A Facebook data centre in Luleå, Sweden.

THE INFORMATION FACTORIES

Data centres are chewing up vast amounts of energy
— so researchers are trying to make them more efficient.

BY NICOLA JONES

Upload your latest holiday photos to Facebook, and there's a chance they'll end up stored in Prineville, Oregon, a small town where the firm has built three giant data centres and is planning two more. Inside these vast factories, bigger than aircraft carriers, tens of thousands of circuit boards are racked row upon row, stretching down windowless halls so long that staff ride through the corridors on scooters.

These huge buildings are the treasuries of the new industrial kings: the information traders. The five biggest global companies by market capitalization this year are currently Apple, Amazon, Alphabet, Microsoft and Facebook, replacing titans such as Shell and ExxonMobil. Although information factories might not spew out black smoke or grind greasy cogs, they are not bereft of environmental impact. As demand for Internet and mobile-phone traffic skyrockets, the information industry could lead to an explosion in energy use (see 'Energy forecast').
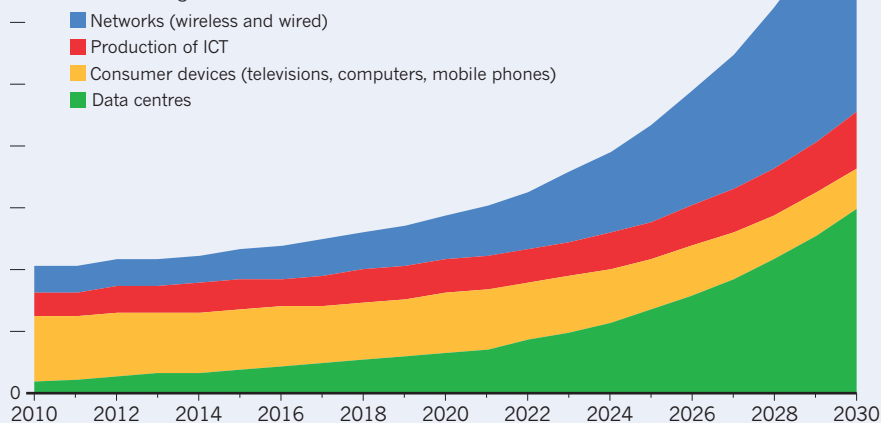
Already, data centres use an estimated 200 terawatt hours (TWh) each year. That is more than the national energy consumption of some countries, including Iran, but half of the electricity used for transport worldwide, and just 1% of global electricity demand (see 'Energy scale'). Data centres contribute around
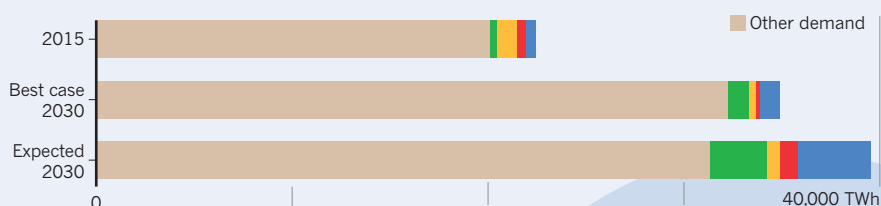
ENERGY FORECAST

Widely cited forecasts suggest that the total electricity demand of information and communications technology (ICT) will accelerate in the 2020s, and that data centres will take a larger slice.

9,000 terawatt hours (TWh)

- Networks (wireless and wired)
- Production of ICT
- Consumer devices (televisions, computers, mobile phones)
- Data centres

20.9% of projected electricity demand

2010 2012 2014 2016 2018 2020 2022 2024 2026 2028 2030

The chart above is an 'expected case' projection from Anders Andrae, a specialist in sustainable ICT. In his 'best case' scenario, ICT grows to only 8% of total electricity demand by 2030, rather than to 21%.

Global electricity demand

Other demand

2015
Best case 2030
Expected 2030

0   40,000 TWh

INTERNET EXPLOSION

Internet traffic* is growing exponentially, and reached more than a zettabyte (ZB, 10^{21} bytes) in 2017.

