

CS250: FOUNDATIONS OF COMPUTER SYSTEMS

[MOORE, DENNARD, AND SILICON]

The end of the party

Have you a plodding program?

Just you wait and soon it will sprint

Miniaturization, Dennard's scaling

working in tandem good times

doubling densities constant power

Moore's Law faster clocks, magical speedups

Alas the party's come to a crashing end

In the wee hours of the millennium

Running headlong into quantum effects

Laid waste by heat

SHRIDEEP PALLICKARA

Computer Science

Colorado State University

COMPUTER SCIENCE DEPARTMENT



COLORADO STATE UNIVERSITY

1

Topics covered in this lecture

- Why miniaturization matters in hardware
- Moore's Law
- Dennard's scaling
- Silicon's Odyssey



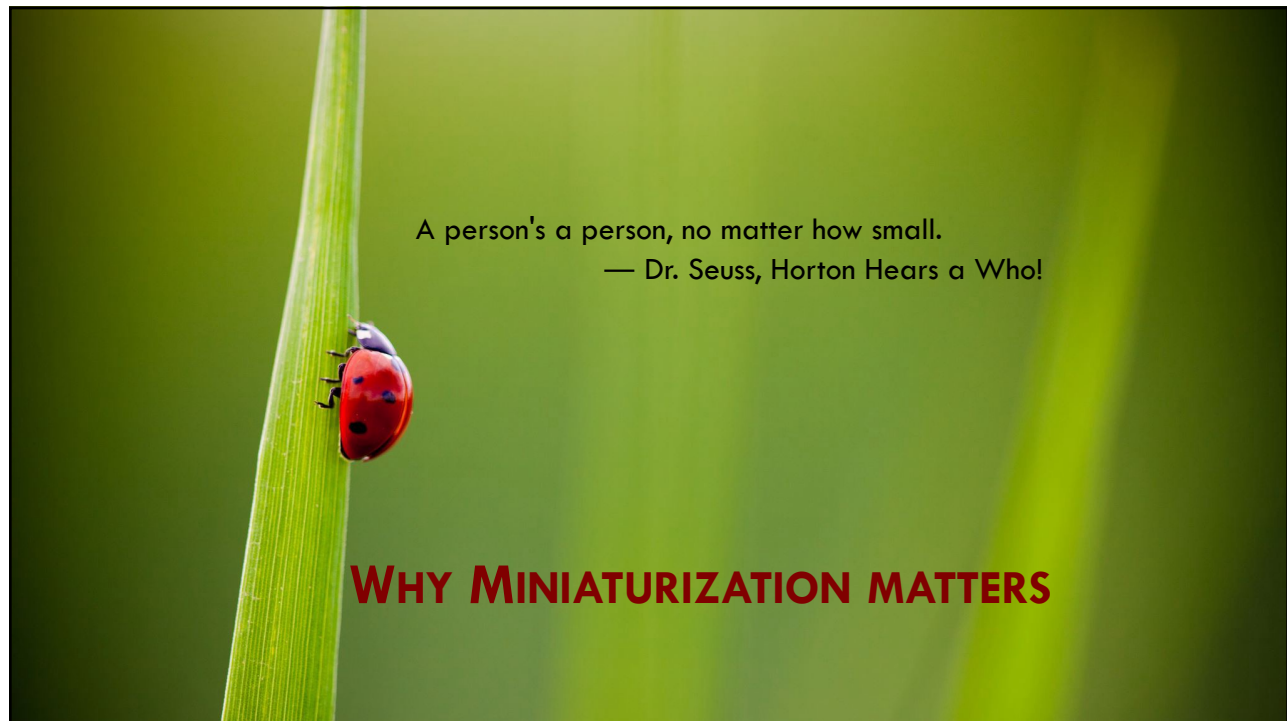
COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.2


2



3

Why Miniaturization Matters in Hardware: An analogy

- Imagine you have to drive your kids to and from school, which is 10 miles away, at an average speed of 40 miles per hour
- The combination of distance and speed means that only two round trips per hour are possible
- You can't complete the trip more quickly
 - ▣ Without either driving faster or moving closer to school

