A Biography of the PIXEL

Alvy Ray Smith
Co-Founder of Pixar

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by New Best-set Typesetters Ltd.

Library of Congress Cataloging-in-Publication Data

Names: Smith, Alvy Ray, 1943– author.
Title: A biography of the pixel / Alvy Ray Smith.
Description: Cambridge, Massachusetts : The MIT Press, [2021] | Series: Leonardo | Includes bibliographical references and index.
Identifiers: LCCN 2020029699 | ISBN 9780262542456 (paperback)
Subjects: LCSH: Computer animation—History.
Classification: LCC TR897.7 .S52 2021 | DDC 777/.7—dc23
LC record available at https://lccn.loc.gov/2020029699

10 9 8 7 6 5 4 3 2 1

Contents

Series Foreword

Leonardo/International Society for the Arts, Sciences, and Technology (ISAST)

Leonardo, the International Society for the Arts, Sciences, and Technology, and the affiliated French organization Association Leonardo, have some very simple goals:

1. To advocate, document, and make known the work of artists, researchers, and scholars developing the new ways in which the contemporary arts interact with science, technology, and society.
2. To create a forum and meeting places where artists, scientists, and engineers can meet, exchange ideas, and, when appropriate, collaborate.
3. To contribute, through the interaction of the arts and sciences, to the creation of the new culture that will be needed to transition to a sustainable planetary society.

When the journal *Leonardo* was started some fifty years ago, these creative disciplines usually existed in segregated institutional and social networks, a situation dramatized at that time by the "Two Cultures" debates initiated by C. P. Snow. Today we live in a different time of cross-disciplinary ferment, collaboration, and intellectual confrontation enabled by new hybrid organizations, new funding sponsors, and the shared tools of computers and the internet. Sometimes captured in the "STEM to STEAM" movement, new forms of collaboration seem to integrate the arts, humanities, and design with science and engineering practices. Above all, new generations of artist-researchers and researcher-artists are now at work individually and collaboratively bridging the art, science, and technology disciplines. For some of the hard problems in our society, we have no choice but to find new ways to couple the arts and sciences. Perhaps in our lifetime we will see the emergence of "new Leonardos," hybrid creative individuals or teams that will not only develop a meaningful art for our times but also drive new agendas in science and stimulate technological innovation that addresses today's human needs.

For more information on the activities of the Leonardo organizations and networks, please visit our websites at http://www.leonardo.info/ and http://www.olats.org/.

The Arizona State University–Leonardo knowledge enterprise provides leadership to advance ASU's transdisciplinary art-science research, creative practice, and international profile in the field: https://leonardo.asu.edu/.

Roger F. Malina

Advising Editor, Leonardo Publications

Beginnings: A Signal Event

Fi_
Walking boar, Altamira, ca. 20,000 BCE.

> Thou shalt not make unto thee any graven image, or any likeness of any thing that is in heaven above, or that is in the earth beneath, or that is in the water under the earth.
> —Exodus 20:4, King James Version

In the beginning—predating scriptural taboo, surely—a flame-lit graven image appeared to move. A picture on the walls of Spain's Altamira Cave was a walking boar *AND* the rock on which an ancient artist painted it *AND* the charcoal and ochre pigments used (figure 0.1). For 20 millennia or so there was no other place in the world to see that pre-biblical pig. Only there in the flickering firelight of a Paleolithic movie theater could you see its legs move and head bob.[1]

Even as late as 1800, just two centuries ago, a picture of Napoleon crossing the Alps *AND* the canvas on which Jacques-Louis David painted it *AND* the oil paints he used were an integral unit. Imagine that you wanted to share Napoleon's delightfully hagiographic picture (figure 0.2) in Europe with a friend in New York. There were no cellphones or video cameras yet, not even such a thing as a photograph. The only way to display him in New York was to transport the one physical painting there—if you dared. An engraving, etching, or sketch might help, but those were only better or worse copies—new images that would never fully and faithfully capture the original.[2]

Through all that time a painting and its medium of creation were inseparable. No one even conceived of separating the two. What could a picture be, independent of its medium?

Then in the early nineteenth century photography was invented, ushering in the world of what we now call "the media." Faithful reproduction was upon us. Movies followed in the late nineteenth century and television in the early twentieth. All media then were analog—smooth and continuous. And a picture could be transferred from one medium to another—a hint that something, after all, about a picture floats separate from its medium.

The notion of the digital—discrete and spiky—didn't fully exist until 1933. At midcentury, 1950, there were only a couple of digital pictures in existence. The few people who knew about them actually thought they were frivolous distractions from the more serious projects of digital computers. All the other pictures in the world were made and viewed with analog means—oil on canvas, ink on paper, and chemical emulsion on photographic film, to name a few.

But at the millennium, 2000, there was an unheralded event—the Great Digital Convergence: a single new digital medium—the all-encompassing bit—replaced nearly all analog media. The bit became the universal medium, and the pixel—a particular packaging of bits—conquered the world. It became possible to remove a painting, so to speak, from its canvas. As a consequence, most pictures in the world are now digital. Analog pictures have all but vanished relative to ubiquitous digital imagery. Museums and kindergartens are among the few reliable places to find the analog.

This book heralds that signal millennial event by celebrating Digital Light—the vast realm that includes any picture, for any purpose, made from pixels. It extends from parking meters to virtual reality, from dashboards to digital movies and television, from CAT scans to videogames to cellphone displays, and many, many more—*anything* mediated by pixels.

What's puzzling about the new medium is that you can't see it. Bits, and pixels made of bits, are invisible. Pixels are not to be confused, as they often are, with the little is the most consequential engineering miracle of all time. With that speed, computers amplify what we puny humans can do by unimaginable amounts.

But all of the mind-boggling, world-changing power of computers really reduces to a careful flipping between two states, often named 0 and 1. Computation is all bits. That may sound trivial, but I hope to inspire you with the unexpected beauty—and mystery—inherent in computation. Again, no mathematics required.

The most important but least known of the three fundamental ideas is the underlying theme of this book: you can pass back and forth between waves and bits—between the analog and digital worlds. The idea dates back only to 1933 when the Russian Vladimir Kotelnikov established it as we know it today. Its formal name is the Sampling Theorem. This entire book—being a biography of the pixel, with *pixel* being our name for a sample of the visual world—is about sampling. Pixels are invisible bits that represent visible waves. My fervent intent is that you understand this piece of magic and be amazed by how it works. No mathematics is required here either.

Now that I've said *no math required* three times in two pages, you might be thinking: *But what if some of us care about the math?* For you—but really for all my readers—I provide an online annotations site at http://alvyray.com/DigitalLight. There you will find additional details about people, places, and events that would have made this book too unwieldy to fit between its physical covers—and you will find as well mathematical equations to support the magic of Digital Light and the pixels that make it possible.[3]

There's a common misconception that a pixel is a little square of color. But in fact, the pixel is a profound and abstract concept that binds our modern media world together. It's the organizing principle of Digital Light.

A visual scene consists of an infinite number of points of color. Infinite is, by definition, too large to deal with. So how can we replace a smooth visual scene with only a finite number of discrete bits—the pixels—and not lose an infinity of information between them? The Sampling Theorem tells us how to do it. It's the secret that makes the modern media world work.

Sampling, which depends on Fourier's waves, was created almost simultaneously with computation in the mid-1930s. Sampling met computation and conceived a child, Digital Light, the subject of this book.

Contributions: Two High Technologies

Part II of the book is devoted to the history of two high technologies that shaped Digital Light: computers and movies. As in part I, I present each technology intuitively and give the history of its invention—debunking some common myths. The true stories are invariably more intriguing, inspiring, and complicated than the myths.

In Digital Light we can *take* pixels from the real world—from cameras on the International Space Station, say, tracking the latest hurricane. But more importantly for this book we can also *make* pixels from scratch. That's where computers come in and why I present the development of that high technology in such detail.

One of many surprises from my research for this book was the discovery that the first pixels occurred on the first computers. They were born together. So by establishing which computers were the first in the world, we also learn who had the first pixels in the world. That's why the chapter on computers in part II is called "Dawn of Digital Light: The Quickening." It features the first picture made with pixels, in 1947. And it introduces a driving concept of enormous force, called Moore's Law:

Anything good about computers gets better by an order of magnitude every five years.

Despite its simplicity, that statement is revolutionary. What was a factor of 1 in 1965, when Gordon Moore announced it, is about 100 billion now, and will reach 1 trillion by 2025. It's an exploding supernova. Moore's Law is the astonishingly powerful dynamo behind every development of computers for over 50 years—including Digital Light.

The digital movie part of Digital Light derives from classic film. The chapter titled "Movies and Animation: Sampling Time" reviews this pre-digital moving-image technology. It also helps to further illustrate sampling: the familiar "frame" of flickering film is actually a sample.

The verbs *take* and *make* apply to film movies too. We take a classic film movie from real-world sets with a camera. We make a classic animated movie from drawings of a non-real world. The central mystery of both kinds of movies is why they should work at all. How does a sequence of still frames deliver both motion and emotion? The Sampling Theorem helps explain motion, at least. The first digital movies—such as Pixar's *Toy Story*—are digital heirs to classic animated films.

The Rise and Shine of Digital Light

The Digital Light story is too large to cover in a single book, so this one necessarily has to focus. It covers Digital Light from the first pixels at midcentury to the first digital movies at the millennium. Unsurprisingly, I have chosen to write about the particular technologies, people, and history I know best from personal experience. I was born before computers—and pixels—and my career was principally devoted to creating the first digital movies. But that chosen path can help explain how Digital Light came to be more generally. Videogames and virtual reality aren't that far removed.

The three chapters of part III pick up the story of Digital Light after its dawn (told in chapter 4) and those first midcentury pixels. Moore's Law of 1965 cleanly separates them into Epoch 1 of Digital Light, before Moore's Law, and Epoch 2, after. The chapter titled "Shapes of Things to Come" is devoted to Epoch 1. The other two chapters, "Shades of Meaning" and "The Millennium and The Movie," cover the massive changes wrought by Moore's Law in Epoch 2.

Epoch 1 was the era of monstrous, but ponderously slow, computers. A few lucky individuals got access to those expensive beasts. From this period came the Central Dogma of computer graphics: a fictitious world is described inside a computer with three-dimensional Euclidean geometry and Newtonian physics. Then it's observed by a virtual camera that renders its view of the world into two dimensions in Renaissance perspective for display.

I cover Epoch 2 until 2000, more or less. The culmination at the millennium was the rise of three great digital movie studios, Pixar, DreamWorks, and Blue Sky. Their stories are intertwined.

I don't expect you to be able to create a digital movie after reading this book, but I hope you might understand how it's possible. It's like a music-appreciation course: After you learn the principles of music, you won't be able yourself to compose a Bach suite for cello, but you'll better understand the components and so love Bach even more. Understanding just how a modern movie like *Toy Story* is made can have the same effect.

How to Talk about High Technology

Several common themes emerged from my research on the technological histories in this book:

Theme 1. Conditions for progress: An idea, chaos, and a tyrant The conditions for the progress of a new technology appear to be an excellent idea, a disruptive chaos that demands and drives the idea's development, and a tyrant or tyranny that—often unknowingly—protects the creators while they advance the idea.

To give just one example, Joseph Fourier had the great idea that a continuum, such as a visual field, is all music. It's just a sum of waves. The chaos of the French Revolution ushered him onto the world stage in Paris, and the rise of Napoleon gave him his chance to work. Napoleon himself was the tyrant who exiled Fourier to the countryside to keep him out of Paris. In this protected place for a protracted time, Fourier developed his idea into the theory of waves that eventually returned him to Paris. His idea has influenced all subsequent science and technology—and Digital Light in particular.

Theme 2: The "high" in a high technology guarantees its history cannot be a simple narrative I've been dismayed repeatedly to discover that the received history of a technology almost always has it wrong. And it's not because there isn't evidence about what actually happened. The culprit seems to be our human love of the simple narrative based on the trials and tribulations—and ultimate triumph—of a single creative genius hero. There are departments of the history of science in many universities, but the history of technology seldom receives such scholarly scrutiny. This leaves the received stories to the individuals, companies, or countries that profit from self-aggrandizing versions. The tyrant typically claims the credit. There are numerous examples in this book.

To avoid this simplistic narrative pitfall, I rely on genealogy. For the history of each technology, I've designed a family flow chart—a graphical device, which seems fitting—for the persons, places, ideas, and machines involved. The chart shows who got what from whom (whether by hook or by crook) and the often-dense interplay of the cast of players. There's hardly ever a single person from whom all else derives. The flows from one chart proceed to another and then to another. We watch as the different strands braid together in different ways to influence the next generation. Each chapter thus becomes an extended caption for its flow chart, enhanced with detailed stories of the people involved and intuitive presentations of their ideas.

Theme 3: Technologies come from the interaction of different kinds of creativity Two classic errors in telling the story of high technologies are to pit scientists against engineers, and technically creative people against artistically creative ones. I call the former the battle between the (ivory) *tower* and the (chemical) *stinks*. Theory does differ from engineering, but the level of creativity isn't higher in one than the other. It's just different. Digital Light wouldn't have happened without the idea of the stored-program computer, a mathematical concept *and* Moore's Law, an engineering miracle. Equally pernicious is the belief that artists are creative but technical people aren't—or vice versa. Again and again we find that it's the interactions between equally creative scientists, engineers, and artists that lead to breakthroughs.

Let's now commence our two-century tour, at about the time commemorated by David's famous painting—but with a substantially less flattering portrait of Napoleon.

Foundations: Three Great Ideas

It's All Music

Fourier's great idea was this: The world is music. It's all waves.

This musical insight led to radio, which isn't a surprise perhaps. But it also led to television. In fact, among its many progeny are all media technologies—all the varied media that recently coalesced during the Great Digital Convergence. In short, Fourier's great idea stormed the world and fostered all the thunder and lightning of modern media.

But it's even more pervasive than that, reaching far beyond just media. Hardly a branch of science and technology is untouched by it—electricity and magnetism, optics, X-ray diffraction, probability theory, earthquake analysis, and quantum mechanics. The list goes on and on. It's no exaggeration to say that Fourier changed how we understand the world.

