MICROCOSM

THE
PROVOCATIVE
NATIONAL
BESTSELLER
BY THE
AUTHOR OF
WEALTH AND
POVERTY

A prescient look inside the expanding universe of economic, social and technological possibilities within the world of the silicon chip

“Reading this book is like an entrance exam into the exciting 1990s and the glorious 21st century beyond.”

— *Los Angeles Times Book Review*

GEORGE GILDER

GEORGE
GILDER

The Quantum Revolution in
MICROCOSM
Economics and Technology

A TOUCHSTONE BOOK
Published by Simon & Schuster
New York London Toronto Sydney Tokyo Singapore

Touchstone

Rockefeller Center
1230 Avenue of the Americas
New York, New York 10020

First Touchstone Edition, 1990
TOUCHSTONE and colophon are registered trademarks
of Simon & Schuster Inc.
Designed by Levavi & Levavi/Nina D'Amario
Manufactured in the United States of America

1 3 5 7 9 10 8 6 4 2

7 9 10 8 Pbk.

Library of Congress Cataloging in Publication Data
Gilder, George F., date
Microcosm: the quantum revolution in economics and technology.
Bibliography: p.
Includes index.
1. Microelectronics industry. 2. Microelectronics—
Social aspects. I. Title.
HD9696.A2G55 1989 338.4'7621381 89-11476
ISBN 0-671-50969-1
ISBN 0-671-70592-X Pbk.

CONTENTS

Preface 11

PART ONE: *THE OVERTHROW OF MATTER* *15*

Chapter 1: The Message from the Microcosm 17

Chapter 2: The Prophet 32

Chapter 3: Wires and Switches 46

PART TWO: *THE TECHNOLOGY OF MIND* *59*

Chapter 4: The Silicon Imperative 61

Chapter 5: The Monolithic Idea 73

Chapter 6: Flight Capital 82

Chapter 7: Intel Memories 91

Chapter 8: Intel Minds 101

Chapter 9: The Curve of Declining Entropy 113

Chapter 10: Japan's Microcosm 127

Chapter 11: The CMOS Slip 139

Chapter 12: Mountain of Memories 150

PART THREE: *THE TRANSFORMATION OF CAPITAL* 163

Chapter 13: Mead's Theory 165
Chapter 14: From Boltzmann to Conway 180
Chapter 15: The Silicon Compiler 193
Chapter 16: Competing Visions 208
Chapter 17: The New Balance of Power 220

PART FOUR: *THE IMPERIAL COMPUTER* 233

Chapter 18: Patterns and Analogies 235
Chapter 19: Analog People 247
Chapter 20: The Age of Intelligent Machines 262
Chapter 21: The IBM Machine 280
Chapter 22: The Neural Computer 290
Chapter 23: The Death of Television 307

PART FIVE: *THE QUANTUM ECONOMY* 317

Chapter 24: The New American Challenge 319
Chapter 25: Revolt against the Microcosm 331
Chapter 26: The Law of the Microcosm 345
Chapter 27: The Eclipse of Geopolitics 353
Chapter 28: Triumph over Materialism 371

Bibliographical Notes 385

Acknowledgments 403

Index 407

PREFACE

Listen to the technology and find out what it is telling you.
—CARVER MEAD

This book is an exploration of the meaning and future of modern technology. With its origins in quantum physics and its embodiment in the microchip, the exemplary product of this technology is the computer. Still in the infancy of its influence and power, the computer in its current form is the product of some twenty-five years of development. In all its manifestations—reaching from the microprocessor in a car's fuel injection system to the microprocessor at the heart of a personal computer, from a supercomputer modeling the weather to a worldwide ganglion of fiber-optic threads of glass and light—computer technology epitomizes the fruits of the microcosm of quantum physics. It was quantum theory early this century that revealed the inner structure of matter for the first time and made modern computers possible.

Broadly considered, the computer is the most important product of the quantum era. By exploring this central machine of the age, however, we discover not the centrality of machines and things but the primacy of human thought and creativity.

The quantum era is still unfolding in a fourfold transformation of the world—in science, technology, business, politics—and even in philosophy. But all the changes converge in one epochal event: the overthrow of matter.

The change originates in the microcosm of quantum theory itself —the new physics launched in Europe early this century—which overthrew matter in the physical sciences. At the foundation of the universe, Isaac Newton's hard, inert, and indivisible solids gave way to a rich panoply of paradoxical sparks, comprising waves and particles that violate every principle of Newtonian solidity. At the root of all the cascading changes of modern economic life—devaluing material resources in technology, business, and geopolitics—is this original overthrow of material solidity in the science of matter itself.

The second step in the overthrow of matter came in the use of quantum theory to overcome the material limits of weight, heat, and force in the creation of new machines. The industrial age essentially managed and manipulated matter from the outside, lifting it against gravity, moving it against friction, melting or burning it to change its form. The quantum era manipulates matter from the inside, adapting its inner structure to human purposes.

In the microchip, combining millions of components operating in billionths of seconds in a space the size of the wing of a fly, human beings built a machine that overcame all the conventional limits of mechanical time and space. Made essentially of the silicon in sand— one of the most common substances in earth—microchips find their value not in their substance but in their intellectual content: their design or software.

The third great manifestation of the overthrow of matter is the impact of this technology on the world of business. By overcoming the constraints of material resources, the microchip has devalued most large accumulations of physical capital and made possible the launching of global economic enterprises by one entrepreneur at a workstation. With the overthrow of the constraints of material scarcity, gravity, and friction, large bureaucracies in government and business lose their power over individual creators and entrepreneurs.

The fourth phase of the overthrow of matter is the collapse of the value of natural resources and territory in determining the distribution of power among nations. The microcosms of science, technology, and enterprise have converged in a global quantum economy that transcends all the usual measures of national power and wealth. The most valuable capital is now the capital of human mind and spirit. Intellectual capital can transform any physical environment, even a few small, cold, stony islands off the eastern coast of Asia, into a center of production and wealth. And the lack of such capital, or its abuse, can turn the greatest empire into a hollow shell reverberating with the frustration of tyrants.

In overthrowing the thrones of matter, this new epoch—the quantum era—also overthrows the great superstitions of materialism. Worship of things—whether in a Marxist dialectic or a Midas's hoard—collapses in a world in which thought is paramount even at the heart of matter itself.

CHAPTER 1

The Message from the Microcosm

> *However much discussed and however promising this atomic theory might appear, it was, until recently, regarded merely as a brilliant hypothesis, since it appeared to many far sighted workers too risky to take the enormous step from the visible and directly controllable to the invisible sphere, from the macrocosm to the microcosm.*
>
> —MAX PLANCK, inventor of quantum theory

The central event of the twentieth century is the overthrow of matter. In technology, economics, and the politics of nations, wealth in the form of physical resources is steadily declining in value and significance. The powers of mind are everywhere ascendant over the brute force of things.

This change marks a great historic divide. Dominating previous human history was the movement and manipulation of massive objects against friction and gravity. In the classic image of humanity, Atlas bears the globe on stooped shoulders, or Sisyphus wrestles a huge rock up an endless slope. For long centuries, humans grew rich chiefly by winning control over territory and treasure, slaves and armies. Even the Industrial Revolution depended on regimented physical labor, natural resources, crude energy sources, and massive transport facilities. Wealth and power came mainly to the possessor of material things or to the ruler of military forces capable of conquering the physical means of production: land, labor, and capital.

Today, the ascendant nations and corporations are masters not of land and material resources but of ideas and technologies. Japan and other barren Asian islands have become the world's fastest-growing economies. Electronics is the world's fastest-growing major industry.

Computer software, a pure product of mind, is the chief source of added value in world commerce. The global network of telecommunications carries more valuable goods than all the world's supertankers. Today, wealth comes not to the rulers of slave labor but to the liberators of human creativity, not to the conquerors of land but to the emancipators of mind.

Impelled by an accelerating surge of innovation, this trend will transform man's relations with nature in the twenty-first century. The overthrow of matter will reach beyond technology and impel the overthrow of matter in business organization. Devaluing large accumulations of fixed physical capital, the change will favor entrepreneurs over large bureaucracies of all kinds. The overthrow of matter in business will reverberate through geopolitics and exalt the nations in command of creative minds over the nations in command over land and resources. Military power will accrue more and more to the masters of information technology. Finally, the overthrow of matter will stultify all materialist philosophy and open new vistas of human imagination and moral revival.

The exemplary technology of this era is the microchip—the computer inscribed on a tiny piece of processed material. More than any other invention, this device epitomizes the overthrow of matter. Consider a parable of the microchip once told by Gordon Moore, chairman of Intel and a founding father of Silicon Valley:

"We needed a substrate for our chip. So we looked at the substrate of the earth itself. It was mostly sand. So we used that.

"We needed a metal conductor for the wires and switches on the chip. We looked at all the metals in the earth and found aluminum was the most abundant. So we used that.

"We needed an insulator and we saw that the silicon in sand mixed with the oxygen in the air to form silicon dioxide—a kind of glass. The perfect insulator to protect the chip. So we used that."