 **COLORADO STATE UNIVERSITY** Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT MACHINE LANGUAGE L20.4

4

Modern computers drive electrons around instead of kids

- Electricity travels at the **speed of light**, which is about 300 million meters per second
 - Except in the US, where it goes about ~186,000 miles/second or a billion feet per second
- Because we haven't yet discovered a way around this physical limitation
 - The only way we can minimize travel time in computers is to **have the parts close together**



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.5

5

Miniaturization and clock speeds

- Computers today can have clock speeds around 4 GHz, which means they can do four billion things per second
- Electricity only travels about **75 millimeters in four-billionth** of a second
- A typical CPU that measures about 18 millimeters on each side
 - There's just enough time to make **two complete round trips** across this CPU in four-billionths of a second
 - It follows that making things small permits higher performance



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.6

6

Miniaturization: The energy angle

[1/2]

- When driving kids to and from school coffee alone is insufficient
 - ▣ It takes energy to travel!
- Making things small reduces the amount of travel needed, which **reduces the amount of energy** needed



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.7

7

Miniaturization: The energy angle

[2/2]

- **Reduced energy requirements** translates into **lower power consumption** and **less heat** generation
 - ▣ Reduced heat keeps your phone from burning a hole in your pocket!
- This is one of the reasons why the history of computing devices has been characterized by efforts to make hardware smaller
 - ▣ But making things very small introduces other problems



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.8

8



9

Moore's Law

[1/2]

- **Empirical observation** linked to gains from experience in production
- The number of transistors in a dense integrated circuit (IC) doubles about every two years
 - ▣ For most of the last fifty years ... equivalent to saying that computers did get **twice as fast**
- Not a law of physics
 - ▣ An observation and projection of a historical trend



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.10

10

Moore's Law

[2/2]

- Nearly two decades since what has been called “the breakdown of Moore’s law”
 - And the switch to multicore processors instead of ever faster single chips
- But again, this is wrong!
 - Moore’s law has not broken down at all – transistor numbers are continuing to increase
 - What has happened is that it is **no longer possible to keep running these transistors at ever faster speeds**



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.11

11



DENNARD SCALING
& THE FREQUENCY WALL

12

Dennard scaling

- Named after Robert Dennard
 - Led the IBM research team that first described this effect in a 1974 paper
- As transistors got smaller the **power density was constant**
 - Power use stays in proportion with area; both voltage and current scale (downward) with length
 - If a transistor's linear size reduced by 2?
 - The power it used fell by 4!
 - With voltage and current both halving



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.13

13

Dennard's Observations

[1/3]

- With every technology generation, transistor dimensions could be scaled by -30% ($0.7\times$)
- This has the following effects **simultaneously**:
 - Chip area reduces by 50%
 - To keep the electric field constant, the voltage, V , is reduced by 30% ($0.7\times$)
 - Voltage is field times length
 - Circuit delays reduce by 30% ($0.7\times$)
 - Because time is length over velocity



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.14

14

These in turn results in changes to capacitance and chosen frequency

- The 30% reduction in delay allows an **increase in operating frequency** by about 40% (1.4×)
 - ▣ Frequency varies with one over delay $1/0.7 \sim 1.4$
- The 30% reduction in all distances and related 50% drop in area leads to a decrease in capacitance, C, by 30% (0.7×)
- **Power consumption decreases** by 50%, because active power is CV^2f
 - ▣ $0.7 \times 0.7^2 \times 1.4 \sim 0.5$



Dennard's Observations: Takeaway

- In every technology generation:
 - ▣ Area halves and ...
 - ▣ Power consumption halves!
- In other words, if the transistor density doubles?
 - ▣ **Power consumption stays the same with twice the number of transistors**



What has broken down is not the ability to etch smaller transistors

- We are **unable to drop the voltage and the current** needed to operate reliably
- In the run-up to hitting this wall, the reduction in the size of the transistors ran slightly ahead
 - ▣ We ended up getting to 3GHz a little faster than expected



So what's the problem?

