Sustainable Computer Systems

Lieven Eeckhout

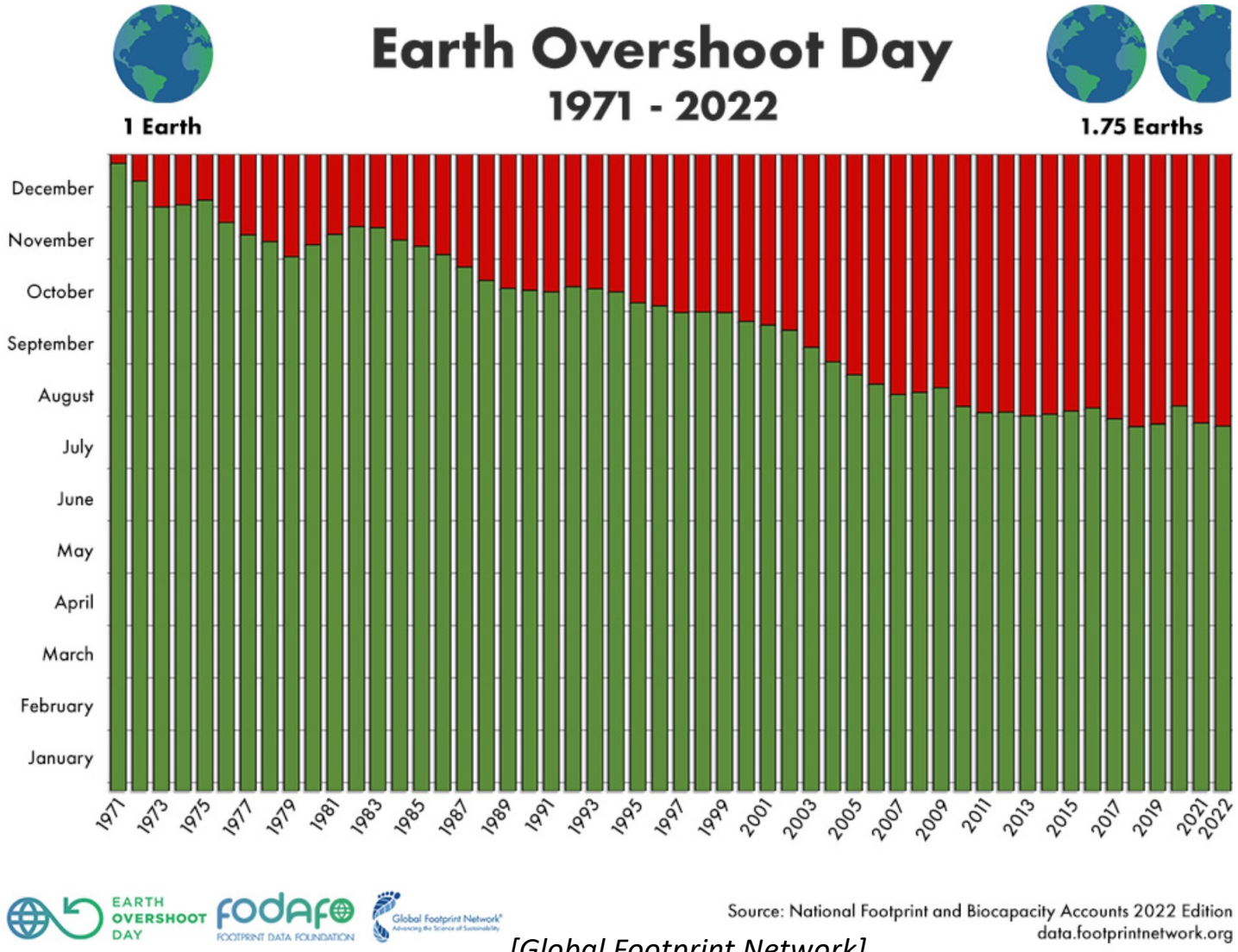
Ghent University, Belgium

IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP) – Porto, Portugal – July 21, 2023

“*Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs.*”

[The Brundtland Report of the World Council on Economic Development, 1987]

How Are We Doing?