Regardless of which side of the culture aisle you sit on, you probably know Isaac Newton and Albert Einstein. The world recognized Newton's theory of gravitation and Einstein's theory of relativity in their lifetimes. But for the past 200 years, only physicists and engineers have kept Fourier's flame alive. They knew that Fourier was paterfamilias and celebrated him accordingly. And their work, in many diverse fields, demonstrated the greatness and universality of Fourier's wave idea.

Fourier himself took only the crucial first step. He was the first to formulate the idea mathematically and test it experimentally. Although he planted the explanation from which thousands of solutions would grow in nearly all fields of science, he himself cultivated only the first flower, his solution to heat flow in solids.

That unromantic specialty is a good reason for his delayed recognition. He was the "celebrated Fourier at the Academy of Science" for his theory of how heat flows. It doesn't ring of poetry like Einstein's concept that gravity is a warp in spacetime.

Yet Fourier's great idea is far more fundamental to the modern experience than Einstein's. Stated in its musical form, it's as lovely as spacetime warps—and more intuitive. There's no reason for it to remain hidden behind its usual cloak of impenetrable mathematics.

It's high time to reverse Hugo's assessment and celebrate both the man and his great idea. The ubiquitous modern technology of Digital Light is the right vehicle to finally give Fourier his due.

A Lust for Immortality

Jean Joseph Fourier was born on March 21, 1768, in Auxerre, an ancient provincial capital about a hundred miles southeast of Paris. Within ten years both his parents were dead, leaving him and his fourteen siblings orphaned. Revolution was in the air. The new American country was only a year old and Benjamin Franklin was charming them in Paris with his coonskin cap and flirtatious ways.[3]

There was something special about the orphan Fourier, and the good people of Auxerre made sure that the talented child was educated. They placed him in a school run by Joseph Pallais, who was noted for having been Jean-Jacques Rousseau's music teacher. Alas, there's no evidence that the man who discovered the music of the world was musical himself.

Fourier then attended the École royale militaire in his hometown Auxerre (there were 11 such branches of the école in all of France). Local supporters helped subsidize him again. These military schools uniformly stressed science and mathematics, and Fourier especially took to math—maniacally so.

At 13 he collected candle stubs to light a large "cupboard" and extend his study of mathematics into the wee hours, past "lights out." The suffocating—and surely cold—cupboard compromised his health for the rest of his life. Perhaps that cupboard first gave him his peculiar interest in heat.

But the candlelit extracurricular training in the cupboard soon paid off. He won a student prize in math. This skill launched his scientific career and ultimately made him immortal. And yet he also won a first prize in rhetoric, enabling his political life and leading directly to an early brush with mortality. His speaking skill almost killed him before the mathematical skill had a chance to develop.

The danger wasn't obvious at first. Further military school training might have led him directly to the army, but he certainly wasn't the military type. He was sickly for one thing, and a math geek. So, after completing his studies at the École royale militaire, he entered the church. He became a novice at an abbey in Auxerre and taught mathematics to the other novices there. He sanctified his name at about this time to Jean Baptiste Joseph Fourier, the form that he used thereafter.

Fourier disappeared into the cloisters on the very eve of the French Revolution. The few letters of his that survive show he was vaguely aware of it, but indifferent. He was more worried about his fame and the state of an algebra paper than the state of France. "Yesterday was my 21st birthday. At that age," he anguished in a March 1789 letter, "Newton had acquired many claims to immortality."

In September he wrote another letter, bemoaning the fate of his algebra paper. In the time between the two letters, the Revolution had begun. But the September letter said nothing about those tumultuous events.

Nevertheless, Fourier's private world did begin to change after that. He presented a paper on "algebraic equations"—presumably the paper that had worried him so much—before the Academy of Science in Paris in December.

And he departed the abbey without taking his vows. The revolutionary government soon suppressed religious orders anyway.

Yet Fourier wasn't a revolutionary for the next three long years. He taught math in Auxerre instead. In particular, he didn't sign the petition from the revolutionary Popular Society of Auxerre to the National Convention in Paris demanding the trial of King Louis XVI.

But in early 1793, just a month after the king's beheading, we begin to hear from Citizen Fourier.

> Bliss was it in that dawn to be alive,
> But to be young was very heaven.
> —William Wordsworth, *The Prelude*[4]

The dawn of the French Revolution famously enthralled Wordsworth, and young Fourier exulted in that very same heaven once he belatedly embraced it. His words were clumsier than the poet's: "It was possible to conceive the sublime hope of establishing among us a free government exempt from kings and priests." But he was just as passionate: "I readily became enamored of this cause, in my opinion the greatest and the most beautiful which any nation has ever undertaken."[5]

Fourier didn't have to leap far from these sentiments to politics itself. He gave a rousing debut speech in February 1793 to the revolutionary commune in Auxerre. He had a plan for raising local recruits for the Army of the Republic. Favorably impressed, the Popular Society of Auxerre adopted his plan and invited him to join. The Reign of Terror was now in full play—with ultimately over ten thousand beheadings of suspected enemies of the state. He wisely accepted the "invitation."

But naive Fourier—the recent novice—immediately blundered, and his timing was about as bad as could be. That rhetoric-trained tongue of his got him into serious trouble. Unwisely he used it to defend three citizens of Orléans. Unwise because they were already on Robespierre's enemies list, and Robespierre was the Terror's "tsar."

The revolutionaries promptly relieved Fourier of all duties outside Auxerre. Miserable at not being able to further the Republic, he trekked to Paris and met with Robespierre himself to plead his case. This bold tactic seriously backfired. His support for the Orléans prisoners had earned him a spot on Robespierre's short list of enemies. With the contorted logic of the Revolution, and despite protests from the always supportive citizens of Auxerre, the very Terror that Fourier had championed threw him into jail on July 17, 1794. It was effectively a death sentence.

"I experienced every degree of persecution and misfortune," he said. "None of my adversaries have run more dangers, and I am the only one of our compatriots who was condemned to death."[6]

The next stop, within days, would've been the Revolutionary Tribunal in Paris and a rubberstamp to the guillotine. He was justifiably terrified. He couldn't have known that just ten days later—July 27, or 9 Thermidor, French Revolutionary Time—Robespierre would fall hard. He who bade heads fall and never had enough—to channel Wordsworth again—tasted the guillotine himself. Lucky for the future of science, and the pixel in particular, Robespierre's head saved Fourier's.[7]

The Wave

Was that algebra paper of 1789—the one that pushed the Revolution from Fourier's mind—the first public glimpse of the great scientist? Did it contain keys to his great idea? Surely it sharpened his mathematical skills and increased his "mathematical maturity," as mathematicians say. But we don't know what was in it.

We don't know, in particular, when Fourier first began to use the *wave*, the fundamental shape in his great idea. But by 1807 he had gone public with it. It's the shape you get by unwinding a perfect circle, so it's a revolutionary form. And it's elegantly simple. The pixel has a noble heritage.

To get a picture of Fourier's wave, start with a circle (figure 1.1). An old-fashioned analog clock face serves the purpose. The tip of its second hand moves steadily around that circle, from a single minute mark to the next one, second by second. The lower picture shows the upper clockface, three seconds later.

The larger, red dot traces out the wave. An animated movie would be especially instructive here, but short of that, imagine the passage of time as movement to the right, as shown by time's arrow. There's a tick mark along the arrow for each tick mark along the edge of the clock. Imagine that the red dot is always tethered to the tip of the second hand by a taut horizontal band. The band lengthens toward the right as time passes, and the path traced by that red dot is the wave.

The important points here are, simply, what the wave looks like and, intuitively, that it's intimately related to the circle. The details of that intimate relation aren't as important as the intuition, but a few more details might help to register the intuition more memorably.

Consider the centerline of the clock—the line that connects 9 o'clock to 3 o'clock. The red dot always marks the current height of the tip of the second hand above or below that centerline. At the (purely arbitrary) moment chosen for the upper illustration, the

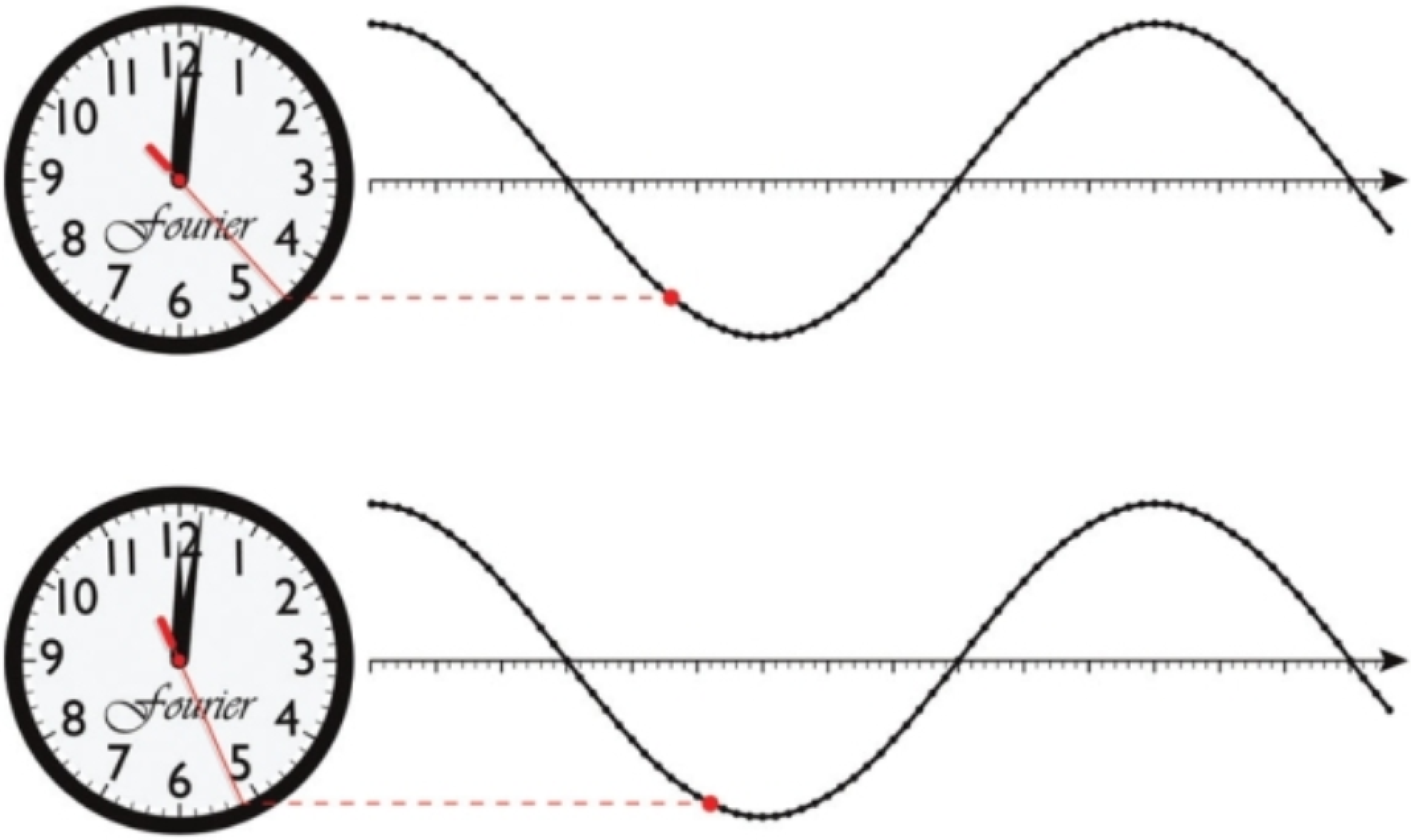

Figure 1.1

red dot has traced through twenty-three positions since the first position recorded, because twenty-three seconds have elapsed since then. During the next tick the red dot will move right to the next dot on the wave. And two more ticks after that yields the lower illustration. So as the second hand ticks its way around the clock, again and again, its tip—actually the red dot attached to the tip—traces out the wavy path shown, up and down, up and down . . .

The second hand makes one revolution after another every minute, so the wave traced by the red dot heads to the right forever. It extends to the left forever too. The wave in the figure appears to start one minute after noon—another purely arbitrary choice—but obviously the clock's been ticking off the seconds as far back as you might care to trace.

The wave in the picture is one of Fourier's waves. He didn't invent them, but he made profound use of them. These are the components of his music. Mathematicians call this particularly lovely circle-made wave a *sine* wave. Since this is the only kind of wave we need, I'll usually refer to it simply as a *wave*.

Everything about Fourier's wave is simple, beautiful, graceful, perfect. Scientists and engineers have a formal term for Fourier's solution for heat flow, and for his great musical idea generally. They describe it as *harmonic*.

let's start with music itself, a familiar physical reality that helps make intuitive sense of Fourier's profound idea.

Sound

Music is composed of—and only of—waves of different frequencies, called sound waves of course. The strings on a violin vibrate at different frequencies, and the same for a piano. Compared to a clock's pokey second-hand frequency of one cycle per minute, a piano's middle C is a speed demon of a sound wave from a string vibrating at 262 cycles per second. A clarinet or a flute resonates at certain frequencies, as does each of a pipe organ's pipes. Lyric coloratura sopranos sing at a higher frequency than do altos and much higher than baritones or basses. We say sopranos sing at a higher *pitch* rather than frequency, referring to our brain's perception of the music rather than to the physics of its creation, but it amounts to the same thing. A chord is essentially several waves—say three or four—played simultaneously, or added together, as we say. A choir consists of many voices at different pitches, and an orchestra comprises many instruments at different frequencies, from string basses to piccolos.

The dynamics of music, from pianissimo to fortissimo, reflect the amplitudes of the waves in the music. The higher the amplitude, the louder the sound. The massive double diapason organ pipe—with the foot pedal maxed to the floor—shakes a cathedral with the awfulness of the Almighty. Strike the piano key harder or turn up the volume on a radio to increase the amplitude of the waves. It's no surprise that the principal component of a radio or a stereo system is its amplifier.

Fourier's idea seems only natural when you're describing music, but the power of his idea starts coming into focus when you consider that all sound—not just music—is composed of audio waves. We speak of low-frequency rumbles and high-pitched whistles. Dogs famously hear higher pitches than we do. Our most visceral intuition of Fourier's big idea is that any sound or music is composed of audio waves of different frequencies all added together and interpreted by our ears and brain as a Stravinsky rite, a beloved child's voice, or even construction site noise.