The result was a technology—metal oxide silicon (MOS), made from metal, sand, and air—in which materials costs are less than 1 percent of total expense. Combining millions of components on a single chip, operating in billionths of seconds, these devices transcend most of the previous constraints of matter. The most valuable substance in this, the fundamental product of the era, is the idea for the design.

The microchip not only epitomizes but also impels the worldwide shift of the worth of goods from materials to ideas. These transvaluations are not mere luck, to be annulled by some new scarcity. Nor do they reflect only the foresight of one industry, summed up in Moore's parable. The rise of mind as the source of wealth spans all industries and reflects the most profound findings of modern physics and philosophy. The overthrow of matter in economics is made possible by the previous overthrow of matter in physics. All the cascading devaluations of matter in the global economy and society originate with the fundamental transfiguration of matter in quantum science.

Max Planck, the discoverer of the quantum, offered the key when he asserted that the new science entailed a movement from the "visible and directly controllable to the invisible sphere, from the macrocosm to the microcosm." The macrocosm may be defined as the visible domain of matter, seen from the outside and ruled by the laws of classical physics. The microcosm is the invisible domain, ruled and revealed by the laws of modern physics.

The borders of these rival domains are not set by size or location. Some of the largest phenomena of the age—from global telecommunications to nuclear explosions—follow the quantum laws of the microcosm. Indeed, because atoms observe the microcosmic rules, the microcosm is ubiquitous and comprises all objects large and small. But because the microcosm is invisible, it has been relatively uninfluential in shaping public views and philosophies.

It is understandable that humans resist the microcosm and even rebel against it. Quantum theory is an abstruse and difficult set of ideas. It baffles many of its leading exponents and it perplexed Albert Einstein to his grave. Defying the testimony of the human senses, the new physics is contrary to all human intuition and metaphor. In the quantum domain, all conventional analogies of physics—such as tops, springs, and billiard balls—are radically misleading. Therefore, we cannot "understand" quantum theory in the way we can comprehend classical physics. Quantum theory simply does not make sense.

The reason the new physics does not make sense to most humans, however, is that prevailing common sense is wrong. Common sense serves the materialist superstition: the belief that we live in a world of solid phenomena, mechanically interconnected in chains of cause and effect. The common wisdom of mankind has yet to absorb the simple truth that in proportion to the size of its nucleus the average atom in one of our cherished solids is as empty as the solar system. Few ponder the fact that an electron—a key to physical solidity—does not occupy any specific position in space; in a famous experiment, a single electron passes simultaneously through two separate holes in a screen. Such quantum images are difficult for human beings to grasp or believe. Humans balk at the basic paradox of a physical theory that defies the testimony of human senses and overthrows matter in the very science of matter itself.

In various grudging ways, physical scientists all recognize this overthrow of matter. As the author of the quantum theory of chemistry, Linus Pauling developed the fundamental model of what we call matter. He knows as much about it as any other man. But at the beginning of his popular text, *Chemistry,* he concedes that "no one really knows how to define matter. . . ." Like virtually all physical scientists, he then proceeds to use the term anyway. Yet the paramount theme of the new physics is that what scientists now call matter is totally unlike what classical physicists used to call matter and completely alien to the solid objects that the term calls to mind.

A key reason that quantum theory is so difficult to understand, so apparently riddled with paradox, is this continued use of the vocabulary of materialism to discuss phenomena that in the usual sense lack all material qualities, such as solidity, location, continuity, and inertia. On the other hand, it is this very transcendance of material limits that explains why quantum technologies far outperform mechanical devices and promise a new era of technology and economics.

Separating the old and new sciences is a nearly unbridgeable gulf. Compare Pauling's agnosticism about the definition of matter with Isaac Newton's materialist confidence. Newton described matter as "solid, massy, hard, impenetrable, movable particles . . . even so very hard, as never to wear or break in pieces; no ordinary power being able to divide what God himself made one in the first creation . . . [to] compose bodies of one and the same nature and texture in all ages." Newton's matter at heart was inert, opaque, and changeless.

Just as important, Newton had no problems with common sense. His science was anthropomorphic. He assumed that matter at its fundamental levels behaved like the material objects that we perceive. "We no other way know [the characteristics of matter] than by our senses. . . ." Its impenetrability, for example, "we gather not from reason but from sensation."

For some two hundred years nearly all leading scientists shared these materialist assumptions, based on sensory models and determinist logic. Newton's great eighteenth-century countryman Adam Smith extended the materialist metaphor to society, contending that the economy itself is a clockwork, a "great machine." Later Karl Marx applied materialism to the very fabric of political ideas, which were deemed mere figments of ownership in physical capital, or of alienation from it. Sigmund Freud and his followers developed a psychological theory of forces and pressures, inhibited or released, built up or fed back, much like the classical mechanics of steam engines.

Today most sophisticated people imagine that they have transcended Newton and have come to terms with the findings of modern science. But they have not. As an intellectual faith, materialist logic still prevails. Even theologians and philosophers who spurn materialism in defining the meaning of life accept it as the lesson of science. They still believe that the solid world they see and feel—governed by determinate chains of cause and effect, rooted in Newtonian masses and forces—is real and in some sense definitive.

The atom may not be ultimate, but they assume some other particle is, perhaps the quark. It has become a cliché to call the quark "the fundamental building block of nature." At the foundations of the physical world, so it is supposed, are physical solids—"building blocks"—that resemble in some way the solids we see. They link together in causal chains of mechanical logic like a set of cogs and levers. These solids are deemed to comprise all matter, from atoms and billiard balls to bricks and the human brain.

Such beliefs are manifest in how we think. MIT mathematician Gian Carlo Rota has written: "The naive prejudice that physical objects are somehow more 'real' than ideal objects remains one of the most deeply rooted in Western culture. . . . A consequence of this belief—which until recently was not even perceived as such—is that our logic is patterned exclusively on the structure of physical objects." A logical argument, for example, works very much like a material machine. Many of us still assume that like material things, causes and effects are directly and mechanically related in time and space.

Since most visible objects and machinery seem to observe the macrocosmic rules most of the time, most human beings are comfortable in a macrocosmic world. Like primitive tribes, they worship things they can see and feel. They think in logic based on the behavior of these things. In all these attitudes they cling to the materialist superstition: the belief that mechanical and mindless interactions of inert and impenetrable matter are the ultimate foundation of reality. They resist the apparently murky and paradoxical message from the microcosm: the overthrow of matter in a quantum world.

Understood as episodes in the dismantling of Newtonian matter, however, the historic discoveries of the new physics follow a coherent and powerful logic. The move from macrocosm to microcosm can be seen as a progress from a material world composed of blank and inert mally disturbed materials release distinctive and identifiable wave patterns. Every solid has a specific frequency. In the prevailing theory of a kind of subatomic solar system, the frequencies of these waves were assumed to signify the rate at which the electrons were rotating around the nucleus. But like the waves shed from a furnace in Planck's theories, these radiations persistently violated classical predictions.

An eminent Danish physicist, Bohr suggested that rather than measuring the frequency of electron orbits, these waves signified the energy emitted by electrons in jumping from one orbit to another. Dropping from a higher to a lower energy state, an electron released a photon of light of its resonant frequency. Absorbing a photon of its resonant frequency, an electron would move to a higher energy level. In quantum theory, an atom was not a system of continuous orbits but a hierarchy of distinct energy states.

Announced in 1913 and proved for the single electron of the hydrogen atom, the Bohr model was the first great vindication of quantum theory. One test of scientific advance is whether it extends the realms of human understanding and control. Bohr's breakthrough ultimately opened the microcosm as an industrial site.

The established physics could not explain the effectiveness of chemistry, let alone extend it to atoms. Unlike a solar system, atoms do not exist in majestic isolation. Ceaselessly in movement, they endlessly jiggle together in what is called Brownian motion. People even step on them. In a world of Newtonian continuities, electron orbits would vary continually as atoms collided with one another. Constantly knocked loose in these collisions, electrons in a conductor should flow far more copiously and respond to heat more massively than experiments showed.

In this realm of readily flowing and colliding particles, chemical elements would not be stable. Elements would ooze together rather than assume the fixed properties of the periodic table, the basis of all chemistry. Negative electrons, in fact, would rapidly spin into the positive nucleus, collapsing the matter of the universe. Far from enabling the vastly precise functions of microelectronics, Newtonian physics could not even explain why you do not fall through the floor, and then through the world.

Reunifying chemistry and physics in the microcosm, the new model of the atom explained the apparent solidity of the physical world. Establishing a gap, called a band gap, between an electron in its ground state and an electron excited to a higher energy level, the new physics showed why the constant collisions of atoms do not cause the atomic structure to collapse. A small collision will not affect an atom. Just as a musical string can resound only in its fundamental and resonant frequencies (all others canceled out by the rules of wave interference), an electron will not respond to any small disturbance. It will react only if it receives its necessary quantum of energy, defined by its resonant frequency times Planck's constant.

In this way the quantum concept specified precisely the energy needed to change an atom or to connect it with another atom. Thus the new physics defined a grammar of chemical combinations. By restricting electron energies to specific quanta and by restricting the number of electrons in each energy state through the laws of quantum wave interference, Bohr's model brought an intelligible order to the atom. Revealed as a fixed mathematical construct, one atom could connect to another by fixed laws.