[Global Footprint Network]

Agenda: Key Questions

Sustainable computing: What are the challenges?

How does the environmental impact of ICT scale? What are the contributing factors?

What can we, as computer architects, do to reduce the environmental impact of microprocessor chips?

Challenges when Doing Research in Computer System Sustainability

1. ***Sustainability is multi-faceted problem***
   - ***Global warming, raw materials, e-waste, water consumption, biodiversity, etc.***

2. ***Inherent data uncertainty***
   - ***Many unknowns, data limitations, industry secrecy***

3. ***Need to account for entire lifecycle***
   - ***Embodied footprint: raw material extraction, manufacturing, end-of-life recycling***
   - ***Operational footprint: usage of device during lifetime***

GHG Emissions lead to Global Warming

bush fire

drought

biodiversity loss

flooding

hurricanes

Contribution of ICT to global greenhouse gas (GHG) emissions estimated to be around 2.1–3.9%, and rising

…on par with aviation industry…

[Freitag et al., 2020]

Sustainability is a Multi-Faceted Challenge

Sustainability is much more than combating global warming

It is also about

- **Raw material extraction**

  World Bank projects that demand for metals and minerals will increase rapidly with climate ambition

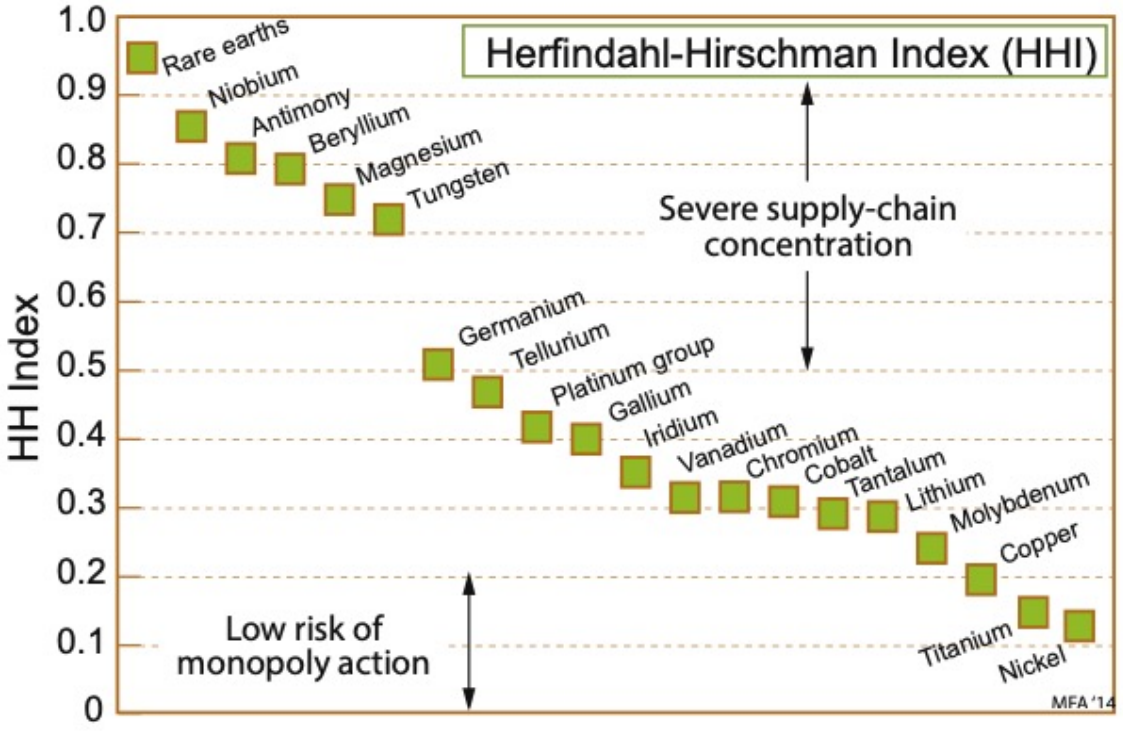
  - Electric storage batteries: 10x more metals (aluminum, cobalt, iron, lead, lithium, manganese and nickel) needed by 2050 under a 2℃ scenario

