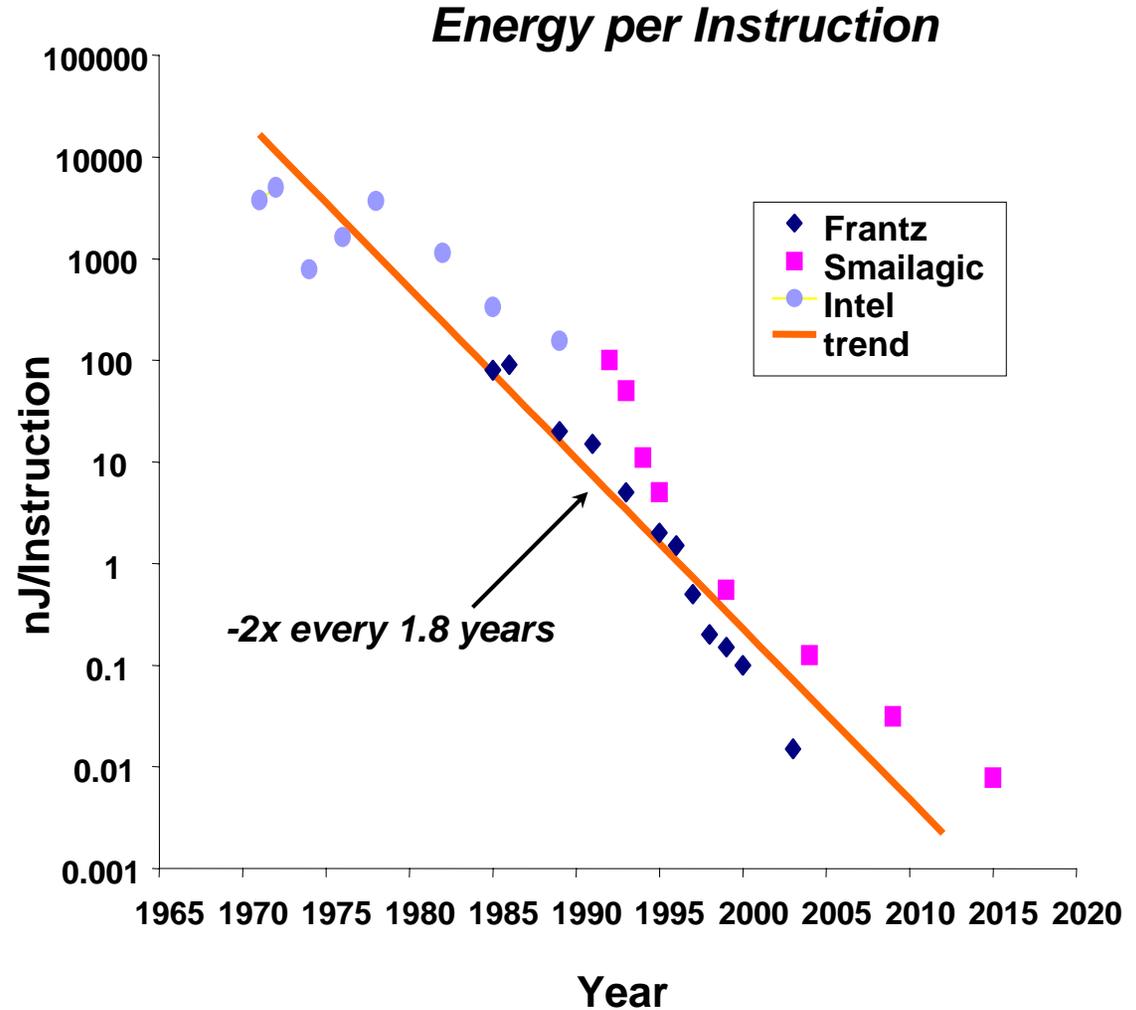
- **Historical perspective - fulfillment of Moore's Law**
- **Impediments to continued IC density growth**
 - Heat dissipation
 - Quantum effects
 - Production technology
- **Algorithms – the softer side of exponential growth**
- **Implications regarding sustaining exponential growth**
- **Summary and Conclusions**



