SAMPLE

Building the Electronics of Tomorrow

The Laws, the Limits, and the Honest Economics of the Silicon Decade

Aiman JoeHanz

Free sample: Introduction & Chapter 1

putrabooks.github.io

Buy the full book:

Carousell carousell.sg/p/building-the-electronics-of-tomorrow-ebook-pdf-1471312719

Ko-fi

Payhip

Contents

Introduction: The Two Expirations

Chapter 1: The Doubling Law

Chapter 2: The Power Wall

Chapter 3: The Roadmap's Own Words

Chapter 4: The Shrinking Itself

Chapter 5: The Package Revolution

Chapter 6: The Light in the Machine

Chapter 7: The Materials Beyond Silicon

Chapter 8: The Fab as Strategy

Chapter 9: Beyond the Switch

Chapter 10: The Honest Decade

Introduction: The Two Expirations

Every industry that grows up on an exponential eventually has to grow up, and the industry that grew the modern world out of sand has now spent a decade doing exactly that. Electronics is the discipline of teaching inanimate matter to switch, and for roughly fifty of its years the discipline ran on two promises it did not have to keep, because physics and economics kept them on its behalf. The first promise said that the number of components you could afford to put on a sliver of silicon would double on a schedule you could plan factories around. The second promise said that as the components shrank, they would shrink their appetite for power in the same proportion, so that each generation would be faster, smaller, and cheaper all at once, a three-way win that no other industry in history has ever been handed. The first promise carried an engineer's name, Gordon Moore's, and became the most quoted observation in the history of technology. The second promise carried a quieter name, Robert Dennard's, and did most of the actual work. Both promises have now expired, and the interesting question, the question this book is built to answer, is what an industry does the morning after.

The answer turns out not to be a single new law. What replaced the two expirations is a portfolio, and the portfolio is what the next decade of electronics actually looks like: new transistor shapes, because the old flat sheet of silicon stopped cooperating; new packages, because the cheapest way to add transistors is no longer to shrink them but to stack and stitch them; new materials, because silicon is a magnificent all-rounder and a poor specialist; new carriers, because wires are losing the argument to light at exactly the scale where wires used to be unbeatable; new physics, because the switch itself is not the only way to compute; and, strangest of all, new money with flags on it, because the leading-edge factory crossed a price threshold that turned it from a corporate bet into a national strategy. None of these alone is a successor to Moore. Together they are something more honest and harder to use, which is a set of choices, and the practitioners who internalize the shift, from the engineer picking a substrate to the investor pricing a fab, are the ones who will get the decade right.

This book is the forty-fourth in a series about the technologies that are building tomorrow, and it keeps the series' habits: the tour is written in the technology's own grammar rather than in marketing's, every number is either verified against the public record or recomputed in the text and labeled as illustrative arithmetic, and each chapter ends with a stealable page, a compact working summary you can lift and use before you have read the

chapter it belongs to. The research behind the book rests on a registered set of sources, cited by name in the text, from Gordon Moore's 1965 paper, which started the entire conversation, to the IEEE's current device roadmap, to the design-automation surveys now charting the packaging revolution, to the public record of the industrial-policy era the industry has just entered.

A note on what the reader is assumed to know, because the book is deliberately built to serve two audiences at once and to shortchange neither. The working engineer will find no step-by-step recipes here, and should not, because recipes age faster than the physics under them; what she will find is the reasoning that decides which recipe to trust, the cost curves and yield arguments and energy accountings that her vendors quote selectively or not at all. The interested outsider, the investor, the student, the curious neighbor, will find the technical content explained on first use, in plain clauses, with the numbers that matter left standing in the open rather than buried in an appendix, because the industry's real secrets are not in its physics, which is published, but in its economics, which is usually half-quoted. Between the two readers sits the book's recurring method, worth naming once: every claim is priced. A technology that cannot say what it costs, in watts or dollars or years, is not being described, it is being sold, and the difference between description and salesmanship is the difference between a chapter and a brochure. The decade the electronics industry has entered is expensive, public, and well documented, which is the best possible condition for honest writing, and the book takes the condition at its word.