  Under EU's climate-neutrality scenarios for 2050, the EU needs

  - 18x more lithium in 2030, and almost 60x more in 2050
  - 5x more cobalt in 2030, and almost 15x more in 2050
  - 10x more **Rare Earth Elements (REEs)** in 2050
    - REEs for permanent magnets: Dysprosium, Neodymium, Praseodymium, Samarium; The remaining rare earths are Yttrium, Lanthanum, Cerium, Promethium, Europium, Gadolinium, Terbium, Holmium, Erbium, Thulium, Ytterbium, Lutetium

[World Bank (2017): The Growing Role of Minerals and Metals for a Low Carbon Future]
[European Commission 2020: Critical Raw Materials Resilience: Charting a Path towards greater Security and Sustainability]

Supply Chain Risk



Herfindahl-Hirschman Index (HHI):

$$HHI = \sum_{i=1}^{n} f_i^2$$

f_i is fraction of market sourced by nation *i*, and *n* is total number of source-nations.

One nation is monopoly: HHI = 1

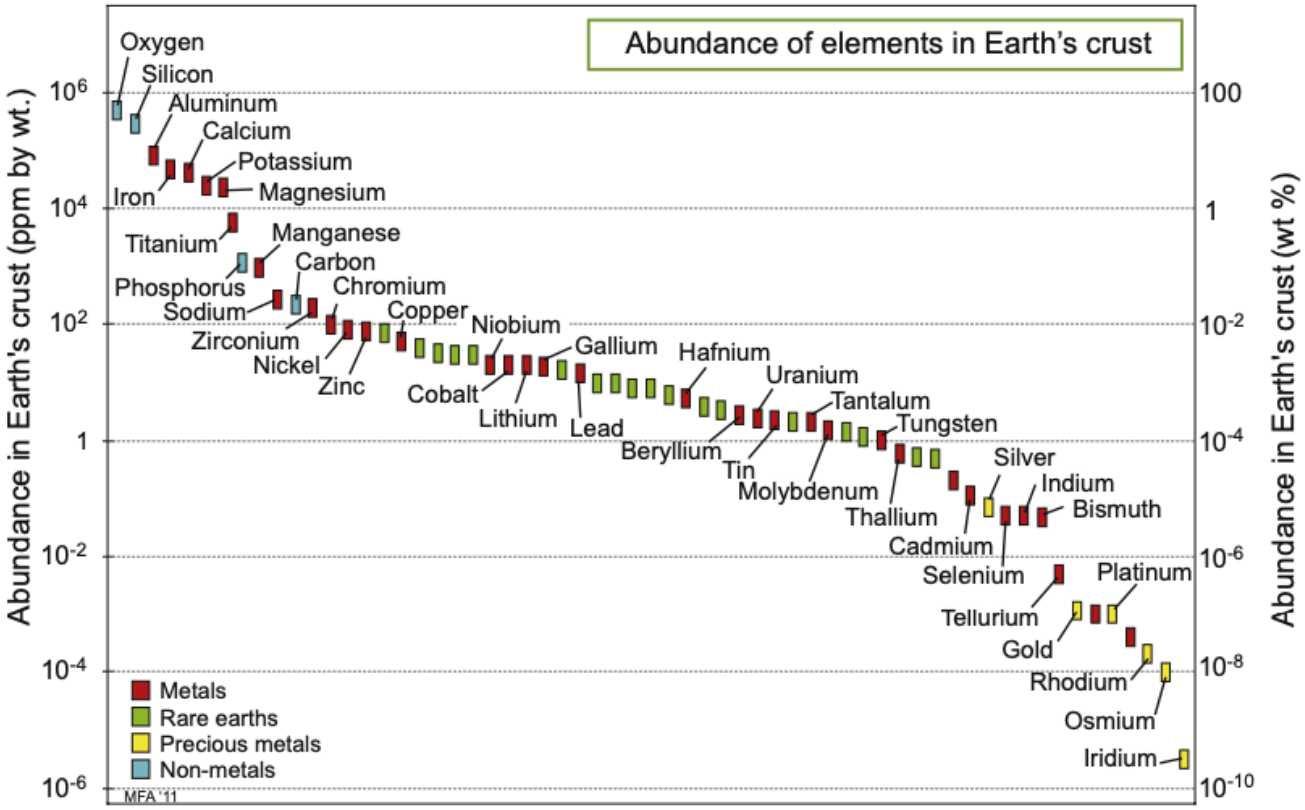
Two nations with equal share: HHI = $0.5^2 + 0.5^2 = 0.5$