1987
2 TB[†]

1997
60 PB

2007
50 EB

2017
1.1 ZB

*Traffic to and from data centres. [†]TB, terabyte (10^{12} bytes); PB, petabyte (10^{15} bytes); EB, exabyte (10^{18} bytes).

With the spectre of an energy-hungry future looming, scientists in academic labs and engineers at some of the world's wealthiest companies are exploring ways to keep the industry's environmental impact in check. They are streamlining computing processes, switching to renewables and investigating better ways to cool data centres and to recycle their waste heat. ICT's energy use must be "vigilantly managed", says Eric Masanet, an engineer at Northeastern University in Evanston, Illinois, who co-authored an International Energy Agency (IEA) report[2] last year on digitalization and energy — but if we stay on top of it, he says, we should keep future energy demand in check.

SHIFT TO HIGH GEAR

Perhaps the most startling forecast of ICT's future energy demand comes from Anders Andrae, who works on sustainable ICT at Huawei Technologies Sweden in Kista; he predicts that data-centre electricity use is likely to increase about 15-fold by 2030, to 8% of projected global demand[1]. Such dire numbers are controversial. "There have been many alarmist predictions of growing ICT energy use over the years, and all have proven to be bunk," says Masanet. Last year's IEA report estimated that although data-centre workloads will shoot up — tripling 2014 levels by 2020 — efficiency gains mean that their electricity demand might sneak up only by 3% (ref. 2). ICT's carbon footprint as a whole might even drift downwards by 2020, as smartphones take over from larger devices, researchers have suggested[3].

Data-centre electricity demand has remained roughly level over the past half-decade, in part because of the 'hyperscale shift' — the rise of super-efficient information factories that use an organized, uniform computing architecture that easily scales up to hundreds of thousands of servers. Hyperscale data centres emerged about a decade ago when companies such as Amazon and Google began to need fleets of a quarter of a million servers or more, says Bill Carter, chief technical officer at the Open Compute Project. It was started by Facebook in 2011 to share hardware and software solutions to make computing more energy-efficient. At that point, it made no sense to use off-the-shelf hardware from a computing firm, as companies had typically done.

"You had the opportunity to strip things down to just what you need, and make it specific to your application," Carter says. The new hyperscalers made bare-bones servers designed for purpose. "We stripped out video connectors, because there's no video monitor. There's no blinking lights because there's no one walking the racks. There's no screws," says Carter. On average, one server in a hyperscale centre is said to be able to replace 3.75 servers in a conventional centre.

The savings made by hyperscale centres can be seen in their power usage efficiency (PUE), defined as the total energy needed for everything, including lights and cooling, divided

---

0.3% to overall carbon emissions, whereas the information and communications technology (ICT) ecosystem as a whole — under a sweeping definition that encompasses personal digital devices, mobile-phone networks and televisions — accounts for more than 2% of global emissions. That puts ICT's carbon footprint on a par with the aviation industry's emissions from fuel. What could happen in the future is hard to forecast. But one of the most worrying models predicts that electricity use by ICT could exceed 20% of the global total by the time a child born today reaches her teens, with data centres using more than one-third of that (see 'Energy forecast')[1]. If the computationally intensive cryptocurrency Bitcoin

continues to grow, a sharp rise in energy demand could come sooner rather than later (see 'The Bitcoin bite').

For now, despite rising demand for data, ICT's electricity consumption is staying nearly flat, as increased Internet traffic and data loads are countered by increased efficiencies — including shuttering older facilities in favour of ultra-efficient centres such as Prineville's. But those easy wins could end within a decade. "The trend is good right now, but it's questionable what it's going to look like in 5–10 years," says Dale Sartor, who oversees the Center of Expertise for Energy Efficiency in Data Centers at the US Department of Energy's Lawrence Berkeley National Laboratory in Berkeley, California.

by the energy used for computing (a PUE of 1.0 would be a perfect score). Conventional data centres typically have a PUE of about 2.0; for hyperscale facilities, that's been whittled down to about 1.2. Google, for one, boasts a PUE of 1.12 on average for all its centres.