Figure 1.3 shows the word *yes* as Fourier waves of various frequencies and amplitudes (time proceeds to the right). The "y" part at the left contains the lowest frequencies and the highest amplitudes—it's the accented part of the word. The "e" part in the middle has lowest amplitudes and mixed frequencies. The "s" part at the right has lower amplitudes and highest frequencies—the hiss of the "s."

Sound waves are actually made of—their "stuff" is—rhythmic compressions of air, or pressure waves. Consider the woofer in your speaker system—the big one that you can actually feel vibrate on loud bass tones. It's easy to imagine the woofer's fast-moving

Figu[illegible]

membrane shaking the air in front of it. This pulsing air moves from the woofer's surface—pushed along by it, so to speak. A louder sound causes higher compression of the air in each, larger vibration.

We understand these pressure waves directly. Think of a lowrider humping slowly down Crenshaw Boulevard in Los Angeles with its massive boombox shaking nearby windowpanes.

A very loud and low sound wave converts those high-pressure waves into physical vibrations, seemingly of the earth itself. But actually, those earth-shaking waves are exactly the same as the waves that enter our ears and make our eardrums vibrate in unison with the woofer, boombox, or double deep organ pipe. Then a clever system of little bones—the delightfully named hammer, anvil, and stirrup—pass these vibrations to the inner ear where thousands of tiny hair cells respond to the different frequencies. They pass the frequency information directly into the brain.

The normal young human ear can hear all frequencies between 20 cycles per second and 20 thousand cycles per second. Modern usage shortens "cycles per second" to "Hertz" (abbreviated Hz), but I'll stick with the longer phrase to preserve its intuitiveness. There are other sounds outside the human ear's capability, like the ultrasonic whistle the dog can hear but we cannot.

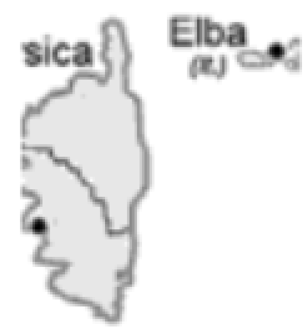

Figure 1.4

But what are the waves in vision—the ones that lead to pixels? At what frequencies does vision vibrate?

Napoleon Bonaparte

Just over a year after Fourier's birth, Nabolione Buonaparte was born on Corsica, an island in line with Paris and Auxerre but about a hundred miles southeast of mainland France (figure 1.4). Corsica had become French in the interval between the births of the two boys, so Buonaparte was just barely French. His name caught up only later, in his twenties, when he Frankified it to Napoleon Bonaparte.[8]

Bonaparte attended the École royale militaire in Brienne-le-Château. This was in the same system of schools attended by Fourier, who was at the one in Auxerre about 60 miles away. So Bonaparte received essentially the same training in science and mathematics as Fourier. It was enough to give Bonaparte an abiding interest in math. He was to interact with mathematicians all his life, and not just Fourier. There's even a geometry theorem, Napoleon's Theorem, named for him.

Bonaparte moved deeper into the military establishment with further schooling at the elite École militaire in Paris. Details of his training there don't concern us, but his final exams do. His examiner described Bonaparte as having "a thorough knowledge of mathematics." He was Pierre-Simon Laplace, sometimes called France's Isaac Newton.

In a move that boggles the mind of a contemporary American, Bonaparte was later to make Laplace a member of the French Senate. A noted mathematician and physicist as senator!

Fourier Meets Bonaparte

After Robespierre's fall, the revolutionary French government not only freed Fourier from jail but awarded him a plum job as professor at the newly formed École polytechnique in Paris—now France's MIT and Caltech combined. Finally, he was professing mathematics in Paris and meeting its leading lights, such as Laplace. This positioned him nicely to catch the eye of the ambitious Bonaparte who was looking for "savants" to accompany him on a trip to Egypt. And to catch a lot more heat—physically and politically.

Bonaparte had recently conquered Italy in a fight against the Austrians and returned to Paris a hero. With such acclaim, and an army in tow, the French government rightly perceived him as a threat. They were not unhappy to see him (never mind his army) set off in 1798 for Egypt and a fresh conquest. This departure finally brought Fourier and Bonaparte together.

In emulation of his hero Alexander the Great, Bonaparte took a team of French intellectuals (his savants) on the invasion, including the young professor Fourier from the École polytechnique. Again, the contrast to our times is striking. Imagine a team of leading mathematicians and archaeologists on the first plane into Iraq or Afghanistan.

The Egyptian expedition was a military flop, but a soaring intellectual success. The discovery of the Rosetta Stone was a famous high point. Egyptology sprang from those beginnings. Bonaparte himself founded the Institut d'Egypte while there and was its vice president. Fourier soon became its permanent secretary. He contributed to the lasting scholarly reputation of the Egyptian campaign by writing, over the next decade or so, the introduction to the massive (twenty-plus volumes) *Déscription de l'Egypte*—with overly fulsome praise for Bonaparte. Bonaparte even dropped in a few edits here and there in an effort to snatch an intellectual victory from the jaws of military defeat.

After several initial French victories—at Alexandria, Cairo, and the Great Pyramids—Admiral Nelson and the British destroyed the French fleet at harbor in Egypt. Despite British control of the Mediterranean, Bonaparte managed to sneak across it to France, after a little more than a year in Egypt, leaving nearly everybody else—and much unfinished military business—behind.

His motivation was to take a shot at controlling France—which of course he accomplished. He made himself First Consul of a new government just before the end of the eighteenth century—the first step on his march to become emperor.

When his savants attempted to follow him home, they weren't so lucky. The British allowed them through the naval blockade but kept the Rosetta Stone. It's still a highly visited treasure in the British Museum.

Bonaparte's hasty and opportunistic departure from Egypt left Jean Baptiste Kléber in the lurch with a military mess. Bonaparte made him general in charge of the army in Egypt, informing him of this via mail so he couldn't refuse. With this trick Bonaparte gained the scorn of Kléber, who was left to mop up.

Now Fourier made his second major political mistake. He became too closely identified with the unhappy general. Kléber made him a bureau director in Egypt. Then when a student assassinated Kléber in Cairo, Fourier gave the general's eulogy. That gifted, but troublesome, tongue was wagging again: first he'd offended Robespierre, now Bonaparte.

Bonaparte clearly didn't want Fourier propounding Kléber's views in the capital. He didn't want France to know the less-than-noble military details of his Egyptian campaign. Fourier had hoped that after Egypt he could resume his prestigious position in Paris, at the center of the intellectual action. Instead, Bonaparte exiled the outspoken Fourier to Grenoble.

But in doing so, he chose his words euphemistically, "requesting" that Fourier take the job as prefect of the Isère department, governed from Grenoble. In other words, Bonaparte "asked" him to become the provincial governor (prefect) of a province that was closer to Corsica than Paris—a long way from the action.

Fourier accepted. Bonaparte was now the most powerful man in France. It *was* exile, however, and that's how Fourier perceived it. He was the only leading mathematician and physicist in France who wasn't in Paris in the decade after Egypt.

Vision

You are probably comfortable thinking about what you hear as a sum of waves, but you don't think about what you see that way. This next step, then, is going to take a bit

Berkeley, California. There are spatial frequencies everywhere in it. Once you begin to see them, here or anywhere, then it isn't so difficult to make the leap that Fourier made—that waves and their various frequencies are *all* you need to completely describe a visual scene.

Consider the terracotta pots, some of the manmade constituents of this scene. Observe how they are on a loose grid. They aren't regularly spaced *exactly*, but they do have a repeat pattern that's loosely regular—say a pot every one inch or so. In the Fourier equivalent of this part of the picture there is a group of waves with slight departures from that frequency of pot placement to account for the slight departures from strict regularity.

The patterning on the pots is regular too. Consider only a single line through the picture, say a line aligned with the pots and indicated by the lower pair of arrows, line a. In other words, constrain your attention to the regular frequencies along one line at a time—as if the light intensities along that line were the sound amplitudes in a passage of music. Can you imagine a yes spoken by your garden's light intensities?

Starting from the left on line a, in detail, there is the medium frequency of the spiraling grooves on the pot there, then the high frequency (but low amplitude) dirt between the pots, then a medium frequency of the grooves on the second pot, then the low frequency of the wave defining the curve of the largest pot, then two slightly higher frequencies for the two other pots, and so forth.

The plants too have frequencies. The different cactus and succulent species have various frequencies of branch or leaf placement. To see this, constrain your attention to the almost regular frequencies at places along the line b. The line cuts through (the image of) two cacti of starburst-like structure. Along that line the "leaves" of each species occur at an almost regular frequency. (The cactus spines are actually its leaves. The "leaves" are its branches.)

The next line up, line c, intersects the large balls of hydrangea flowers, each composed of hundreds of tiny florets. The florets intersect the line at a higher frequency than any other plant yet discussed. The next line up (line d) intersects the leaves of the hydrangea. They occur along that line at a lower frequency than the florets do along line c. That's to say, there are about two leaf widths in the same space as there are dozens of floret widths. The florets are denser than the leaves or, stated in frequencyspeak, they occur at a higher spatial frequency.

The four lines through the picture are shown parallel to one another, but Fourier doesn't require that. You might exercise your intuition by looking for repeating patterns in lines through the picture at other angles, even perpendicular to the ones shown. For example, you can probably see the regular structure of the bumps on the large flat oval "leaf" of the prickly pear cactus near the center of the photo. Those bumps are where its spines are located. You could draw two waves more or less perpendicular to one another through the bumps in the two dimensions. That's more or less what the two-dimensional Fourier idea does.

Back to the manmade parts: The brick patio has repeat frequencies in width and breadth. The fence in the background (difficult to see) has its frequency of board repeats, and so does the trellis weaving above it.

We could continue this analysis for every part of the photograph. A line through the tree bark would produce a wave with high frequency detail because of the roughness. The pea gravel in the pots occurs at very high spatial frequency. And so forth. Gardens are symphonies of spatial music.

Fourier's legacy is that *everything* we see—the visual world presented to our retinas, whether it appears to have repeated patterns or not—is a symphony of spatial music. It can be represented with two-dimensional space waves of all frequencies and amplitudes. It's all music. It works just like music but in two dimensions for our eyes instead of one for our ears. We need this intuition about the wave nature of the visible world in order to explain pixels in the next chapter.

Rosetta Stone

Jean-François Champollion was a citizen of Grenoble while Fourier was prefect there, at Bonaparte's "request." Fourier introduced young Champollion to the Rosetta Stone (figure 1.7) and the mystery of the ancient Egyptian hieroglyphs that occupy its top third. Over the next two decades Champollion deciphered the hieroglyphs using the ancient Greek equivalent in the bottom third of the stone as an important source.

Fourier's special relationship with the First Consul made this possible. Champollion was drafted several times, but in every case Fourier successfully appealed directly to the chief Egyptophile at the time—Bonaparte—to exempt the young man. Thus free to pursue his passion, Champollion cracked the Rosetta Stone's code and established the field of Egyptology.[9]

Fourier's great idea—*the world is music, it's all waves*—is the Rosetta Stone of science, and Fourier was its Champollion. Scientists, engineers, and technologists of all ilk speak Fourier's frequency language today. It's the lingua franca of sound, image, video, and on and on—including many of the most common physical processes, and particularly the media technologies. Fourier showed how to translate back and forth between his frequency language and the ordinary languages of, say, colors across space or sound amplitudes through time.

Photo by Claudio Divizia.

You've perhaps heard someone say, "I haven't got enough bandwidth," in casual conversation. If you're technical, you'll immediately know what it means. If you're not, you've probably deduced that it means something like "capacity," but you don't have a clue why. It's a Fourier frequency word, that's why.

Bandwidth is, technically, a measure of the capacity of a communications channel. The young human ear, for example, can hear sounds of 20,000 cycles per second down to 20 cycles per second. So its bandwidth is the difference of the two, the width of the band of frequencies that are meaningful to the ear. In casual speech bandwidth is used metaphorically to mean something like your capacity to process information.

Frequencyspeak like this is one of the separators of the two cultures. It's Greek to the arts, so by the Rosetta Stone metaphor, frequencyspeak (Fourier's wave language) must be the hieroglyphs. Now let's see how Fourier's hieroglyphs—his waves in two-dimensional form—can represent pictures.

Corrugations and Furrows

What exactly is a two-dimensional wave? We've only looked so far at one-dimensional waves, like those used for sound. Fourier's idea applied to the visual world needs a two-dimensional version of his wave.

A one-dimensional wave is an unwound circle. A two-dimensional wave is an unwound cylinder. To visualize a wave in two dimensions, imagine the clock's second-hand wave extruded perpendicular to the page. What you get is a corrugated surface, like a furrowed field. A corrugated metal building is composed of sheets of two-dimensional space waves. Some porch roofs are made from corrugated plastic panels (figure 1.8). Certain ruffled potato chips are corrugated. The wavy layer between the flat surfaces of what we call corrugated cardboard is another space wave. A Mediterranean-style red tiled roof is another.

So, a corrugation is a two-dimensional space wave. Look at its edge—a cross-section—to see a one-dimensional wave. In fact, any straight cross-section of it is also a one-dimensional wave.

Fourier's big idea is that all of the visual world can be represented as a sum of—only—corrugated space waves of all frequencies and amplitudes. The only added twist, to handle the second dimension, is that the waves can be rotated to any degree of orientation. The furrows can run north and south, or east and west, or northeast by southwest, or at any other angle, especially important for the natural world. The "stuff" of the space waves might be galvanized iron or, as the figure here suggests, plastic.

point on a cellphone or television screen displays red, green, and blue light intensities, then our brain takes in the relative strengths of these three signals from our cones and perceives a color from them. Emitted light from electronic displays mixes colors in a different way than does reflected light from an oil painting, let's say, or a printed page of colored inks. For emitted light to appear yellow, we must turn on the red and green parts and turn off the blue part. But for light reflected off a white page to appear yellow, we must block the blue part of the white light and let only the red and green through. Regardless of the display technology, it all amounts to the same thing at the eye. I'll nearly always use *emitted* (not *reflected*) light terminology.

