OPEN POSSIBILITIES.

Understanding and Optimizing the Environmental Footprint of Computing

Carole-Jean Wu & Udit Gupta
Meta AI

Meta (formerly Facebook), Arizona State University, Harvard University: Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin S. Lee, David Brooks, Gu-Yeon Wei, Bilge Acun, Kim Hazelwood

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Computing Industry is Growing

Growing # of devices and data center capacity

Emerging applications demanding more compute resources

Further efficiency improvements challenging

Global AI Market Growth
29.86 billions (2020) to 299.64 billions (2026)

https://www.fnfresearch.com/artificial-intelligence-market

Growing Importance in the Research Community

Architecting a Sustainable Planet

Srilatha (Bobbie) Manne
Principal Hardware Engineer
Microsoft

DTCO including Sustainability: Power-Performance-Area-Cost-Environmental score (PPACE) Analysis for Logic Technologies

M. Garcia Bardon[1], P. Wuytens[1], L.-Å. Ragnarsson[1], G. Mirabelli[1], D. Jang[1], G. Willems[1], A. Mallik[1], A. Spessot[1], J. Ryckaert[1], B. Parvais[1,2]
[1] imec, Kapeldreef 75, 3001 Leuven, Belgium; [2] Vrije Universiteit Brussel (VUB), Brussels, Belgium
marie.garciabardon@imec.be

Climate Change AI

Zero Carbon Cloud
Sustainable Computing
University of Chicago

Green AI

Roy Schwartz*◊ Jesse Dodge*◊♣ Noah A. Smith◊♡ Oren Etzioni◊
◊Allen Institute for AI, Seattle, Washington, USA
♣ Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
♡ University of Washington, Seattle, Washington, USA

July 2019

Abstract

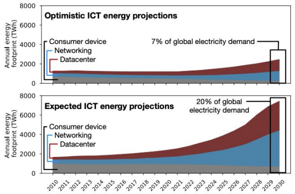
The computations required for deep learning research have been doubling every few months, resulting in an estimated 300,000x increase from 2012 to 2018 [2]. These computations have a surprisingly large carbon footprint [40]. Ironically, deep learning was inspired by the human brain, which is remarkably energy efficient. Moreover, the financial cost of the computations can make it difficult for academics, students, and researchers, in particular those from emerging economies, to engage in deep learning research.

This position paper advocates a practical solution by making **efficiency** an evaluation criterion for research alongside accuracy and related measures. In addition, we propose reporting the financial cost or "price tag" of developing, training, and running models to provide baselines for the investigation of increasingly efficient methods. Our goal is to make AI both greener and more inclusive—enabling any inspired undergraduate with a laptop to write high-quality research papers. Green AI is an emerging focus at the Allen Institute for AI.

Chasing Carbon: The Elusive Environmental Footprint of Computing

Udit Gupta[1,2], Young Geun Kim[3], Sylvia Lee[2], Jordan Tse[2], Hsien-Hsin S. Lee[2], Gu-Yeon Wei[1], David Brooks[1], Carole-Jean Wu[2]
[1]Harvard University, [2]Facebook Inc., [3]Arizona State University
ugupta@g.harvard.edu carolejeanwu@fb.com

***Abstract*—Given recent algorithm, software, and hardware innovation, computing has enabled a plethora of new applications. As computing becomes increasingly ubiquitous, however, so does its environmental impact. This paper brings the issue to the attention of computer-systems researchers. Our analysis, built on industry-reported characterization, quantifies the environmental effects of computing in terms of carbon emissions. Broadly, carbon emissions have two sources: operational energy consumption, and hardware manufacturing and infrastructure. Although carbon emissions from the former are decreasing thanks to algorithmic, software, and hardware innovations that boost performance and power efficiency, the overall carbon footprint of computer systems continues to grow. This work quantifies the carbon output of computer systems to show that most emissions related to modern mobile and data-center equipment come from hardware manufacturing and infrastructure. We therefore outline future directions for minimizing the environmental impact**



Energy and Policy Considerations for Deep Learning in NLP

Emma Strubell Ananya Ganesh Andrew McCallum
College of Information and Computer Sciences
University of Massachusetts Amherst
{strubell, aganesh, mccallum}@cs.umass.edu

Abstract

Recent progress in hardware and methodology for training neural networks has ushered in a new generation of large networks trained on abundant data. These models have obtained notable gains in accuracy across many NLP tasks. However, these accuracy improvements depend on the availability of exceptionally large computational resources that necessitate similarly substantial energy consump-

Consumption	**CO_2e (lbs)**
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192