Older or less technologically adept data centres can contain a mix of equipment that is hard to optimize — and some that is even useless. In 2017, Jonathan Koomey, a California-based consultant and leading international expert on IT, surveyed with a colleague more than 16,000 servers tucked into corporate closets and basements and found that about one-quarter of them were "zombies", sucking up power without doing any useful work — perhaps because someone simply forgot to turn them off. "These are servers sitting around doing nothing except using electricity, and that's outrageous," says Koomey.

In a 2016 report, the Lawrence Berkeley National Laboratory estimated that if 80% of servers in small US data centres were moved over to hyperscale facilities, this would result in a 25% drop in energy use[4]. That move is under way. Today, the world has around 400 hyperscale data centres, many of them mopping up services for small corporations or universities that in the past would have had their own servers. Already they account for 20% of the world's data-centre electricity usage. By 2020, hyperscale centres will account for almost half of it, the IEA says (see 'Hyperscale shift').

RUNNING HOT AND COLD

Once the hyperscalers have taken as much of the load as they can, further efficiencies will be harder to find. But corporations are trying. One emerging management technique is to make sure that servers are at full throttle as much of the time as possible, whereas others are turned off rather than being left idle. Facebook invented a system called Autoscale that reduces the number of servers that need to be on during low-traffic hours; in trials, this led to power savings of about 10–15%, the company reported in 2014.
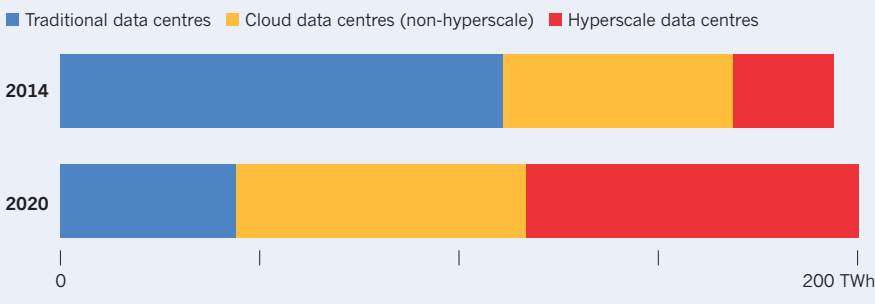
One big way the hyperscalers have whittled down their PUE is by tackling cooling. In a conventional data centre, standard air conditioning can soak up 40% of the energy bill. The use of cooling towers, which evaporate water to drive the cooling of air, causes another environmental problem: US data centres are estimated to have used about 100 billion litres of water in 2014. Getting rid of compression chillers and cooling towers helps to save both energy and water.

One popular solution is to simply locate data centres in cool climes and blow the outside air into them. Such centres don't have to be in icy regions: Prineville is cool enough to take advantage of so-called 'free air cooling', and so are many other data-centre locations, says Ingmar Meijer, a physicist at IBM Research in Zurich, Switzerland.

Piped water is an even better conductor of heat, allowing centres to be cooled using warm water, which is less energy intensive to produce and recycle in the cooling system. Even in temperate climates, warm-water cooling has become the de facto solution for managing high-performance computers that are run fast and hot, including those in US Department of Energy labs and the Bavarian Academy of Sciences' SuperMUC supercomputer in Garching, Germany. Commercial centres in warm climates sometimes also invest in these systems, such as eBay's Project Mercury data centre in Phoenix, Arizona.

For high-density, high-power computing, the most efficient thing to do is to immerse servers in a non-conductive oil or mineral bath. Facebook trialled that in 2012 as a way to run its servers at higher speeds without overheating them. For now, immersion cooling is a specialist area with tricky maintenance, says Meijer.

In 2016, Google turned its DeepMind artificial intelligence (AI) research team to the task of tuning its data centre's cooling system to match the weather and other factors. Google says that, in tests, the team reduced its cooling energy bill by 40% and "produced the lowest PUE the site had ever seen". This August, the company announced it had turned cooling control at some data centres over to its AI algorithm.