So imagine the grayscale wave in the top row of figure 1.9 as a wave of color instead—say a yellow wave. Where the crest of the wave shown is a gray somewhere between mid-gray and full white, the crest of our colored wave is a yellow somewhere between mid-yellow and full yellow. You can think of a colored wave as *three* waves at the same frequency but of different amplitudes, one each for our red, green, and blue receptors. So our yellow wave consists of three waves, one like the grayscale wave shown above for the red receptor, another just like it (at the same amplitude) for the green receptor, and a third like it but with zero amplitude for the blue receptor. With equal parts red and green and no blue, our brain sees a yellow wave.

The pictures in the middle and bottom rows of figure 1.9 show additional two-dimensional waves to familiarize you with what they look like. Just as for the one-dimensional case, the three waves are at three different frequencies with three different amplitudes. For the waves shown, the highest frequency wave (bottom row) has the lowest amplitude (mid-gray), and the lowest frequency wave (top row) has the highest amplitude (near white). But that doesn't have to be the case. A wave can have any frequency at any amplitude.

Here's the intuition to take away from these figures: the fine detail in a picture must come from the high-frequency waves in its Fourier description. They are the only waves that change fast enough. Computer graphics whizzes say things in frequencyspeak like, "There are lots of high frequencies in *that* scene," meaning—without explanation, of course—that it has many tiny details and sharp edges.

The latter remark about edges refers to a handy fact you should know. It's certainly not intuitive. It's a direct consequence of Fourier's math: the sudden transition at a sharp edge requires very high frequencies in Fourier's world; a really sudden change requires a really high frequency.

Here's my formulation of Fourier's punch line again:

> Any visual field—let's call it a picture or a pattern—is a sum of, and only of, gracefully undulating waves such as those shown, all unfurlings of perfect cylinders.

The visual world as waves is no more mysterious than sound—or let's say it's just as wonderfully mysterious. The Gettysburg Address can be described by the intensity of sound pressure at each moment during Lincoln's speech or it can be described as the sum of the sound waves of various frequencies and amplitudes that add up to the same thing. Fourier taught us that both descriptions are equivalent.

Scientists and engineers love Fourier's version because, by using the nonobvious frequency description, they can solve problems in the real world that resist solution in the obvious, equivalent point-by-point description.

Instead of talking about a thousand points of light, we talk in frequencyspeak about the frequencies and amplitudes of the waves of light intensity that add up to the points. Fourier's transformative teaching is that these two descriptions are equivalent. The point-by-point visual field is the Greek text. The Fourier wave musical equivalent is the hieroglyphs. For a geek, hieroglyphs are easier to read than Greek. The Rosetta Stone is Fourier's proof that the two are the same.

Fourier's critics didn't believe he was right, but the math said it was true. That's the magic of the idea—and the awesome, intuition-breaking power of math. Adding up waves of different frequencies really will give you a picture of . . . anything! Of my succulent garden, or the page you're reading, or of your child, for instance. That's Fourier's great and very large idea.

A Need for Heat

Fourier survived the Terror, and he even came to rule in France—if only as a provincial governor—thanks to Bonaparte. But neither his early anti-aristocratic stance nor his association with the emperor were handy credentials when the king returned to power—twice. Nevertheless, Fourier was a survivor in this dance of regency and empire, as he had been in the Terror. He managed, with considerable political skill, to remain prefect of the Isère department for almost thirteen years, and only the last year was dicey.

For 12 years, April 1802 to April 1814, Fourier made the best of his exile as an accomplished governor in Grenoble. He negotiated an agreement with *40* communes to drain the massive swamp of Bourgoin in Isère, quite a coup since all previous attempts at negotiation had failed. He pushed through a new road from Grenoble to Turin. He purchased books for the library, championed young men of his department—most notably Champollion—and worked on *Déscription de l'Egypte*, finally published in 1810. During this interval Fourier somehow found time to develop his wave theory. He didn't get there from listening to Mozart or observing the Alhambra but by thinking of heat in terms of waves. Perhaps his *théorie de la chaleur* (his theory of heat)—and with it his great harmonic idea—was his master plan for readmission to Paris.

Fourier himself said about his extraordinary interest in heat, "The question of terrestrial temperatures has always appeared to me to be one of the greatest objects of cosmological studies, and I have had this subject principally in view in establishing the mathematical theory of heat." He was looking for an idea as big as Newton's.

Victor Cousin, a man who knew Fourier, had a different take on his motivation. He wrote that when Fourier came back to Grenoble from Egypt, he never went outside, even in the hottest weather, without an overcoat and another in reserve. He'd developed a physical need for heat that sounds eccentric at best, but which amounted to a disease.

The French have a word for him that we lack in English: Fourier was *frileux*. He felt the cold terribly. He was very sensitive to cold, perhaps painfully so. The word also carries the connotation that he was incapable of personal warmth. People may have thought that Fourier was chilly in more ways than one.

Nobody knows when Fourier first started his work on heat, but he was engrossed in the theory while in Grenoble, just a couple years after Egypt. A flawed draft version probably completed in 1804 or 1805 contains the first known appearance of his waves. He presented an enormously improved version of that manuscript to the academic public in a late 1807 "memoir" called *Theory of Propagation of Heat in Solids*, now considered to be the paper that established his idea. During the time between writing the two versions, he had become familiar with the actual physics of heat and performed experiments to test and verify his mathematical results.

In Fourier's words, the whole process "contributed to give the theory an authority which one might have been inclined to refuse in a matter still obscure and apparently subject to so many uncertainties." That's classic science. Use theory to hypothesize. Use experiment to verify. It proves that Fourier was both a mathematician and an experimental physicist. He wasn't afraid to get his hands dirty.

Fourier realized that he could describe the complex pattern of heat flowing through a solid object as a sum of his waves. That meant that he could predict how (and how long) heat would flow from, say, the touchpoint of a cannon to its mouth.

So, for perspective, between 1803 and 1807, while he was perfecting his theory of heat and performing the supporting experiments, he was also negotiating the swamp-draining contracts, building roads, mentoring young men in his department, and working on the Egypt publication. Where did Fourier get the bandwidth?

Fourier navigated the politics of empire and revolution. He used his personal knowledge of Napoleon—as Bonaparte the emperor was called—to get the big road project approved based on a one-page proposal, which Napoleon granted in two days. But

Fourier had a harder time with the even fiercer politics of academia, perhaps because Fourier in exile was alienated from the hurly-burly of academic goings on in Paris.

Among the official Academy of Science readers of the 1807 "memoir" was Laplace, Napoleon's senatorial appointee, who had known Fourier since his days at the École polytechnique. Laplace was uncomfortable with the mathematics of Fourier's great idea. The first public attack on it came from Laplace's protégé, Simeon Denis Poisson, who assumed Fourier's chair at the École polytechnique when Fourier became prefect of Isère. The outcome of the long and acrimonious controversy was that the memoir was never published. Eventually, Laplace came to support Fourier, but Poisson—Fourier's implacable enemy—never did.

To posterity's mind, Fourier had established the theory in his memoir of 1807—making 2007 the idea's uncelebrated bicentennial—but he had to pass through another gauntlet before it was truly accepted. Probably because of the prolonged controversy, the academy announced that it would award a grand prize in mathematics in 1811 on the subject of . . . the propagation of heat in solid bodies. Cheeky! Fourier responded with what is known as his Prize Essay, the extended version of the 1807 paper that kept the principal contributions of the earlier memoir in place.[10]

So let's summarize what Fourier did while he was preparing the Prize Essay: Swamps. Roads. Mentoring. *Déscription de l'Egypte*. Again, where did he get the bandwidth?

One of the extensions of the Prize Essay beyond the 1807 memoir speaks to Fourier's Newtonian ambitions. He applied his theory of heat flow to a sphere—a planet-sized one—and so was the first to study the terrestrial phenomenon now known as the greenhouse effect. Sunlight that passes through Earth's atmosphere heats the Earth's surface, and the atmosphere holds some of the heat nearby. It's an imperfect metaphor but the atmosphere resembles the glass of a greenhouse, which allows sunlight in but doesn't let all the heat out. The heat of the natural greenhouse-effect warms the Earth and makes it the place for life as we know it. Fourier, despite his brilliance, couldn't have foreseen the unnatural greenhouse-effect heat that currently threatens life on Earth.[11]

Despite the academy shenanigans, Fourier won the prize with his Prize Essay. But again, he didn't get the full support of the academy commission who read it. In particular, Laplace voted against it.

"The manner in which the author arrives at these equations is not exempt of difficulties," the commission said in its report. "His analysis to integrate them still leaves something to be desired on the score of generality and even rigor."

Lacking rigor in math is the ultimate insult. The academy dragged its heels in publishing the Prize Essay as it had for the memoir.[12]

It wasn't until after Napoleon's final exile, to Saint Helena in 1815, and Fourier's return to Paris, that he managed to get his Prize Essay published. It still came under withering fire from the implacable Poisson. But Fourier found an elementary error in Poisson's writings and a false claim in his competing theory of heat. Fourier demolished both in a letter to Laplace. That was the coup de grâce to his opposition, finally bringing Laplace to his side.

Dancing with Tyrants

Triumphant scientific progress wasn't matched by glorious political success. Fourier's continuing dance with Napoleon upset his thirteenth and final year as prefect of Isère. Napoleon abdicated in April 1814, and Louis XVIII was restored as king. The ex-emperor headed for his first exile in Elba, an island about thirty miles straight east of Corsica and southeast of Paris. It was natural that he would pass through Grenoble on his way there. Under normal circumstances it would also be natural for Fourier to greet the man who had "given" him the job. The prefect would have to greet the ex-emperor. But the times were hardly normal. It promised to be an embarrassing encounter.

The restoration of the king had left Fourier's position provisionally intact, but would Napoleon consider his bow to regency a *volte face*? Fourier wasn't about to hang around to find out. Behind the scenes he engineered a detour in Napoleon's travel route to bypass Grenoble, citing potential harm to the ex-emperor. It worked. Napoleon sidestepped Fourier, and the new royal government proceeded to confirm Fourier as prefect. The king's brother, who would become Charles X, paid him a visit, consummating the deal.

But that wasn't the end of the dance with Napoleon. On March 2, 1815, Fourier received this frightening letter from a neighboring prefect:

> I have the honor to inform you that Bonaparte at the head of 1,700 men disembarked yesterday at Gulf Juan . . . [and] is heading for Lyon by Saint-Vallier, Digne, and Grenoble. No matter how extraordinary this news may seem to you it is entirely true.[13]

Napoleon returned from Elba for his final Hundred Days of power, displacing the king again. He returned along that fateful southeast-northwest corridor joining Elba and Corsica to Grenoble, Auxerre, and Paris. Fourier was going to have to deal with Napoleon again, just when he had managed to pass himself off as a loyal royalist.

He prepared the prefectural residence for occupancy by the returning emperor. He left a nice note there for him, giving him a personal welcome. In it he stated clearly his new loyalty to the king and explained his conflict. But he was hightailing it out the back gate of Grenoble to Lyon as Napoleon entered the front gate of the city.

Figure 1.10
Julien-Léopold Boilly, *Academician Fourier*, 1820.

If Egypt did induce Fourier's obsession with heat, then his monument's symbology celebrates both the cause and effect.

The younger Grenobloise, Jean-François Champollion, the first professor of Egyptology, followed Fourier to Pére Lachaise only two years later. His brother, Jacques-Joseph Champollion-Figeac, wrote one of the earliest biographies of Fourier. And Fourier's dear friend Sophie Germain—he hardly seems *frileux* when it comes to her—was also buried in Pére Lachaise shortly after him.

One of Germain's final acts was contributing to the fund for Fourier's Pére Lachaise monument. Poisson, enemy of them both and recalcitrant to the end, did not contribute.[17]

Fourier's troubles are reflected in the fate of his original 1807 masterpiece, the "memoir" that first announced his great result. The manuscript disappeared for over 160 years, finally rediscovered lying in the library of the École nationale des ponts et chaussées, the National School of Bridges and Roads. This is not as strange as it may first seem. The school is the oldest civil engineering school in the world, older than Fourier.[18]

The Nature of Genius

Why did Fourier's great idea meet with such resistance? The problem, as his critics saw it, was this: How could a highly irregular pattern—like an arbitrary song or painting—be equivalent to a sum of highly regular waves?

In mathematics something is proved or it's not, or it's provably unprovable. Fourier himself didn't attain this pinnacle of truth for his theory, although his intuition was remarkably sound. It fell to young Peter Gustav Lejeune Dirichlet to close the remaining mathematical gaps. He came to Paris in 1826 and met and admired Fourier. With the older man's guidance, Dirichlet firmly established the theory—with full rigor—and published it in 1829, the last year of Fourier's life.[19]

Nevertheless, some mathematicians are still troubled by the esoteric backwaters of Fourier's math. But engineers aren't. In the late 1960s I took a remarkably influential foundation course on Fourier techniques from Ron Bracewell at Stanford. He took special pains to emphasize the mathematical difficulties at the edges of Fourier theory and to teach the strict limits of its applicability. But he also emphasized how those mathematical niceties don't apply to analyses of real-world phenomena. Or rather he made clear that the real world falls within the limits established by mathematicians like Dirichlet.

Mathematicians must deal with all possible patterns, not just the patterns we actually find in the real world. Mathematicians deal in abstracts, but engineers deal in physical realities—heat, light, sound, roads and bridges, and images. To engineers Fourier's frequencies and amplitudes are as "physical" as the physical world they describe. If Mother Nature produces a pattern, then Fourier's great idea almost always works to describe it.[20]

Newton and Einstein knew that they were addressing the universe, and so did their admirers. Fourier didn't or couldn't foresee how universal his great musical idea was—nor did others at the time. We don't have a word for genius accumulated through time, or genius by ramification. Our usual notion of genius is local to a lifetime—both in accomplishment and in recognition.

Yet for two centuries engineers have made successful and massive use of his harmonic idea for our comfort and enjoyment. All of modern media depends on it. The pixel and the story of Digital Light is just one of the latest examples.

Surely Fourier has now arrived at his deserved position, whether we choose to call it genius or not. It only remains to reach across the aisle separating the two cultures and make him known and respected on both sides.