Performance Implications of Shrinking Feature Size

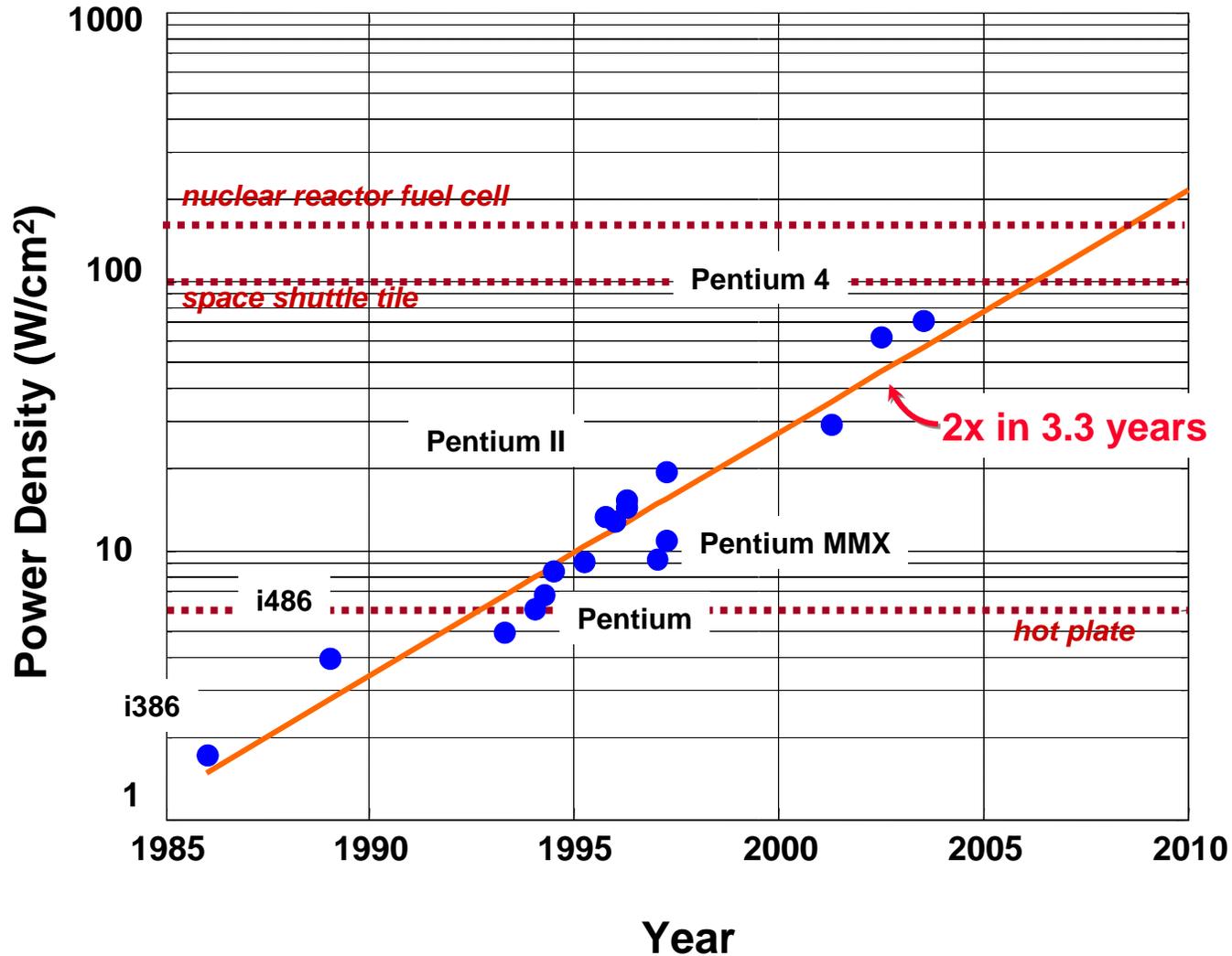
$\Delta \equiv$ feature size

Performance Metric	Geometrical Dependency
Clock Frequency	$1/\Delta$
Transistor Power	Δ
Transistor Density	$1/\Delta^2$
Total Device Power	$1/\Delta$
Power Density	$1/\Delta$
Energy/Instruction	Δ^2





Moore's Law Growth in Power Density

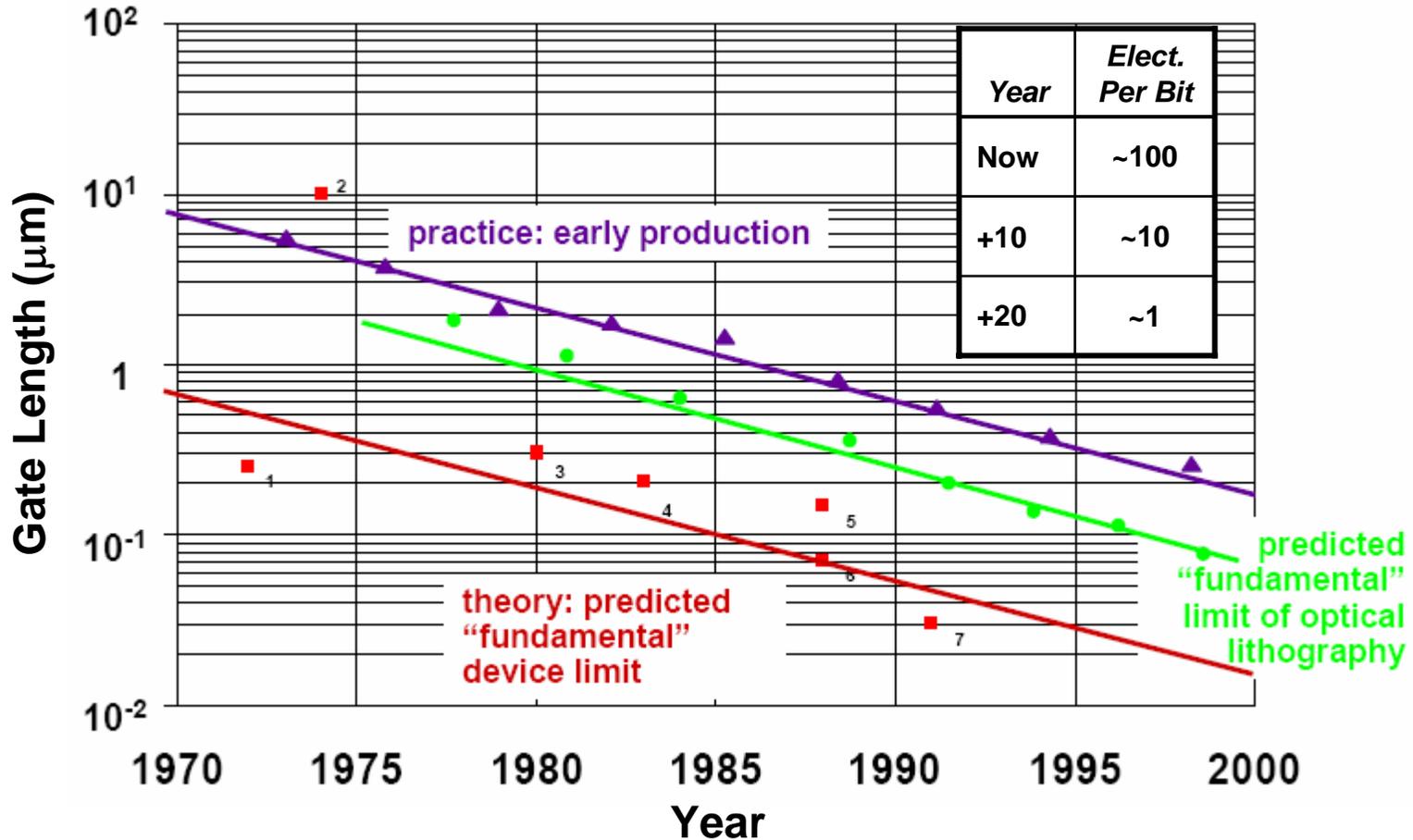




Moore's Law is Dead, Long Live Moore's Law!

Theory & Practice: Feature Size for MOSFET Devices

It's tough to make predictions, especially about the future. - Yogi Berra



Sources:

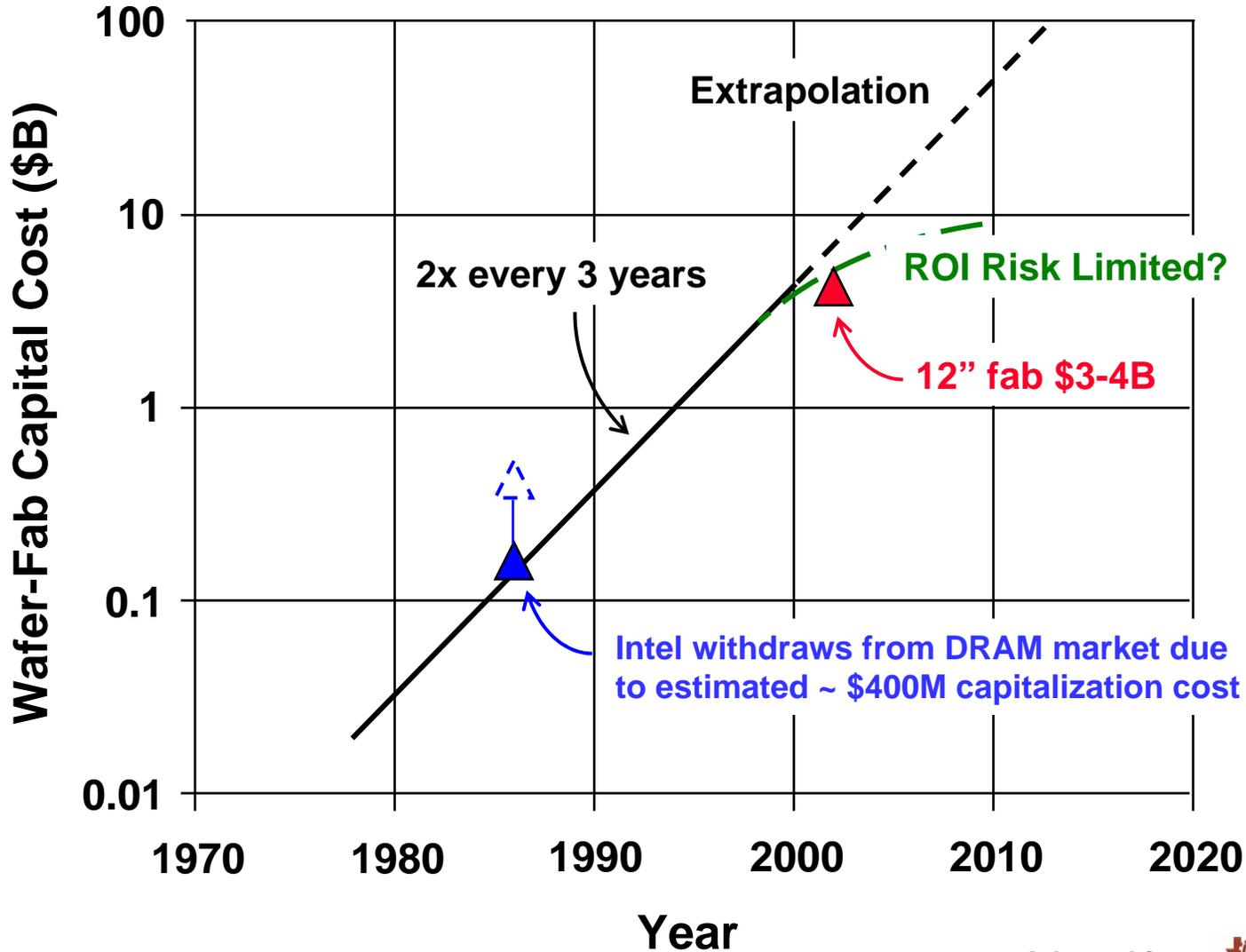
Combined graph and original concept: Lance Glasser, former Director, DARPA/ETO

