

Furnace: oxidation and annealing in the heat

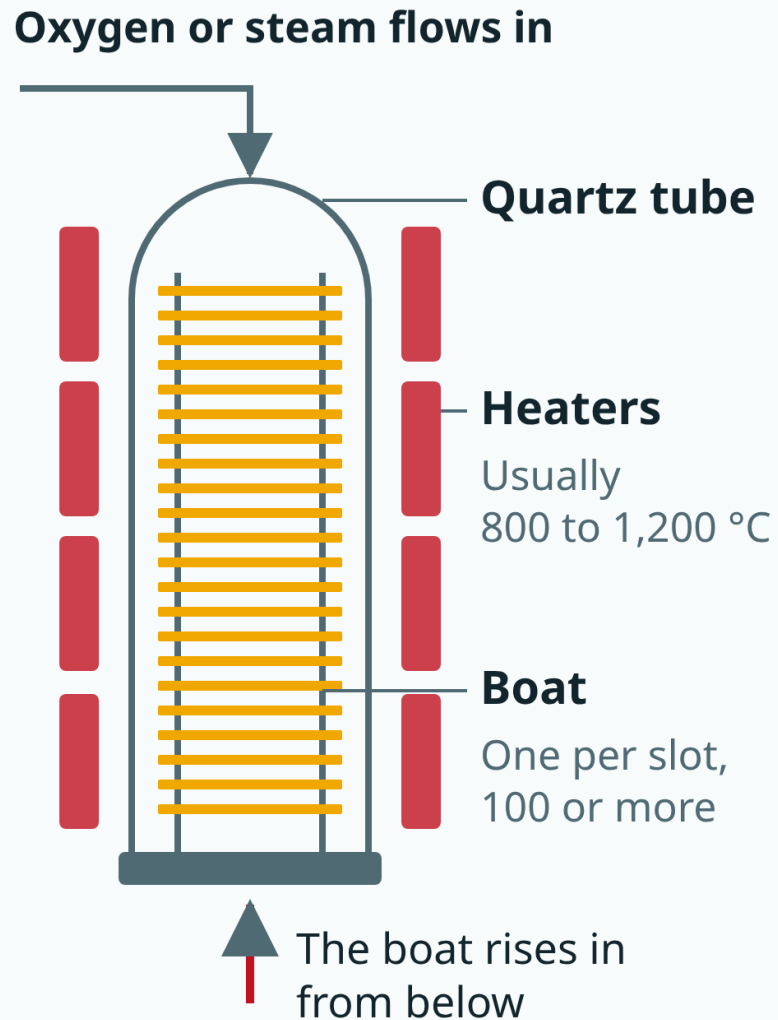
In many processes, the very thin insulating layer under a transistor's gate is grown in a furnace, and implanted atoms only start to conduct after heating. But every heating moves what was made before a little more, so the heat has to be counted from start to finish.

Semiconductors 101: a free course following the life of a chip

Free to use in class and for non-commercial teaching. Please credit fab24hr. Do not sell it or present it as your own work.

Goals

- 1 Explain why thermal oxidation needs high heat, where the new oxide grows, and what dry and wet oxidation are each used for.
- 2 Explain how annealing makes implanted atoms conduct, what too much or too little temperature and time do, and why some anneals now heat for only seconds.
- 3 Describe how a furnace heats a hundred-plus wafers alike (heating zones, dummy wafers) and how that is checked (monitor wafers), and why the thermal budget is planned from start to finish.



Thermal oxidation

Silicon and oxygen grow silicon dioxide in the heat

Oxygen gets through;
new oxide forms below

A furnace heats a hundred or more wafers at once (simplified). Nearly half of the oxide that thermal oxidation grows has eaten its way into the wafer.

Step by step

1 Clean the wafers thoroughly before the furnace.

Why: Metals and organics on the surface get baked in by the heat.

If not: The oxide comes out worse, and metal that gets into the silicon makes transistors leak.

2 Load the boat: dummy wafers at the ends, product in the middle, monitor wafers at the top, middle and bottom.

Why: The very top and bottom of the boat see different temperatures and gas flow from the middle.

If not: Product at the ends comes out different from the product in the middle.

3 Raise the boat into the tube slowly, and heat up slowly.

Why: Heated too fast, a wafer heats unevenly, and its edge and middle expand by different amounts.