2 Kotelnikov's Samples: Something from Nothing

> There was a vague, unverified legend, unconfirmed by anybody, that you might nevertheless hear in camp: that somewhere in this Archipelago were tiny paradise islands . . . and the only work was mental work—and all of it super-supersecret. And so it was that I got to those paradise islands myself (in convict lingo they are called "sharashkas") and spent half my sentence on them. It's to them I owe my survival.
>
> —Aleksandr Solzhenitsyn, *The Gulag Archipelago*[1]

The man who invented the pixel and started the digital revolution was Chairman of the Supreme Soviet of Russia. Not at the same time but the same man, nevertheless. His name was Vladimir Kotelnikov. In 2003, when he was 95, another Vladimir—Putin—knighted him in the Kremlin. By then he sported most of the merit badges of Soviet Russia, including two Stalin Prizes and six Orders of Lenin, and was twice a Hero of Socialist Labor. He had survived the 1917 October Revolution and all the purges and wars that since define modern Russia. He had barely avoided the Gulag—the very island in it where Solzhenitsyn had toiled—protected by the powerful wife of one of Stalin's bloodiest henchmen. He warned Americans about Sputnik and mapped Venus with digital images—pixels from space.

Kotelnikov was honored in America too, with the Alexander Graham Bell Medal in 2000—befittingly as the Great Digital Convergence transpired. Yet he's largely unknown in the States. He rarely gets credit for his greatest discovery—the Sampling Theorem—the idea that lies at the heart of the entire digital media world. That crown usually goes to Claude Shannon, a famous American engineer and mathematician, even though Shannon never claimed it.

Like Fourier's story, Kotelnikov's features three drivers of technological breakthrough: a great scientific idea, the chaos of revolution and war that turns it into a necessary invention, and tyrants who protect the scientists and promote their technology.

Figure 2.1

Kotelnikov's great idea—which led directly to the pixel—is intertwined in the remarkably parallel life stories of obscure Kotelnikov and famous Shannon.

The Spreader

Here's the idea: digital can *faithfully* represent analog. The discrete and separated and spiky can accurately represent the smooth and joined and curved. The broken continual can truthfully represent the unbroken continuous. If this doesn't surprise you now then I hope to provoke your amazement soon, because it appears that we can throw away an astounding amount of information—an infinite amount, in fact—without losing anything. This is the key idea that makes Digital Light (as well as Digital Sound) possible. It's the fundamental truth that enabled the Great Digital Convergence and hence the modern world.

Just as the wave is the shape that represents Fourier's frequencies, there's a shape that signifies Kotelnikov's samples (see figure 2.1). We'll soon see that it's intimately related to the "shape" of a pixel. Mathematicians call it a *sinc*, and engineers a *reconstruction filter*. Because both those names are nonintuitive, I call this lovely shape a *spreader*. You'll see why shortly.

Notice that a spreader resembles one of Fourier's waves with its crests and troughs progressively squeezed until they're ultimately reduced to nothing in either direction. In fact, that's exactly what it is. The associated wave has an amplitude everywhere as high as the central hump, and a frequency the same as the spreader's frequency of up and down wiggles (figure 2.2).

The spreader comes from mathematics, not the real world, but it might remind you of a pebble dropped in still water, with ripples radiating outward, their height decreasing with increasing distance. Like a wave, a spreader continues to wiggle in each direction forever. But the crests at some distance from the central hump are so low that they don't matter. That's important in the real world, as we'll see. The earliest picture of the

His next two wars were the October Revolution of 1917 and the subsequent civil war between Reds and Whites. Russia was transformed. Young Vladimir was too—but it wasn't the wars that changed him. In the midst of the chaos he heard, for the first time, a radio broadcast.

"How does it work?" he asked his father.

"It's something you can't understand yet."

That challenge focused him, at age 10, on a life in radio. He would spend most of the next nine decades in radio and communications engineering, a career coincident with the rise, turmoil, and fall of the Soviet Union.[8]

"Great was the year and terrible the Year of Our Lord 1918, the second since the Revolution had begun." So begins Mikhail Bulgakov's *White Guard*, his account of the terrible, destructive, anarchical world of Kiev. In their second case of bad timing, the Kotelnikovs picked exactly this unfortunate moment to move to Kiev again and experience Bulgakov's nightmare firsthand. The professor boiled soap, and the children unraveled curtains for the thread—anything to sell to drive away the hunger.[9]

In 1924 Aleksandr moved his family yet again, this time to Moscow. He took a new position at the Moscow Higher Technical School. Part of the school would soon metamorphose into the Moscow Power Engineering Institute (MEI in Russian) and become one of the leading technical universities in the world. Think of it as the Moscow Institute of Technology.[10]

Vladimir, our Kotelnikov, was one of MEI's first graduates. They wouldn't accept him initially because of his intellectual pedigree—he wasn't of worker or peasant stock—but a fortunate rule change allowed him to enter, and he never left. His 1931 diploma was in electrical engineering, with a specialty in radio. He would be with the institution for 75 years.

Then it happened—his annus mirabilis. The year 1932 saw Kotelnikov create two papers on his own, with no apparent supervision, each of which would have established him in the annals of engineering forever. One of them was "The Theory of Non-Linear Filters" and won't further concern us. But the other contained his great idea, the Sampling Theorem. He submitted it in November, and it was published the next year with the unpromising title, "On the Transmission Capacity of 'Ether' and Wire in Electric Communications."

When he submitted it to the MEI faculty, one of them said, "It appears correct, but sounds more like science fiction." Something from nothing. Nevertheless, they approved it, launching him on an academic trajectory toward a deanship at MEI.[11]

In 1933, with a lectureship at MEI, Kotelnikov also began working for the People's Commissariat for Communications (NKS in Russian). Communications are of utmost

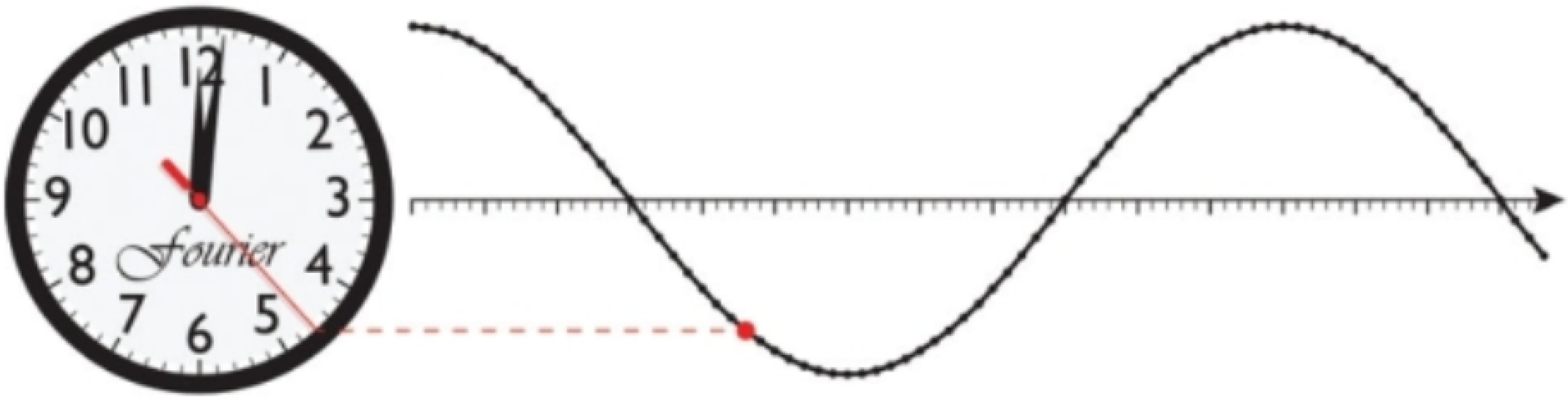

Figure 2.4

importance in war, so it's no surprise that the Bolsheviks had created NKS on the very day they seized power in Red October 1917. Kotelnikov became a communications engineer there and eventually headed his own institute. He would always have one foot in the ivory tower and the other in the real world of politics and war.[12]

He was on his way, with two important papers and positions in two prestigious organizations. He was poised for rapid advancement in both academia and government. On both paths Kotelnikov would have to dance with tyrants, as Fourier had.

Digital and Analog Infinities

Let's not be shy about the word *infinite*. There are actually many different kinds of infinity, but the only two we need here are the digital kind and the analog kind. The familiar diagram (figure 2.4) of the clock's second-hand wave will help make the difference clear.

You'll recall that there's one round black dot on the wave for every minute mark on the clock's face during each and every cycle the second hand makes around the dial. As the second hand progresses around and around the dial, the dots unfurl to the right forever. How many of them are there? Well, you can count them—one, two, three, and so forth—but you'll have to count forever. That's digital infinity. There's always another one. Mathematicians call it countable infinity, for the obvious reason.

The second kind of infinity, analog infinity, isn't so easy. Consider two successive dots on the wave. How many points are there on the wave between the dots? Answer: there are so many you can't even count them. Analog infinity is larger than digital infinity—as strange as that sounds. The mathematician Georg Cantor proved it was true, and here's what he meant:

Between any two points on the wave, there's always another point on it. For example, there's the point—on the wave—halfway between the two points. Now think of

Figure 2.5

that midpoint and the left of the two original points. Is there a point between them? Yes, always—for example, the point halfway between them. Now repeat for this one-quarter point and the left of the two original points. And so on, ad infinitum, as the expression goes. The difficulty is that you can never divide finely enough to reach an end to the dividing process. In other words, you can never get to a place where you can even start counting. Mathematicians like to call it uncountable infinity, but I'll stick with analog infinity. Both work: Smooth things have an analog or uncountable infinity of parts. Discrete things have a digital or countable infinity of parts. In a profound way, digital is lesser than analog—even if you use lots of dots to represent the smooth thing.

But Kotelnikov's great idea appears to be—and here's the surprise—that digital is equivalent to analog. Nothing is lost by going digital. A discrete digital thing can faithfully represent a smooth analog thing. Figure 2.5 shows a snippet of sound, say, or of a visual scene along a horizontal line through it. Kotelnikov's idea works in either case. The straight line along the bottom is the zero loudness or zero brightness level—completely silent or completely dark. The curve is the changing loudness of the sound as time passes, or the changing brightness of the visual scene as you move to the right along one line through it. In either case the big dots highlight equally spaced points along the snippet. We'll build intuition with this one-dimensional example, then gradually extend to the two dimensions that a full visual scene actually requires—just as we did for Fourier waves in the first chapter.

Figure 2.6 is what you get if you omit all the points on the smooth curve except the ones at the big dots. Between these points all we have now is the straight line, the zero loudness or zero brightness level. It's not hard to imagine what the two-dimensional version would look like. Think of an uneven bed of nails, equally spaced horizontally and vertically. Their heights would vary according to the brightnesses on a corresponding smooth surface that was a visual scene. And the bed of nails would be a "surface" that is zero everywhere except at the nails.

Figure 2.5 is analog, and figure 2.6 is digital. The spikes in the latter are called, naturally enough, *samples* of the analog curve. In the bed-of-nails case for two dimensions,

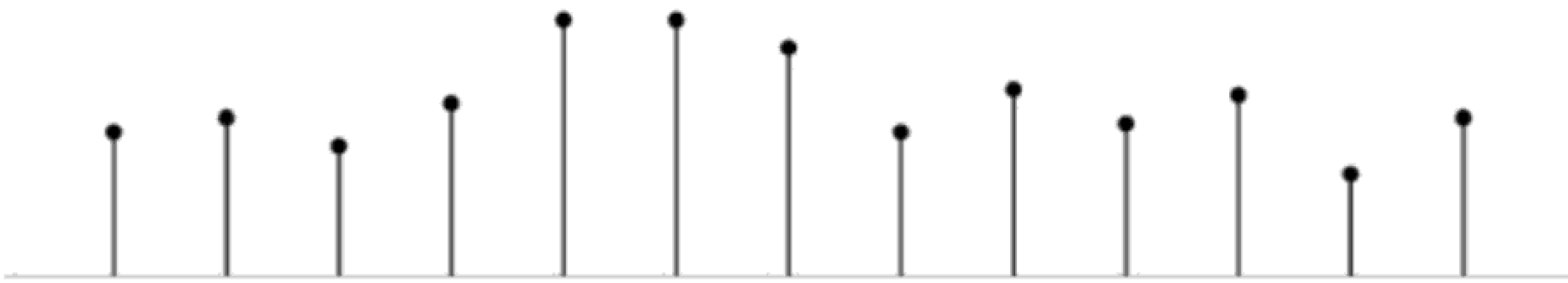

Figure 2.6

the nails are the samples of the corresponding analog surface. Kotelnikov's great idea says that we don't need the full smooth curve to represent a sound or the full smooth surface to represent a visual scene. We need only the samples. In other words, the analog infinity of points between the highlighted points in the first figure can be ignored! He appears to say that nothing *can* represent something. How could that be? The answer lies hidden, of course, in the "appears to."

You might imagine that if you simply used more samples and placed them "close enough" they would *become* the analog sound curve. It's the same intuition many people have that pixels—whatever they are—spaced closely enough together would *become* the visual scene they represent. But this intuition is wrong. You can't get close enough. You can't make digital infinity ever reach analog infinity. You can't count what's uncountable. Yet Kotelnikov appears to say you can. What gives?

Furthermore, his idea says that in the second figure the dots shown *are* close enough—that you gain no advantage, no additional information—by taking more closely spaced samples. Are you puzzled yet? I hope so because we're getting to the crux of the matter—and its elegance.

With these questions quivering in the air, we're almost ready for a first pass at Kotelnikov's great idea. But first let's revisit Fourier's idea, since Kotelnikov's depends on it. Fourier taught us that an analog sound or visual field can be represented by a sum of waves. Figure 2.7 shows one of the waves in the sum of waves for the analog snippet we've been using, repeated at the top (designating the sample locations with dots) for convenience. You can *see* that nothing in the snippet wiggles up and down any faster than this wave, so this wave is the one with the highest frequency. All the other waves in the Fourier sum for the snippet have lower frequencies, otherwise you would see a faster wiggle somewhere in the snippet.