Computing's Environmental Footprint

Applications

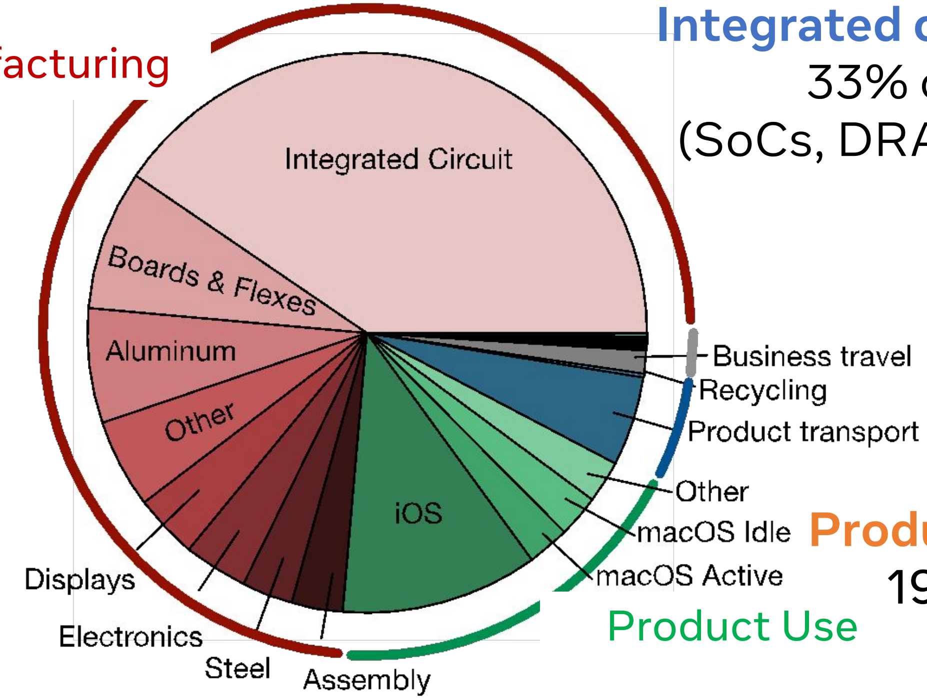
Hardware Manufacturing

Embodied vs. **Operational** CO_2

Chasing Carbon: The Elusive Environmental Footprint of Computing. Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin Lee, Gu-Yeon Wei, David Brooks, Carole-Jean Wu. In Proceedings of the International Symposium on HPCA. 2021

Manufacturing Dominates the Environmental Footprint of Client Computing

Manufacturing accounts for 74% of Apple's end-to-end breakdown in 2019

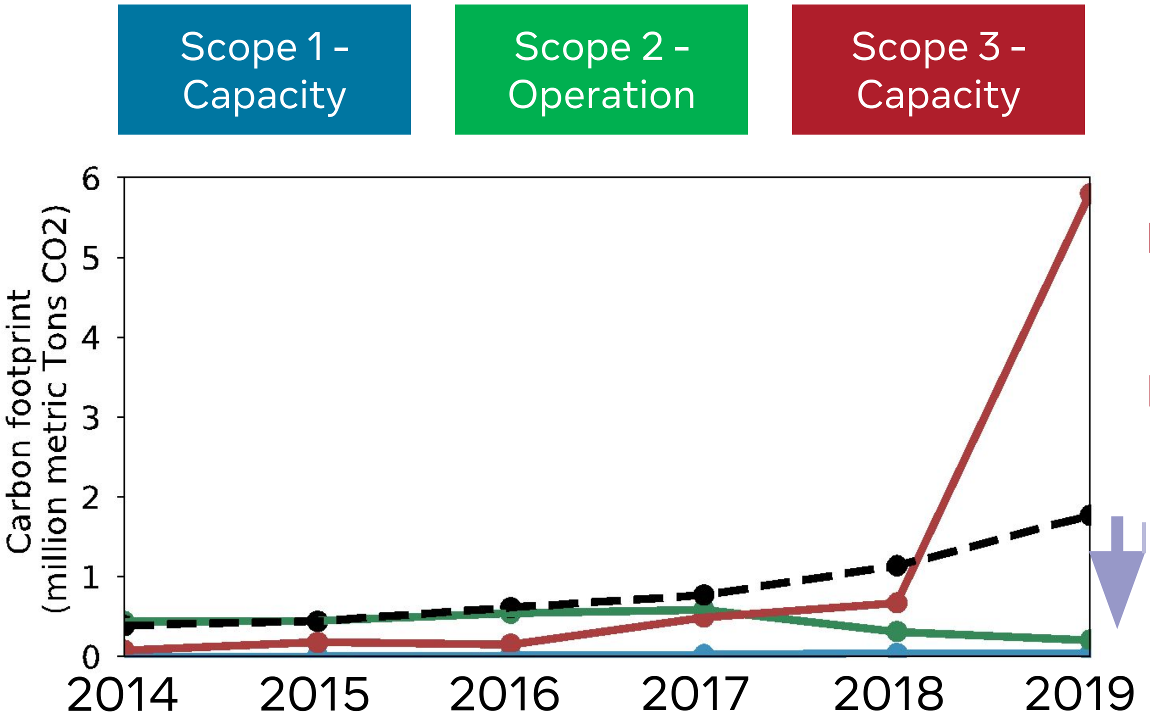


Integrated circuits account for 33% of emissions (SoCs, DRAMs, NAND Flash)