Theory: Provided by Prof. David Ferry, Arizona State University

Practice: *The National Technology Roadmap for Semiconductors* (SIA Publication, 1994)



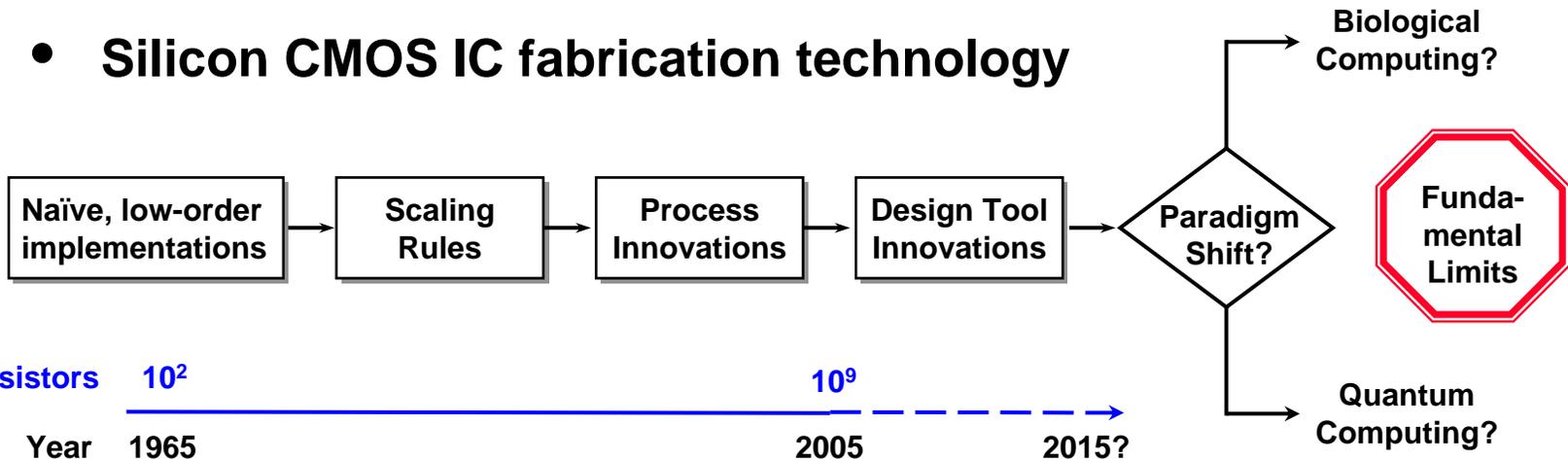
Capitalization Cost Impediments





Fulfillment and Impact of Moore's Prediction

- Silicon CMOS IC fabrication technology



- Examples of far-reaching impact

Altair 8800, 1975



Itanium, 2005



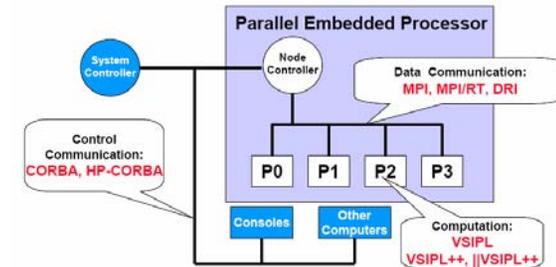
Exponential Improvements In Computing at a Fixed Price Point