Exploring innovative cooling solutions and making existing ones cheaper will become more important in coming years, says Carter. "As we connect the world, there's areas that won't be able to use free air cooling," he notes, pointing to Africa and southern Asia. And other developments will tax IT infrastructure in new ways. If self-driving cars flood onto the roads, for example, small server installations at the base of mobile-phone towers, used to help those cars communicate and process data,

will need high-power devices that can handle real-time AI workloads — and better cooling options. This year, the Open Compute Project launched a project on advanced cooling, with the aim of making efficient cooling systems more accessible. "The hyperscalers have it figured out; they're extremely efficient," says Carter. "We're trying to help the other guys."

Hand-in-hand with better cooling is the idea of using the heat coming off the servers, thus

HYPERSCALE SHIFT

Efficient 'hyperscale' data centres are predicted to swallow up half of data-centre electricity demand by 2020, as smaller, less-efficient centres shut down.

■ Traditional data centres　■ Cloud data centres (non-hyperscale)　■ Hyperscale data centres

"WE'RE UP AGAINST THE LIMITS OF SHRINKAGE."

ENERGY SCALE

Global electricity demand

20,000 TWh

Electricity use by ICT

2,000 TWh

Data-centre electricity demand

200 TWh

Bitcoin use by mid-2018
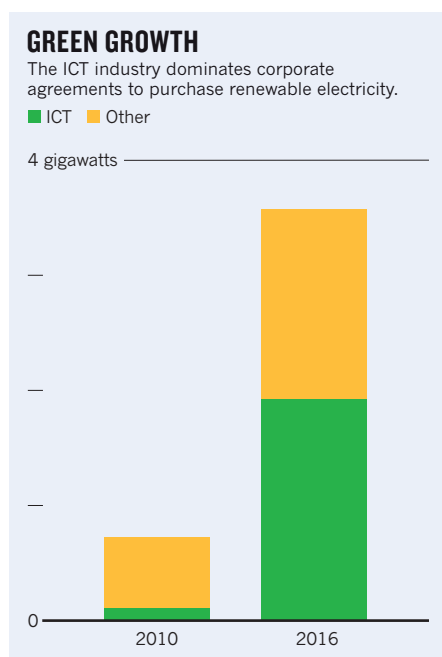
20 TWh

Figures are approximate.

saving electricity demand elsewhere. "It's like a free resource," says IBM researcher Patrick Ruch in Zurich. There are a few examples: the Condorcet data centre in Paris sends its waste heat directly into a neighbouring Climate Change Arboretum, where scientists study the impacts of high temperatures on vegetation. An IBM data centre in Switzerland warms a nearby swimming pool. But heat doesn't travel well, so waste-heat use tends to be limited to data centres located next to a convenient customer, or in a city that already uses piped hot water to heat homes.

A scattering of players are aiming to make waste heat more generally usable, including preliminary efforts to turn it into electricity. Others aim to use the waste heat to run cooling devices — IBM's US$2-million project THRIVE, for instance, is developing new materials that can better soak up water vapour and release it when exposed to heat, to make more efficient 'sorption heat pumps' to keep data centres cool.

POWER PLAY

At their core, data centres are only as good as the processors they're made of — and there's scope for improvement there, too. Since the 1940s, the number of operations a computer can perform with each kilowatt hour (KWh) of energy has doubled about once every 1.6 years for peak performance and every 2.6 years for average performance. That's a 10-billion-fold improvement over 50 years. By some measures, the rate of improvement has slowed since 2000, and the current generation of computing will run up against a physical barrier that limits the function of transistors in just a few decades, according to Koomey's calculations[5].

"We're up against the limits of shrinkage," says Koomey. Making comparable efficiency gains after that, he says, will require

GREEN GROWTH
The ICT industry dominates corporate agreements to purchase renewable electricity.

🟩 ICT 🟧 Other

THE BITCOIN BITE

Cryptocurrency mining could push up energy demand

Since the cryptocurrency Bitcoin was born in 2008, concerns have grown that the energy demand of its production will escalate rapidly. Virtual coins are 'minted' by miners who buy specialized servers to crunch time-intensive computations in a growing blockchain that proves the validity of the new cryptocoins. By mid-2018, says Alex de Vries, a data consultant for international professional-services company PwC in Amsterdam, Bitcoin miners were probably using about 20 terawatt hours of electricity per year globally — less than 10% that of data centres, and less than 0.1% of total electricity use[6]. But estimates of how fast their usage is going up are contentious.