Here's Kotelnikov's great idea: if you sample something smooth at twice its highest Fourier frequency, then you can always *exactly* recover the smooth something using only the samples. The samples are discrete, disjointed, separated from one another—definitely *not* smooth. This is the first part of his idea—the great Sampling

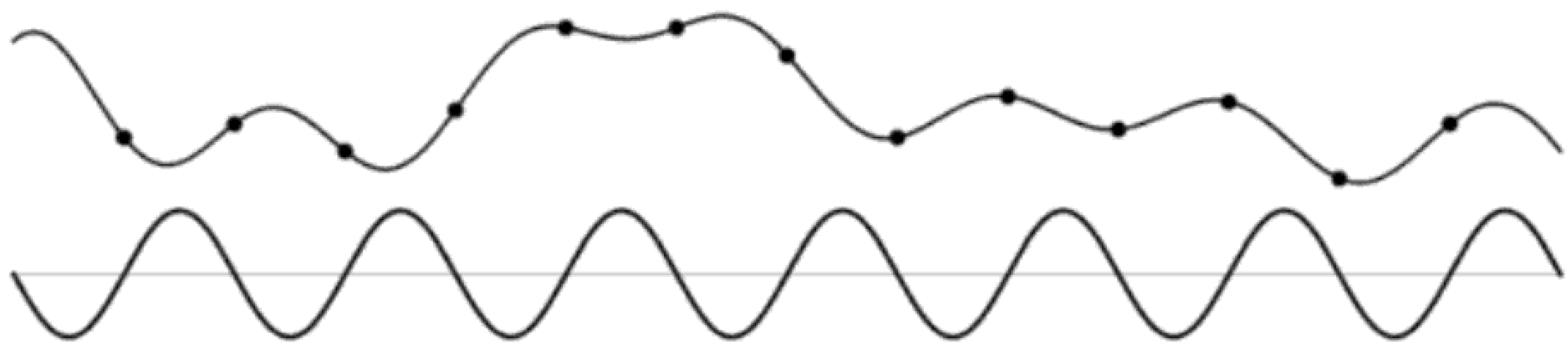

Figure 2.7

Theorem—the part that says it's *possible* to substitute digital disjointedness for analog smoothness. The second part tells us *how* to go about the actual recovery of the original analog from the digital samples.

Kotelnikov stands on the shoulders of the giant Fourier. Fourier's frequencies capture how fast an analog picture changes across a field of view. Then Kotelnikov's great idea tells us how to digitally represent Fourier's waves. Astonishingly, only two samples suffice for each cycle of the fastest changing wave. It takes two, intuitively, because one sample represents the cresting part of that wave and the other the troughing part.

The Pixel

There's a name in the digital world for Kotelnikov's samples of a visual field. We call them *pixels*. There it is! That's the definition of pixel. It's intimately associated with both Fourier and Kotelnikov. Kotelnikov's sampling is what makes Digital Light possible.

Pixels are *not* little squares! This may surprise you because *very* often they're represented that way—so often, in fact, that many people *equate* pixels with little abutting squares of color—which is perhaps the most widespread misunderstanding of the nascent digital age. The word *pixelation* has even institutionalized the misconception.

In fact, pixels have *no* shape. They're just samples taken on a regular grid—the uneven bed of nails. They exist only at a point, so they have no extent, no width, zero dimensions. You can't see them, and they have no visible color. They just have a number *representing* a shade of gray, or three numbers *representing* a color. It's the recovery of analog from digital, using Kotelnikov's idea, that appears to give a pixel shape, as we'll see.

The word *pixel* itself had to fight for existence. Pixels were called many other things at first—*spots*, *point arrays*, *raster elements*, *picture points*, and *picture elements*, for example. *Picture elements* won, but then came a battle over representing the term in shorthand. For many years IBM and AT&T made an effort to contract it to *pel*. But with their candidate *pixel*, the vibrant young image-processing community of the mid-1960s have prevented Shannon either from making his knowledge of the theorem public or from learning about it during those years.

Here are the two giants, side by side (figure 2.9). Kotelnikov is on the left—with the wilder hairdo—in 1932 at age 24, a year before he proved the Sampling Theorem. On the right is the slender Shannon at an undetermined age, but perhaps about 32, near the time of his proof in early 1949. Both men were leaders of digital communications in their respective countries—especially communications in the presence of noise or encryption. Both received the highest awards. And both stated and proved the Sampling Theorem as it's used today in Digital Light.

We can't help but wonder if there was cross-cultural leakage. Curiously, the younger Shannon often paralleled Kotelnikov's intellectual achievements a few years later—but I've found no evidence that Shannon knew Kotelnikov's work. In any case, it's no surprise that Russians call the great idea Kotelnikov's Sampling Theorem. Shouldn't we?[17]

Spread and Add

The second half of Kotelnikov's idea—the great Sampling Theorem—tells us how to reconstruct a smooth picture from non-smooth pixels, and to do it accurately. The

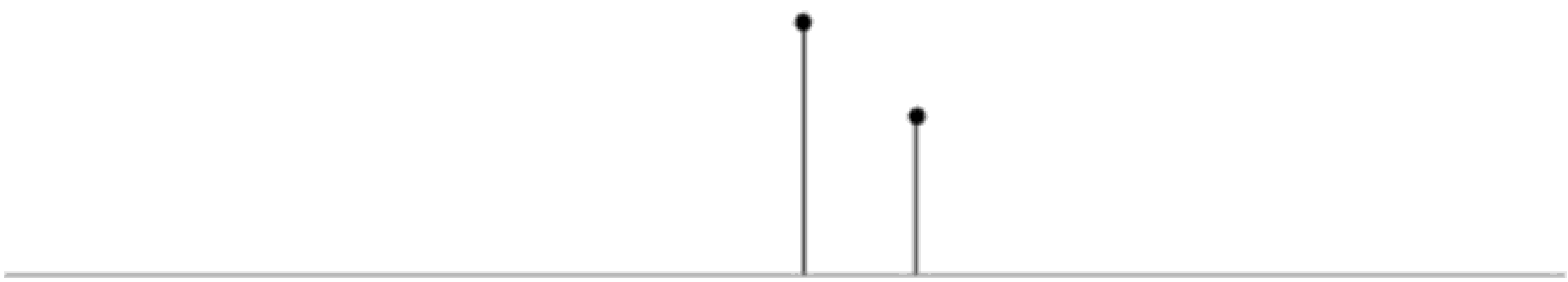

Figure 2.10

surprise about digital is that almost no information seems to exist in a digital image—an analog infinity of points has simply been omitted between each pair of pixels. The same goes for digital audio and each pair of soxels. The second half of the Sampling Theorem tells us where those missing infinities can be found.

Here's how to recover the analog from digital. Spread each pixel with the spreader, the shape fundamental to this chapter. Add up the results. That's it. The Sampling Theorem tells us that this spreading and adding process accurately reproduces the missing infinities between the pixels! As with all great theorems, such deductions aren't at all obvious. We have to trust the math.

The running example will help again, but with only two soxels shown for simplicity (figure 2.10), the two central ones. (We'll extend it to pixels shortly.) Recall that a soxel is a sample of an analog sound curve, where the height of the curve above the zero line represents its loudness. So the height of a soxel represents the loudness of the curve at just that point sampled by the soxel. The soxel at the right is less loud—quieter—than the one on the left.

First we'll *spread* the left soxel with the spreader, which we saw in figure 2.1. Recall that it wiggles at the same frequency as the Fourier wave with the highest frequency of the original snippet. Its maximum amplitude (loudness), at the central hump, represents the volume turned all the way up. To do the spreading replace the left soxel with a copy of the spreader (figure 2.11). I like to say that this spreads the soxel from no shape (nothing) to the shape shown (something). Its highest point—the peak of its central hump—has the same loudness as the soxel it replaces. The two soxels are shown dashed. In particular, this makes clear that the height of the spreader—its maximum loudness—matches that of the left soxel. In this example, that loudness is 80 percent of full volume. Imagine that you have a knob for that loudness adjustment.

Now move another copy of the spreader into place over the right soxel (figure 2.12) and twiddle the knob until its maximum loudness matches that soxel's loudness, in this example 50 percent of full volume—hence spreading the second soxel.

And here's the result (figure 2.13) of adding these two "spread soxels" together. At each horizontal position take the two heights of the spread soxels there (light gray),

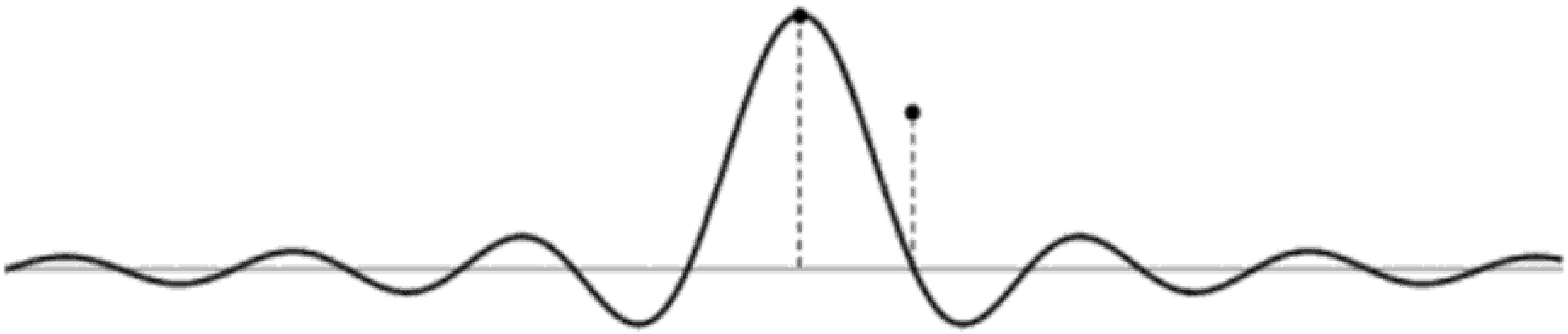

Figure 2.11

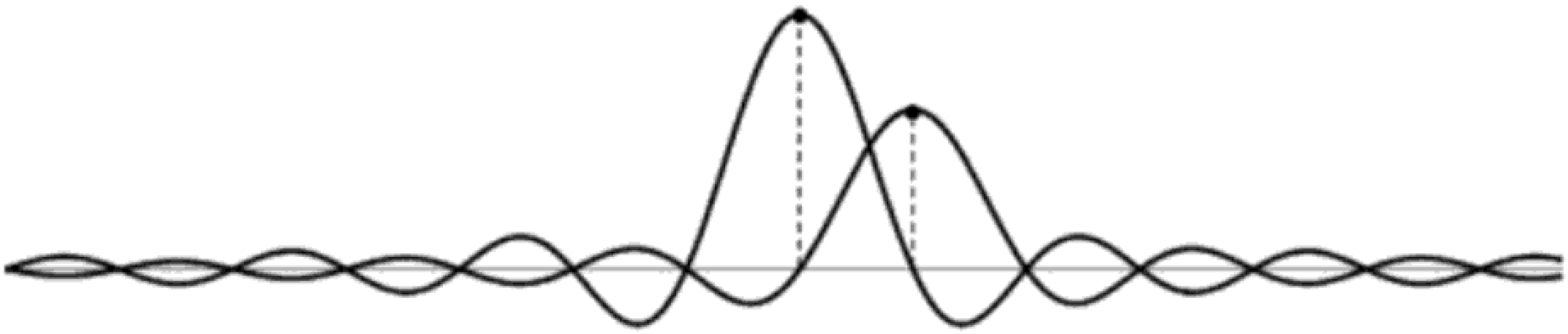

Figure 2.12

Figure 2.13

measured from the zero-loudness line, and add them together to get a point on the bold curve.

I've ignored reality so far. The spreader—this chapter's featured shape—doesn't exist in the real world. It's infinitely broad. It continues to ripple left and right forever. Obviously, no real-world spreader can be that broad, so real-world spreaders by necessity have to be approximations of the ideal spreader.

A popular, practical, and remarkably accurate spreader is called a cubic spreader (figure 2.14). Note how closely it resembles the middle part of the ideal spreader, including the presence of two lobes going negative (below the zero-loudness line). The cubic spreader is zero everywhere beyond two samples left and two right of the central sample—the one spread by the spreader. In other words, it has a finite width, so it can exist in the real world.

Figure 2.14

What I've described so far is one-dimensional spreading. Sound varies in one dimension only (of time) so it's an accurate picture for soxels, but not for pixels. A pixel spreader must spread in two dimensions since images extend in two dimensions (of space), often called horizontal and vertical. A pixel spreader must spread each nail (pixel) in a bed of nails so that each "spread pixel" contributes to a two-dimensional surface that we can see. You can think of the preceding figures as accurately showing the cross section of a pixel spreader in the horizontal dimension, keeping in mind that the cross section in the vertical dimension is exactly the same. But we can do better.