Many source-nations: HHI → 0

Esp. problematic if HHI is high and materials come from politically unstable region(s)

[M. F. Ashby, Materials and Sustainable Development, 2016]

Some Materials are Rare



[M. F. Ashby, Materials and Sustainable Development, 2016]

How much of everything have we got?

Enormous range: some elements are abundant, others are rare

Mining rare elements can become extremely expensive and challenging

Raw Material Mining

- Energy/carbon-intensive industry
- Has significant impact on the environment

For example: copper (Cu)

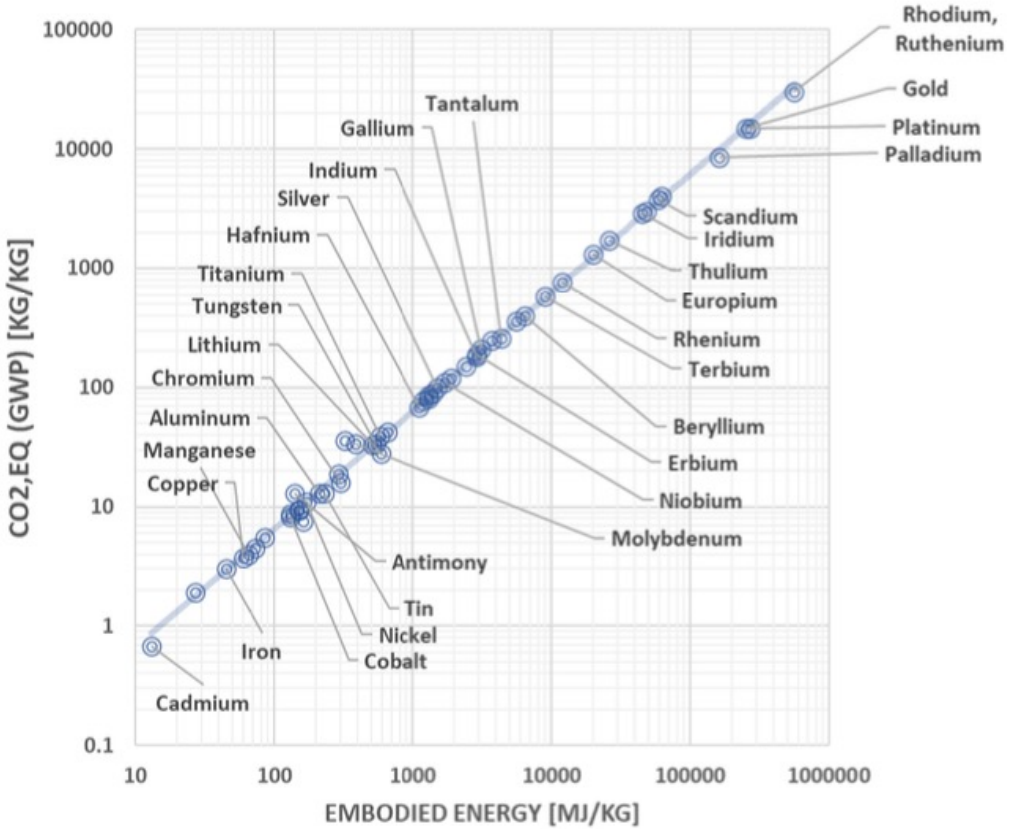
~50 MJ energy for 1 kg of Cu

~4 kg of CO_2 for 1 kg of Cu

For example: gold (Au)

~200 BJ energy for 1kg of Au

~15 tons of CO_2 for 1kg of Au



[M. Ashby (2016): Materials and Sustainable Development]

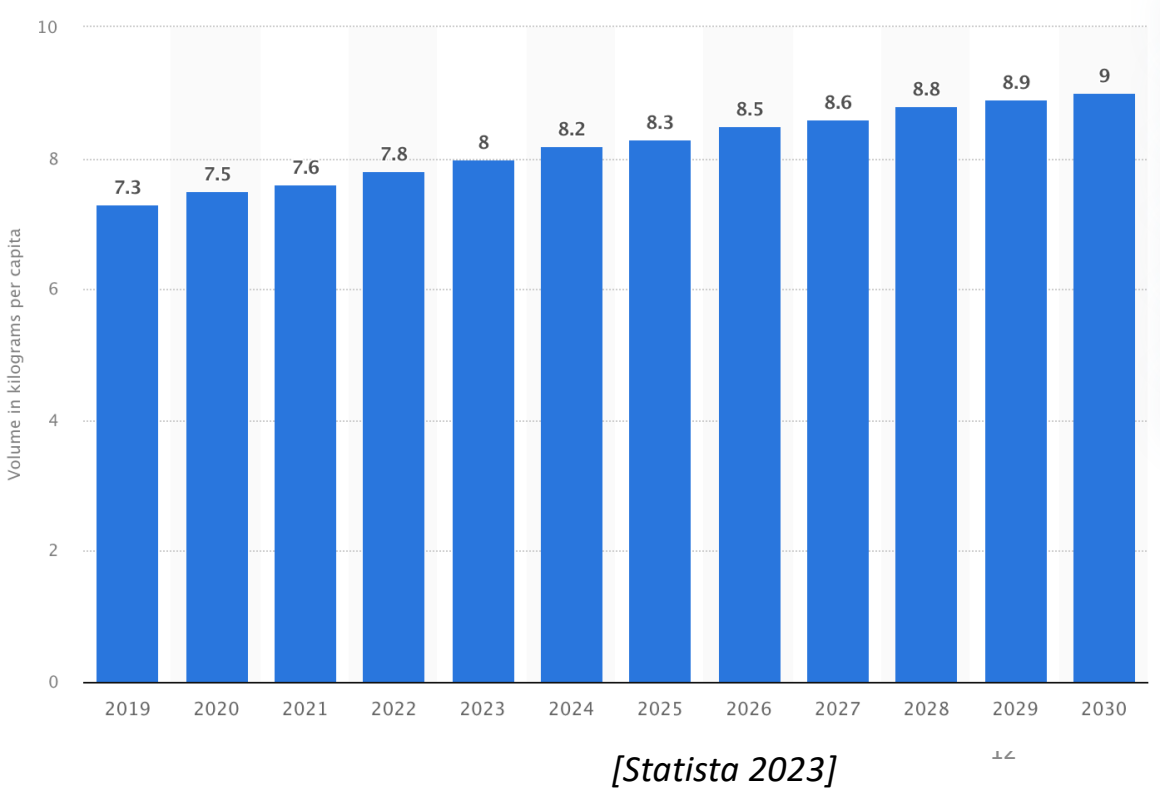
Sustainability is a Multi-Faceted Challenge