One orientation note on the register's citations, since a book that promises priced sources should say how the prices are displayed. Throughout the chapters, a named claim, a measurement, a document, carries its source in the sentence, Moore's paper, the roadmap's edition, the statute's section, and the reader can trace every one of them without leaving the public record; a number that cannot be sourced that cleanly is labeled illustrative at the moment it is used, and the label is the book's honesty mark, not a hedge, because an illustrative number in an arithmetic argument is a placeholder the reader replaces with his own measured value, which the argument's form invites. The distinction matters at the decade's temperature: the industry's public claims have never been more polished or less auditable, and a reader who cannot tell a measured number from a marketing one is not reading the decade, he is being read by it. The chapters therefore end, each of them, with a compact page of the questions a practitioner should keep, and the pages are the book's own register in miniature, dated, sourced, and ready to be re-run

against whatever the industry publishes next year, which is how the author intends the book to age: not as a monument, but as a worked example of the audit the reader will keep performing long after the decade, and this book, have both moved on. The tour route, in one breath, for the reader who wants the shape before the substance: the two laws that made the industry, the wall that ended the free ride, the taxonomy that organizes the response, the shrinking lane and its costs, the packaging revolution, the photonic crossover, the materials that out-specialize silicon, the fabs that became strategy, the physics waiting behind the throne, and the decade's honest accounting, ten chapters, six decisions, one register, read in order or entered by need. The chapters carry no prerequisites beyond arithmetic and patience, every term is defined at first use, and every chapter's compact page can be read alone on a commute, which is the book's design and its invitation: keep the pages, run the audits, and let the decade, for once, be the well-documented one.

The route the book takes is deliberately the industry's own route, in the industry's own words. The first chapters establish what the two expirations actually were, because the popular account of Moore's law's death is mostly wrong in the details and right in the direction, and the details are where the decisions live. The middle chapters walk the portfolio lane by lane: the shrinking that continues without pretending to be exponential, the package that turned the printed circuit board into a battleground, the light that is quietly annexing the wiring, the materials that out-specialized silicon in power and in light emission, and the factory that became too expensive for any single company's balance sheet to carry alone. The final chapters face the horizon honestly: the switch candidates that do not yet deserve the word roadmap, the atom-scale imagination that Feynman set out in 1959 and that remains the discipline's longest-term North Star, and the arithmetic, energy first among all the numbers, that will judge every choice in the portfolio.

A word on tone, because the subject invites two kinds of dishonesty. The first is the marketing kind, the claim that some single technology, quantum computing or photonics or a miracle material, is about to resume the free ride, and this book will have none of it: every candidate gets priced, and most of them are expensive in ways their boosters do not mention. The second is the elegiac kind, the essayist's sigh that the great age has passed and everything since is bookkeeping, and the book will have none of that either, because it is false. The industry's own roadmap, published annually by the engineers who have to live inside it, lists its three escape lanes in language that is neither breathless nor resigned, and the third lane is literally named Beyond CMOS,

which is the roadmap of an industry that plans to outlive its own best invention. The free ride is over. The discipline is not, and the difference between those two sentences is the whole subject of this book.

Chapter 1: The Doubling Law

The law began as arithmetic in a trade magazine, and the arithmetic was about money, not about transistors. In April 1965, Gordon Moore, then the research director at Fairchild Semiconductor and two years from co-founding Intel, wrote a piece for the industry weekly Electronics whose modest title, "Cramming more components onto integrated circuits," gave no warning of what the piece would become. Moore had sat down to answer a question any customer would ask a young company selling a new kind of part: if I design my product around your integrated circuits this year, what will I be able to buy in five, in ten? He plotted the number of components, transistors and resistors and diodes together, that his industry had managed to place on its most complex production chip each year since the planar process arrived, and the points fell on a line. The line said the component count was doubling every year, and Moore, with the caution of an experimentalist extrapolating from five or six data points, hedged twice in the paper's own sentences: certainly over the short term, he wrote, the doubling rate could be expected to continue, and he projected it for at least ten years. The paper's central figure, the one every subsequent decade reprinted, was not a count of components at all but a family of cost curves, and that figure is the part everyone quoting the law has forgotten.

The cost curves carried the real claim. As you put more components on a chip, the cost of manufacturing each component falls, because the expensive steps, the masking and the etching and the handling, are charged per wafer and not per component; but the fall is not forever, because as complexity rises, so does the fraction of chips killed by a single defect, and the yield loss claws the cost curve back up. For any given manufacturing year there is a minimum-cost complexity, Moore argued: a chip size below which you are paying for silicon you did not need, and a chip size above which defects are destroying your economics faster than integration is rescuing them. His observation, the one that made the law an economic law rather than a curiosity, was that the minimum-cost point moves. Every manufacturing year, the trough of the cost curve slides toward higher complexity, and it slides fast enough that the doubling of affordable complexity becomes an annual event. This is the law in its original form, and it is worth stating plainly because the twentieth-century version everyone remembers, transistors double every two years, is the law's later, public, simplified restatement; the engineer's original was always about the affordable minimum, about the point where a customer gets the most function for the money, and the industry that forgot the trough while memorizing the doubling spent the 2010s relearning it.