If not: The silicon crystal is pulled out of line, leaving slip lines at the edge.

Step by step

4 At temperature, do the load's job: grow an insulating layer by thermal oxidation, or anneal after ion implant.

Why: Heat lets oxygen get through the oxide already grown so it keeps growing; annealing lets jumbled atoms settle back.

If not: Without annealing, the implanted atoms sit stuck in the gaps and do nothing for conduction.

5 After the run, measure the top, middle and bottom monitor wafers: thickness for oxide, sheet resistance for an anneal.

Why: A little difference in temperature changes what grows; the more atoms at their posts, the lower the sheet resistance.

If not: Without measuring, no one knows whether the top and bottom came out alike, and a bad run is the whole batch.

6 Count every heating in the thermal budget (the total heat a wafer takes in its life), planned from start to finish.

Why: Each heating makes the atoms implanted earlier move a little more, and the shapes already made slowly blur.

If not: Without planning it from the start, you find out late that too much heat was spent; the shapes made earlier have already blurred, and what's spent can't be taken back.

What you adjust, what can go wrong

Furnace temperature

Changes: A little hotter grows oxide a little faster, and in an anneal puts more atoms at their posts.

If it is off: A small difference changes the result: if the top and bottom of the tube differ, wafers in the same load come out different thicknesses; and the hotter it runs, the more heavily it counts against the thermal budget.

Time in the heat

Changes: The longer the wafers stay, the thicker the oxide grows and the more complete the anneal.

If it is off: The longer the heat lasts, the more implanted atoms spread sideways and blur what was built, and the thermal budget can't be taken back.

Dry or wet oxidation

Changes: Dry is slow and makes a denser layer; wet is much faster, for layers that need to be thick.

If it is off: A little extra moisture in the tube speeds oxidation, and the oxide comes out thicker than planned.

What you adjust, what can go wrong

Heating and cooling rate

Changes: Slow heating and cooling warm the edge and the middle of the wafer more alike.

If it is off: Too fast, and the edge and the middle expand differently, leaving slip lines at the edge.

How to anneal (a furnace batch, or rapid heating one wafer at a time)

Changes: A furnace does a whole batch, and a run takes hours; rapid heating does one wafer at a time for only seconds, so the implanted atoms have less time to spread sideways.

If it is off: The smaller the structures, the less they can take a long heat: anneal them in a slow furnace run and even a tiny bit of movement visibly blurs the shapes; with only seconds of heat, the anneal still has to be enough for the atoms to take their posts.

What a furnace does: heating a hundred-plus wafers at once

A furnace is a fab's high-temperature oven: a hundred or more wafers stand in a quartz rack and slide into a long tube, usually at 800 to 1,200 °C. It has two main jobs. Thermal oxidation makes silicon and oxygen grow a layer of silicon dioxide that doesn't conduct; in many processes, the most important one is the very thin insulating layer under a transistor's gate (newer processes often grow it one wafer at a time, and the most advanced add a different insulator by atomic layer deposition; see the thin-film unit). Annealing lets the atoms knocked out of place during implantation settle back, and only then do the implanted atoms start to conduct. Some thin films are grown in furnaces too. One run, from heating up to cooling down, takes hours and can approach a full day, so loading many wafers at once pays off; the price is that when a run goes wrong, the whole batch goes wrong.

Oxygen has to get through the oxide already there, so add heat

Silicon left in air grows a very thin layer of silicon dioxide on its own, a bit like iron rusting, except that this "rust" is clear glass. At room temperature it stops at one or two nanometers: oxygen has to get through the layer already there to reach the silicon, and it doesn't have enough energy. With heat, the oxygen gets through, and new silicon dioxide keeps forming underneath, at the boundary with the silicon, using up the wafer's own silicon. There are two ways to supply the oxygen: dry oxidation uses pure oxygen gas, grows slowly and makes a denser layer, for the thinnest, most critical layers; wet oxidation uses water vapor and is much faster, for layers that need to be thick. That's why even a tiny bit of extra moisture in the tube has to be kept out, or the oxide comes out thicker than planned.

