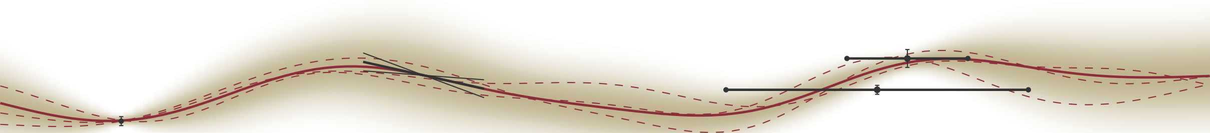
NUMERICS OF MACHINE LEARNING
LECTURE 08
SPECIAL LECTURE:
THE ENERGY IMPACT OF COMPUTING AND AI/ML

Philipp Hennig
18 June 2019

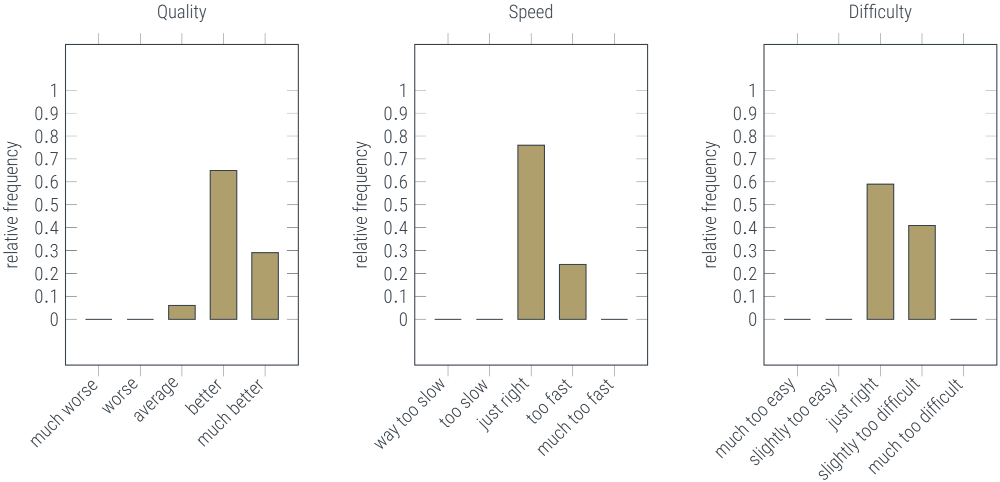
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

FACULTY OF SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
CHAIR FOR THE METHODS OF MACHINE LEARNING



Last Lecture: Debrief

Feedback dashboard



Last Lecture: Debrief

Detailed Feedback

Things you did not like:

- this lecture should have come earlier

Things you did not understand:

- **from LU to Cholesky**
- complexity of various solves

Things you enjoyed:

- **`scipy.linalg` references**
- **LAPACK intro**
- **behind-the-scenes of `linalg.solve`**
- different types of matrices and their properties
- blackboard
- history

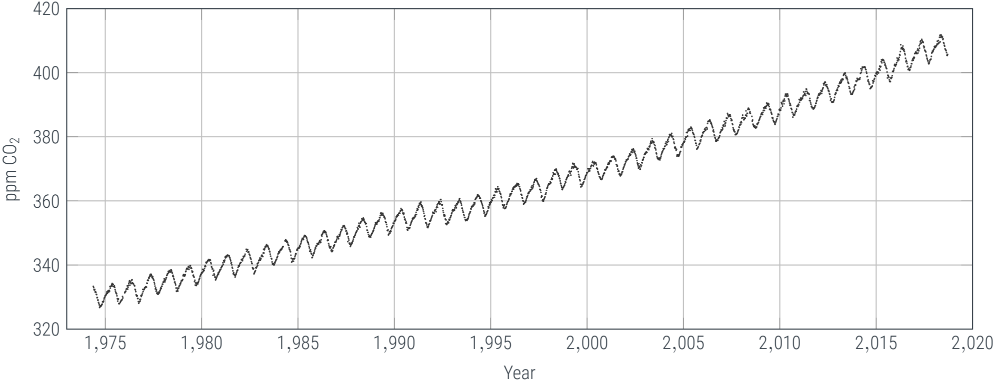
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