The context that produced the paper is worth restoring, because the law reads differently when the industry around it is remembered as the young, contested, unproven thing it was. Integrated circuits were barely seven years old in 1965, and they were losing the argument: the dominant engineering opinion held that wiring transistors together on a single piece of silicon was an expensive stunt, useful for missiles where weight and reliability mattered more than money, and wasteful for everything else, because the individual components were cheap and the yield losses of printing them together were ruinous. Moore's employer, Fairchild Semiconductor, was the young firm whose entire business rested on the opposite opinion, and Moore's paper is best read as the argument for that opinion, published in the trade press where the customers could see it. The planar process, Fairchild's own invention, had made the difference: by building the transistor's structure in layers rather than sculpting it, the process allowed the oxide that protects the surface to be patterned photographically, which allowed the components and their wiring to be defined by a photographic mask rather than by hand assembly, and once definition is photographic, complexity becomes cheap, because a mask with more components costs no more to expose than a mask with fewer. That is the mechanical heart of the law, and the chapter's reader should carry it in one sentence: the reason complexity doubled was that photographic exposure did not charge by the component, and the reason the exposure kept improving was that the optics, the resists, and the masks were all on improvement curves of their own, riding an industry that had discovered it could print its products rather than assemble them.

The printing metaphor deserves one more paragraph, because it explains the yield arithmetic that the cost curves formalized, and the yield arithmetic is the book's first working encounter with defect economics, a term that returns in the packaging chapter with higher stakes. A wafer is printed with many copies of a chip at once, and the copies are not all born equal: dust, a resist flaw, a mask defect, a stray ion, any of them kills the copy it lands on, and the probability that a given copy survives falls exponentially with its area, because a bigger target collects more shots.

The defect arithmetic rewards one worked number, illustrative but honest to the industry's own textbooks, because it converts the chapter's abstractions into a feel for the stakes. Suppose a process running a defect density of one fatal flaw per square centimeter, a crude but defensible round number for a mature generation, and consider two products: a small die a tenth of a centimeter on a side, and a large die ten times the area; the small die survives

with probability roughly ninety percent, while the large die survives at roughly thirty-seven percent, the exponential at work, so a wafer of large dies yields barely a third of its silicon while the same wafer cut into small dies yields nearly all of it, and the difference, in a factory running tens of thousands of wafers a year, is the difference between a product line and a write-off. The worked number explains three industry behaviors at once: why the leading-edge chips of the machine-learning era, whose areas ran to the reticle's edge, drove the chiplet movement of the fifth chapter, why the mature nodes stay commercially immortal, their defect densities worked down to almost nothing, and why the industry's engineers speak of defect density with the reverence soldiers reserve for weather, because it is the one enemy that never surrenders, it only agrees to be outrun, and the outrunning, node after node, was the unheralded labor that kept Moore's trough sliding while the world watched the transistor counts. The mathematics, worked out in the industry's early years and printed in every process textbook since, says the surviving fraction of a die of area A on a process with defect density D falls as the exponent of the product, so doubling a chip's area on a fixed process cuts its yield far worse than in half, and the only rescue is a falling defect density, which is what the fabs actually worked on between the nodes the public was quoting. Moore's minimum-cost trough, then, was the meeting of two curves moving in opposite directions: complexity made each component cheaper by amortizing the wafer's fixed cost over more of them, while area made each wafer worth less by multiplying the defect targets, and the trough's annual slide toward more complexity meant the defect fighters were winning their fight slightly faster than the area tax was rising. The reader who holds that picture holds the economics of every decision this book will price, from the chiplet partition of the fifth chapter, which is the trough logic replayed at the level of dies, to the fab economics of the eighth, which is the trough logic replayed at the level of nations.