De Vries estimates that, by now, Bitcoin is gobbling up at least 0.33% of global electricity. Including other cryptocurrencies such as Ethereum bumps that up to 0.5%. "I think it's shocking," he says. But others, including Marc Bevand, a cryptocurrency researcher in San Diego, California, say that those numbers are inflated and based on gross assumptions. Bevand estimates

that energy use might be as low as half of de Vries' current numbers by January 2019. "There is growth, but people are hyping this," says Jonathan Koomey, a California-based IT consultant, who is gathering data on cryptocurrency electricity consumption.

For now, Bitcoin mining is only profitable in places where electricity is cheap (about half the global average, says Bevand), which include China, Iceland and areas along the Columbia River in North America where hydroelectric power is plentiful. As Bitcoin miners dig in to an area and stress the grid, energy companies respond by raising their fees. That might prompt miners either to shut down or to take steps to drastically improve the energy efficiency of their hardware or system cooling.

Bitcoin could possibly be migrated to a less-energy-intensive blockchain system, says Bevand (as Ethereum is planning). Or, notes Koomey, "Say Bitcoin collapses for some reason; all those facilities will just go away." **N.J.**

---

a revolution in how hardware is built and computing is done: perhaps through a switch to quantum computing. "It's basically impossible to predict," he says.

Although the core focus is on reducing ICT's energy usage, it's worth remembering that the information industry could also make our power use elsewhere smarter and more efficient. The IEA notes that if all vehicles become automated, for example, there is a utopian possibility that smoother traffic flow and easier carpooling would reduce the total energy demand of the transport industry by 60%. Buildings, which have accounted for 60% of the increased global electricity demand over the past 25 years, have huge scope for energy-efficiency improvements: smart heating and cooling, hooked up to building sensors and weather reports, could save 10% of their future energy demand. Chiara Venturini, director of the Global e-Sustainability Initiative, a Brussels-based industry association, estimates that the IT industry currently abates 1.5 times its own carbon footprint, and that could go up to almost 10 times by 2030.

ICT might also help to reduce global emissions by giving renewables a leg-up over fossil fuels. In 2010, the environmental group Greenpeace published its first ClickClean report, which ranked major companies and threw a spotlight on the environmental burden of IT. In 2011, Facebook made a commitment to using 100% renewable energy. Google and Apple followed in 2012. As of 2017, nearly 20 Internet companies had done the same. (Chinese

Internet giants such as Baidu, Tencent and Alibaba, however, have not followed suit.) Back in 2010, IT companies were a negligible contributor to renewable-power purchase agreements with energy companies; by 2015, they accounted for more than half of such agreements (see 'Green growth'). Google is the largest corporate purchaser of renewable energy on the planet.

Cutting back our thirst for data might be the ultimate way to prevent energy use going into hyperdrive. But it's hard to see anyone agreeing to, say, limit their Netflix use, which accounts for more than one-third of Internet traffic in the United States. Banning high-definition colour cameras on phones alone could reduce data traffic in Europe by 40%, says Ian Bitterlin, a consulting engineer and data-centre expert in Cheltenham, UK. But, he adds, no one seems likely to dare to institute such rules. "We can't put the lid back on Pandora's box," he says. "But we could reduce data-centre power." ■

Nicola Jones *is a freelance journalist based in Pemberton, British Columbia, Canada.*

1. Andrae, A. & Edler, T. *Challenges* **6**, 117–157 (2015).
2. International Energy Agency. *Digitalization and Energy* (IEA, 2017).
3. Malmodin, J. & Lundén, D. *Sustainability* **10**, 3027 (2018).
4. Shehabi, A. *et al. United States Data Center Energy Usage Report* (LBNL, 2016); available at https://go.nature.com/ejg7sr
5. Koomey, J. G. *et al. IEEE Ann. Hist. Comp.* **33**, 46–54 (2011).
6. de Vries, A. *Joule* **2**, 801–805 (2018).

SOURCE: IEA