0	Introduction	Learning is Computation, Computation is Learning
1	Mathematical Background	Gaussian and Least-Squares Inference
2	Integration — Quadrature	Integration is Regression
3	Integration — Bayesian Quadrature	Regression for Integration
4	Integration — Monte Carlo I	Randomness is a flawed concept
5	Integration — Monte Carlo II	Markov Chains to Explore and Exploit
6	Integration – Monte Carlo III	Efficient Markov Chains
7	Linear Algebra — Direct Methods	Solving Linear Systems by Bookkeeping
8	Special Lecture	The Climate Impact of Computing and AI
9	Linear Algebra — Iterative Methods	Solving Linear Systems as Optimization
10	Optimization — Basic Methods	Minimizing Smooth Multivariate Functions
11	Optimization — Quasi-Newton	Curvature can be Learnt
12	Bayesian Optimization	Optimization of Empirical Functions
13	Revision	

Why this Lecture?

https://lecturesforfuture.org

Data: Mauna Loa Observatory, National Oceanic and Atmospheric Administration, 26 September 2018

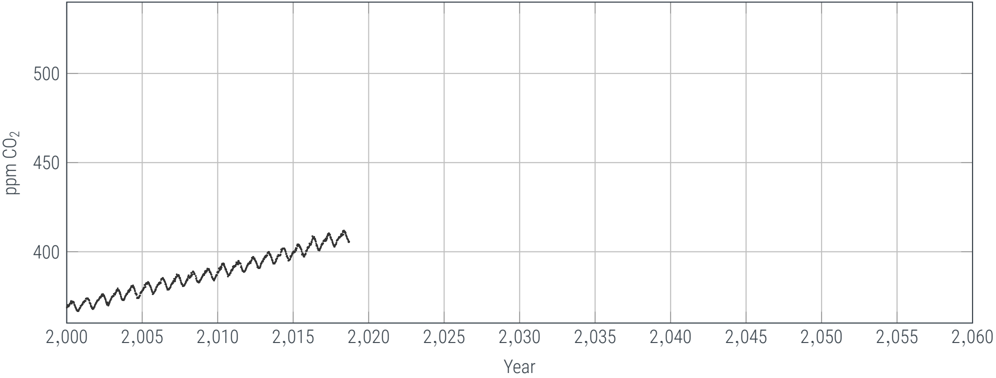


`ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_weekly_mlo.txt`

Why this Lecture?

https://lecturesforfuture.org

Data: Mauna Loa Observatory, National Oceanic and Atmospheric Administration, 26 September 2018

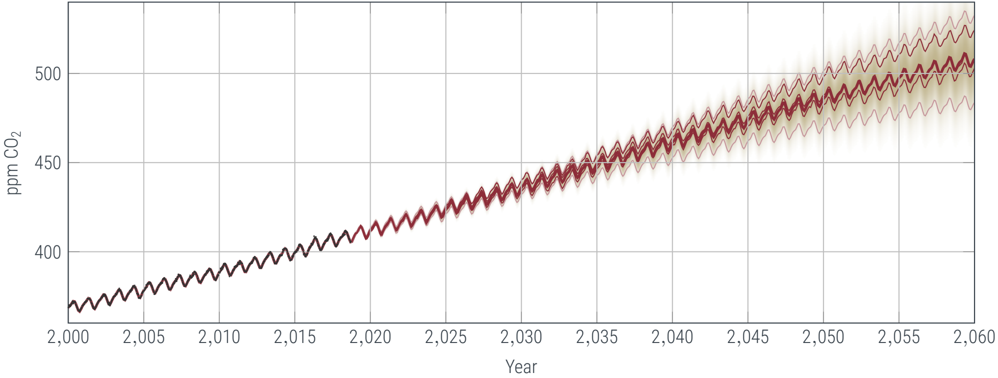


`ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_weekly_mlo.txt`

Why this Lecture?