The paper's last page is the one that reads today like a catalog written sixty years early, and it is the part of Moore that people who quote the law have also mostly forgotten. One item in the catalog deserves the spotlight, because it is the paper's most famous prophecy and its least believed at the time. Moore closed by telling his readers that integrated electronics would supply the logic for the home computer, in an era when the word computer meant a machine the size of a room, owned by institutions, priced like buildings, and the trade audience of 1965, engineers who knew exactly what computers cost, needed persuading that such a thing could ever be cheap enough for a home;

six decades later the persuasion reads as understatement, the home contains dozens of them, most inside objects their owners do not call computers at all. The prophecy's method is the chapter's last lesson in forecasting: Moore did not predict the home computer from any vision of domestic life, he priced it, extending his cost curve forward until the machine's function fell into the price range of a consumer durable, and the objects his method kept pricing into affordability kept arriving for the rest of the century, which is why the chapter tells the reader to distrust prophecies and trust extrapolated costs, a rule the remaining chapters apply to photonic links, wide-bandgap converters, and quantum machines alike. The catalog's other entries, the digital watches, the medical instruments, the telephone switching, reward the same reading, each priced rather than prophesied, and the set's completeness is the paper's deepest property: Moore was not predicting products, he was announcing a pricing mechanism, and every product the century built on it was his catalog, extended. Having established that the economics would deliver enormous component counts at trivial cost per component, Moore asked what the world would do with the surplus, and his answers, offered as speculation, included electronic brains for home computers, put in his own scare quotes because the phrase sounded absurd in 1965; personal portable communications equipment, which is a description of the telephone in nearly every pocket on the planet; and automatic electronic controls for the automobile, which is a description of every car built in this decade. The guesses were not clairvoyance. They were the direct product of the cost curves, and that is the lesson for the builder: when a component's cost falls by orders of magnitude, the applications do not arrive as a surprise to the people who did the falling. They arrive first to the people who watched the curve and asked the follow-up question Moore asked, which is what becomes practical when the impractical thing gets cheap.

The decade after the paper tested the paper, and the test's score is the reason the law hardened from observation to institution. The ten years Moore had projected, hedged twice, ran from 1965 to 1975, and the component count did roughly what the line said it would, with the industry's most complex production parts rising from dozens of components toward the tens of thousands; the projection's clean arc, though, hid the messy mechanism, which was that the rate itself was being renegotiated as the industry aged. In 1975, delivering the address the industry remembers as the law's first revision, Moore himself stepped the doubling period from his original year toward two, attributing the slowdown to the removal of the easy gains, the

tricks of circuit cleverness that had piggybacked complexity on top of the pure component count, and to the rising cost of the physical shrink itself; the two-year cadence ratified that day is the number most people mean when they quote Moore's law, and the quotation's layered history, one year in the paper, two years in the revision, roughly maintained by the industry for the following three decades, is a working lesson in how an industry and an author negotiate a law together. The institution-building came next: Intel, the company Moore co-founded, built its product cadence around the law openly, alternating a process shrink with an architecture change on a schedule the press summarized as tick-tock, and the schedule's public reliability across the 2000s did as much as any physics to convince the industry's customers that the doubling was a deliverable and not a forecast.

The forecast's sociological life inside the industry is its own study, and the chapter closes on it because it set the pattern every later technology roadmap, the third chapter's subject, would follow. Once the cadence was believed, it became a planning input everywhere at once: the equipment makers wrote their product roadmaps against the years the node was expected to need their optics and their deposition tools; the design-automation firms scheduled their software releases against the transistor counts their users would face; the system builders wrote their next product's compute budget against the chips the cadence promised, and, most consequentially of all, the fabs' own engineering organizations were managed against the cadence's dates, with whole generations of process engineers hired, trained, and judged by their node's arrival time. An economy in which thousands of independent firms coordinate their five-year plans by reading the same public sentence is an economy that has invented something new, and the thing it invented is what this book's third chapter calls the roadmap genre; that it worked, that the forecasts kept arriving because everyone worked to make them arrive, is the reason the law's decades read in hindsight less like physics and more like a promise the industry kept to itself. The promise's moral for the next decade is double-edged, and the reader should hold both edges: the same coordination that can keep a hard schedule honest when the physics is difficult can also keep a schedule alive past the point where the economics has quietly died, and the decade of this book opens with the industry, for the first time in its history, publishing its own doubts about which of the two is happening.