- **Static power losses have increased** every more rapidly as a proportion of overall power supplied as voltages have dropped
 - ▣ Static power loss: Power consumed by a circuit when it's **idle**; due to **leakage** of current flowing through transistors though they are "off"!
- Notably, **static power losses heat the chip**
 - ▣ *Further* increasing static power loss and threatening **thermal runaway** – and complete breakdown
 - This is a vicious cycle not a virtuous one



Reasons for increased static power loss

- Complex **quantum effects** due to
 - ▣ Component sizes being reduced
 - ▣ Modifications to chemical composition of chips to handle miniaturizations
- There seems to be no way out!
 - ▣ “Moore’s law” ... in the sense of ever faster chips, is dead



Consequences of the breakdown on Dennard scaling

- **Unable to increase clock frequencies** significantly
- Most CPU manufacturers focus on multicore processors to improve performance
 - ▣ An increased core count benefits many workloads
 - ▣ The increase in active switching elements from having multiple cores *still results in increased overall power consumption*
 - Worsens CPU power dissipation issues



The problems of dissipation persist to some extent in multicores as well

- **Only a fraction** of an integrated circuit can actually be active at any given point in time without violating power constraints
- The remaining (inactive) area is referred to as **dark silicon**



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.21

21

ANOTHER ARCHITECTURE

COMPUTER SCIENCE DEPARTMENT



COLORADO STATE UNIVERSITY

22

The Harvard architecture

- Named after the Harvard Mark I computer a joint effort between IBM and Harvard



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

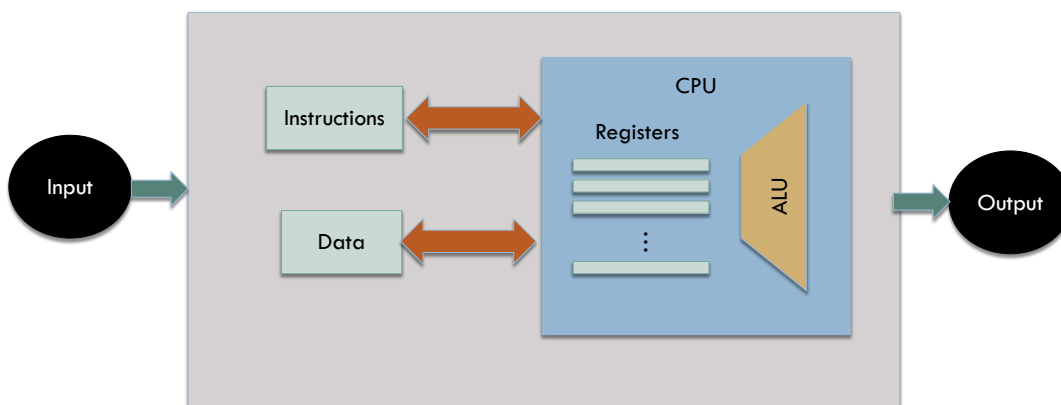
MACHINE LANGUAGE

L20.23

23

Depiction of the architecture

- The only difference between them is the way the memory is arranged



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.24

24

Some comparison points

- All else being equal, the von Neumann architecture is slightly slower
 - ▣ Because it can't access instructions and data at the same time, since there's only one memory bus
- The Harvard architecture gets around that but requires additional hardware for the second memory bus



THE ODYSSEY

"Out of sight, out of mind"
Homer, The Odyssey

Even in devices that don't have “smart” in their name

- Mechanical linkages have long since given way to a network of semiconductors
- They are not just the **world's brain** or even its nervous system but, increasingly, its sinews, veins and receptors



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.27

27

At first glance, a gleaming silicon wafer looks nothing like the silicon we think we know

- Silicon is the primary ingredient in sand, stone or concrete
- Silicon is remarkable because it combines optical, structural, and electrical versatility in one element.
 - ▣ It has unique properties that allow it to form glass
 - ▣ In concrete form, it is strong enough to help support buildings
 - ▣ Electrically, it stands apart from most elements because it is a **semiconductor**



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.28

28

No one could quite think of what to do with semiconductors

- But eventually they discovered that they worked brilliantly as a kind of switch
- The first such switch, or **transistor**, as it was named, was made a couple of days before Christmas 1947
 - By Walter Brattain and John Bardeen, two physicists working under William Shockley at Bell Labs in the U.S.