Embedded Processors For Real-time Digital Signal Processing



Low-power Wireless Applications



Loosely-Coupled Hardware & Software Design Methodologies

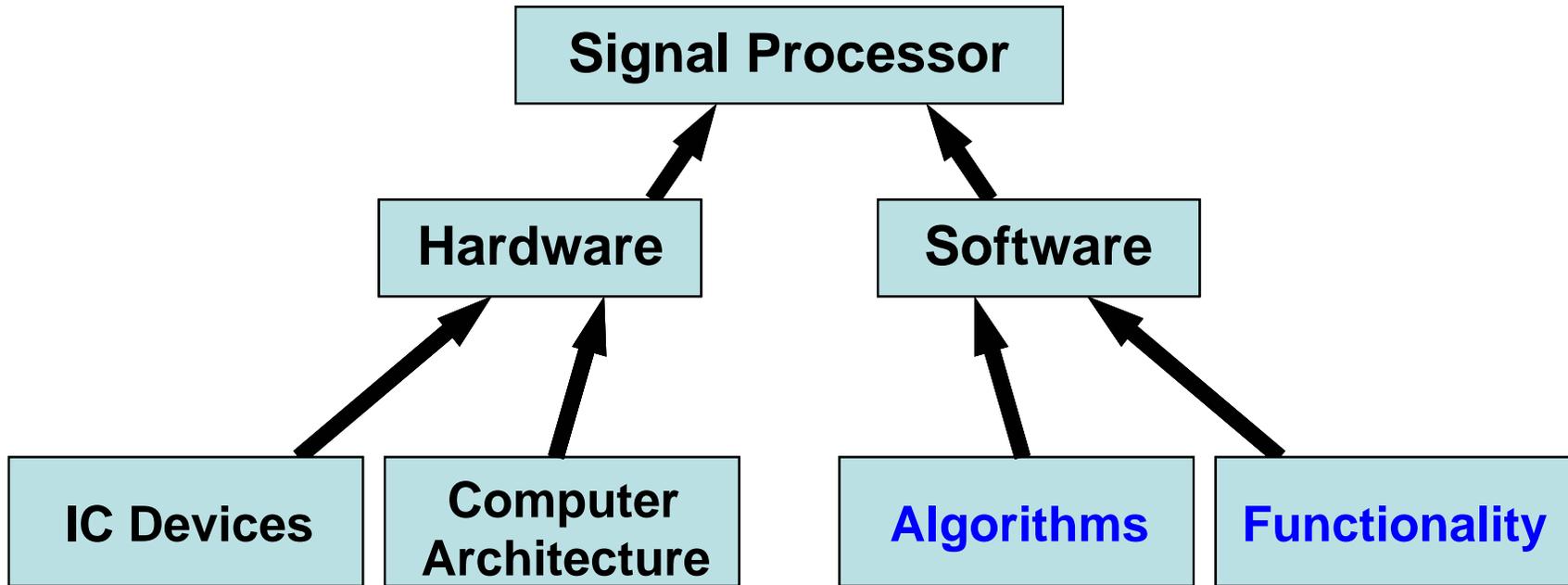


Outline

- Historical perspective - fulfillment of Moore's Law
- Impediments to continued IC density growth
- **Algorithms – the softer side of exponential growth**
- Implications regarding sustaining exponential growth
- Summary and Conclusions



Elements Contributing to Embedded Processor Performance



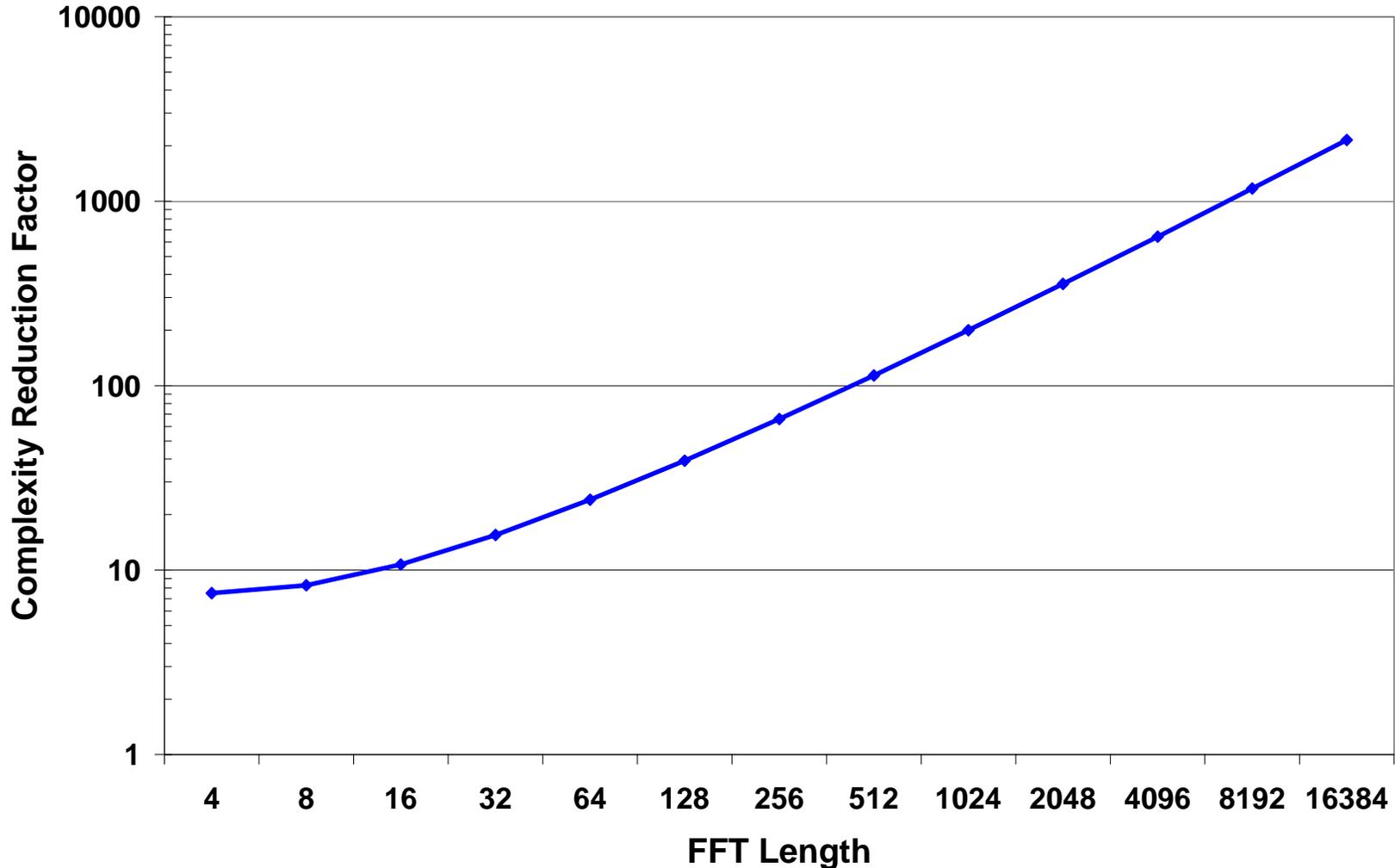


Different Character of Hardware (IC) Vs. Algorithm Improvements

<i>Improvement Metrics</i>	<i>Hardware</i>	<i>Algorithms</i>
Regularity	Predictable	Unpredictable
Dependent variable	Time	Order complexity
Impact on applications	Incremental	Leap-ahead
Useful lifetime	3 years or less	10 years or more
R&D Cost growth	2x in 3 years	1.11x in 3 years