Bohr and his followers thus opened the way to understand the forces of attraction and repulsion among atoms. Only atoms with a missing electron in their outermost energy level are chemically active. Electrons tied to their atoms create stability; electrons freed of their atoms create electricity; electrons reaching out resonantly to other atoms or switching back and forth between atoms tie together the elements of chemistry, such as silicon, oxygen, and aluminum. By regulating the atomic binding and unbinding of electrons, the new theory illumined the laws of the chemical bond and the nature of electricity.

Without this new understanding, we could never adequately comprehend the flow of electricity through a solid. Nor could we see how to make a kind of semi-conducting sand become a conductor under certain precise conditions—thus rendering it an excellent pathway for information. These and thousands of other precise manipulations would be impossible without quantum mechanics.

Knowledge is power. Quantum physics gave humans access to the inner structure of matter. As Mitchell Feigenbaum, a pioneer of chaos theory, put it, quantum theory "tells you how you can make computers from dirt." But in overcoming Newtonian solidity, quantum scientists failed to offer a coherent or intelligible alternative image of matter.

As a result, the most profound discoveries of the twentieth century failed to penetrate the public consciousness, intellectual life, or natural philosophy. Most people continued in the old materialist idiom, forcing the new concepts into the old frame. Others used the quantum as a warrant for a murky subjectivism totally alien to the spirit of the science, which is rigorously objective and supremely practical. Like all fundamental shifts in the scientific view of reality, the quantum perspective has been slow in working its way through the culture. Bound by the necessary conservatism of their trade, scientists hesitate to traffic in the unprovable domains of meaning and non-scientists demur at entering the abstruse domains of science.

Such problems of the microcosm are evident in the history of the theory of the electron, the basic entity of electronics. Most interested people understand much of what electrons do. But very few have any clear idea of what an electron actually is, or its implications for the concept of matter and its overthrow in the world economy.

From the telephone to the human brain, from the television set to the computer, information mostly flows in the form of electrons. This function of electrons has quantum roots. As in Planck's black body radiation, electrons do not respond to applied energy in a continuous, proportional, or linear way. They are non-linear; they have quantum thresholds and resonances. These quantum functions shape their electrical properties. In order to move through a solid, electrons must be freed from their atoms, jumping from one energy state to a free state across measurable energy "band gaps" in strict accordance with quantum rules. These rules give electrons identifiable and controllable features that can be used to convey information.

With controlled pulses of electrons down wires, computers could be interconnected around the world. With controlled flows of electrons in and out of tiny capacitors, computer memories could be constantly read, written, and restored. With minute charges of electrons in silicon, computer transistors could be switched on and off. The most studied phenomenon in physics, electrons are constantly measured, manipulated, traced, aimed, and projected. Yet throughout the history of science, the electron has remained a humbling perplexity. Let us listen, and find out what it is telling us about the bizarre abundance of the domains beyond matter.

At the time of Bohr's breakthrough, the electron was assumed to be matter. By contrast with the photon, the source of its energy, it showed measurable mass. But in 1923 Louis de Broglie developed a wave theory of the electron and Einstein endorsed it. In 1926, Erwin Schrödinger contrived a famous equation, still widely used by engineers and physicists today, that describes the electron's wave behavior. The following year, Werner Heisenberg offered a set of strange mathematical matrices derived from particle concepts that explained electron behavior as well as Schrödinger's wave equation did. Schrödinger then strengthened the theory immeasurably by showing that Heisenberg's particle paradoxes were mathematically convertible into his own wave equation. Like light, electricity proved to be a flow of wave-particles.

As if to banish any chance of a revival of Newtonian theory at the heart of the atom, Heisenberg then presented a radical new concept called the Uncertainty Principle. He asserted the intrinsic impossibility of ascertaining at once both the momentum and location of an electron. Thus Heisenberg showed that the necessary parameters of Newtonian equations were necessarily unmeasurable at the atomic level.

In subsequent years, the hugely more massive protons and neutrons of the atomic nucleus were also shown to exhibit wave action. Wave behavior was clearly essential to the quantum domain. Many investigators began to see the world as consisting of simple waves of energy that could consolidate into mass. But this precarious trench of materialism also became untenable.

Crossing decisively into the microcosm, Heisenberg declared that the waves which Bohr had examined in recreating the atom were not conventional waves at all. Designated "probability amplitudes," they were waves or fields that defined the statistical likelihood of finding an electron at any particular location.

This was a climactic step in the overthrow of materialism in physics. With the electron itself depicted as a wave and the wave depicted as a probability field, the specific particle in this theory had disappeared into a cloud. With it disappeared the last shreds of Newtonian logic and mechanistic solidity. As Bohr put it, quantum theory required "a final renunciation of the classical idea of causality and a radical revision of our attitude toward the problem of physical reality."

Sixty years later, the quantum cloud has still not been dispersed. Nobel Laureate Steven Weinberg recently summed up the current state of the argument:

> The inhabitants of the universe were conceived to be a set of fields —an electron field, a proton field, an electromagnetic field—and particles were reduced to mere epiphenomena. In its essentials, this point of view has survived to the present day and forms the central dogma of the quantum field theory: *the essential reality is a set of fields* subject to the rules of special relativity and quantum mechanics; all

> else is derived as a consequence of the quantum dynamics of those fields.

Einstein made the point even more strongly: "There is no place in this new kind of physics both for the field and matter, for the field is the only reality."

What is the field? As Einstein explains, "Instead of a model of actual space-time events, it gives the probability distributions for possible measurements as functions of time." Although actual waves measured by physical apparatus, the fields of atoms are less fields of force than fields of information.

This idea of information fields cannot be translated into a physical analogy. A probability wave is a wave—a detectable force—but it is also an idea, an index of the likelihood an electron will appear at one place rather than another. It is a field of information.

The information remains full of paradox. But the paradoxes all derive from the materialist superstition. All analysts want to retain the last purchase of solidity, the concept of a particle. But at the most fundamental level, there is no such thing as a particle.

What are called elementary particles are neither elementary nor particles. If elementary is taken in the Newtonian sense of indestructible matter, Heisenberg points out, electrons are not elementary because they dissolve on contact with their antimatter twin, the positron. Envisaged as more intense points moving through a wave of probability, electrons do not even retain any particular substance.

The most famous of many efforts to resolve such perplexities came in experiments suggested by Niels Bohr, based on Thomas Young's original two-hole scheme for showing the wave nature of light. Once again, researchers shoot an electronic "gun" at two tiny holes in a screen in front of an electrically sensitive target. Scientists first assumed that wave interference arose when the "waves" of electrons collided in some way. But the assemblage of hits on the target can show Young's wave interference pattern even when electrons are shot one by one minutes or hours apart. The waves are not "crowd effects"; they are intrinsic to each quantum particle.

In some experiments, a single quantum particle passes through both holes at the same time, thus exhibiting wavelike non-locality. In some crucial microelectronic products, from diodes to non-volatile memories, electrons use probability waves to "tunnel" through barriers that they are incapable of either penetrating or surmounting as physical particles.

more predictable than the waves in your bathtub. Such macrocosmic phenomena as turbulence in the sky or in a furnace or in a tear running down your cheek defy measurement and prediction by present human powers. The new approaches of chaos science at last are opening useful avenues of understanding. But the quantum physics of microcosmic phenomena is far more accurate and reliable than any macrocosmic process.

Thoughts can be incomparably more accurate than things. Quantum physics, as Feynman points out, achieves a greater correspondence between theory and experiment than any previous science. He estimates that such quantum computations as the magnetic charge of an electron are comparable in accuracy to a measurement of the distance between New York and Los Angeles correct to the width of a human hair. Indeed, it is the immense precision of microcosmic data compared to macrocosmic chaos that accounts for the increasing efficiency of electronic components as they become even smaller and more densely packed together, moving ever deeper into the quantum domain. It is this precision of the microcosm that makes possible the overthrow of matter in technology and economics.

That microcosmic devices would be faster and better and more reliable than macrocosmic machines was little understood in the early years of the electronic era. The unreasonable effectiveness of microcosmic technology was asserted and largely explained in the 1960s by Carver Mead, a man who deserves the title of prophet of the microcosm.

CHAPTER

2

The Prophet

The information industry entered the microcosm almost as an accident. The seeds were sown in 1959 in Pasadena, at California Institute of Technology, home of Feynman and Pauling . . . and Carver Mead.

In the 1950s and 1960s, as Mead remembers, "Computers were still big ugly things." Rococo cathedrals of refrigerated wires and tubes, costing millions of dollars, they demanded the fealty of thousands of servants to the most minute whims of their machine languages. Using this technology, human beings had first to line up for access, then bow to a rigid mechanism, following the intricate liturgy of a priesthood of experts. Rather than tools or chips, computers were shrines.

During the 1960s, protests against this priesthood shook Berkeley, the computerized campus to the north. Lampooning the IBM punchcards then used to feed computers, students wore pins declaring: "I am a human being. Do not spindle, fold, or mutilate." But student protests were more successful against the Vietnam War than against the mainframe computer. For decades giant computers seemed a permanent fact of life.