https://lecturesforfuture.org

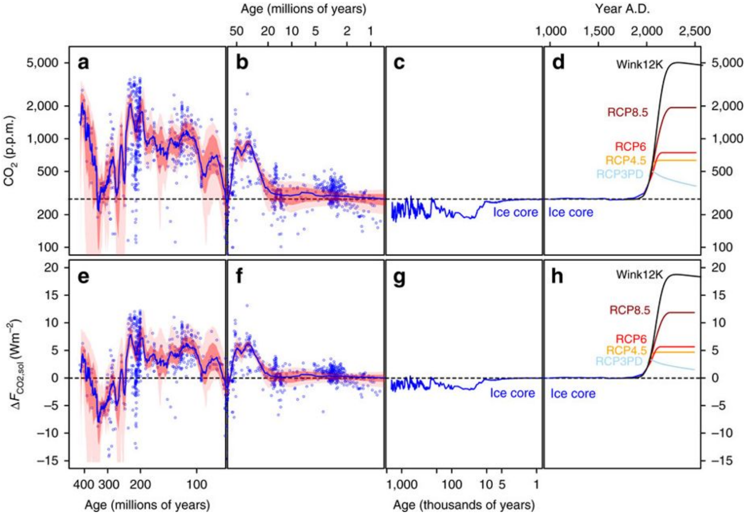
Data: Mauna Loa Observatory, National Oceanic and Atmospheric Administration, 26 September 2018



`ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_weekly_mlo.txt`

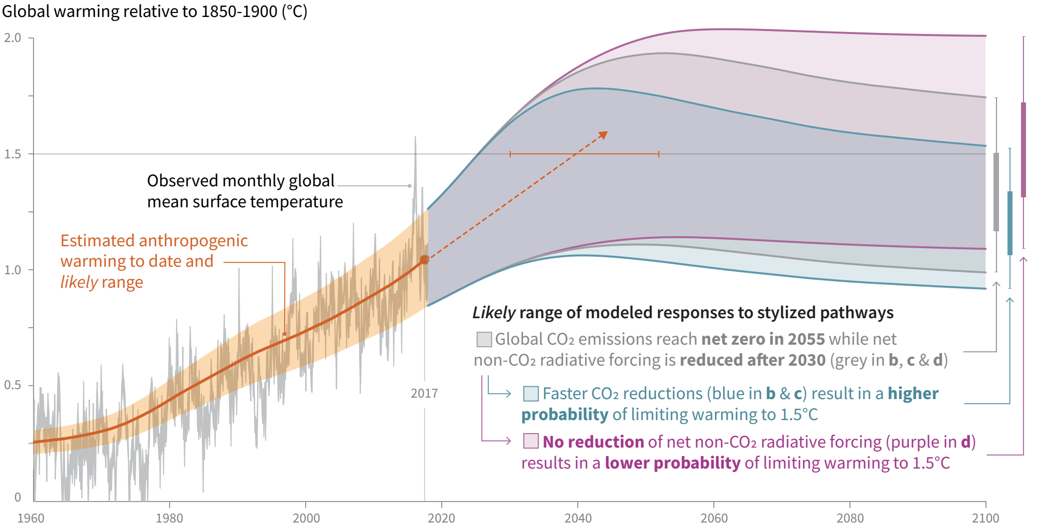
This Is Unprecedented

Source: Foster, Royer & Lunt, *Nat. Comms.* 14845 (2017)



This Can Not Go On

Source: IPCC 2018, Summary for Policymakers, https://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf



Global warming relative to 1850-1900 (°C)