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.29

29

These transistors have been getting smaller ...

- That little period in a printed book may look insignificant, but on a modern chip it is room enough for about 15 million transistors
- The transistors in today's smartphones are not just smaller than a red blood cell (about a 1000 times smaller, as it happens)
 - They are smaller than the COVID-19 virus
 - Actually, you could fit 4 of them inside a coronavirus



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.30

30

An Apple chip, plus half the industrialized world

- Chips the are Apple-branded are actually made by *Taiwan Semiconductor Manufacturing Company*
 - ▣ It's better known, TSMC
- TSMC in turn was only able to make the chip with the help of machines made by another, even more obscure company, ASML
 - ▣ And at the heart of ASML's machines are critical components made by other companies
 - The lenses are made by Zeiss, with glass from Schott
 - The lasers are made by a German company, Trumpf

* ASML: Advanced Semiconductor Materials Lithography



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.31

31



32

Our odyssey begins ...

- In a forest about 15 miles south of *Santiago de Compostela*
- This part of Galicia is best known for the pilgrimage that draws visitors from around the world each year
 - ▣ To visit the tomb of St. James the Apostle in Santiago's cathedral
- The place is serene and surrounded by hills
- The most distinctive of those hills is **Pico Sacro**, an outcrop you can see from miles around, a pyramid jutting over the fertile fields below



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.33

33

The geological explanation for how this hill came to be ...

- Around 350 million years ago this was where two supercontinents (**Laurasia** and **Gondwana**) collided
 - ▣ Thrusting up the land and forcing a quartz outcrop hundreds of meters into the sky
- Walk up the hill and you can see the white quartz stone alongside granite rocks ...
- And eventually you come across a sign marked "**Serrabal**"



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.34

34

Serrabal is a quartz mine

- The vein of rock beneath Pico Sacro and its surrounding hills is an exceptionally pure quartz deposit
- The quartz is strikingly white
- Its purity makes it highly prized and sought after from far away



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.35

35

Where the silicon story leaves the hill

- The company that owns the mine is **Ferroglobe**
 - ▣ A Spanish business which is the world's biggest silicon metal producer
- After they are blasted out of the ground, cleaned and emptied into trucks
 - ▣ The chunks are driven an hour or so north towards an industrial park just outside the port of *A Coruña*



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE



36

Where the quartz becomes something else

- The rocks from Serrabal are emptied on to the floor outside the warehouses
 - ▣ A pile of white stone on the grey concrete
- They are mixed with coking coal (a baked form of coal) and woodchips and tipped into a furnace, heated up above 1,800°C
- What happens in that furnace, where an electrical current is run into the mixture of quartz and coal, remains something of a mystery



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.37

37

The yield is not for the faint of heart

- In the furnace, the molten silicon separates from its oxygen
- The silicon sinks to the bottom, where it is drawn off through a tap
- The conversion is expensive and inefficient by ordinary standards
 - ▣ Roughly 6 tons of quartz, coal, and woodchips produce only about 1 ton of silicon metal



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.38

38

After the silicon metal produced by the furnaces here at Sabón is poured out of the furnace

- It solidifies
- And then smashed up into a kind of granulated metal
- At this stage, the silicon is around 98 to 99% pure
 - ▣ Which sounds pretty good to most of us, but is a long way from the purity you need for a silicon chip or a solar panel



The metallurgical-grade silicon leaves Ferroglobe for the next stage of processing

- It is sent to **Wacker**, where it is refined into a much purer form called polysilicon
- Wacker is a German company and a polysilicon behemoth
- Its main plant is in Burghausen, about 90 minutes east of Munich near the Austrian border
- The chemical plant at Burghausen is so vast it takes up about the same footprint as the town it sits alongside.