Still based on the architecture first propounded in a memo by John von Neumann in 1945, the computer's central processing units

(CPUs) worked on programs and data kept in a separate storage system. Proceeding through its tasks step by step, fetching data and instructions one at a time from relatively remote memories, the entire system was bound together with miles of copper wire.

Through the 1960s and 1970s, the makers of new computers would slowly replace the Von Neumann components, one for one, with solid state switches. But the Von Neumann architecture remained supreme, on its pedestal in the central processing room, still mostly an electromechanical device, wrapped in wire and refrigerated, and still essentially a part of the macrocosm.

Nonetheless, working as surely as Planck's constant on the foundations of nineteenth-century physics, microcosmic ideas were infiltrating the fabric of twentieth-century technology. A Japanese physicist named Leo Esaki passed through Pasadena in 1959 and left behind him a quantum concept. It led to a flawed product, a failed research project, and a new vision for microelectronics. Through a long struggle, still under way, that vision would transform the industry, then the global economy, and finally the politics of nations.

The flawed product was called a tunnel diode. Invented at Sony in Japan in the mid-1950s, the tunnel diode won a Nobel Prize for Esaki, its inventor. The prize was for demonstrating a rare quantum effect in a seemingly practical device. As it turned out, the device was not practical. But it dramatically announced to the electronics world that the quantum era had begun.

The failed research project was Carver Mead's. In response to Esaki's Pasadena visit, Mead resolved to perfect the tunnel diode. He failed. But his failure bore fruit more important than any prize. Still in his early twenties, Mead found in this flawed device the secrets of the quantum era and led the way into it.

Named from the Greek words meaning two roads, an ordinary diode is one of the simplest and most useful of tools. It is a tiny block of silicon made positive on one side and negative on the other. At each end it has a terminal or electrode (route for electrons). In the middle of the silicon block, the positive side meets the negative side in an electrically complex zone called a positive-negative, or p-n, junction.

Because a diode is positive on one side and negative on the other, it normally conducts current only in one direction. Thus diodes play an indispensable role as rectifiers. That is, they can take alternating current (AC) from your wall and convert it into direct current (DC) to run your computer.

In this role, diodes demonstrate a prime law of electrons. Negatively charged, electrons flow only toward a positive voltage. They cannot flow back against the grain. Like water pressure, which impels current only in the direction of the pressure, voltage impels electrical current only in the direction of the voltage. To attempt to run current against a voltage is a little like attaching a gushing hose to a running faucet.

It had long been known, however, that if you apply a strong enough voltage against the grain of a diode, the p-n wall or junction will burst. Under this contrary pressure, or reverse bias, the diode will eventually suffer what is called avalanche breakdown. Negative electrons will overcome the p-n barrier by brute force of numbers and flood "uphill" from the positive side to the negative side. In erasable programmable read-only memories (EPROMs), this effect is used in programming computer chips used to store permanent software, such as the Microsoft operating system in your personal computer (MS-DOS). Avalanche breakdown is also used in Zener diodes to provide a stable source of voltage unaffected by changes in current.

Esaki, however, took diode breakdown into the quantum domain. Applying less than one fifth of the avalanche voltage, he made electrons move through the junction at some five times the speed. In an appropriately prepared device, the Esaki electrons would violate all the rules of electricity and simply bore through the barrier, tunneling from the positive to the negative side. Relatively large currents would flow well before any avalanche breakdown began.

Indeed, running the diode through all voltages up to the point of breakdown, researchers showed that the Esaki current was entirely separate from the avalanche current. Following the burst of Esaki current and before the avalanche, the voltage could double and triple. But the tunneling flow would *drop* some 80 percent. Flouting Ohm's Law, the most basic rule of electricity, this drop in current with a rise in voltage added yet a further mystery to Esaki's magic tunneling.

Magic or not, however, the tunnel diode was both scientifically intriguing and commercially exciting. Because electrons tunnel far more quickly than they normally flow in semiconductors, the Esaki burst effect promised extremely fast switches, approaching the speed of light. All things being equal, the faster the switch the better the computer, which uses vast arrays of on-off switches to perform its high-speed calculations. For a long period in the fifties and sixties, Esaki was the toast of the industry.

Defying the usual rules of matter, tunneling was also a prime interest of quantum physicists. It was as if a high jumper with a normal seven-foot limit had a certain restricted but statistically reliable probability of being able to leap over a twenty-foot barrier, provided you didn't get him too excited. Or perhaps a better analogy is a football player with a statistically dependable likelihood of being able to bull through a brick wall without breaking it. But in the quantum domain, all such analogies founder.

In essence, the electrons tunneled through the wall in the form of quantum probability waves before they slowed down, assumed mass, and began their flow as particles which the barrier could stop. Showing that the probability theory of quantum effects could countervail the most sanctified rules of electricity—and create a superfast device to boot—Esaki had achieved an important vindication of the new physics. It was the kind of surprising phenomenon that fascinated the physicist in Carver Mead.

Despite Esaki's brilliant papers, the taciturn foreigner had little to say at Caltech. He left Mead's laboratory suffused with the fumes from his chain-smoked cigarettes and spurned Mead's effort to recruit him. Instead, he took a fellowship in the research labs at IBM, where he remained until his retirement more than a quarter century later. But his visit in 1959 had a profound effect on Mead, prompting him to launch an intense investigation of tunneling effects.

Mead pursued these efforts for close to a decade and wrote some important papers. At first his interest focused on the possibility of making more effective diodes or other devices. He tried to determine how far an electron could be made to tunnel. But soon this work led him to a more general line of exploration.

Tunneling only occurs at dimensions measured in angstroms. An angstrom is truly a quantum measure: one ten thousandth of a micron. The usual yardstick of the microcosm, a micron is one millionth of a meter, or between one seventy-fifth and one hundredth of the width of a human hair. An angstrom thus is close to one millionth the width of a human hair.

Plunged so deep into the microcosm, Mead gained an intuitive sense of the quantum domain and reached an amazing conclusion for all semiconductor electronics. The industrial world might be telling him to invent a faster diode. But the technology was telling him a way to transform the industrial world.

During that period in the mid-1960s, most of the leading electronic laboratories—from Bell and RCA in New Jersey to IBM in Yorktown Heights and the research centers of Silicon Valley—were seeking new ductor performance, combining switching delays with heat emission —drops a thousandfold. With the power-delay product collapsing, reliability soars. With transisitor size dropping, production yields mount and costs plummet.

All these benefits were available with just a tenfold reduction in the size of the features on the chip. But it was clear to Mead that a fortyfold drop could be achieved, meaning a drop by a factor of hundreds of thousands in the power-delay product. The conclusion was monstrous and irresistible. Electronic circuits made of untold thousands or even millions of transistors on single silicon chips would rule the world. All the development efforts on tunnel diodes and other exotic devices became suddenly irrelevant. Other acclaimed technologies, such as magnetic bubble memories and Josephson Junctions, celebrated at IBM and Bell Laboratories, would never achieve any widespread markets.

The industry would thrive with soaring improvements in cost-performance ratios. The problem would not be heat or speed or reliability or cost; it would be designs. How would it be possible to design chips with hundreds of thousands or even millions of transistors on them? What would they do? One thing they would do, it was clear to Mead, was transform forever the computer industry.

As Mead foresaw in 1968—three years before it was done— it would soon be possible to put an entire computer on a chip and sell it for a few dollars. The semiconductor industry would have to convert itself from a manufacturer of components to a maker of entire systems on single chips. This would mean an entirely new kind of microchip, moving the industry to new domains beyond matter, into complex challenges of systems design and microcosmic manufacturing.

Born in 1934, near Bakersfield in Southern California, Mead is a descendant of the first settlers to run cattle into the Kern River Valley. Son of the manager of a remote electrical power station for Los Angeles, Mead as a boy earned money fishing and trapping in the nearby rivers and forests. But as a teenager, afflicted with a bad complexion, disdainful toward the sports and enthusiasms of his peers, spurned by the svelte California girls in his classes, he retreated to the microcosm. For modern pioneers it was a frontier as spacious and intriguing as the West had been for earlier American pioneers.

As a high school freshman he learned calculus and began befriending members of the electronics club at the local community college.

Eventually he entered Caltech and studied with Pauling, Feynman, Max Delbrück, and other giants of the quantum era. As a Caltech freshman, Mead learned from Pauling the elegant permutations of the chemical bond, the quantum sources of solid state, the foundations of his career. But his most important personal lesson came from Feynman.

At the time, Mead was suffering a small but persistent canker of doubt about his scientific prowess. Surrounded on all sides at Caltech by mathematical wizards—making chalk cluck and squeak and leaving chicken tracks all over the blackboard—he wondered whether there was room in science for less poultry, more poetry. Early in Feynman's own career, Enrico Fermi and Robert Oppenheimer had provoked the same doubts in the young man from the Bronx. But as a Caltech professor, Feynman had found that his own intuitive method was often indispensable to achieving new truth.

Mead says simply today: "Feynman gave me confidence to go ahead and do science. . . . He taught me that the first and most crucial step was usually to think about what the problem is trying to tell you, rather than tell it what you know already about the math."