Observed monthly global mean surface temperature

Estimated anthropogenic warming to date and *likely* range

***Likely* range of modeled responses to stylized pathways**

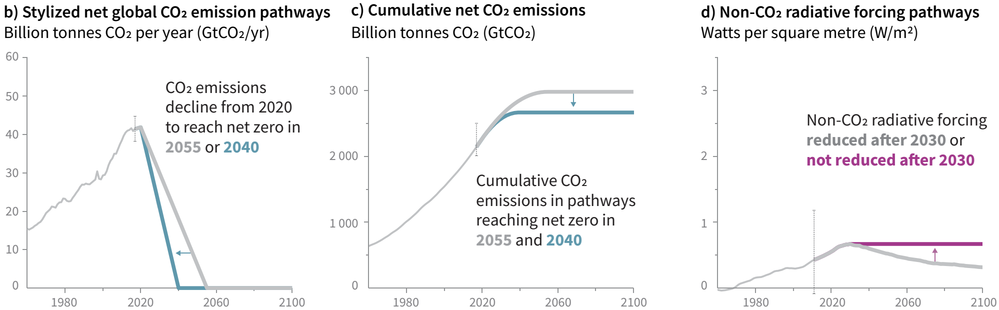
☐ Global CO_2 emissions reach **net zero in 2055** while net non-CO_2 radiative forcing is **reduced after 2030** (grey in **b**, **c** & **d**)

→ ☐ Faster CO_2 reductions (blue in **b** & **c**) result in a **higher probability** of limiting warming to 1.5°C

→ ☐ **No reduction** of net non-CO_2 radiative forcing (purple in **d**) results in a **lower probability** of limiting warming to 1.5°C

Action must be Ubiquitious and Universal

Source: IPCC 2018, Summary for Policymakers, https://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf



b) Stylized net global CO_2 emission pathways
Billion tonnes CO_2 per year ($GtCO_2$/yr)

CO_2 emissions decline from 2020 to reach net zero in **2055** or **2040**

c) Cumulative net CO_2 emissions
Billion tonnes CO_2 ($GtCO_2$)

Cumulative CO_2 emissions in pathways reaching net zero in **2055** and **2040**

d) Non-CO_2 radiative forcing pathways
Watts per square metre (W/m^2)

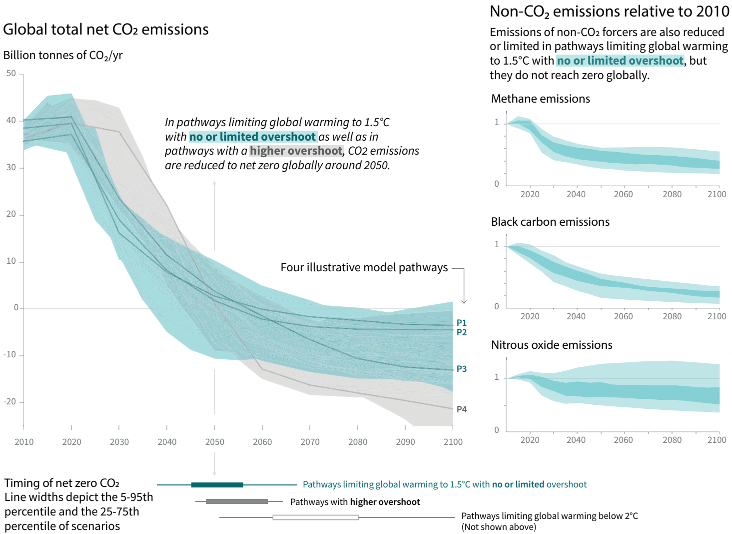
Non-CO_2 radiative forcing **reduced after 2030** or **not reduced after 2030**

Faster immediate CO_2 emission reductions limit cumulative CO_2 emissions shown in panel **(c)**.

Maximum temperature rise is determined by cumulative net CO_2 emissions and net non-CO_2 radiative forcing due to methane, nitrous oxide, aerosols and other anthropogenic forcing agents.