Product use account for 19% of emissions

Manufacturing → Product Transport → Product Use → Recycling

Historical analysis of Facebook's carbon footprint



Scope 3 dominates Facebook's carbon emissions.

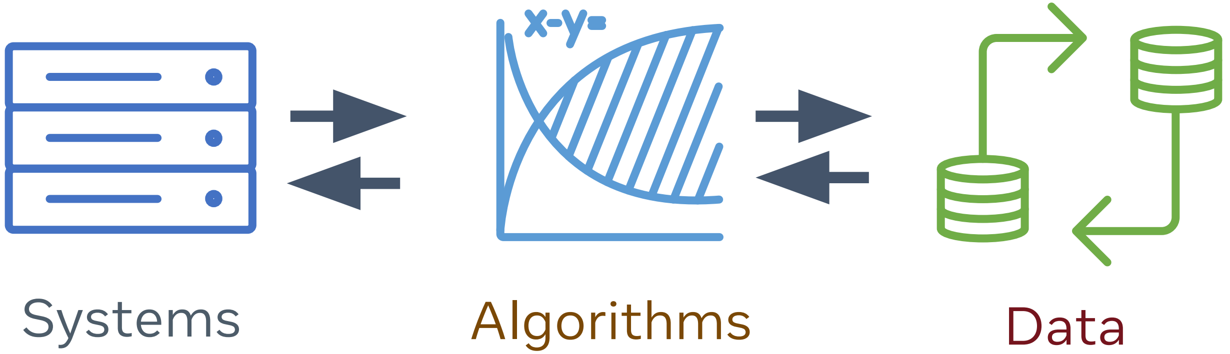
49% of Scope 3 comes from hardware, infrastructure, data center construction

Impact of renewable energy in data centers

OPEN POSSIBILITIES.

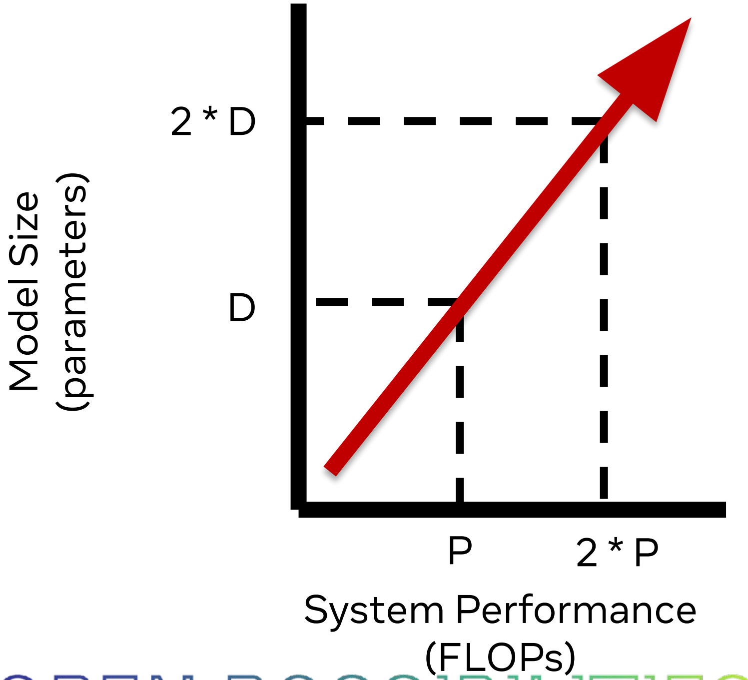
OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Efficiency optimization is a must



But, efficiency alone is not enough!

Benefits of higher efficiency overshadowed by higher application demands



Size of state-of-the-art NLP models has grown by 3 orders of magnitude in 2 years [1]

Size of Facebook's production recommendations models has grown by an order of magnitude in 3 years [2]

[1] "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" E. Bender et. al. 2021.
[2] "Understanding Capacity-Driven Scale-Out Neural Recommendation Inference" M. Lui et. al. ISPASS-2021.