This approach impelled Mead to a lifelong effort to escape the momentary claims and crises of his field and to see the thing whole: to transcend the common sense of the day, the dense traffic of convention, the ways of the wealthy wisemen of the Valley, and uncover the deeper meanings of the technology: "to sort out what the thing was really trying to tell me from the things that the experts were telling me."

In 1968, when Mead got back to Pasadena from his presentation at Lake of the Ozarks, he immediately assigned his most brilliant available graduate student to get to work calculating more precisely all these marvels of the microcosm. Mead himself began to enrich his vision.

In the decade since Esaki's visit, Mead had done more than investigate tunnel diodes. He had invented the first commercially viable transistor made of gallium arsenide, a compound material that offered speed advantages over silicon. He also had been introduced to Fairchild Semiconductor by Gordon Moore and had become a noted consultant in Silicon Valley. There he gained a valuable perspective on the commercial aspects of the technology.

Broadening his academic horizons as well, he had studied biology with Max Delbrück. Intrigued by the recognition that the nerve membrane works like a transistor, Mead spent a summer at Coblenz in

Germany exploring the links between physics and biological systems. In 1968 he had also embarked on a study of computer systems design, a field most other solid state theorists had regarded to be beyond their domain.

From all these experiences, he began to weave together a powerful prophecy for the future of the computer industry. A gnome of quiet voice, pointed beard, and kindly smile, he would bring his vision—tenaciously, trenchantly—to prestigious men who disdained it, to old friends who bitterly opposed it, to allies who betrayed it. His role is recognized in no current history of the field. But carried by hundreds of his students throughout the industry, by thousands of his readers, and by many unconscious followers of his logic, his word became law. In the end the microelectronics industry would transform itself—against the resistance of some of its most powerful leaders in both the United States and Japan—to conform with his message.

Writing in 1968, he described "the dilemma of the computer industry": "It has an enormous investment in big machines and big software programs, and the only thing the industry can do right now is to use the new microelectronics as it fits into the existing system. . . ." To use chips, that is, as replacement parts that add power and speed to the existing big and ugly machines.

Mead continued, "We have computer power coming out of our ears. What we need is the kind of systems we would like to have in our automobiles, in our telephones, in our typewriters—where people now spend vast amounts of time on the repetitive and mundane operations involved in keeping track of a lot of little things. . . ."

Mead contrasted the lordly computer to the car. "A person doesn't feel dehumanized by such a machine—one that frees him from routine tasks and is under his control. . . . An automobile is a machine that gives you a lot of power, yet it leaves you as much in control as if you were walking." He predicted that the new silicon technologies would allow creation of small computers that brought power to the people as automobiles did, rather than bringing it to large institutions and sophisticated programmers, as mainframe computers did.

Mead offered the analogy of the electric motor, "invented at a time when most industries had a big steam engine out in back driving a big shaft the length of the factory. . . . Even though it was perfectly clear that the way this innovation should have been used was to put electric motors on each machine . . . the most that could be done economically was to replace the big steam engine with a big electric motor."

Small electric motors eventually came to inject power everywhere into the world's tools, its appliances, and even its toys. Mead predicted that small computers could similarly inject human intelligence into the entire environment of human beings.

Beginning his argument with an analysis of the angstrom dimensions of electron tunneling, Mead had proceeded to a radical case for the transformation of the computer industry. In the autumn of 1968, he was summoned to the living room of his friend Gordon Moore to offer advice to Moore and Robert Noyce, who were then preparing to form Intel Corporation. To them, too, he made his argument about the obsolescence of the large computer and the need to escape the predicament of design complexity.

As a portent of the future, Mead pointed to Hewlett Packard's new programmable calculator—the HP 9100. A machine the size of a typewriter, it would soon find its way onto nearly every engineer's desk and later as a hand-held device into the pocket of every student of engineering. It both performed an array of scientific calculations and could be used to control HP's line of scientific instruments. Mead urged Noyce and Moore to develop chips and design techniques for the coming era of dispersed computing power.

The two men agreed with Mead about the future of computing. But they were initially wedded to a contrary strategy for their new company, replacing the current electronics of the mainframe with faster and cheaper semiconductor devices operating within the old design. The two entrepreneurs wanted to begin by exploiting a clear existing market. But the future would not wait. Within three years, Intel would become the leading instrument of the new market predicted by Mead, providing all the key components of the personal computer.

As a result, in most companies today, all information workers now have computers on their desks. To accomplish a computing task, no longer is the user forced to join a queue outside the doors of the data-processing center waiting for access to the lordly mainframe on a pedestal inside. In the end, the 1960s Berkeley revolt against the mainframe largely triumphed. The microcomputer revolution has redistributed computing power onto the desks of nearly every student, office worker, and corporate manager.

Mead celebrates this achievement. But he believes it represents only a first step in a more far-reaching agenda for information technology. For while computer power is now dispersed throughout most companies and around the world, there still thrives in miniature, in the inner structure of the machine itself, the old tyranny of the data-processing center.

It is a secret scandal of computing. Enshrined inside nearly every computer, large or small, is its own central processing room. Data and instructions queue up endlessly in registers, buffers, caches, and stacks waiting for access to a single lordly central processing unit (CPU). Most of every computer, most of the time, is in a wait state, scuttlebutting and flipflopping, taking turns at access to the CPU. Modern CPUs are extremely fast but nevertheless constitute a drastic bottleneck.

It is the original computer architecture of John von Neumann, far more pervasive today than ever before. In 1947, when he proposed it, and in the 1960s when it provoked protest at Berkeley, the bottleneck was a necessary economy. The industry operated in the macrocosm of electromechanics where switches were expensive vacuum tubes and wires were nearly free. Von Neumann's computer architecture reflected these relative costs of switches and wires in the macrocosm. He economized on vacuum tube switches for processing and was lavish with interconnections.

Moving to the microcosm, however, the Von Neumann components—storage, interconnect, and processor—all can be made of the same sliver of silicon. The computer on a chip outperforms the computer on a pedestal and costs one millionth as much. Switches are now virtually free—microscopic transistors inscribed on a chip. Today, as Mead declares, it is wire—once free—that has become costly in every way, clogging the chip with complex metals hard to lay down, subject to deterioration from heat, and difficult and expensive to link to the world.

The microcosm of free switches and expensive wires dictates a reversal of current practices. Rather than wiring up one superfast processor to lots of supporting chips, the new technology dictates combining lots of processors on each chip. The microcosm favors innovation designed to channel data to large numbers of switches processing in parallel.

Yet the Von Neumann architecture still widely prevails. The computer remains the same lopsided technology it was when information workers lined up in long queues outside the central processing rooms of large corporations. Nearly all the processing power is devoted to manipulating data rather than producing or acquiring it and getting it into the machine. Yet the effectiveness of any system is determined by the speed and accuracy of input and output as much as by its ment of this essential prophecy. But its roots were in the past, at Bell Laboratories, where executive vice president Mervin Kelley decided after World War II to "take advantage of the understanding of solids made possible by quantum mechanics," and launched a huge quantum web of wires and switches.

CHAPTER
3

Wires and Switches

William Shockley worked in the midst of the world's widest mesh of wires. They reached across the continent, spanned the Atlantic Ocean, and spread in trunks and branches to nearly all points of the industrial world.

The intellectual center of this miraculous web, spread out to catch the voices of nations, was Shockley's employer, Bell Laboratories. Moving intelligible signals through these wires and around the globe represented an immense feat of engineering science.

In 1947, however, most of the switching of these signals was still done by human operators plugging patch cords into switchboards. The global telephone network still depended for intelligence on the wet wires and switches, axons and neurons of the human brain.

Like the first computers emerging in that era, telephones in 1947 were still mostly electromechanical systems, entrenched in an industrial macrocosm of heavy machinery attended by hundreds of thousands of regimented workers, reciting by rote: "Number please." Statisticians calculated that at the current pace of expansion of telephone use, all the women in the country would soon be telephone switchers.

By the early 1950s, the mechanical switches were rapidly giving way to electromechanical crossbar and vacuum tube systems that could automatically respond to dialed numbers. Sure enough, all the nation's citizens did become operators . . . of their own phones' dialing devices. But as the number of telephones increased additively, the number of possible connections multiplied. The problem shifted out from taking calls to routing them to their destinations.

The challenge of finding the best routes resembled the classic traveling salesman problem. Not only did determining the shortest among increasing millions or even billions of possible paths confound human beings but even an approximate solution strained the computational resources of computers. Soon AT&T's labyrinth of wires was served by the world's most formidable computer-switching systems.

The computers were themselves new, more compact ganglia of wires and switches. In the telephone firm, central switching systems captured the signals from dialed numbers, sought their addresses, defined the routes between them, reconciled conflicts, checked line performance, and collected billing data, in much the way that a business computer might capture the signals from a keyboard, search memory addresses for data, process it in accordance with spreadsheet formulas, and transmit results to screens, printers, and remote memory addresses, using a variety of networks, including probably the phone network itself.

Both systems are logic machines. In the telephone network, each switch is a decision point that sums up all previous decisions in the chain that links one caller to another. Like a syllogism in logic, it reaches the one right answer through a concatenation of other right answers. Is the handset off the hook, yes or no? If yes, connect switch for dial tone. Then follow the number, digit by digit, each opening a new level of the network as it branches through the logic chain, summing up previous switching choices in the form of an open channel and actuating new choices or switching decisions that open a further channel, step by step to the destination.