Annealing: getting implanted atoms to take their posts

Silicon atoms sit in a neat arrangement, like oranges packed in a crate. Ion implantation fires other atoms in at high speed, the neat rows get knocked out of place, and most implanted atoms end up wedged in the gaps, where they don't help conduct. With heat, the damage slowly sorts itself out and the implanted atoms take the place of silicon atoms; only then do they conduct. This is called activation. Whether an anneal did its job is often checked with sheet resistance: the more atoms at their posts, the better the layer conducts and the lower the sheet resistance. Annealing is a trade-off: higher temperatures and longer times put more atoms at their posts, but the longer the heat lasts, the more the atoms spread sideways and blur what was built. So some anneals are still done in a furnace, a whole batch at once, and others have moved to rapid heating, one wafer at a time, for only seconds.

Inside the furnace, step by step: from loading the boat to unloading

Before the furnace, wafers get a thorough clean, because metals and organics on the surface get baked in by the heat. Then they go one per slot into a quartz rack called a boat. The very top and bottom of the boat see different temperatures and gas flow from the middle, so a few dummy wafers that never ship fill the ends, the product stays in the middle, and monitor wafers (wafers that never ship, used only to check the tool) go in at the top, middle and bottom. The boat rises slowly into the tube and heats up slowly: heat it too fast and the wafer heats unevenly, the edge and the middle expand by different amounts, and fine straight marks called slip lines are left at the edge. At temperature, the process gas goes in, for example oxygen for oxidation, and is held for a time; then the furnace cools slowly. The temperature, gas and time of every step are written in one recipe (the settings the tool follows).

Did this load come out right? Check the top, middle and bottom monitor wafers

How fast oxide grows depends a lot on temperature: a little hotter, a little faster. A hundred-plus wafers form a long stack, and if the top runs hotter than the bottom, the top wafers come out thicker. So a furnace splits its heating into several zones, each with its own heater and thermometer (a thermocouple), adjusted separately. After the run, the oxide on the monitor wafers from the top, middle and bottom is measured; a clear oxide can be measured with light. If the top and bottom differ more than usual, someone has to find out whether the temperature or the gas flow is off. For an anneal, sheet resistance is measured instead. The numbers go on a control chart to catch slow drift.

Thermal budget: every heating is counted, from start to finish

In the heat, atoms slowly move, the way a drop of ink spreads faster in hot water than in cold. The atoms implanted earlier were placed at a planned position and depth, and every later heating step, whether oxidation, annealing or film growth, moves them a little more, so the shapes slowly blur. So engineers keep an account called the thermal budget: the total of how hot a wafer has been, and for how long, over its whole life. Once it's spent, it can't be taken back. The account is planned from start to finish: the conditions for an early step must already allow for all the heat that comes later, and the smaller the structures, the tighter it gets. Once metal wiring is on the wafer, it can never go back into a furnace at over a thousand degrees; only a few hundred degrees are allowed.