What happens next is known as the Siemens process

- The next step is the **Siemens process**, where silicon is essentially taken apart and built back up again
- First, the purified silicon metal is ground into a powder and mixed with hydrogen chloride
- That mixture is then distilled and heated inside a bell jar to around $1,150^{\circ}\text{C}$
- Out of this comes long silicon rods that look a bit like the heating elements in an old kettle
- Except here the prize is not limescale, but **ultra-pure silicon** made through an extremely energy-hungry process



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.41

41

Clean silicon but with a dirty secret

- Ultra-pure silicon is **astoundingly energy-intensive** to make
- Its energy cost is more than 3,000 times that of cement
- It is also about 1,000 times higher than turning iron into steel
- The total volumes may be smaller, but the process is still demanding, costly, and often dirty



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.42

42

An aside ... gold

- Gold doesn't come out of the earth in nuggets or as a rich seam forged by Mother Nature
- For a standard gold bar (400 troy ounces) mining companies dig about 5,000 tons of earth
 - That's nearly the same weight as 10 fully laden Airbus A380 super-jumbos, the world's largest passenger planes ... for one bar of gold!
- Gold extracted from excavated earth is the product of a chemical reaction involving one of the most toxic cocktails known to humankind



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.43

43

Getting back to the Siemens process ... now we start counting nines

- At this stage, the material is pure enough to be called **polysilicon**
- Its quality is measured by how many nines appear in the purity number
- **Solar-grade polysilicon** for multicrystalline cells can reach eight nines of purity: 99.999999%
- **Monocrystalline polysilicon** goes even further, reaching 9 nines: 99.99999999%



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.44

44

The Silicon aristocracy

- The *pièce de résistance* of the silicon world is **semiconductor grade polysilicon**
- It can reach 10 nines of purity: 99.99999999%
- At that level, impurities become almost unimaginably rare
- For roughly every 1 impure atom, there are about 10 billion pure silicon atoms



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.45

45

Where are we in our odyssey?

- We are only at the midpoint of the journey, and the silicon is not yet ready to become a chip



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.46

46

THE PUREST SUBSTANCE IN THE WORLD

COMPUTER SCIENCE DEPARTMENT



COLORADO STATE UNIVERSITY

47

One atom can ruin the party

- At this stage you could be forgiven for asking:
 - Why all the fuss over a few decimal points?
 - Is it honestly worth the effort?
- A lone, rogue atom in an otherwise pure silicon matrix is enough to disrupt the flow of electrons in a transistor



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.48

48

Why order matters

- Defects in the material, called **grain boundaries**, can interrupt the flow of electrons
- The more grain boundaries there are, the greater the chance the semiconductor will fail
- Silicon works best when its internal structure is orderly and uniform
- Think of eggs packed neatly in a carton, not scattered higgledy-piggledy across the floor



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.49

49

Next step: Perfection ...

- The next challenge is to reorganize the pure but disorderly atoms into a near-perfect crystal matrix
- To do that, the polysilicon is sent across the planet to the north-west coast of the U.S., near Portland, Oregon
- This stage is about turning purity alone into the structural order that chips require
- Waiting there is **Shin-Etsu**, another little-known company that is in fact a giant of the material world



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.50

50

A river runs through it ...



- Our journey has reached the banks of the **Columbia River**, the unlikely heart of the 21st century U.S. silicon industry
- That location is no accident: the river runs from the Rockies to the Pacific and powers 14 hydroelectric dams along the way
- All that energy is needed for the next transformation, where polysilicon is turned into the **pure crystalline wafers** that foundries require
- By this point, a pattern is becoming clear:
 - At every stage of silicon's journey, astonishing precision is matched by astonishing appetite for energy



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.51

51

The crystal is pulled from the fire

- At Shin-Etsu, engineers use the **Czochralski technique** (or simply CZ) to begin one of the most mesmerizing steps in the whole journey
- The polysilicon is loaded into an **ultra-pure quartz crucible** and heated to just under 1,500°C
- A **tiny seed crystal**, no bigger than a pencil-sized rod, is dipped into the molten silicon and slowly drawn upward while rotating
- As it rises, atom by atom, the melt organizes itself around that seed and a **flawless** solid cylinder begins to emerge
- Out of the furnace comes a **boule**: a large single crystal that will eventually be sliced into wafers