Figure 2.15 shows a full two-dimensional pixel spreader. Drop a guillotine through its brightest point—so that the blade slices the little mountain in half, from its peak to its base as shown. Then the bleeding edge revealed is exactly the cubic spreader above. You'd see the same edge if you cut the little mountain in half the other way. Since this pixel spreader is a cubic spreader in each dimension, it's called a *bi*cubic spreader. It's

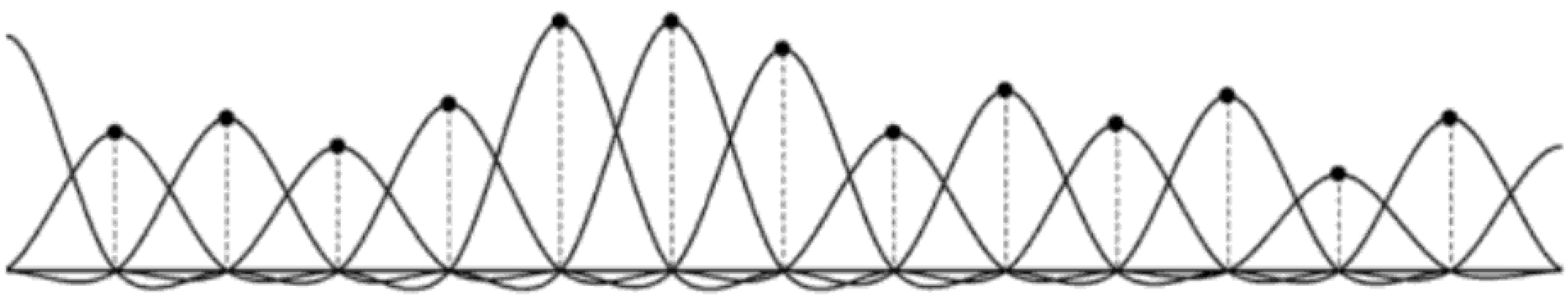

Figure 2.16

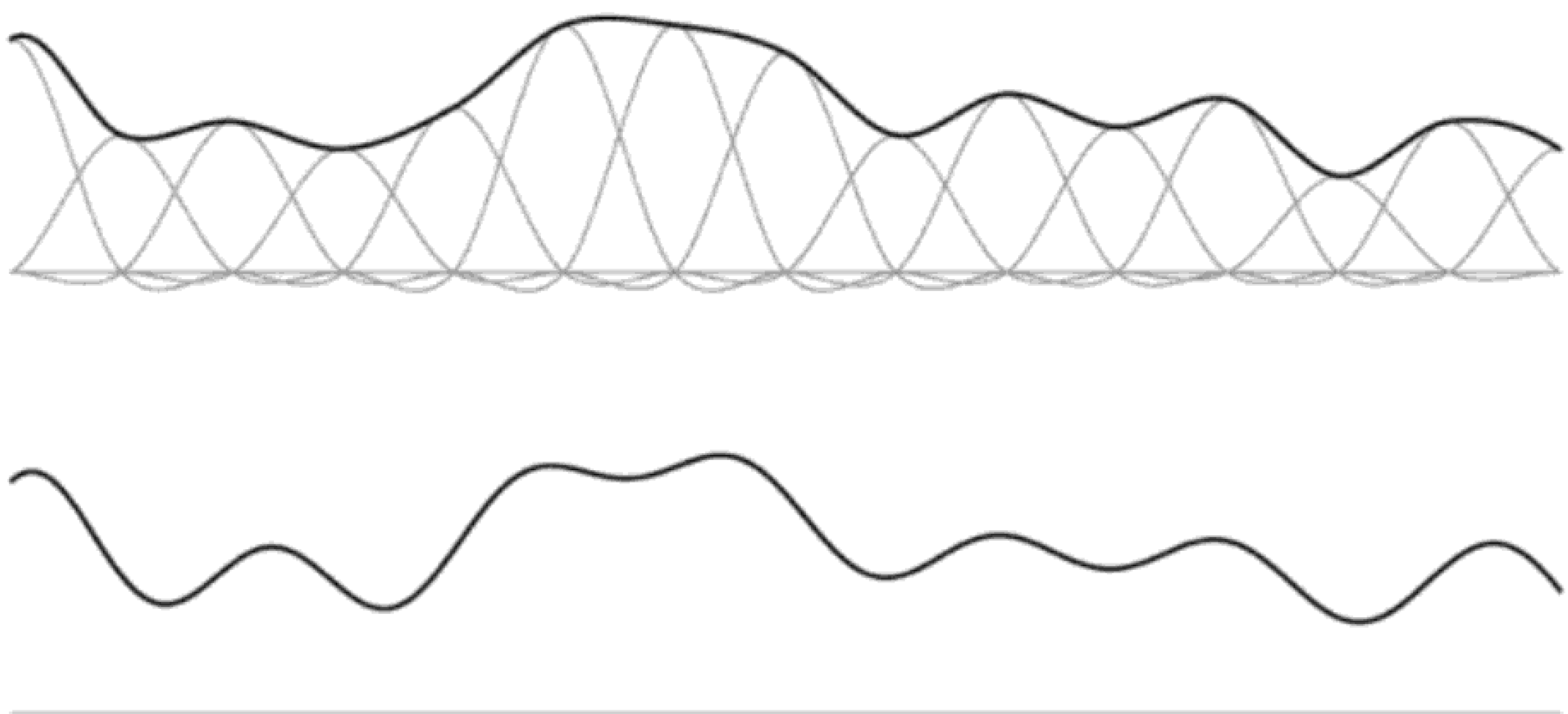

Figure 2.17

the kind built into Adobe's Photoshop (one of the oldest and most popular pixel apps) for changing the size of a picture.

So now let's spread and add a row of pixels, but let's do it by using this bicubic spreader instead of the ideal one. For simplicity and consistency, we'll do this in cross section. But imagine that each spreader is actually a little mountain like the one we've just seen—and that the uneven bed of nails extends in the other dimension with a little mountain at each of its nails. First, we spread the pixels (figure 2.16).

Then we add them all up to get a resulting brightness everywhere. Figure 2.17 shows that result along one horizontal line through a visual scene. The bold brightness curve reconstructs the original analog snippet, which I repeat just below it for ease of comparison. It's not perfect—just an approximation—since we've used a less than ideal spreader, but it's close. It's like a worn or duplicate key compared to the original. The teeth of either open the lock.

You can imagine that these figures—already complex for just a dozen or so pixels—would become hopelessly complex for, say, a million pixels. The important point is

Figure 2.19

Surely then the pixels must be little squares, right? Not at all. Zoom uses a quick-and-dirty—and inaccurate—trick to make you think it's "magnifying." Zoom simply repeats each pixel 16 times horizontally, and then repeats each row of 16 identical copies 16 times vertically. The result is that each original pixel is replaced with a 16 by 16 square array of pixels of the same color. So each array looks—what a surprise—like a square of that color. But each square is definitely *not* a picture of one of the original pixels close up. It's a picture of more than *200* (spread) pixels for each original pixel, arranged in a square. The Zoom trick was useful back in the bad old days of slow computers, but in the modern world it is simply misleading. It's time to scotch the notion that pixels are little squares. It never was true.

Figure 2.19 (right) shows you what the original picture looks like really, when Photoshop does a proper magnification by 16 (with a feature called Image Size). The second half of the Sampling Theorem explains why you see blobs, each with the original gray in the center but softening rapidly in every direction to whiter shades of gray, partially overlapping neighboring blobs.

The Great Terror and Its Tyrants

Stalin's Great Purge—the Great Terror—lasted from 1934 to 1939. It was implemented by the notorious NKVD (People's Commissariat for Internal Affairs), predecessor of the KGB. Many were condemned to death in the infamous show trials, and hundreds of thousands, perhaps millions, of Russians were sent to the prison camps—the islands—of the Gulag Archipelago. It was Solzhenitsyn who taught us about the Gulag and the unfolding tragedy that "was unprecedented in remembered history."[18]

Two tyrants for Kotelnikov emerged from this horror. Robert Conquest, the definitive historian of the Great Terror, described four powerful young Stalinists who led the purge as "particularly murderous." Among them were Kotelnikov's tyrants Georgi Malenkov and Lavrenti Beria.[19]

Malenkov's report to the Eighteenth Party Conference in Moscow in February 1941 kicked off yet another purge, Stalin's Secret Purge, second only to the Great Purge. Malenkov revealed that industrial capacity was plummeting, and incompetence was rampant. So just before Hitler invaded Russia in June, dozens of the People's Commissariats (NKs in Russian—what we might call Federal Bureaus or National Departments) were decimated, including in particular the People's Commissariat for Communications (NKS). The entire NKS was disbanded—except for Kotelnikov's lab, which was left untouched because of its wartime importance.[20]

During World War II, Stalin headed a committee of five that ran the country: Malenkov and Beria were both members. Beria also ran the NKVD with the assistance of a deputy, Viktor Abakumov, a third Kotelnikov tyrant. When Stalin created SMERSH (yes, you Ian Fleming fans, this counterintelligence organization really existed), he put Abakumov in charge.

After World War II Malenkov became powerful, second only to Stalin at first. Stalin's technique for keeping Malenkov and Beria under control was to raise Abakumov to a still higher position. So Beria and Abakumov ran the state apparatus that controlled the Gulag slave labor camps. They organized massive executions of their rivals in Leningrad and had thousands more sent to the camps. And when Stalin died in 1953, Malenkov became Premier of the Soviet Union.

But Malenkov's importance for Kotelnikov lay principally in his wife, Valeriya Golubtsova. She would be the defining force in Kotelnikov's career—his protectress.[21]

Golubtsova had a first-rate revolutionary pedigree. Her mother, Olga, was one of the Nevzorov sisters—Zinaida, Sophia, Avgusta, and Olga—who were part of Lenin's inner circle when they were all young, long before the October Revolution. Two of the sisters married men with connections: Zinaida and her husband Gleb Krzhizhanovsky had spent exile in Siberia with Lenin and his wife during their courtships and early marriage years. Gleb was Lenin's friend and one of the oldest figures in the revolutionary movement, a Party member since 1893. Sophia's husband, Sergei Shesternin, was the man Lenin trusted with top-secret Bolshevik finances.[22]

Golubtsova was brilliant—a student at Moscow Power Engineering Institute (MEI) herself when women engineers were almost unheard of anywhere. She was also powerful, difficult, and cool. A determined organizer, she became MEI's longtime director and turned it into an academic powerhouse. When she spotted an intellectual star for her institution—such as Kotelnikov—she could and would make sure he stayed there and not in the Gulag. If not exactly a tyrant herself, she used her tyrant-in-law status very effectively to get what she wanted.[23]

How Digital Light Works

Kotelnikov worked out the theory of sampling, and he designed the ideal way to spread a pixel into a picture. But he had no idea at the time of the way those ideas would actually come to be used in practice. For one thing, there's a huge practical assumption in the modern digital world: an actual device, and not just a theoretical wave, will be at the other end to do the spreading and adding necessary to restore the analog light your eye requires.

Digital Light works because it assumes that the missing information will magically appear—in the form of an appropriate pixel spreader—somewhere in the process, at the moment of display. Digital Sound works the same way. It assumes there'll be a device somewhere later that will spread the soxels into the original analog sound that your ear requires. The display of sound is a speaker system and the amplifier that feeds it. Display, in fact, *is* the act of reconstructing the analog original—the form your brain needs—from samples.

Display is a fundamental assumption of the modern world. The Great Digital Convergence depends on the fact that the same pixels—the same abstract representation of an image—can be displayed in myriad ways by myriad technologies.

If you take a picture with your cellphone and then look at it on your computer screen, chances are that you're seeing two different pixel spreads. If you view it on paper, there's no question that you're witnessing a different spread. The only thing that's constant is the actual pixel behind the display, the invisible one that's reconstructed by the display, a point with a number attached—or three numbers if it's colored. One of the unsung accomplishments of Digital Light is the reduction of the vagaries of different display technologies—and many more to come, no doubt—to a thin layer of hardware called the display and its software called a display driver. Each manufacturer makes its own displays and drivers. We users never have to worry about anything but the pixels themselves. Our displays do the spreading. This is the Great Digital Convergence at work.

It's a convergence that divides. It drives a wedge between the creation and display of pixels. In Creative Space, pixels are simple, pure, and universal. In Display Space, the rubber meets the road. All the nastiness of actual real-world technology comes to bear on the problem of spreading the pixels that are supplied from Creative Space. Here's where the engineering expertise of thousands of manufacturers is focused on high-quality approximations of the ideal spreader within the real-world limitations of actual physical stuff—plasmas, liquid crystals, phosphors, inks. All we users need to worry about are the platonic pixels of Creative Space. That's why I've urged that the spread pixel, which can be perceived, should not be confused with the actual pixel, which cannot. Let's not conflate thin Display Space with vast Creative Space.

Often display manufacturers call the little glowing elements in their displays "pixels." But this confuses the spread pixel with the pixel. These should be called display elements, not pixels. The glow—the emitted illumination, high at the center, diminishing to zero away from center—of one of these display elements is the spreader for a particular display device. A pixel goes into the display element, a spread pixel comes out, and that's what you actually see. Generally, the glowing shape is unique to the manufacturer and can even vary from product to product at the same company.

So, here's how Digital Light works: It extracts pixels from the real world, even from Mars or Venus—or from unreal worlds inside computers, as we'll see—and sends them all over. These pixels carry only discrete position and intensity (or color) information. The actual light that seems to come from a pixel is only produced by a display device at the last possible moment before you see it with your eyes—not at any of those intermediate places around the world or among our sister planets. The smoothness of the original picture only reappears in the display. If the original creation of the pixels is done well, as per the Sampling Theorem, and if the reconstruction of those pixels into an analog display is done well, also per the Sampling Theorem, then what you see is an accurate representation of the original. We *do* get something from nothing. The pixel is the nothing—it has no dimension—and the displayed colored blob of light of the display is the something. The Sampling Theorem enables this scheme. The digital world wouldn't exist without it.

Spies and Scramblers

> "The idea of the scrambler is to reproduce the human voice by artificial means . . . to reproduce it by adding together at least the main harmonics, each transmitted by a separate set of impulses. You are familiar, of course, with Cartesian rectangular coordinates—every schoolboy is—but what about Fourier's Series?"
>
> —Aleksandr Solzhenitsyn, *In the First Circle*[24]

Kotelnikov's paper—the one with the Sampling Theorem—was rejected in 1936 by the *Electrichestvo* (*Electricity*) journal when he tried to get it into wider circulation. But there was a far more intriguing event that year: Kotelnikov visited the United States! I picked up a hint of this and was able to find an image of his disembarkation record. It reveals that he had a 60-day visa, his trip was paid for by the Soviet government, and his destination was the Amtorg Trading Corporation in New York City.[25]

Armand Hammer had formed Amtorg in 1924 to facilitate trade between the USSR and the US. He was a fascinating man, son of a founder of the Communist Party USA, who named him after the "arm and hammer" logo of the Socialist Labor Party of America. Armand was a noted businessman (head of Occidental Petroleum), philanthropist, and art collector.

But Amtorg was more than a trading company. It was a hotbed of industrial and military espionage. In fact, a declassified NSA (National Security Agency) report reveals that the United States expended major effort in 1931 trying to crack the codes that were used between Amtorg and Moscow. The US effort failed because "the Russians were using a one-time pad for their encipherment."[26]

This is particularly interesting because the one-time pad encryption system, which depends on a one-time, pre-shared key, wasn't *proved* effective for another decade. Amtorg spies obviously trusted it without proof for that long at least. Although we don't know exactly what Kotelnikov was doing there in 1936, creating and breaking codes seems a lot more likely than trade.