Action must be Ubiquitious and Universal

Source: IPCC 2018, Summary for Policymakers, https://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf



Global total net CO₂ emissions

Non-CO$_2$ emissions relative to 2010

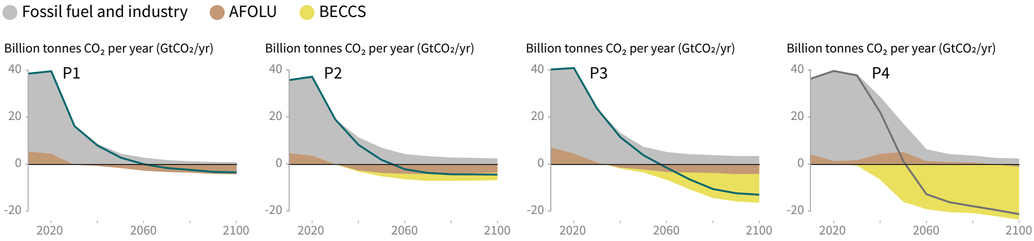
Emissions of non-CO₂ forcers are also reduced or limited in pathways limiting global warming to 1.5°C with **no or limited overshoot**, but they do not reach zero globally.

In pathways limiting global warming to 1.5°C with ***no or limited overshoot*** *as well as in pathways with a* ***higher overshoot****, CO2 emissions are reduced to net zero globally around 2050.*

Timing of net zero CO₂
Line widths depict the 5-95th percentile and the 25-75th percentile of scenarios

Action must be Ubiquitious and Universal

Source: IPCC 2018, Summary for Policymakers, https://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf



P1: A scenario in which social, business and technological innovations result in lower energy demand up to 2050 while living standards rise, especially in the global South. A downsized energy system enables rapid decarbonization of energy supply. Afforestation is the only CDR option considered; neither fossil fuels with CCS nor BECCS are used.

P2: A scenario with a broad focus on sustainability including energy intensity, human development, economic convergence and international cooperation, as well as shifts towards sustainable and healthy consumption patterns, low-carbon technology innovation, and well-managed land systems with limited societal acceptability for BECCS.

P3: A middle-of-the-road scenario in which societal as well as technological development follows historical patterns. Emissions reductions are mainly achieved by changing the way in which energy and products are produced, and to a lesser degree by reductions in demand.

P4: A resource- and energy-intensive scenario in which economic growth and globalization lead to widespread adoption of greenhouse-gas-intensive lifestyles, including high demand for transportation fuels and livestock products. Emissions reductions are mainly achieved through technological means, making strong use of CDR through the deployment of BECCS.

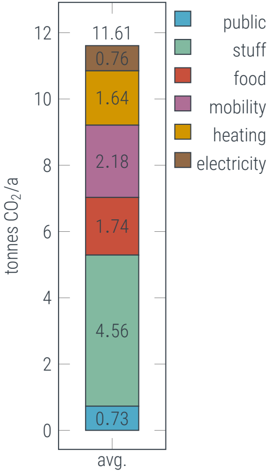
Today:

1. Where is this CO_2 coming from?
2. How much of it is caused by computing, and where?
3. What can we do to reduce the CO_2 emissions caused by computing and AI?
4. What can AI & ML do to mitigate the climate crisis?

What is the status quo?

Annual CO_2 production per person in Germany

source: Umweltbundesamt — `https://uba.co2-rechner.de/`

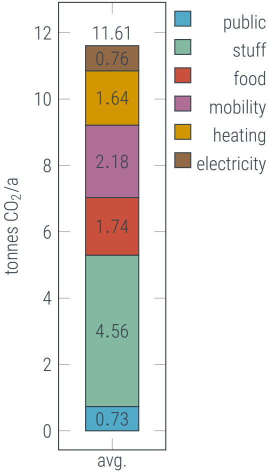


- public expenditure includes infrastructure (incl. waste and water), **education**, social services, defense.
- 4.56 kWh for stuff ≈ 370 €/month (but averaged over household)
- **stuff** is essentially CO_2 produced by others during their work time
- 0.76t CO_2 **electricity** $=$ 1450 kWh/a $=$ 166W
- Current electricity mix in Germany: $\sim$ 500 g CO_2 / kWh
- **heating** assumes a mix of energy sources. Specific heat of water 0.00117kWh/l/°C. Showering at 40°C ·9 l/min = 0.3 kWh/min $\triangleq$ 10 kg CO_2/h
- **mobility:** for intercontinental flights: 0,24t CO_2 / h

What is the status quo?