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.52

52

Like candyfloss minus the joy

- The motion is a little like pulling candyfloss onto a stick
- But this is the opposite of a carnival: no sugar, no joyfully screaming children, and no fun
- The process takes place inside a chamber filled with **argon** gas
- Same basic idea, but vastly higher precision and stakes



The wafer after a great deal of fuss

- The finished silicon boule stands roughly 2-3 meters (~6-10 feet) tall
- A **silicon-carbide wire** saw then slices it into wafers less than a millimeter thick
- Each wafer is about the size of a small pizza, then polished and chemically cleaned until it is perfectly flat
- At that point, the quartz that came out of the ground has become a silicon wafer



That tidy summary from the previous slide hides a great deal

- In reality, the material spends **months** inside Shin-Etsu's labs
 - Being sliced, smoothed, cleaned, tested, and passed from one exacting machine to the next
- You or I won't ever be allowed into this holy of manufacturing holies
 - For fear of industrial espionage is ever persistent



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.55

55

The world depends on a town that you have probably never heard of

- If you want the ultra-pure quartz needed to make the crucibles for silicon wafers, there is one place that matters:
 - **Spruce Pine, North Carolina**
- This small town on the Blue Ridge escarpment sits at the root of a surprisingly large part of the silicon supply chain
- For many years, the mine there (and effectively much of the world's supply of high-purity quartz) is run by a single company
- That company is **Sibelco**, a secretive Belgian firm with an outsized role in modern chipmaking



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.56

56

Spruce Pine is in a league of its own

- Quartz as pure as the snow-white stone at Serrabal is already hard to find
- Quartz as pure as Spruce Pine's is harder still ... almost impossible!
- A few tiny producers exist in India and Siberia
 - ▣ But none match the consistency and quality of the two mines at Spruce Pine



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.57

57

Only now does the chip story really begin

- Our silicon began its life on the side of a mountain in Galicia
- Since then, it has been transformed again and again: solid to liquid to solid to vapor to solid to liquid to solid
- After this **long industrial pilgrimage**, it now sits inside a sealed canister on the other side of the world
- Only now is it finally approaching the stage where it can become something recognizably like a chip



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.58

58

The middle of nowhere, at the center of everything

- In Tainan, the old capital of Taiwan, the city soon gives way to fields of sugarcane and cabbage
- The air is thick and sticky, and for a moment it feels as though you are heading away from the modern world
- But that impression is gloriously misleading
- For anyone in computing, this landscape leads straight to the center of the universe



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.59

59

A very unassuming road to the most advanced factory on earth

- Rising out of the fields is a guarded entrance and, beyond it, a cluster of gleaming silver buildings
- The site is formally called the Southern Taiwan Science and Technology Park
- But what matters is the name written in red across the buildings: **TSMC**
- This is **Fab 18**, widely regarded as the most advanced factory in the world



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.60

60

The company you've probably never heard of is possibly running the modern world

- TSMC exists to manufacture the chips designed by companies like Apple, Tesla, Nvidia, and Qualcomm
- It is the powerhouse behind the **fabless model**, where firms design chips but do not build them themselves
- Though little known outside computing, TSMC has pushed manufacturing to the edge of physics and become one of the world's most important companies
- That dominance comes at a staggering price: over three years starting in 2021, TSMC invested \$135 billion
 - ▣ The equivalent of 10 U.S. Gerald R. Ford-class aircraft carriers!!!



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.61

61

When a fab costs more than a tunnel

- Fab-18 is not one building but a sprawling complex of six interconnected units
- Cost about \$17 billion
- That makes this “factory” a little more expensive than the Channel Tunnel, even after adjusting for inflation



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.62

62

The cleanest place you would not want to linger

- Peel back the silver shell of Fab 18 and much of the space is not offices or work areas, but vast filtration and air-conditioning systems
- All of it exists to protect the **cleanroom**, the part of the factory where even tiny specks can ruin a chip
- In an ordinary office, air might be filtered a handful of times each hour; in this class-1 cleanroom, it is replaced about **600 times an hour**
- Taken together, the cleanroom complex spans roughly the area of **25 football pitches**, making it one of the cleanest places on Earth
- Beneath it sits the sub-fab, where aggressive chemical mixtures churn below the floor and are pumped upward to the machines waiting above