Sustainability is much more than combating global warming

It is also about

- Raw material extraction
- **E-waste**

- due to linear economy
- 8 kg per capita per annum
  - this includes small to large appliances
- only 17% gets recycled



[Statista 2023]

Sustainability is a Multi-Faceted Challenge

Sustainability is much more than combating global warming

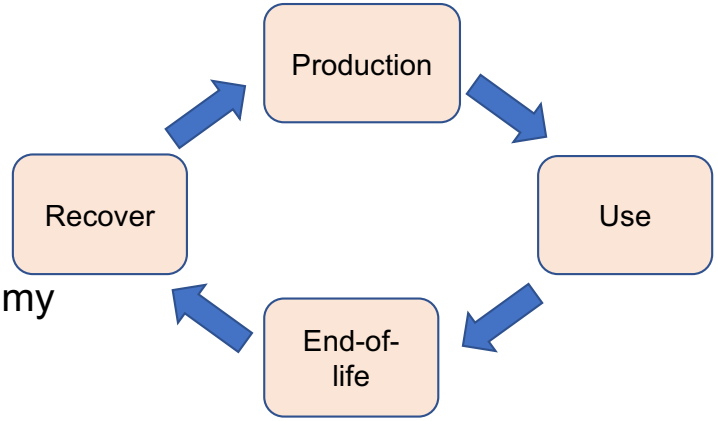
It is also about

- **Raw material extraction**
- **E-waste**
- **New business models**

Key motivation for circular (rather than linear) economy

Keep materials in the economy longer

- Fewer raw materials are needed
- Less impact on climate
- Avoid (e-)waste
- Improved security of material supply
  - Be less depending on third-party countries
- Design for repairability



Selling services instead of goods

Consumer wants (societal needs)

Light, not lamps

Mobility, not cars

Connectivity, not smartphone

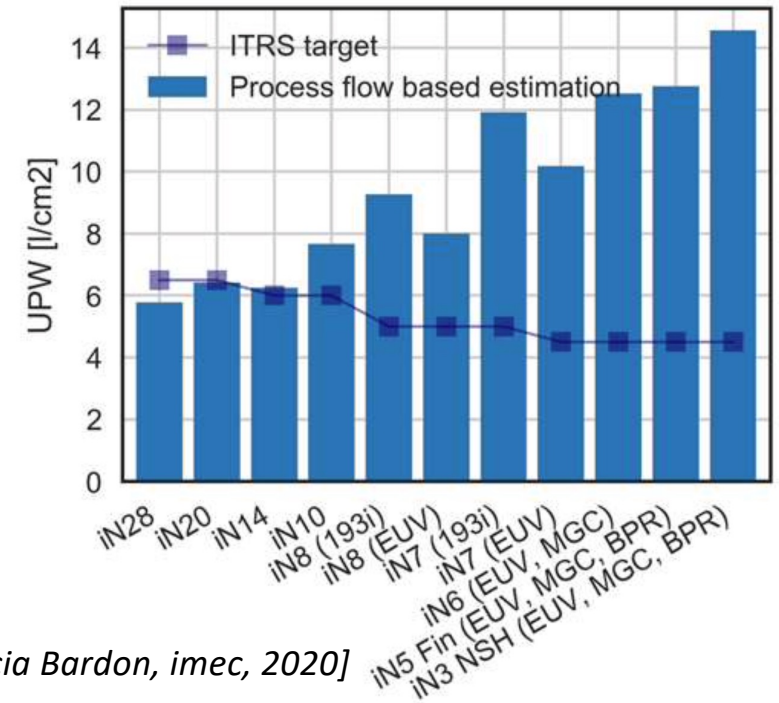
Sustainability is a Multi-Faceted Challenge

Sustainability is much more than combating global warming

It is also about

- Raw material extraction
- E-waste
- New business models
- **Ultra pure water**
- **Biodiversity**
- **Etc.**