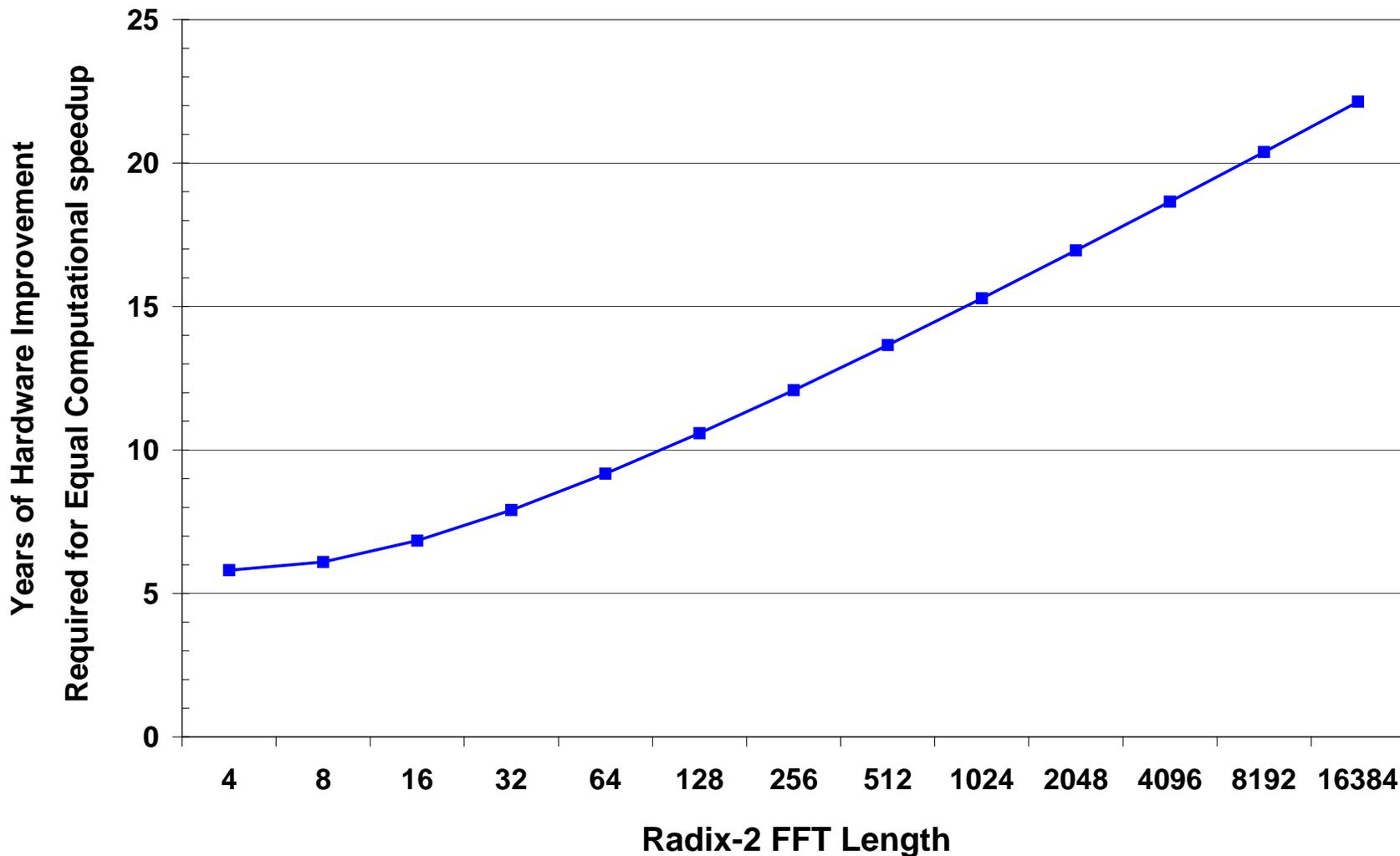


Computational Complexity Reduction Afforded by the FFT Over a Sum-of-Products DFT