Similarly, the millions of switches inside a computer are decision points for the complex calculations that it performs. In a computer the on-off positions of these switches represent one and zero (and thus on a base-two system all numbers), or true and false and yes and no in the binary logic of the machine. Although these are minimal bits of information, each can be assigned a specific meaning of any needed complexity by a consistent software language. Then the accumulation in a logical code of millions of such decisions can figure your sales tax and estimate it for the next five years, pore through *War and*

Peace to find how often Andrei talks to Natasha or Pierre talks to God, or manage your voice through thousands of switches from your phone in Queens to the Queen's phone in London.

The more numerous, the faster, and more efficient the switches, the more powerful the logic of the system. In 1947, however, the only available switches were electromechanical gadgets or vacuum tubes that collectively needed the power of a locomotive to pull a voice around the world or tug a differential equation to its solution. The size, expense, and power requirements of such switches placed an oppressive upper limit on the capacity of all logical systems, from telecommunications to computers.

Switches are the substance of the artificial mind. As long as the switches were macrocosmic, the millions required in any complex logical process, from word processing to telephone routing, would bulk too large and allow too many compounding defects to function efficiently. In overthrowing matter and creating a new technology of mind, engineers would have to dissolve this substance into the microcosm.

Today, across the entire telephone labyrinth, from its supercomputer central switches to its microcomputer desktop autodialers, voice mailers, modems, and facsimile boards, nearly all the switching is performed by sophisticated computer systems consisting chiefly of integrated circuits and advanced microprocessors. Each of these integrated circuits and microprocessors, moreover, is itself a compact and intricate microsystem of tiny wires and switches.

From a far-flung but finite geometrical system of wires, interconnected by human operators, the telephone system has exploded into a gigantic fractal array. That is, it is self-similar at every level. It repeats the same essential pattern of wires and switches, lines and nodes, on every scale of size and operation from continental waveguides to microchip connections.

The worldwide pattern of cables and huge ESS-5 central switching computers is repeated in the pattern of open wires and smaller switches at local end offices, and again in the twisted wires and Private Branch Exchanges in commercial buildings, and again on the printed circuit boards inside the PBXs, and again in the repeated patterns of logic gates on the microprocessors on the printed circuit boards, and finally in the complex vias, contacts, gates, and channels of the physical chip layout.

All the way down the hierarchy of scale, the system becomes more dense with repeated structural details of wires and switches. In accord with the rule of such fractals and the law of the microcosm, the richness of informative detail increases far faster than the drop in size as the system moves toward the quantum domain. Thus the power of the logic and the features of the system improve far faster than the switches multiply. And nearly every one of the infinitesimal switches inscribed by the many trillions throughout the telephone system—on integrated circuits, microprocessors, memory chips, signal processors, controllers, converters, multiplexers, demultiplexers, clocks, counters, and thousands of other devices—is a *transistor.* The transistor is the microcosmic switch. Crucial to the overthrow of matter in both computers and telecommunications, it was defined and impelled by the researches of William Shockley at Bell Labs in the late 1940s. Now informing a global system of wires and switches, transistors control an endless quantum interplay of electrons and photons comprising the essential fabric of modern information technology.

Combined in branching nets of logic spread across minute slivers of silicon, millions of these wires and switches comprise a computer. Stretched across the mostly silicon surfaces of continents, the wires and switches comprise a telecommunications system. Fused into a global ganglion of interconnected tools, the wires and switches of both computers and telecommunications join to form the central nervous system for a new world economy. Eventually, in the form of fiber-optic cables, it will become chiefly a web of light and glass, in which messages flash around the world with no material embodiment at all.

The invention of the transistor was the critical first step in the invention of this new technology of mind. Understanding this historic achievement, you will doubtless have no trouble identifying the man, with a last name beginning with "*Sh*" and ending "*ley,*" who was featured in *The New York Times* in its issue of July 1, 1948, the day after the invention of the transistor was announced by Bell Laboratories.

The answer, of course, is George Shackley, composer of "Anthem for Brotherhood," which was to be performed the following Sunday on the New York television station WPIX.

Below the Shackley story, the *Times* did tell of the transistor announcement. But it failed to mention the inventors of the device, William Shockley, Walter Brattain, and John Bardeen. This oversight was not peculiar to the *Times.* Most of the American press failed to cover the Bell announcement at all.

The imperious young scientist with the stern black spectacles and the spare handsome features beat his fist on his desk at Bell and declared: "I am a better physicist than anyone working for me."

Well, yes, Dr. Shockley, that may be true. But it is John Bardeen, the quiet and shy theoretician, and Walter Brattain, the kindly silver-haired veteran at Bell, who have invented the transistor. Moreover, it is a design that has virtually nothing in common with the field effect (or voltage-controlled) transistor you have been trying to build for many years, following the inspiration of Julius Lilienfeld, who patented such a device in 1926.

True, as far as anyone knows, Lilienfeld never built one that worked. But neither have you, Dr. Shockley. In fact, Dr. Bardeen has shown that contrary to your theory, it is impossible to switch or amplify a current by applying a voltage or field effect to a semiconductor. Electrical pressure will not do it alone; you need an actual current. The surface states—all the dangling chemical bonds on the interface between the voltage source and the semiconductor—prevent the voltage from controlling the flow of electricity.

Shockley knew all that. He had ordered the program to investigate the surface phenomena. The invention of the transistor, he wrote, was an example of "creative failure methodology." The failure of the underlings in their effort to create a field effect transistor would lead to the creation of some other splendid device by Shockley, the brilliant manager of the research program. In fact, Shockley would go on to create a grown junction transistor.

But Bardeen and Brattain were supposed to be creatively failing in field effect devices, not sneaking off to make epochal inventions unbeknown to their brilliant young boss. For a while it looked as if Bardeen and Brattain might receive most of the credit—even the Nobel Prize—and Shockley remain an obscure manager at Bell Labs. In that case, the *Times* would not even notice if he came up with bizarre theories of white supremacy.

Indeed, John Bardeen, later willing to go back and start again in a new research program in superconductivity, would eventually become the only man to receive two Nobel Prizes in the same field. Some observers might reasonably question whether Shockley was in fact a better physicist than any of his subordinates in 1948. Superconductivity, moreover, might prove as important in the end as all these little switches. History can surprise you. Better get to work, Bill.

Shockley did get to work and history would notice. In addition, he had a legitimate complaint. The prize-winning point contact transistor invented by Bardeen and Brattain was to a considerable degree a retrograde development.

Two of the three electrodes (routes for electrons) on the point contact transistor were conventional metal wires, their points delicately in contact with a semiconductor base; it was only partly a solid state device. In fact, it hearkened back to the whisker diode, used in crystal sets for crude radios, more than forward into the age of the integrated circuit.

Shockley was partly right that this transistor, announced by Bell in 1948 and responsible for Brattain's and Bardeen's Nobel Prize, set back the cause of electronics. Difficult to control, hard to explain, and balky to manufacture, the point contact device was just menacing enough to prompt a massive new effort to improve vacuum tubes but not sufficiently superior to pull the industry into the new era.

Shockley's concept of a field effect device was simpler and easier to miniaturize. It was unipolar; it would operate entirely on one semiconductor block. Thus for most purposes it was preferable even to subsequent Shockley bipolar transistors. Rather than running two flows of *current* into a complex arrangement of three semiconductor materials joined in two p-n junctions, the switch worked by applying *voltage* to an insulated gate on top of the block. The electrical pressure or voltage—the field effect—opened a channel for electrons below the insulated gate.

Years later, engineers necessarily turned to Shockley's field effect concept when they rid themselves of all excess baggage for their descent deep into the microcosm. Today, Shockley's dream dominates the industry. But in the late 1940s, Shockley could not build it. So he turned to second best: an effort to create what was called a "grown junction" transistor.

The grown junction transistor was a bipolar device. This meant—in Shockley's new and entirely solid state concept—that the base was of a different polarity (positive or negative) than the emitter and collector on each side of it. All these elements would have to be grown on or "doped" into the crystal. This would take some doing. Although the materials specialists at Bell had made great gains in growing crystals, the process was still unreliable; essentially they lined up the suspects, tested them, and tried to find substances with the appropriate "dirt." If Shockley's grown junction device was to supplant the point contact transistor as he intended—and make him once again the solid state champion of Bell Labs—he would have to come up with a way to mass-produce the grown junctions. But he was making little progress. It appeared that Shockley, this intellectual titan widely regarded as the most brilliant scientist in solid state electronics, would fail to get a product out the door and would be eclipsed in history by two of his subordinates.

Almost two years later, however, in 1950, Shockley's project was saved by a stubborn young solid state chemist from Texas named Gordon Teal. Working with molten tubs of semiconductor glop, Teal had persisted doggedly in a long effort to create pure single crystals of material. Most of his colleagues, including Shockley, wondered at first why it was worth the trouble, since semiconductor devices could be made perfectly well from polycrystalline or amorphous substances.