According to NPR [April 2023]:
"*Epic drought in Taiwan pits farmers against high-tech factories for water*"



[M. Garcia Bardon, imec, 2020]

Challenges when Doing Research in Computer System Sustainability

1. ***Sustainability is multi-faceted problem***
   - ***Global warming, raw materials, e-waste, water consumption, biodiversity, etc.***

2. ***Inherent data uncertainty***
   - ***Many unknowns, data limitations, industry secrecy***

3. ***Need to account for entire lifecycle***
   - ***Embodied footprint: raw material extraction, manufacturing, end-of-life recycling***
   - ***Operational footprint: usage of device during lifetime***

Life Cycle Assessment (LCA) iPhone 12

Global Warming Potential (GWP 100 years) in CO_2

Carbon footprint: Estimated emissions are calculated in accordance with guidelines and requirements as specified by ISO 14040 and ISO 14044. There is inherent uncertainty in modeling carbon emissions due primarily to data limitations. For the top component contributors to Apple's carbon emissions, Apple addresses this uncertainty by developing detailed process-based environmental models with Apple-specific parameters. For the remaining elements of Apple's carbon footprint, we rely on industry average data and assumptions. Calculation includes emissions for the following life cycle phases contributing to Global Warming Potential (GWP 100 years) in CO_2 equivalency factors (CO_2e):

- **Production:** Includes the extraction, production, and transportation of raw materials, as well as the manufacture, transport, and assembly of all parts and product packaging.
- **Transport:** Includes air and sea transportation of the finished product and its associated packaging from manufacturing site to regional distribution hubs. Transport of products from distribution hubs to end customers is modeled using average distances based on regional geography.

customers is modeled using average distances based on regional geography.

- **Use:** Apple assumes a three- or four-year period for power use by first owners based on the product type. Product use scenarios are based on historical customer use data for similar products. Geographic differences in the power grid mix have been accounted for at a regional level.
- **End-of-life processing:** Includes transportation from collection hubs to recycling centers and the energy used in mechanical separation and shredding of parts. For more information on the carbon footprint, visit apple.com/environment/answers.

Recent Work: ACT Model

1 while **ACT node 2** is representative of the actual hardware node of the existing devices.

We summarize the latest efforts on carbon footprint analysis for server systems and for consumer electronics using state-of-the-art LCA and the proposed ACT framework. It is evident that the gap between the LCA-based carbon cost estimates and ACT's is non-negligible. We hypothesize the
of up-to-date carbon emission d
ory, and storage technologies. T

available carbon characterization for SoC's, memories, and storage devices from industry. The modeling parameters can be configured for varying use cases to study the carbon footprint of different systems and hardware. In addition to taking a crucial first step in providing architectural carbon models, we hope ACT encourages industry to publish more detailed carbon characterizations to standardize carbon footprint accounting. In particular, we imagine additional details on energy, gas, raw material consumption, and packaging overheads for a wider array of processing technologies will extend ACT.

[Gupta et al., ISCA 2022]