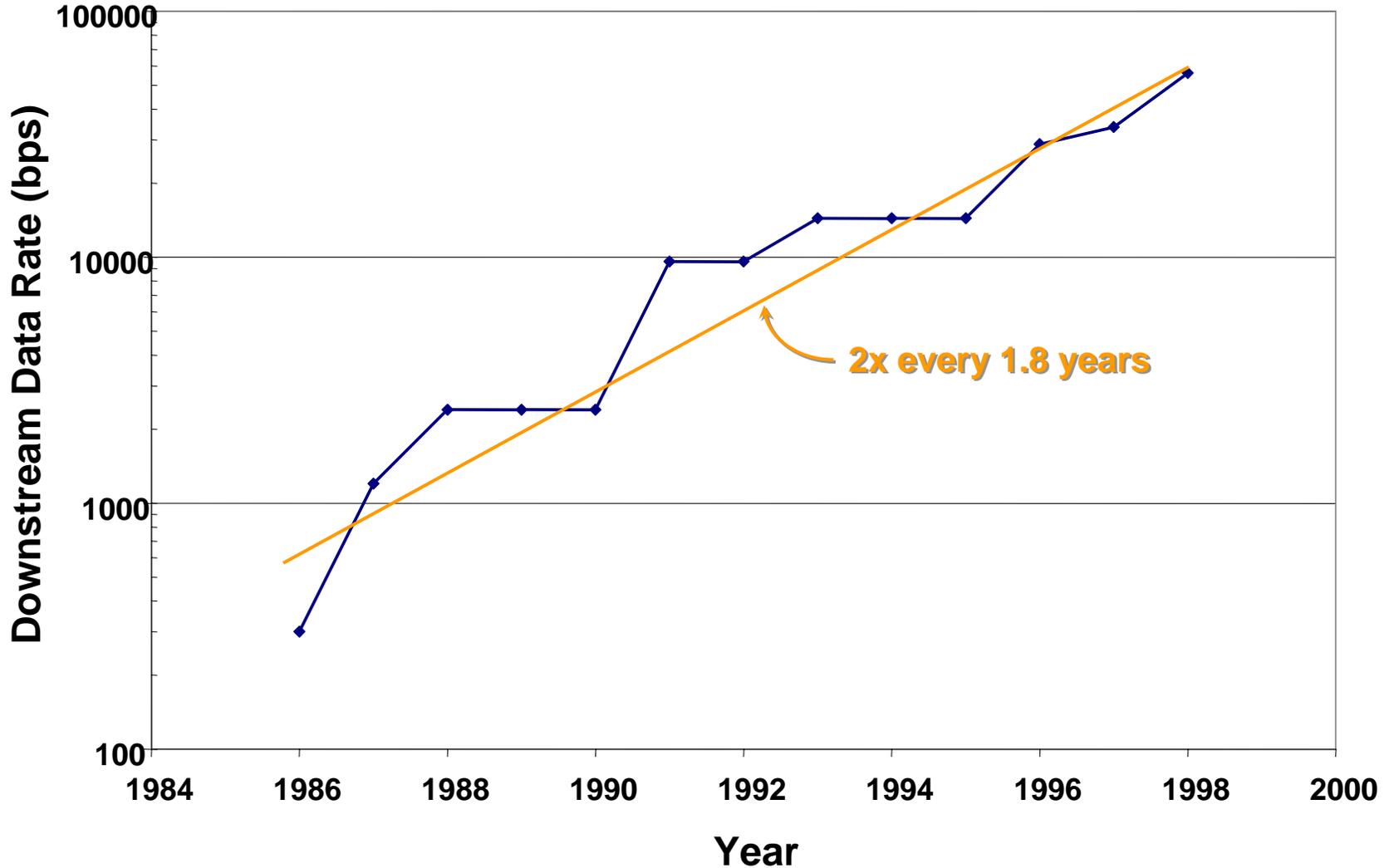


Moore's-Law Equivalent Years Required to Match FFT Computational Speedup



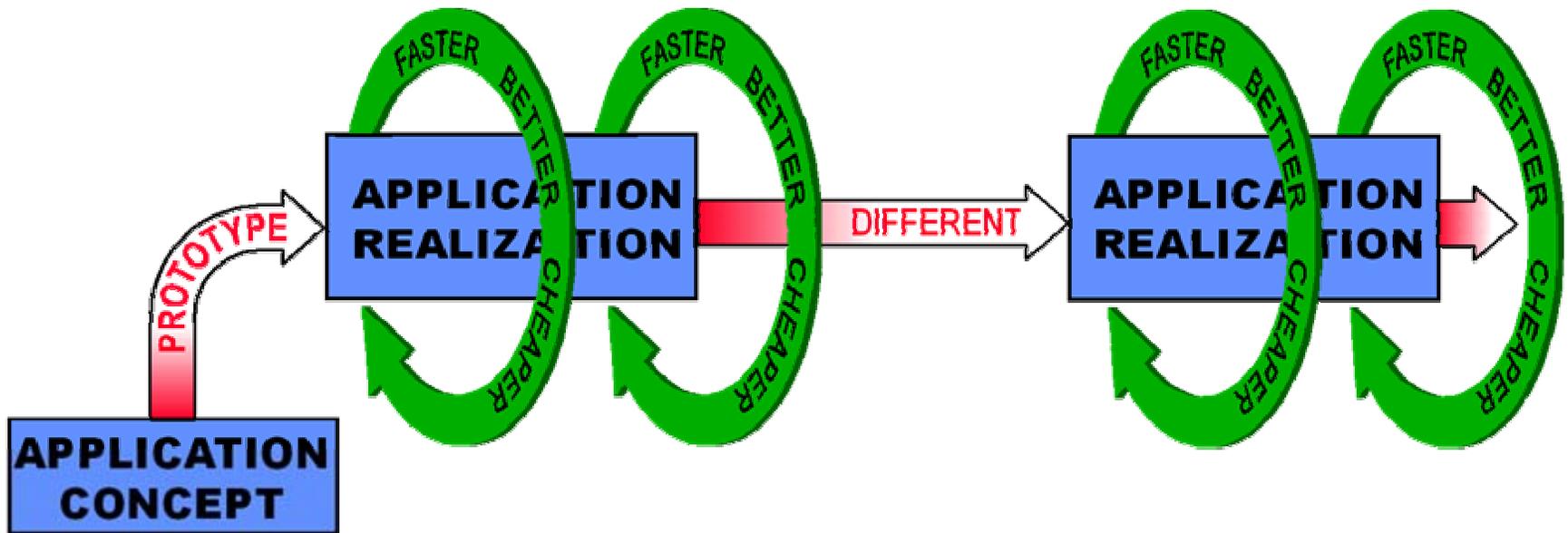


Exponential Improvement in Modem Rates





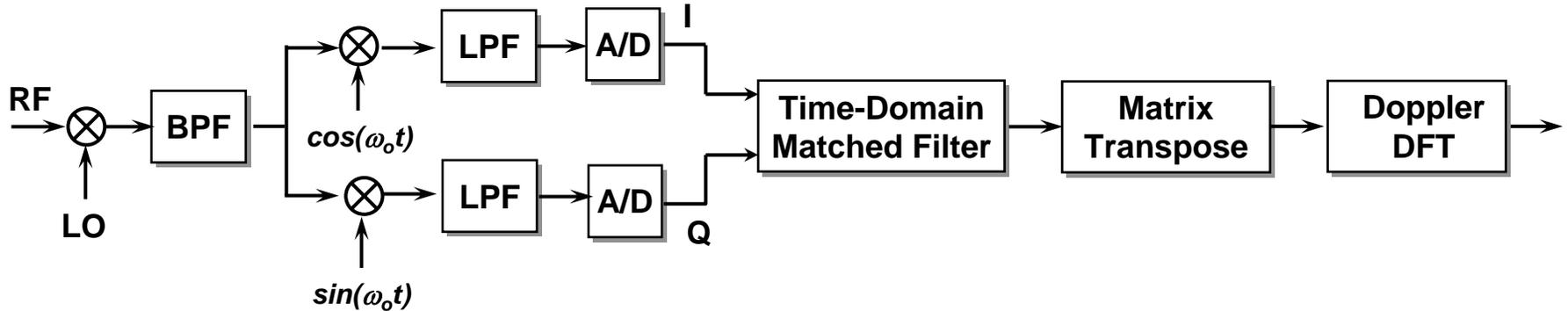
Application Maturation Cycle



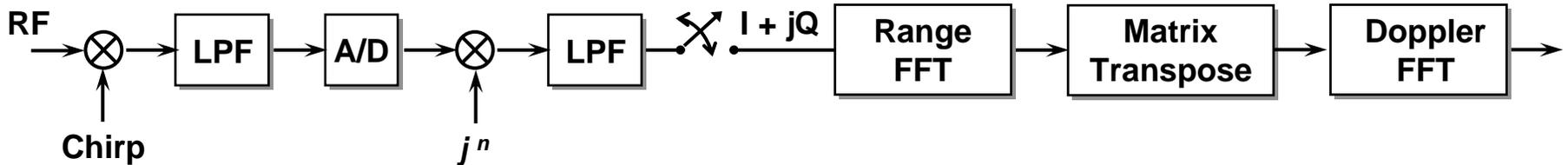


Pulse-Doppler Radar Example

- Algorithmically naïve implementation

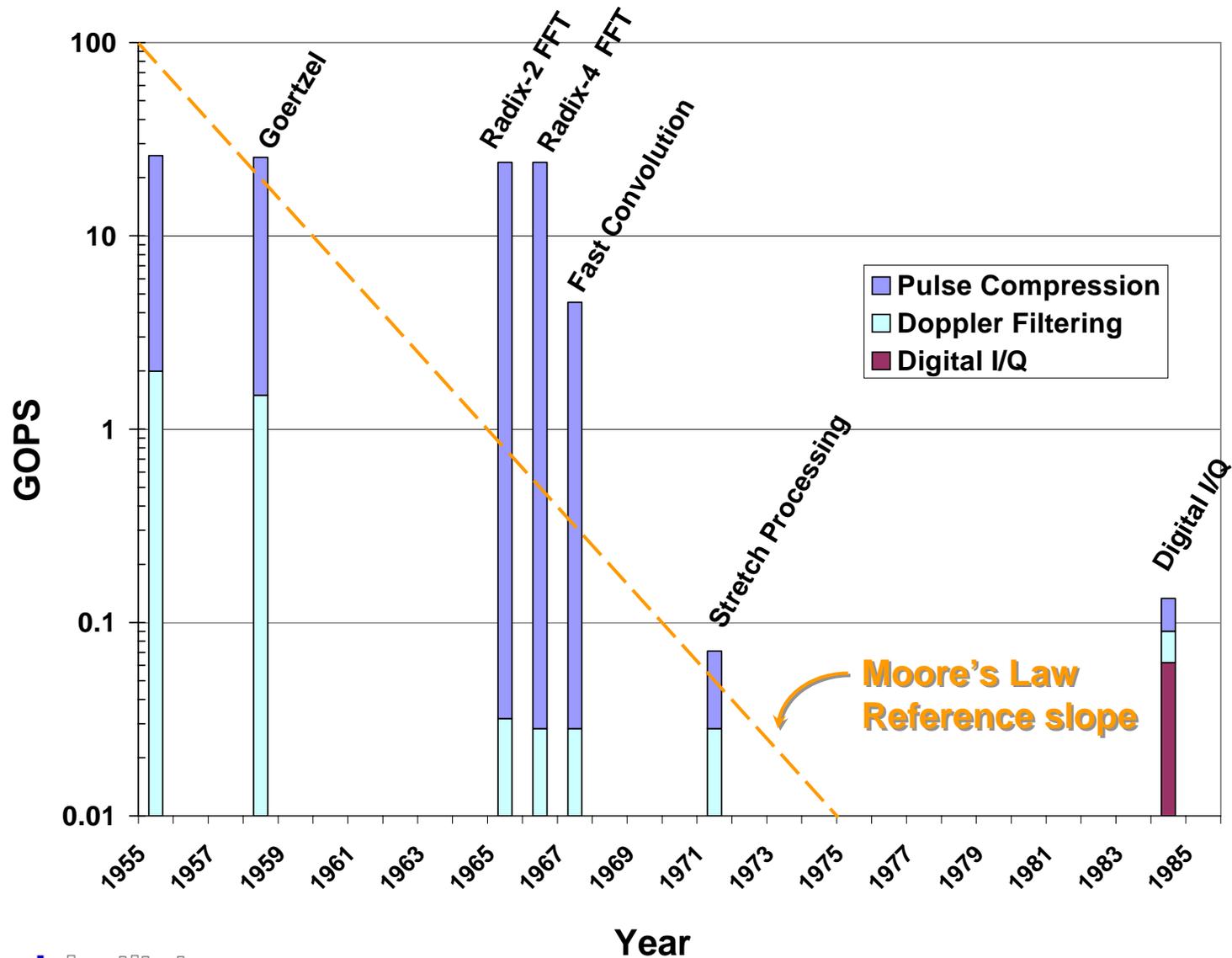


- Reduced-order implementation with digital I/Q