Or perhaps he was gathering data on one-time pad usage—because the man who proved the reliability of the one-time pad system was Kotelnikov. He submitted his proof just three days before Hitler's June 22, 1941, invasion of the Soviet Union. Perhaps not surprisingly, Shannon also proved the reliability of the one-time pad system; it appeared in a classified publication in 1945 and publicly in 1949.[27]

Kotelnikov must have done his one-pad work in his lab—the only remnant of the NKS communications commissariat after Stalin's Secret Purge. The lab had survived because its classified work on radio communication was urgently needed by the military. But as the Germans fast approached Moscow in late 1941, Kotelnikov took no chances that they might break through. He evacuated his lab to Ufa, about 300 miles further east even than Kazan, so 800 miles east of Moscow.[28]

During the battle of Stalingrad, the Russian front had depended on wire communications and had suffered gravely when, inevitably, the wires failed. Kotelnikov's lab produced secure devices that communicated via radio rather than wire, and by 1943 armies in the field were using the equipment. He received his first Stalin Prize that same year for its development. Then later, in May 1945, authorities used the lab's equipment to connect Moscow to the Soviet delegation at the signing of Germany's capitulation. Kotelnikov received his second Stalin Prize the following year.[29]

When the lab returned to Moscow from Ufa, it fell under the control of the NKVD. At this dangerous moment Golubtsova, the new director of Moscow Power Engineering Institute (MEI), stepped forward and proposed that Kotelnikov might like to return to

"Well, what do you yourself want?"

"To work at MEI."

"Then continue to work calmly as before," she advised.[36]

She had saved him again.

Developments in Russian Rocket Science

What Kotelnikov actually did instead of—or in addition to—Stalin's scrambler follows from an account of a 1947 meeting in Golubtsova's MEI office. Boris Chertok, a famed Russian rocket scientist, reported to her about the needs of his missile development sharashka.

"Within a short time, the results of this meeting exceeded our most optimistic expectations. Thirty-nine-year-old Professor Vladimir Kotelnikov was in charge of developing the ideas I had posed," he remembered. "Literally about ten days after my meeting with the MEI scientists, Golubtsova's office issued a governmental decree signed by Stalin on the creation of a special operations sector at MEI. A year later, the collective that had rallied around Kotelnikov was already developing the Indikator-D system, which we used during the flight tests of the first R-1 domestic missiles in 1948. Beginning with this development, all subsequent missiles were equipped with MEI radio systems during test flights."[37]

The R-1 was the Russian version of the V-2, Wernher von Braun's rocket used by Germany to pound London.

"In 1951, the MEI collective entered a competition for the creation of telemetry systems, and the first R-7 intercontinental missile was equipped with its now legendary Tral system."

The R-7 was their first intercontinental ballistic missile (ICBM).

The MEI "collective that rallied around Kotelnikov" was the sharashka known formally by the Russian acronym OKB MEI, presumably led by him from his protected position just as Chertok ran his sharashka, NII-88. Thus began Kotelnikov's contributions to the space race.[38]

Getting Digital Representation Right

So far I've talked as if pixels could represent any brightness level whatsoever. But a point on an analog curve is an analog value, meaning it can have any of an analog infinity of values. Strictly speaking, an analog sample of a picture's brightness isn't a pixel. It only becomes a pixel when it's converted to bits and only then will it be spread to become the glow you see on a display. A digital sample, such as a pixel, can take only certain discrete values, a finite number of them.

Some "computerspeak" is handy here, before we get to computers in the next chapters. Famously a bit can have two values, usually called 1 and 0. Think of a bit as a light switch. It has two positions, often called up and down. Consider two light switches. How many positions can they be in? Well, they can both be up, or both down, or one up and the other down in two different ways. So the answer is four. In other words, two bits (light switches) can have four values. Three can have eight values. The third switch can be up or down, and the other two switches can be in four different positions, as we've already noticed. Two times four is eight. In general, as the number of light switches increases by one, the number of different positions the switches can assume doubles. To save you the trouble, I'll simply tell you that 8 bits (light switches) can hold 256 values (positions), 10 bits can hold 1,024 values, and 16 bits can hold 65,536 values.

In the first decades of computer graphics it was common to limit a pixel of a black-and-white image to 8 bits. It meant that a pixel could represent one of only 256 shades of gray. For example, values 0 and 255 could represent black and white, respectively, with 254 other grays distributed in between. But the actual value on the analog curve might be 49.673. What's to be done? Well, the closest available gray value, 50, would be attached to the pixel for that point, hence introducing a small error into the picture when it's reconstructed later onto a display. How bad is that tiny round-off error in brightness? Is 50 "close enough" to 49.673 to not matter? What is "close enough" for pixel brightnesses?

Here's a case where we really can choose values that are "close enough." Doctors, particularly radiologists, read CAT scans and MRI scans on displays and base their diagnoses on what they see. Display manufacturers for the medical profession, therefore, have measured human grayscale perception precisely. A chart they provide shows that normal humans can distinguish 630 shades of gray on my particular desktop display. So two things are clear. First, the 8-bit pixel of yore with only 256 possible values isn't good enough. But, second, a 10-bit pixel, with over a thousand possible values is more than good enough. The human eye can't see the error that's introduced when you round off to the nearest value among a thousand-plus choices. Modern digital images now use 16 bits for a grayscale pixel, with over 65 thousand shades of gray. That's easily more than enough to fool the human brain—any human for any display. The same goes for 16-bit soxels in the display of sound.[39]

Have you heard audiophiles say that LPs (or vinyl) are better than CDs? In the visual world the claim is that digital photography will never match the subtlety of film photography. Such claims are often held with great passion. Is there substance to them? Well, of course there is. There are several ways that digital sampling can be done incorrectly—not enough values per pixel or soxel, a lousy spreader, or insufficiently high sampling frequency. But many adherents of these claims believe they are criticizing something *intrinsically* wrong with digital, when what they're really criticizing is poor execution. If an engineer set her mind on making a CD that replicated exactly what was on a vinyl long-playing (LP) album, the theory says that she could do it. Similarly, if another engineer set his mind to making a print from a digital image that duplicated a filmed photographic print, the theory says that he could do it. It wouldn't be easy in either case. The theory would have to be practiced as near ideally as possible, but it would be possible.

Some audiophiles claim to be able to distinguish recordings made at higher sampling rates than the 44.1 thousand soxels-per-second CD (compact disc) standard which features 16 bits per soxel. Recall that the human ear can hear up to 20 thousand cycles per second—the highest Fourier frequency—so the Kotelnikov (or Nyquist) sampling rate is 40 thousand soxels per second. CDs exceed the minimum required by Kotelnikov for perfect analog sound reconstruction from soxels. So, if the theory is to be believed, then nobody should be able to detect an improvement over a CD with a system using a higher sampling rate.

The other kind of error that an expert might be able to detect is round-off error—an insufficient number of values of loudness for a soxel. But as already mentioned, no human ear can make finer distinctions than those provided by the 16-bit soxel of the standard CD—with about 65,000 possible loudness levels per soxel. So nobody should be able to detect an improvement over a CD with a system using more bits per soxel.

Two standard digital audio systems that exceed the CD in both sampling rate and bits per soxel are Super Audio CD (SACD) and DVD-Audio (DVD-A). A carefully controlled scientific test came to this conclusion:

> We have analyzed all of the test data by type of music and specific program; type of high-resolution technology; age of recording; and listener age, gender, experience, and hearing bandwidth. None of these variables have shown any correlation with the results, or any difference between the answers and coin-flip results.[40]

In other words, even audiophiles couldn't distinguish CDs from SACDs or DVD-As. But—and this is telling—the same experimenters reported that "virtually all of the SACD and DVD-A recordings sounded better than most CDs." After searching conversations with the recording engineers, they found that this was because they used the Sampling Theorem much more correctly than the engineers who create the typical CD recording. The lesson is that digital can be an accurate representation of analog, but it has to be done correctly. Or stated another way, there is nothing *intrinsically* inferior about digital.

Spearheading the Space Race

> Yesterday evening we forgot to arrange with a friend for a trip into the country. We pick up from the bedside table a small object resembling a cigarette holder. It is a receiving-cum-transmission television set for private use such as every inhabitant of our planet possesses. We give the call signal of our friend and press a button. . . . We shall certainly have these tiny television sets which will go in the waistcoat pocket.
>
> —Kotelnikov, predicting the mobile phone, 1957[41]

Kotelnikov visited the United States a second time, on August 27, 1957, just before the launch of Sputnik—an event that shocked America to the bone. It was the International Geophysical Year, marking the end of the Cold War interruptions in scientific exchange. Kotelnikov was now a full member of the USSR Academy of Sciences, and he was carrying information from Russia for the Americans. He told a conference in Boulder, Colorado, that the Soviets would soon launch a satellite that would broadcast at frequencies of about 20 and 40 million (mega) cycles per second. He was ignored. Americans wouldn't believe that Soviet science had progressed so far. But on October 4, Sputnik launched and broadcast at 20.005 and 40.002 megacycles per second. The space race was on in earnest, and Kotelnikov was there.[42]

He was there too when it ended, almost two decades later. At a 1971 meeting in Moscow, Kotelnikov, the newly appointed acting head of the USSR Academy of Sciences, announced to surprised US diplomats that the proposed Apollo-Salyut Test Mission wouldn't work for technical reasons, but that the USSR was ready to pursue instead the renamed Apollo-Soyuz Test Project. The Apollo-Soyuz joint spaceflight project, which began in July 1975, was one of the most powerful symbols of the détente between the two countries.[43]

Spearheading the space race was just one of Kotelnikov's achievements. He accomplished the radio location of the planets Venus, Mercury, Mars, and Jupiter and got a Lenin Prize for it. Then, MEI-equipped missions to space, Veneras 15 and 16, mapped the northern part of Venus for the first time—sending pixels across the solar system. An asteroid, *2726 Kotelnikov*, honors his many contributions to the Russian space effort.[44]

Getting Rid of High Frequencies

Kotelnikov's great sampling idea tells us what it means to have pixels spaced "closely enough" together. Bad things happen if this rule is violated. You've probably seen the results: stair-stepped edges, wagon wheels rotating backward, moiré patterns on striped neckties, or unpleasant background shimmer in videogames. These artifacts are the bane of the early days of the Great Digital Convergence. "Are," not "were." I recently watched DVDs of the director Michelangelo Antonioni's great films *l'Avventura* and *La Notte*. In *l'Avventura* the stunning Monica Vitti's polka-dot dress was marred because of incorrect sampling. Her dress blinked on and off irregularly as if the polka dots were little lights toggled randomly. And almost every frame of *La Notte* glittered at the edges of windows and buildings that were incorrectly sampled. Not only are such digital artifacts unnecessary, they're unacceptable. They're not intrinsic to the digital world, but they are signs of uninformed use of the Sampling Theorem. Or, in the case of videogames, of game computers without enough power to do sampling correctly.

These unpleasant artifacts occur if you don't sample at a high enough rate. The Sampling Theorem says you have to sample a visual scene at twice the highest Fourier frequency in it. So, either you have to sample at a higher rate to be rid of the ugly artifacts, or you have to get rid of the too-high frequencies in the scene before you sample. Practicality usually dictates the latter path.

The frequencyspeak intuition is that sharp edges mean high frequencies. Very sharp edges mean very high frequencies. How high? Infinitely high for perfectly sharp edges—that is, frequencies so high you can't possibly deal with them. Certainly you couldn't sample often enough to accurately represent them. So practicality steps in again. In general, to represent a scene with pixels, you first have to get rid of its too high frequencies, its too sharp edges. Window edges are good examples. The producers of the flawed *La Notte* DVD clearly hadn't followed the Sampling Theorem. They didn't get rid of the too high frequencies before they sampled, and the result is ugly—or at least distracting.

There's an easy way to remove too high frequencies from a visual scene. Simply defocus the picture slightly—but not so much that you can actually perceive that the picture is going out of focus. This subtlety "schmudges" out all the edges. The Sampling Theorem tells you just how much defocusing is enough.

Dancing around Tyrants

The most disgraceful episode in Kotelnikov's life—at least to many Western eyes—was the Andrei Dmitrievich Sakharov affair of 1975. Sakharov, one of the fathers of the Soviet H-bomb, had renounced the use of nuclear weapons and campaigned actively for their nonproliferation. Both he and Solzhenitsyn, another dissident, were vilified in the Soviet media. But Sakharov received the Nobel Peace Prize for 1975. In response members of the USSR Academy of Sciences issued a statement denouncing him and

In the biography of the pixel we'll repeatedly find that the received stories aren't necessarily the right ones. And the right ones are often better than the received ones—in addition to being true. Vladimir Kotelnikov, not America's giant Claude Shannon, first brought the Sampling Theorem to the world.

The parallels between the two men are uncanny. Kotelnikov was head of the Institute of Radio Engineering and Electronics of the USSR (today Russian) Academy of Sciences for many years, and it's now named for him. America's immense Institute of Electronic and Electrical Engineers (IEEE) was originally the similarly named Institute of Radio Engineers. It gave its Medal of Honor to Shannon, and it marked the millennium and the Great Digital Convergence by giving its Alexander Graham Bell Medal to Kotelnikov in 2000. The irony is that Alexander Graham Bell was also namesake to Bell Labs—famously Shannon's home.[49]

Although Kotelnikov brought the Sampling Theorem to the world, it's a second irony that it was Shannon who taught it to America. We might invoke the phenomenon of "simultaneous invention" if it weren't for the fact that Shannon never claimed the Sampling Theorem for himself. We'll meet Shannon again in the next chapter, where state security will again be the tyrant—but Western security this time.

On Kotelnikov's ninety-fifth birthday in 2003—on the seventieth anniversary of his proof of the Sampling Theorem—in the Catherine Hall of the Kremlin, President Putin made him a Full Cavalier (alternative translations: Chevalier, Knight) of the Order for Service to the Motherland, only the fourth person to hold the order (figure 2.21).

Fi