Annual CO_2 production per person in Germany

source: Umweltbundesamt — `https://uba.co2-rechner.de/`



- food: for the average German, 1.74 t/a. Changing habits would cause changes to
  - vegetarian: 1.29 t/a (26% reduction)
  - vegan: 1.04 t/a (40% reduction)
  - mixed food, but organic/regional/seasonal: 1.45 t/a (17% reduction)
- IPCC goal: zero **net** emissions by 2050. *“Net emissions are defined as anthropogenic emissions reduced by anthropogenic removals.”*
- **On the individual level, this might translate to about 3 t/a.** Some this can and must be achieved by political and societal changes and regulation. But, voluntary or not, lifestyles **will** change.

Today:

1. Where is this CO_2 coming from?
2. How much of it is caused by computing, and where?
3. What can we do to reduce the CO_2 emissions caused by computing and AI?
4. What can AI & ML do to mitigate the climate crisis?

Some Misconceptions

before we even get to the numbers

Computing just needs electricity, which can be produced sustainably.

In the *near* future, virtually *everything* will have to run essentially on electricity. Reduction of existing electricity usage is just as crucial as transition of currently non-electric power consumption.

“Professional” CO_2 production doesn’t count

The CO_2 produced in a professional setting effectively form the customers’ CO_2 footprint under “stuff”.

Corporations are responsible for the majority of emissions.

There is a meaningful debate to be had about most effective regulation (carbon tax). But in the end, CO_2 is produced by humans. The required changes will affect our lifestyles, drastically.

Google runs very efficient data centers

It’s great when technology companies invest in clean electricity.
If they needed *less* energy, those ressources could still be used elsewhere, though.

Numbers in Context I: Running your Devices at Home

sources: apple, nvidia, sysGen, Dell, google, own measurements

At home:

- 30” LED screen: 50W
- 65” LED TV (Samsung LS03): 143W
- Mac mini: 6W (inactive) - 85W (max)
- intel Core i7 CPU $\sim$ 100W (thermal design power)
- nvidia GeForce RTX 2080 Ti: 17W (idle) – 280W (max);
- o2 standard-issue wifi router: 10W
- Fridge/Freezer (Miele KF 37233 iD, A+++): 156 kWh/a = 18W

Mobile:

- Apple USB phone charger, no load: 0.012W.
- Apple iPhone X Battery capacity: 10,35 Wh (typical draw $<$ 1W)

HPC:

- Tesla V100, NVlink: 300W (max)
- GPU hypervisor (8$\times$ Tesla V100, 36-core intel Xeon Skylake, 384 GB RAM, 2 SSD’s): $\sim$ 3kW (max)

Numbers in Context I: Running your Devices at Home

sources: apple, nvidia, sysGen, Dell, google, own measurements

Thus

- forget about 'unplugging your charger'. Actually, just forget about your phone's energy use. Mobile devices are among the most efficient users of electricity out there.
- if you're running a desktop computer as a home server all day ($\sim$ 60W = 525 kWh/a), that may be about 30% of your total electricity consumption and adds about 0.26t CO_2 to your annual footprint. That's roughly the difference between a standard diet and one exclusively on local, organic, seasonal food. If you've got a monitor running on it continuously, double that.
- running your gaming PC at full load (600W, incl. 2 monitors) for 2h a day is similar to the above.
- your wifi router (10W) likely produces about 44kG of CO_2 per year, a similar amount to your fridge!
- if you're training a deep network on imagenet (1 hypervisor, 1 week), that's about 0.25 t CO_2, too.

Computing Devices at home are not the biggest source of your carbon footprint. But especially devices that are **plugged into mains and running for significant times of the day** have a non-negligible impact. Much of it can relatively easily be reduced with some discipline. Switch off your router when you're not home!