Pulse-Doppler Radar Algorithm Improvements





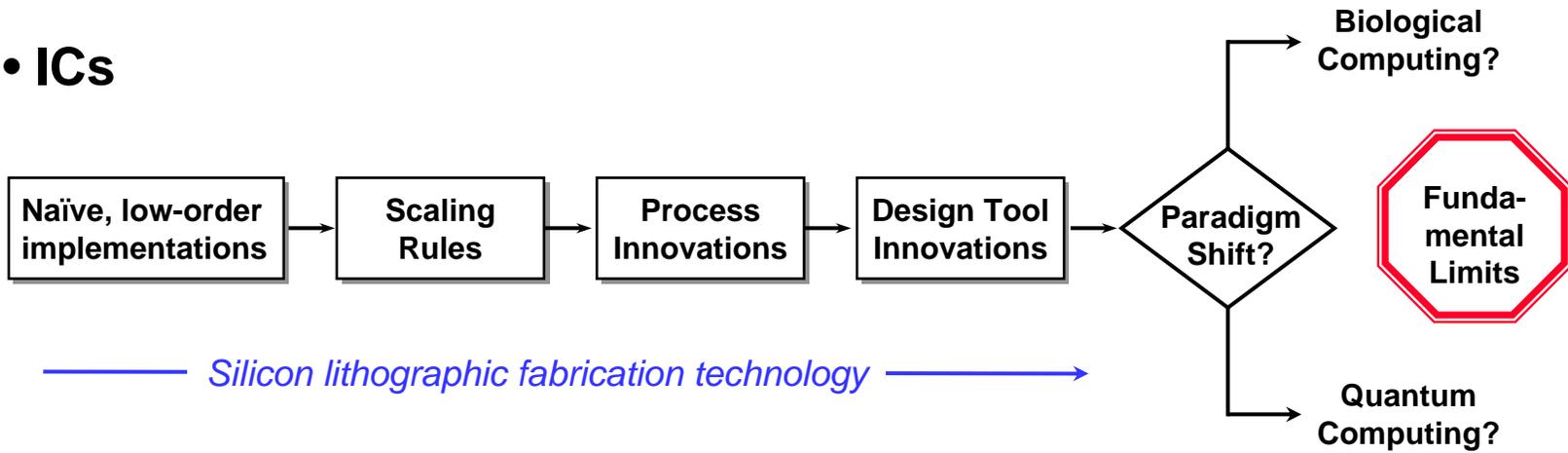
Outline

- **Historical perspective - fulfillment of Moore's Law**
- **Impediments to continued IC density growth**
- **Algorithms – the softer side of exponential growth**
- **Implications regarding sustaining exponential growth**
- **Summary and Conclusions**

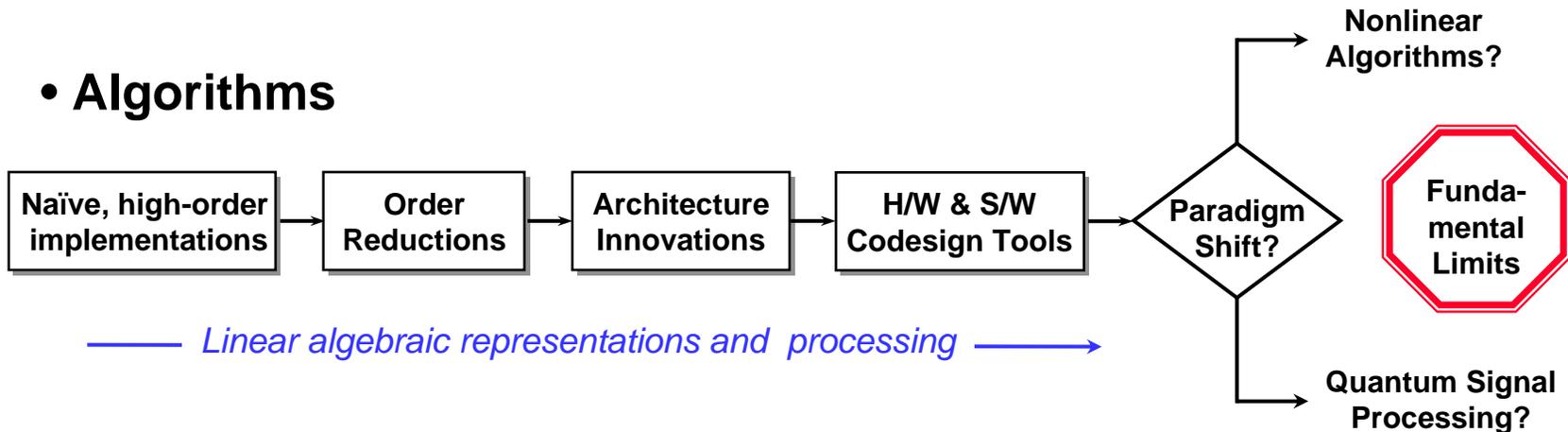


IC Vs. Algorithm Development (A Contrived but Useful Analogy)