Environmentally-sustainable computing infrastructures

Metrics and Accounting

Optimization at Scale

CO_2 Footprint Amortization

Metrics and Accounting

Carbon cost modeling

End-to-end product life cycle analyses

Component-level carbon costs

Inputs
- Workload
- HW design
- Renewable energy
- Fab characteristics
- Recycling
- HW Lifetime

↓

Carbon model

↓

Carbon emissions

Open-source models and tools

Metrics and Accounting

- Carbon cost modeling
- Carbon cost accounting/reporting

OPEN Compute Project

MLPerf

TOWARDS THE SYSTEMATIC REPORTING OF THE ENERGY AND CARBON FOOTPRINTS OF MACHINE LEARNING

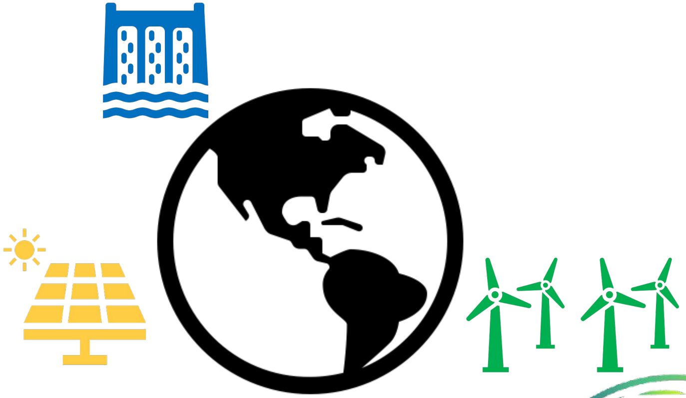
A WORKING PAPER

Peter Henderson[†], **Jieru Hu**[‡], **Joshua Romoff**[◇]
Emma Brunskill[†], **Dan Jurafsky**[†], **Joelle Pineau**[‡,◇]
[†]Stanford University, [‡]Facebook, [◇]Mila, McGill University

Optimization at Scale

Datacenter scale load shaping

Cloud-edge computation scheduling



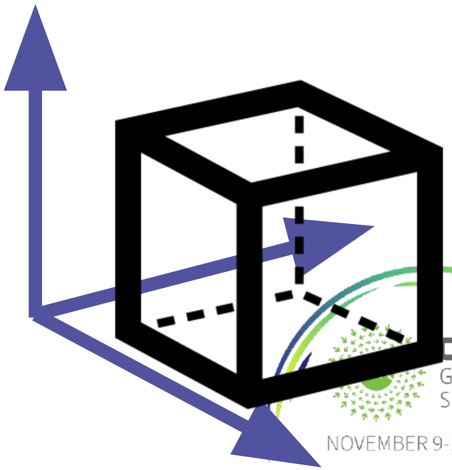
Carbon Footprint Amortization

Utilization ↑

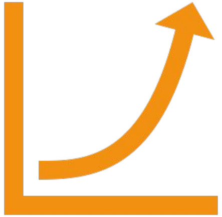
- Virtualization, multi-tenancy, workload consolidation

Hardware Lifetime ↑

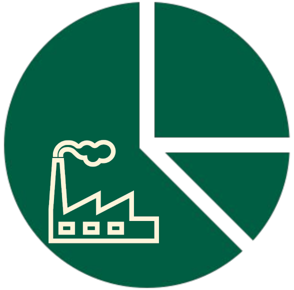
- Modular infrastructure design
- Disaggregated infrastructures
- Resilient hardware/software systems



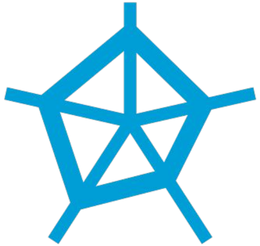
Call to Action



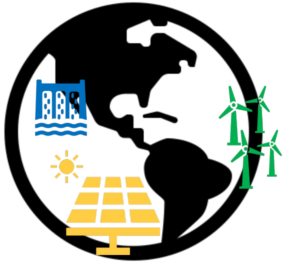
Growing Compute Demands



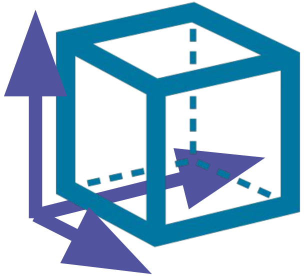
Embodied vs. Operational CO2



Metrics and Accounting



Optimization at Scale



CO2 Footprint Amortization

OPEN POSSIBILITIES.

Understanding and Optimizing the Environmental Footprint of Computing

Carole-Jean Wu & Udit Gupta
Meta AI

Chasing Carbon: The Elusive Environmental Footprint of Computing
Intl. Symp. on High Performance Computer Architecture, 2021.

Meta (formerly Facebook), Arizona State University, Harvard University: Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin S. Lee, David Brooks, Gu-Yeon Wei, Bilge Acun, Kim Hazelwood

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021