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.63

63

Built to ignore the earth beneath it

- Beneath the sub-fab sits an extraordinarily sophisticated system of **dampers**
 - A damper is a device that absorbs and reduces vibrations or shaking
- These dampers largely *isolate* the building from the ground below
- That matters because even tiny vibrations can disrupt the precision machinery inside
- It is also why advanced fabs are **rarely built near airports or major motorways**, and why seismic stability matters so much in Taiwan



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.64

64

Writing chips with light

- Transistors are etched onto silicon wafers not by hand and not by cutting tools, but by **light**
- The basic idea is like a **movie projector in reverse**
- A projector takes a tiny image and blows it up onto a giant screen
- **Photolithography** starts with a large chip blueprint and shrinks it down to microscopic scale
- That reduced pattern is then projected onto the wafer



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.65

65

This is not just light in a box

- The machine that does this, ASML's TWINSCAN NXE:3600D, costs 100s of millions of dollars
- That sounds excessive until you realize the whole game is **precision at almost absurd scale**
- To carve the smallest features, you need light with an exceptionally short wavelength
- That is why this process uses **extreme ultraviolet light, or EUV**
- In chipmaking, smaller wavelength means finer detail



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.66

66

How to make EUV

- Inside the machine, tin is melted into liquid and released as a stream of tiny droplets
- Each droplet is **struck twice** by powerful pulse lasers
- The lasers heat the tin to around a **million degrees**, turning it into *plasma*
- That plasma emits a burst of EUV light
- This happens about 60,000 times every second!
 - 24 x 7 x 365



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.67

67

The smoothest mirrors in the universe

- EUV light is so difficult to handle that ordinary optics will not do
- ASML relies on **Zeiss** to build extraordinary mirrors called **Bragg reflectors**
- These mirrors are made from layered silicon and molybdenum and polished to astonishing smoothness
- These are probably the smoothest man-made structures
- Scaled up to the size of the United States, their biggest bump would be less than half a millimeter high



COLORADO STATE UNIVERSITY

Professor: SHRIDEEP PALLICKARA
COMPUTER SCIENCE DEPARTMENT

MACHINE LANGUAGE

L20.68

68

At last, the light does the writing

- After bouncing across this staircase of mirrors, the **13.5-nanometre** EUV light reaches the wafer
- There, with the help of chemicals coated on the wafer surface, it burns away tiny channels and grooves
- In effect, the **circuitry is engraved by light**
 - ▣ This is how the intricate transistor patterns of a modern chip are written into silicon
 - ▣ After everything we have seen so far, the actual writing is done not by a blade, but by a beam



The chip is global, whether anyone likes it or not

- Fab 18 may manufacture the world's most advanced chips, but most are designed elsewhere, especially in the United States
- Key intellectual property in that design stack traces back to ARM in Cambridge, England
- TSMC's fabs also depend on machine tools from the Netherlands and Japan
- They rely on chemicals from Germany and countless specialized components from many other countries
- The final lesson is hard to miss: no nation makes a leading-edge chip alone



The contents of this slide-set are based on the following references

- Jonathan E. Steinhart. *The Secret Life of Programs: Understand Computers -- Craft Better Code*. ISBN-10/ ISBN-13 : 1593279701/ 978-1593279707. No Starch Press. [Chapter 5]
- Randall Hyde. *Write Great Code, Volume 1, 2nd Edition: Understanding the Machine 2nd Edition*. ASIN: B07VSC1K8Z. No Starch Press. 2020. [Chapter 11]
- Matthew Justice. *How Computers Really Work: A Hands-On Guide to the Inner Workings of the Machine*. ISBN-10/ISBN-13 : 1718500661/ 978-1718500662. No Starch Press. 2020. [Chapter 7]
- *Material World: The Six Raw Materials That Shape Modern Civilization* by Ed Conway. 2023. Knopf.
- Adrian McMenamin: <https://cartesianproduct.wordpress.com/2013/04/15/the-end-of-dennard-scaling/>
- https://en.wikipedia.org/wiki/Dennard_scaling
- https://en.wikipedia.org/wiki/Moore%27s_law