Common misconceptions

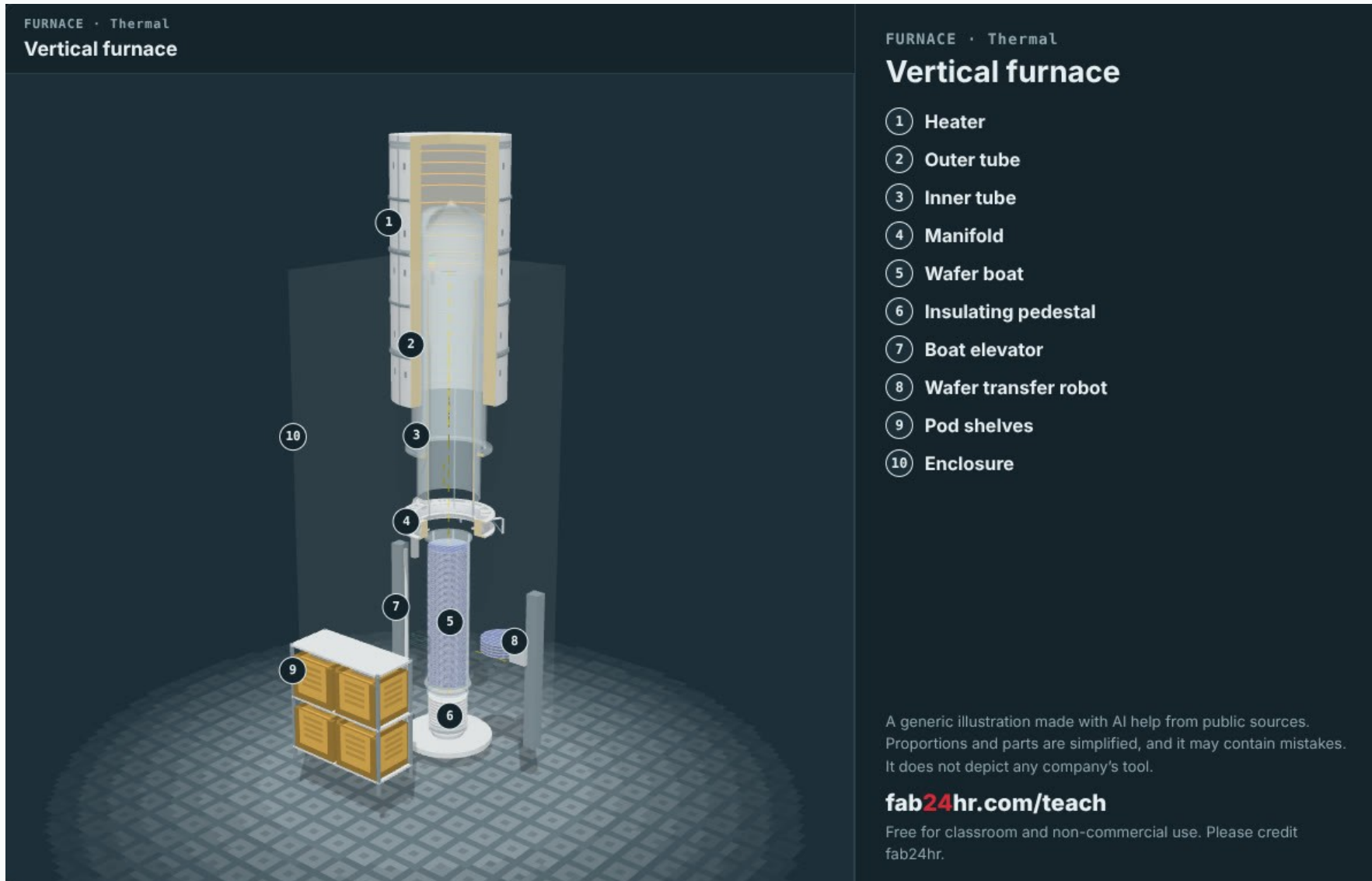
People say: An oxide layer is laid on top of the wafer surface.

In fact: In thermal oxidation, oxygen gets through the layer already there and forms new silicon dioxide underneath, at the boundary with the silicon, using up the wafer's own silicon: of 100 nanometers of oxide, about 44 used to be silicon. Bringing material in from outside and laying it on top is deposition, mostly done in the thin film area; a furnace can also lay down silicon nitride or polysilicon from gases this way.

People say: For annealing, hotter and longer is always better.

In fact: Higher temperatures and longer times do help more atoms take their posts, but the longer the heat lasts, the more the implanted atoms spread sideways and blur what was built, and every heating counts against the thermal budget, which can't be taken back. So an anneal has to be hot enough but not too long, and some have moved to rapid heating, one wafer at a time, for only seconds.

Semiconductors 101, unit 9 · Look: machines to take apart



Play

Play a shift as a furnace process engineer: a hundred-plus wafers a load, and the temperature and time are yours

Scan with a phone to play from here

fab24hr.com/en/?job=pfu&utm_source=slides



Think

- 1 A furnace load of a hundred wafers is efficient, but if that load goes wrong, the loss is large too. In your own life or work, when would you do many at once, and when would you rather do one at a time? Why?
- 2 A furnace run stops halfway. Rerunning the whole step spends a whole extra step's worth of thermal budget; holding the lots, measuring the monitor wafers and running only what's left takes more time. If it were your call, what would you want to know first?

Read

What does a furnace do in a chip fab? Oxidation and annealing explained

fab24hr.com/en/blog/what-is-furnace/

Semiconductors 101: a free course following the life of a chip

<https://fab24hr.com/en/learn/furnace/>

Written from public textbooks, industry standards, university course notes and papers; the companies, people, machine numbers and figures in the game are fictional.

Free to use in class and for non-commercial teaching. Please credit fab24hr. Do not sell it or present it as your own work.

Generic illustration made with AI help from public sources; may contain mistakes; not any company's tool