Numbers in Context II: Making Your Stuff

Dimensions matter

sources: apple, nvidia, sysGen, Dell, google

embodied CO_2:

- Dell Latitude E6400 Laptop: 200kg CO_2 for manufacturing & transport (May 2010, assumes renewable electricity)
- iPhone 8 (report from September 2017): 56kg CO_2 for manufacturing and transport

Thus

- These numbers are probably too small, because they come from the producers. Nevertheless:
- if you're buying a new laptop every 3 years, the embodied CO_2 might be about 20% of the total emissions associated with the device.
- embodied CO_2 is nontrivial, but likely smaller than that produced during use

Numbers in Context III: The Cloud

Consumer devices seem to use the bulk of electric energy for computing.

sources: Google, Deutsche Telekom, US Dept. of Energy

How much CO_2 does your Internet use generate?

- 2009: “a single [Google] search accounts for about 0.2g of carbon”
- 2017: Google uses about 2.6GW of energy for their operations (all renewable). Note: This includes youtube. (German electricity generation: $\sim$28GW). Google makes about 3% of their revenue in Germany, so about $2.6\text{ GW} \cdot 0.03/80\,000\,000 \approx 1\text{W}$ for each German.
- Deutsche Telekom uses about 142kWh / TB in Germany.
- all US Datacentres jointly draw about 8GW of power (source: US Dept. of Energy), using 1,6% of all US electricity (the US produces 480GW electricity on average).

- if you’re using 10GB/month, that’s $\sim$ 17kWh, i.e. 8.5kg CO_2 (kg!) per year for communication.
- If a third of your data comes from Google (youtube!), then producing that data in the cloud probably uses about 26kWh of energy (3W) — less than your wifi router!
- the main energy consumption on the internet is probably not communication but computation. Even that, though, probably amounts to only a few kG CO_2 per person and year.

Takeaways:

- your *personal* CO_2 footprint from computing likely stems primarily from
  - desktop computing
  - embodied CO_2 in devices

  and not so much from data and communication
- but if you are a *professional* in charge of significant computing power, then computing efficiency may be one of the most significant ways *you* can reduce CO_2 emission.

At Google Deepmind, each Developer has personal access to about 8 GPUs. If you're in control of a $8\times$ V100 hypervisor and keep it busy ($3\text{kW} = 13.14$ t CO_2 / a), then thinking hard about how you train your neural networks might be your biggest opportunity to reduce CO_2 emissions.

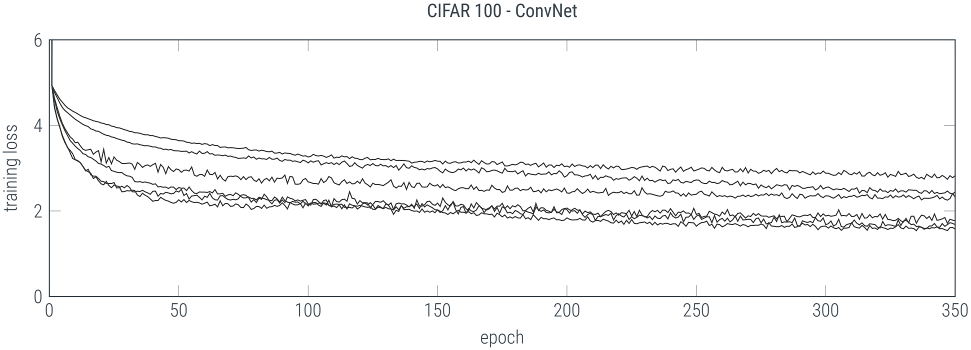
Today:

1. Where is this CO_2 coming from?
2. How much of it is caused by computing, and where?
3. What can we do to reduce the CO_2 emissions caused by computing and AI?
4. What can AI & ML do to mitigate the climate crisis?

Machine Learning is the Most Energy-Inefficient Kind of Computing

a typical workflow

[data from Schneider, Balles, Hennig, ICLR 2019]



In contrast to more established CS areas (like information retrieval, networks, compression, os), machine learning, data analysis and AI are resource inefficient, and the market is currently willing to allow such wastefulness. An individual developer can make a significant difference.