The law's most important property was not that it was true but that it was believed, and this is the part of the chapter that matters most to a reader who

has to plan anything. A prediction printed in a trade magazine became the industry's coordination signal: every company's product roadmap, every equipment maker's development plan, every customer's procurement schedule was written against the assumption that the doubling would arrive on time, and the collective belief made the belief true, because the industry staffed and funded its way to keeping its own promise. The doubling period itself was revised by history rather than by Moore alone, from his annual original to the roughly two-year cadence the industry ratified in the 1970s, and Moore himself spent later decades gently correcting both the people who credited him with more precision than he claimed and the people who predicted the law's death too early. The law survived every scheduled funeral from the 1970s onward not because the physics was easy but because the economics kept paying everyone who kept the faith. The lesson cuts both ways, and the decade this book describes is living on the other side of the cut: a law that was kept alive by coordinated belief can slow when the coordination thins, and the slowing of Moore's law in the 2010s, which the second chapter prices precisely, is as much a story about costs nobody would pay as about limits nobody could cross.

Two corollaries complete the law's honest biography, one about marketing and one about money. The marketing corollary is the node name, the industry's label for each process generation, descended from the day the name measured the transistor's drawn dimensions in nanometers and kept the numbering after the measurement expired; the labels have been drifting from physical geometry since the middle of the 2000s, so that a modern node number is a generation marker and a marketing position rather than any single measurable distance, and the reader who compares chips by node name alone is comparing model numbers across manufacturers who no longer share a yardstick.

The yardstick's restoration is the chapter's practical gift to the reader, because comparisons must be made and the names will not make them honestly. The robust cross-generation metrics are few and public: transistor density per square millimeter, quoted by the manufacturers in their own disclosures and comparable within a class of device; the chip's die area, known for most products from the teardown community's measurements; the achieved clock and the thermal design power, printed on the public datasheets; and the energy per operation on a named workload, the number that survives every era, quoted by the serious vendors and dodged by the rest. A reader who compares two chips on those four numbers, density, area, power, and energy

per operation, has replaced the marketing node with a private yardstick that no renaming can inflate, and the practice takes minutes once habitual; the chapter commends it because the first law's final lesson is not about silicon at all, it is about metrics, which is that a shared yardstick is a public good, the industry enjoyed one for forty years, and its erosion returns the burden of measurement to the buyer, which is a burden this book has been teaching the reader to carry since its first page. The money corollary is the law's less quoted twin, the observation that the factories themselves double in cost on roughly the schedule the products double in density, so that each generation of fabrication plant costs about twice the last, which the press long ago baptized as Moore's second law and the industry quietly budgets as capital expenditure; the twin's arithmetic, colliding with the first law's slowing returns, is the single best explanation for the industry's consolidation from a dozen leading-edge manufacturers toward three, and the eighth chapter of this book, on the fab as an instrument of national strategy, is essentially the second law's political aftermath. The builder who carries one sentence from this chapter should carry this one: the doubling was real, was revised, was institutionalized, and was finally priced out, and the pricing out is not a failure of nerve but the industry's arithmetic maturing into the open. The biography's last act belongs to the author, whose own relationship to the law is the chapter's gentlest lesson. Moore spent his long retirement correcting the record in both directions, reminding interviewers that he had never claimed a law of physics, only an economic observation that happened to hold, and declining the more extravagant prophecies the public hung on his line; the humility was genuine and it was also, characteristically, accurate, because the observation was always a summary of what the industry could afford, and the man who made it never confused affordability with destiny. The industry that mourned the law's slowing in the 2010s did so while shipping the most capable electronics in history, a contradiction the founder's own framing resolves: the law described a bargain, the bargain was renegotiated rather than broken, and the renegotiation, whose terms this book's remaining chapters read clause by clause, is not the law's death but its maturity, the moment an industry stops being carried by a curve and starts paying its own way.

The Trough Page

The stealable page, compact: the law was always about the trough, not the doubling. Find the complexity at which cost per component is minimized for your process generation; expect that trough to slide toward higher complexity each generation, and plan designs to land just above it, not at the frontier for

its own sake. Quote Moore's restatements carefully: his 1965 paper claimed annual doubling of affordable complexity for at least ten years, hedged twice; the two-year form came later, and the industry, not the paper, ratified it. When you hear the law invoked in a pitch, ask which version is being invoked, the affordable-complexity law or the headline doubling, because they fail on different schedules. And when you plan a product around a curve, remember that the curve is a coordination artifact: it held while everyone believed and paid together, and it bends the moment the collective bill arrives.

ABOUT THE AUTHOR

Aiman JoeHanz writes the Building of Tomorrow series -- structural, vendor-neutral guides to the technologies reshaping industry, labor, and daily life. Each volume treats one technology the way an engineer treats a machine: definition first, then anatomy, failure modes, principles, build, deployment, and the honest decade ahead.