Yet Teal's effort was indispensable to Shockley's success with the grown junction transistor. Through a process of alternately doping the material with p- and n-type impurities as it was being pulled from the melt, Teal was able systematically to produce an ingot of germanium already appropriately layered for a junction transistor. Shockley took it from there.

By slicing the crystal across the junctions, attaching wire leads to the positive regions at both ends and a control circuit to the n-type base in the middle, Shockley showed the way to manufacture grown junction transistors. His first devices—the p-n-p's—were based partly on the use of positive "holes" as carriers of charge. In electronics, a hole is a positive charge left by the absence of an electron in an atom of the semiconductor crystal. Shockley showed that positive holes—moving like bubbles through water—could pass through a semiconductor nearly as fast as free electrons. Western Electric produced these grown junction devices for a decade; they completely eclipsed point contact transistors and laid the foundations for a prosperous and important semiconductor industry.

Also in 1950, Shockley published his book *Electrons and Holes in Semiconductors,* which stood for many years as the definitive work in the field and confirmed his credentials for the Nobel Prize that he shared with Brattain and Bardeen in 1956. The fact was that for his theory of the field effect transistor that later dominated the industry and for the junction transistor that was dominating it at the time, Shockley deserved the prize alone. He had at last made his point.

Yet Shockley was not satisfied. "Every physicist," he said at the time, "wants two things: glory and money. I have won the glory. Now I want the money." A superb scientist and inventor, with some ninety patents to his name, Shockley gained glory in a research setting. But like many intellectuals, he greatly underestimated the disciplines of money.

Shockley sought a sure thing. He began by simply demanding a million-dollar fee for his consulting services. That approach failed. Then he started the first semiconductor company in Silicon Valley and attracted to it from across the country what would prove to be the leading business and technical talent to emerge in the field over the next thirty years. Included were Robert Noyce and Gordon Moore, later to found both Fairchild Semiconductor and Intel. But like so many of the venturers who would follow him, Shockley wanted a proven market. Proven markets tend to be markets that are already being filled. For providing small improvements in proven markets, the entrepreneur gains small fees at best, not the glorious riches Shockley desired.

As Peter Drucker has pointed out, a new technology cannot displace an established technology—with its installed base of plant, equipment, training, personnel, and satisfied customers—unless the innovation is about ten times more cost-effective than its predecessor. The established technology will have far more than ten times the effort and momentum behind it and will continue to improve fast enough to prevent the success of any marginal invention.

As a result of this mandate for a tenfold gain, the world is full of bitter inventors with conspiracy theories about the hostility of corporations to their terrific ideas—at least "five times as good as the established product." Five times as good is no good if the new device requires drastic changes in an existing industry. As a discrete power device, the Shockley transistor was too small an improvement over existing technologies to succeed except in a few niches.

Although Shockley had worked at the center of the greatest information system in the world, he never clearly saw that the transistor could succeed as a bearer of information but not as a smaller and more efficient power device. At Shockley Semiconductor, he sought to build replacement parts such as four-layer thyristors to "control every light switch" or supplant every telephone relay, rather than to launch logic tools that could digitize entire systems and yield compounding gains of ten and more for every network of wires and switches in the world.

Shockley always wanted to use his unique knowledge of solid state materials to create clever discrete devices that were low in information content but could control ever larger amounts of physical force. The microcosm demanded numerous low-power devices integrated to control ever larger amounts of information.

Shockley and his colleagues had solved a key problem of information technology. The transistor could both represent information and send it to several destinations. It provided both a bit and a boost: a crucial switch and amplifier both to register the ones and zeroes, bits and branches of binary logic, and to transmit them to the world. To go with its binary uses as a switch, the transistor had gain and fan-out; the output of one transistor could activate several other transistors. Moreover, the transistor reset itself, returning to its original state when the power was removed. With gain, fan-out, and automatic reset, information could cascade through the system without complex replenishment or other outside intervention.

This was a major step. Lack of gain and resetting capabilities would ultimately doom many fast devices otherwise supremely attractive for computation. Tunnel diodes, Josephson Junctions, and optical logic devices, for example, all operate at or near the speed of light. Each can represent a bit. But they lack gain, so they can't boost it or send it out to many other points in the machine, and they often cannot reset themselves. Thus these devices cannot function alone in extended computations without awkward supporting systems. So information technology everywhere still relies on Shockley's far slower transistors.

Shockley and the Bell Labs leadership, however, had underestimated the second key requirement of information technology: a pure crystalline substance to be manufactured in volume. To overthrow matter, it would be necessary to convert matter into a medium, to create matter pure and regular enough to serve as a transparent vessel of data, matter so symmetrical that in itself it would constitute a form of information. Then it would be necessary to produce it cheaply in large volumes.

The impulse to achieve this breakthrough would be unlikely to come from laboratory science itself. Pure scientists can thrive on ingenious prototypes, winning fame, wealth, and prizes. Scientists could get by without a mass-produced crystalline medium. But the creation of a new technology of mind would require hundreds of trillions of perfectly replicated wires and switches. It would entail a manufacturing breakthrough.

To turn this elite science into the fractal fabric of a mass technical culture would requre an entrepreneurial compulsion. The transistor would have to reshape the business environment, and be reshaped by

PART

The Technology of Mind

II

CHAPTER
4

The Silicon Imperative

> *The common sand that you tread underfoot, let it be cast into the furnace to boil and melt and it will become a crystal as splendid as that through which Galileo and Newton discovered the stars.*
>
> —VICTOR HUGO, *Les Misérables*

Using electrons to control electrons, Shockley's transistor was the first microcosmic switch. Making sense and syntax from the evanescent entities of the quantum realm, the new invention moved the world well beyond conventional matter. The transistor provided the first major bridge between the quantum revolution against matter in physics and a quantum revolution against matter in technology. Crossing this bridge would be thousands of inventors and entrepreneurs who would ultimately overthrow matter in business structure and in the world economy as well.

Governing this process of creation is a prime truth of classical economic theory called Say's Law: supply creates its own demand. Invented by the Frenchman Jean-Baptiste Say, the law usually refers to the fact that—including wages, rents, and profits—the payments made in producing a good or service create enough buying power to purchase it. As a theory of the cyclical flow of funds, Say's Law is a foundation of the equilibrium in most economic models and a source of the stability of capitalism. But in practice, Say's Law also unleashes powerful forces of disequilibrium. The prime disequilibrium is called economic growth. It is a process greatly enhanced in the microcosm.

As the driving force of economic growth, Say's Law exalts the creativity of suppliers over the wants and needs of demanders or consumers. As entrepreneurs invent new things and learn how to make them more efficiently, unit costs and prices drop and goods become more attractive. As goods become affordable to a wider public, more people work to acquire them by creating goods to exchange. These new suppliers both provide and acquire new wealth at ever lower expense.

The key is not wants and needs. Wants and needs are as ubiquitous as poverty, and as easy to produce. The key to wealth is always the creativity and diligence of ever-widening circles of suppliers, led by inventors and entrepreneurs.

This truth is difficult for many people to accept. An optical illusion leads people to believe that wealth inheres in physical things they can see rather than in ideas that are invisible. This materialist superstition seems especially plausible wherever scarce and immobile matter predominates in an industry, whether coal, land, gold, iron, or oil.

Under these circumstances, the power to seize or sequester material resources seems the route to wealth. Existing establishments of landowners or resource suppliers dominate society and change is realtively slow. Cost reductions seem to come chiefly from rising efficiency in extracting and manipulating materials and exploiting physical labor. As a scarce resource is used up, its price will even rise in accord with the law of diminishing returns.

It is human inventiveness that lends value to raw materials (even oil was useless gunk until 1855). But the rarity and weight of such inputs impose a clear limit on the productive process. People come to revere the substances themselves more than the inventiveness that gave them worth.

Until the end of World War I, support for the materialist view could even be found in the statistics of growth. Increases in labor productivity came chiefly from the substitution of energy for labor. U.S. output grew more slowly than growth in the use of energy. This meant that while labor productivity was increasing, energy productivity was actually declining; it took more and more energy to achieve a specific advance in labor output.

Materialists could claim that industrial progress was in a sense a sham: a zero-sum game, where the gains of the winners are precisely offset by the losses of the losers. The human race was paying for its economic progress by the exhaustion of irreplaceable fossil fuels. Entropy—the inexorable conversion of usable energy into waste and pollution—was the rule of both thermodynamics and economics.

From this point of view, the current generation of humans was merely stealing the real wealth of the earth from future generations. As socialists said, property was theft.

Since World War I, total output in capitalist nations has been rising far faster than the consumption of energy. But materialists could continue to claim that the gains were spurious. Irreplaceable fuels were being underpriced because future generations could not bid for them. Moreover, energy benefits were offset by pollution and other so-called externalities. In the ecological balance sheets, economic growth still seemed a Sisyphean quest.

In the microcosm, however, all these materialist claims collapse. Ideas are not used up as they are used. Where intellect is the decisive source of value, the economic burdens of matter decline and costs can follow Mead's laws of the microcosm. Space and time expand as size and power drop. In the age of the microcosm, the inventive inputs of producers launch a spiral of economic growth and productivity at steadily declining cost in every material domain: land, energy, pollution, and natural resources.