The Dirty Secret of Academic Research

Conference Emissions

- The typical Machine Learning PhD student flies to about one international conference per year.
  - AISTATS 2019: Okinawa, Japan
  - ICLR 2019: New Orleans, USA
  - NeurIPS 2019: Vancouver, Canada
  - ICML 2019: Long Beach, California, USA
- a **single** intercontinental return flight FRA to LAX produces 5,73 t CO_2. That's like running that hypervisor for 160 days! The biggest climate cost of a NeurIPS paper is not the computing, but flying in to present it.

Today:

1. Where is this CO_2 coming from?
2. How much of it is caused by computing, and where?
3. What can we do to reduce the CO_2 emissions caused by computing and AI?
4. What can AI & ML do to mitigate the climate crisis?

Contributing Our Special Skills

`https://arxiv.org/abs/1906.05433v1` (published 10 June 2019)

Tackling Climate Change with Machine Learning

David Rolnick[1,*], Priya L. Donti[2], Lynn H. Kaack[3], Kelly Kochanski[4], Alexandre Lacoste[5], Kris Sankaran[6,7], Andrew Slavin Ross[8], Nikola Milojevic-Dupont[9,10], Natasha Jaques[11], Anna Waldman-Brown[11], Alexandra Luccioni[6,7], Tegan Maharaj[6,7], Evan D. Sherwin[2], S. Karthik Mukkavilli[6,7], Konrad P. Kording[1], Carla Gomes[12], Andrew Y. Ng[13], Demis Hassabis[14], John C. Platt[15], Felix Creutzig[9,10], Jennifer Chayes[16], Yoshua Bengio[6,7]

[1]University of Pennsylvania, [2]Carnegie Mellon University, [3]ETH Zürich, [4]University of Colorado Boulder, [5]Element AI, [6]Mila, [7]Université de Montréal, [8]Harvard University, [9]Mercator Research Institute on Global Commons and Climate Change, [10]Technische Universität Berlin, [11]Massachusetts Institute of Technology, [12]Cornell University, [13]Stanford University, [14]DeepMind, [15]Google AI, [16]Microsoft Research

Just a few Picks

many more ideas in the op.cit.

https://arxiv.org/abs/1906.05433v1

- **Mitigation**, e.g.
  - generation and demand forecasting and control in smart grids
  - enabling control of nuclear fusion
- **Transportation** and **Smart Cities**, e.g.
  - freight routing and consolidation
  - improving shared and low-carbon options (bike sharing, electric scooters, public transport …)
  - smart buildings (automatic shading, heating, appliances)
- **Industry, Farming, Forestry**, e.g.
  - efficient supply chain management
  - lightweight construction
  - remote sensing of emissions
  - automated and smart afforestation and precision agriculture
- **Enabling Science, Society and Individuals**, e.g.
  - improving climate forecasting and biodiversity/ecosystem monitoring and modelling
  - societal modelling for food security, migration, crises, disaster relief
  - providig tools for individual and societal action

Caveats

let's not get ahead of ourselves

- This paper was written by people who don't know ML, and people who don't know the science
- Some of the ideas are very aspirational, high-risk, long-term (and the authos say so)
- ML can not solve these problems alone. In most cases, it is a supporting tool for scientific, technological and societal advances. Computer scientists need to listen.
- The solutions have to be **deployed**, too. We don't just need scientific advances, but people willing to turn climate relief into a business opportunity.

- computing is a significant, but not the dominant consumer of energy
- a large part of computing consumption happens at home, thus individual action matters
- but CS professionals also have significant leverage to affect resource efficiency through careful software design
- AI and ML have the potential to help enable technological and societal change to mitigate the climate crisis — in careful support of the corresponding core communities.

If you are passionate about a particular use of ML to mitigate climate change,
feel invited to propose your own **Masters thesis** topic, apply for a **PhD position**, or **startup seed funding**!

Incidentally: There's an open PhD position in CO_2 soil-transport modeling available (with Thomas Scholten, soil science) in my lab!