• ICs

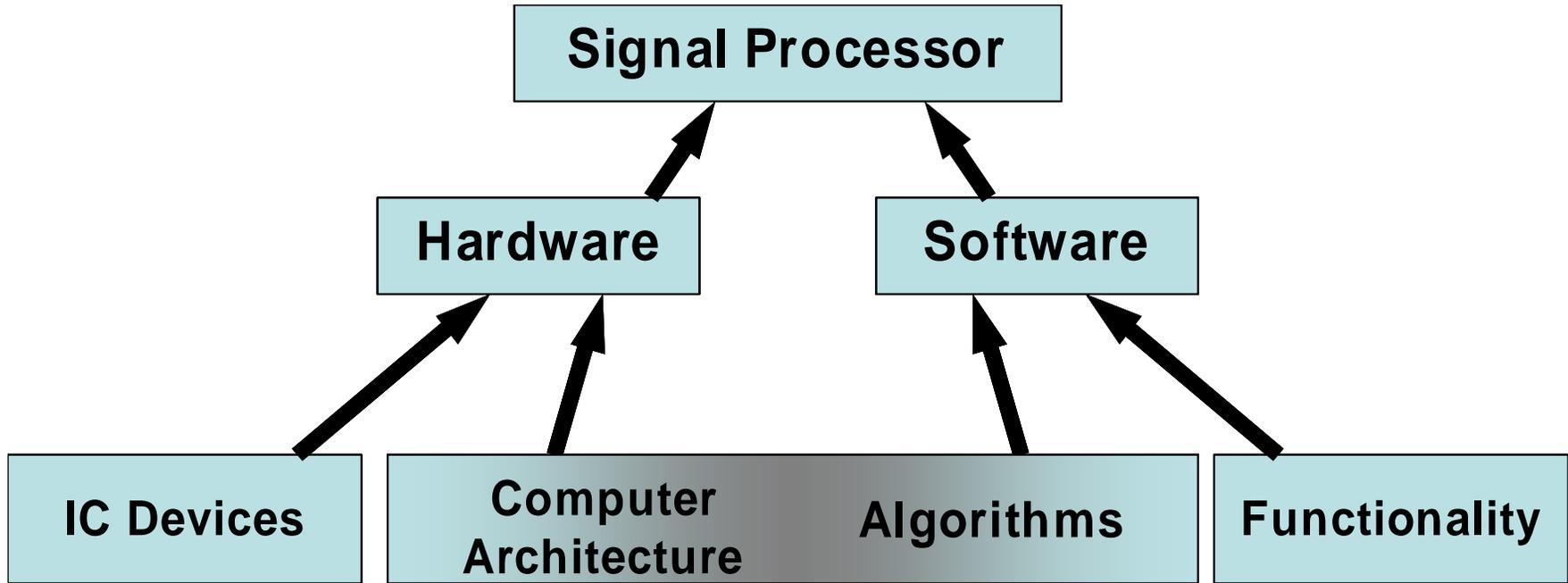


• Algorithms



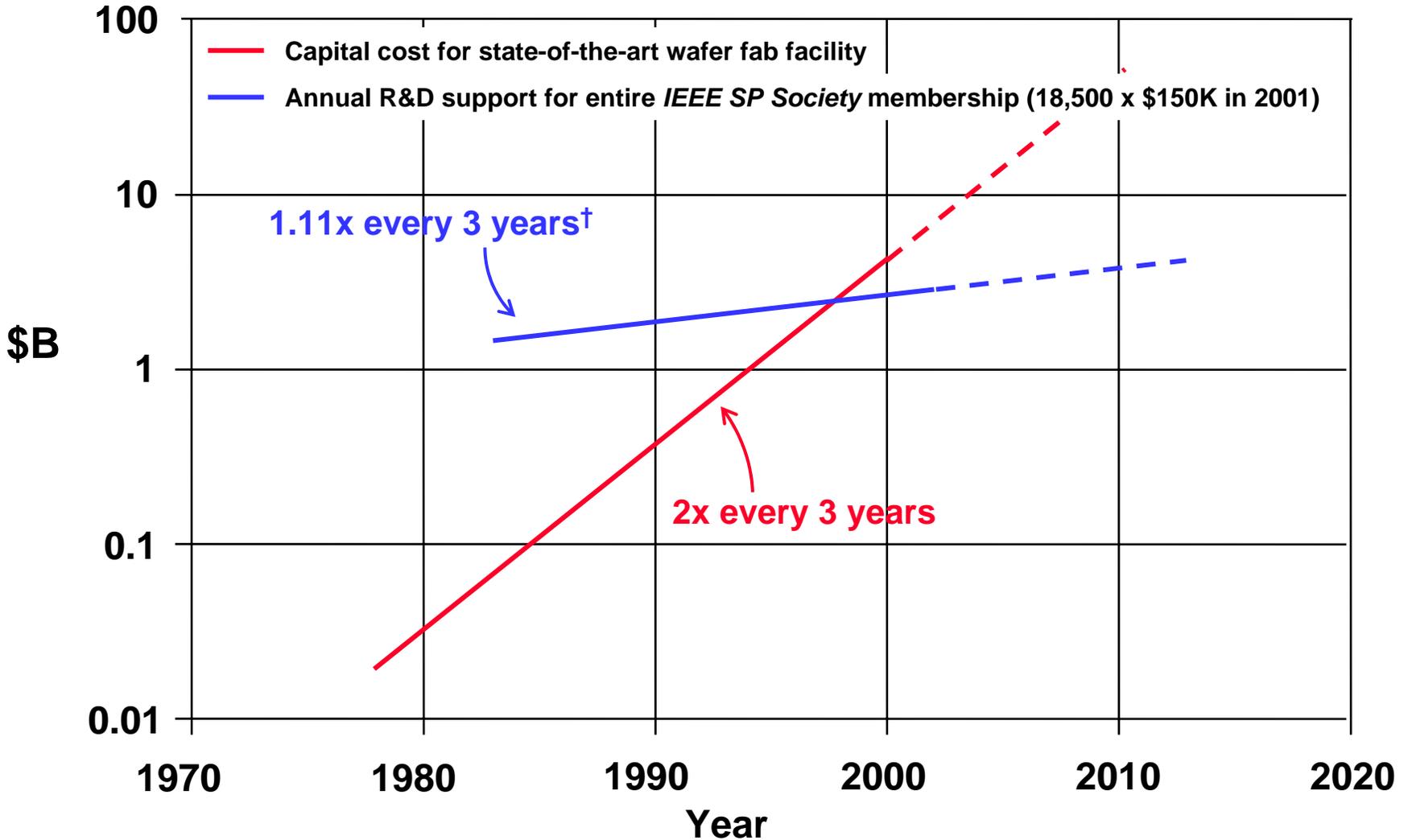


Increased Emphasis on Codesign Methodologies





Wafer-Fab Capitalization Cost Compared to Annual DSP Algorithm R&D Costs



† Salary inflation rate based on US Bureau of Labor and Statistics Median Engineering Salaries 1983-2003



Summary and Conclusions

- **Fulfilling Moore's Law**
 - Enabled by diverse, innovative R&D aimed at realizing a common vision (ITRS semiconductor roadmap)
 - Continued improvements may be impeded by a combination of thermal, quantum, and capital cost limits
- **Taking up the slack**
 - Over same 40-year time frame as Moore's Law, algorithm innovation has yielded exponentially improving performance as well
 - Algorithm innovation also enabled by diverse R&D, but without as clear of an industry-wide common vision
 - Algorithm R&D cost growth significantly lower than fab capital cost growth (1.1x vs. 2x every 3 years)
- **Increasing the effectiveness of algorithm R&D**
 - Develop better methods for quantifying the return on investment for algorithm R&D
 - Consider mechanisms for developing a broader industry vision and commitment to a long-term R&D roadmap
- **Hardware/software codesign methods increasingly important**