Say's Law returns with ever-increasing effect. Liberated from the burdens of matter, Atlas bears the globe not on stooped shoulders but on the crests of creative thought. Sisyphus hurls aside his boulder and plunges beyond gravity, entropy, and friction into the vast inner space of matter. Shifting from the manipulation of materials to the generation of ideas, entrepreneurs launch a spate of novelties, radically diminishing every material cost—from weight and space to toxic waste.

Entrepreneurship is a creative process, and by its very nature, creativity comes as a surprise to us. To foresee an innovation is in effect to make it. If creativity were not unexpected, customers could demand it and expert planners could supply it by rote. An economy could be run by demand. But an economy of mind is necessarily impelled by Say's Law, driven by the unforced surprises of human intellect.

It is a spontaneous process of discovery, of "creative failure," as Shockley put it, of "creative destruction," according to economist Joseph Schumpeter. Under any name, however, innovation tends to devalue the materials of the established system and create a new means of production with a higher content of intellect and ideas. This displacement of materials with ideas is the essence of all real economic growth.

On the manufacturing level, the growth of intellect and ideas in the means of production is often measured in terms of learning or expehe explained, "I was sure that this was a volume business. . . . The silicon transistor would take care of this military, environmental problem. We could probably get started that way, technically, in small quantities. . . . Still unless you found a way to do it in volume," to make it a commodity, "you weren't going to stay in business. It was a volume business, and you were going to be forced into volume processes, doing things at low cost."

In the Navy, Haggerty had encountered the concept of the learning curve. Lower costs, Haggerty reasoned, meant lower prices and expanding market share in a spiral of benefits. Haggerty was seeking a high-volume, low-cost product that could bring the firm down the learning curve in semiconductor electronics, gaining experience that would be invaluable wherever the sphere might move.

He named the product confidently: a portable radio that would use fast and tiny germanium transistors in place of vacuum tubes. Haggerty thus was committing his small squad of engineers, few boasting any credentials in solid state electronics, to mass-produce transistor devices from two radically different semiconductor substances, slow silicon and fast germanium. Moreover, one of his products, the transistor radio, would be TI's first consumer product.

To compound TI's problems, the military began giving financial aid to the competition. The Signal Corps inaugurated a grant program of several million dollars annually to the established firms to improve transistor manufacturing techniques. The amount seems small today but it was significant in the 1950s electronic business.

Yet TI overcame all obstacles. As the years passed, the vacuum tube giants—RCA, Sylvania, GE, Westinghouse, Philco-Ford, Raytheon —with their long head starts, their growing government subsidies, and their large military contracts, all would leave the semiconductor trade. In all the annals of business history, there are few triumphs so grand and improbable as TI's domination of this central electronic technology.

Mark Shepherd, the Dallas policeman's son who later took over the firm from Haggerty, offered as good an explanation as any: "Those companies all knew the things that weren't possible. We didn't. We were stupid."

In the wee hours of the morning of September 25, 1954, two of the heroes of the story, together with an increasingly distracted onlooker, joined in one of the more improbable scenes of the drama. On one side of the room was the mechanical hero. Six feet high, three feet wide at the girth, it looked a lot like a popcorn vending machine left over from the lobby of the Lemon Avenue Bowling Alley. This was Gordon Teal's crystal-growing device, a unique machine, perfected at TI, which Haggerty believes gave the firm a two- to three-year lead in process technology. It had a small motor on top running a pulley down the broad transparent neck into a quartz crucible, glowing like an illuminated jar of melted butter. From this glowing vessel were pulled the pure molten crystals of germanium or silicon at the heart of microcosmic technology.

Sitting nearby, also some three feet wide, and looking uncomfortable, was Mrs. Mark Shepherd, who the next day would bear her second child, Mary K. Crouching attentively was her husband, his face creased with a near-parental concern for the condition of the crystal-pulling machine. Many hours and some one thousand tries before—in fact, on his third attempt—he had managed a major breakthrough in crystal processing, crucially needed for the creation of an adequate radio. Shepherd had made what is termed a rate grown germanium transistor. But if he could not grow another one, it might be written off as a fluke. For the last several hours he had promised that at any moment he would pull another rate grown device, if he could "hit the right set of parameters. . . ."

Shepherd would have given up and gone home except that the future of the company in consumer markets was hanging in the balance, in that glowing womb of quartz, and he had had such a tantalizing success on his third try. At 3:00 A.M., on try 1,005, he finally found the clue to the combination, finally figured out the problem. Attempt 1,006 produced another perfect rate grown transistor. All the next day, Shepherd was as proud and happy and tired as any attentive young father.

Although TI eventually lost the transistor radio business to Sony, the Texas company sold several hundred thousand portable radios at a time when all the leading radio firms had decided that portables were unmarketable. Most important, TI honed its semiconductor fabrication skills and demonstrated its powerful technology. The radio had been designed as much to send as to receive a signal, and its key message, said Haggerty, was the announcement of a technology and a company: "That transistors were usable now, not some years in the future, and that we were ready, willing, and able to supply them."

Up in Armonk, New York, Tom Watson, Jr., leader of IBM, received the message, bought a hundred or so radios to distribute to his staff, and remarked to an assistant: "If that little outfit in Texas can make these radios work for this kind of money, they can make transistors that will run our computers too." IBM eventually became TI's biggest customer and the world's biggest producer of semiconductors.

For all the effect and portent of Haggerty's rash rush into germanium transistor radios, though, the most important event at TI in 1954 was Teal's success—also considered unlikely for several years—in producing dependable silicon transistors in volume. Like the radio, the silicon transistor gave TI an important monopoly position. Unlike transistor radios, however, silicon semiconductors would remain a TI bastion for decades.

Silicon was adopted because it was less sensitive to heat than germanium. But it ultimately prevailed because, unlike oxides of other substances, silicon dioxide was both electrically and chemically inert and thus could both protect and insulate the devices of near-micron dimensions that would appear within the next two decades. The substance of sand, silicon was made in large quantities in polysilicon form by firms such as DuPont and later TI. Then, through methods perfected by Teal, it could be rendered in perfectly regular and variable crystals, suitable to bear and process information.

Without silicon, the move to the large-scale integrated circuits and microprocessors that shape the information age and run your Macintosh or IBM personal computer would have lagged for decades. As DNA was the crystalline substance bearing the informative codes of life, silicon was the crystalline medium of the microcosm.

Bell Labs had worked with silicon, but it lacked the entrepreneurial compulsion to develop the tools to manufacture pure silicon crystals in volume. Metallurgists at the firm believed that such single crystals were unnecessary and their production much too expensive for manufacturing. Yet it was no accident that Teal had come from Bell Labs. He was a leading fundamental scientist from Brown who exemplifies the utter dependence of information technology on the quantum revolution.

Without the light shined into the atomic structure by the quantum physicists, the world would never have achieved the understanding of the microstructure of materials on which Haggerty based his business plan. Without Teal's deep command of the quantum properties of silicon and germanium single crystals, TI could not have mastered their mass production. Teal was following closely in the footsteps of Pauling and other giants who provided the rigorous quantum calculus at the heart of modern materials science and thus made possible the industrial overthrow of matter.

Teal, however, was not only a master of solid state chemistry, he also was mastered by a vision of mass production and profits. Over a year of frenzied effort at TI, he worked out the concepts for mixing selected impurities into the molten silicon and for growing crystals suitable for slicing into transistor blocks.

Gordon Teal launched this new medium in a famous announcement, offhandedly astonishing the crowd, in Dayton, Ohio, on May 10, 1954. The nominal event was the National Conference on Airborne Electronics, but the real event, so Teal resolved, was the sudden consummation of a silicon coup.

Gordon Teal was the next to last speaker on the program. A quiet man who speaks in a low voice, he followed several other reasonably soporific scientists. It is likely that any flickers of excitement that rippled through the crowd as the former Bell Labs luminary approached the podium were extinguished as he began a droning rendition of his thirty-one-page address. Teal was very tired, having stayed up until 5:00 A.M. rewriting the final sections to include his announcement. But the title, "Some Recent Developments in Silicon and Germanium Materials and Devices"—though mildly provocative in the precedence of silicon—gave no hint that the man with the rumpled suit and the bulge in his right pocket had anything extraordinary to divulge.

To some, it might seem strange that a man would travel to Dayton, Ohio, to declare the overthrow of the establishment. But Teal's audience, assembled at Dayton's Engineering Club Auditorium, comprised a group of the unacknowledged legislators of our age. The leading American media would not notice them unless they protested the military budget or defected to Russia, yet they constituted a perfect group for Teal's theatric purposes. Technical press, electronics engineers, Pentagon officials, all increasingly conscious of the rising promise of transistors, they were ready to nod knowingly at the conventional wisdom and then stiffen abruptly at the sudden announcement of revolution. They could be depended upon to publish it to the relevant world.

As the earlier speeches proceeded, Teal made a thoroughly exhilarating discovery about his audience. Significant parts of it were vocally impatient with the limitations of germanium devices. They had been taught to expect miracles from semiconductors, but miracles at room temperature alone would not suffice for anyone whose products had to perform under more extreme conditions.

For example, at this conference on airborne electronics, many of the engineers were dealing with airplane engines and avionics, and