Challenges when Doing Research in Computer System Sustainability

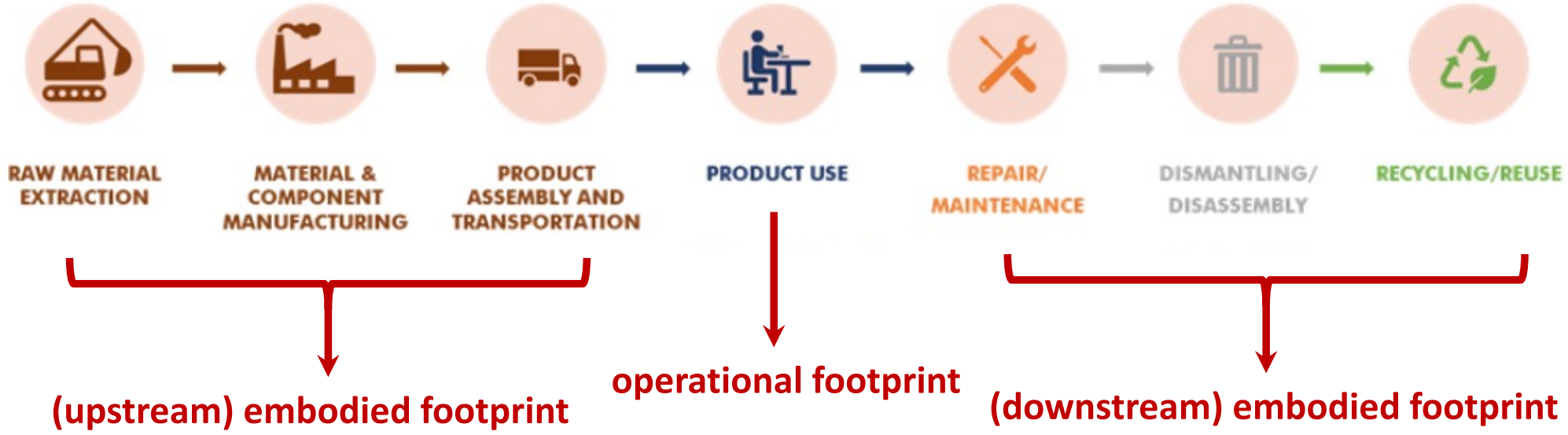
1. ***Sustainability is multi-faceted problem***
    - ***A variety of metrics: GHG (CO2eq) emissions, raw material use, biodiversity, supply chain risk management, etc.***

2. ***Inherent data uncertainty***
    - ***Many unknowns, industry secrecy***

3. ***Need to account for entire lifecycle***
    - ***Embodied footprint: raw material extraction, semiconductor manufacturing***
    - ***Operational footprint: usage of device during lifetime***

A Day in the Life of a Computer Device

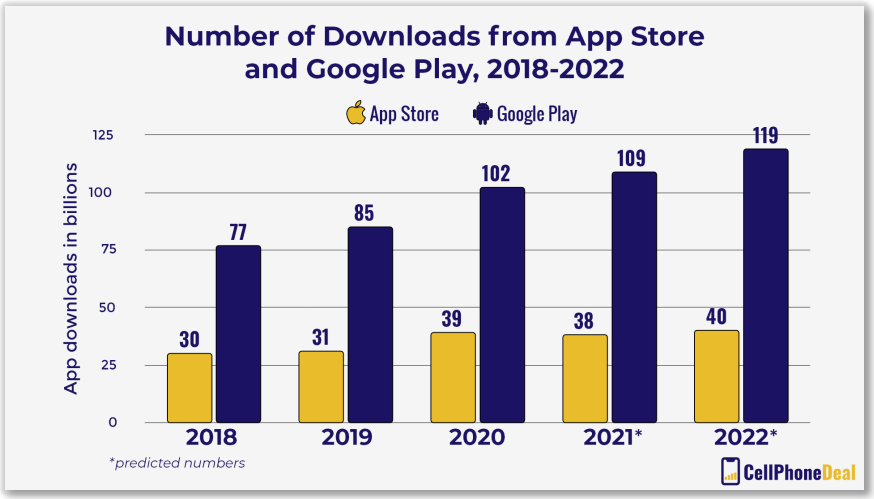
How to weigh embodied vs operational footprint?



[Global Economic Council, 2021: State of Sustainability Research -- Climate Change Mitigation]

Embodied vs Operational Footprint

… but how predictive is historical data?



on regional geography.

- **Use:** Apple assumes a three- or four-year period for power use by first owners based on the product type. Product use scenarios are based on historical customer use data for similar products. Geographic differences in the power grid mix have been accounted for at a regional level.

visit apple.com/environment/answers.

The Truth About Smartphone Addiction, And How To Beat It

[Forbes, 2019]

Do Not Disturb: How I Ditched My Phone and Unbroke My Brain

[New York Times, 2019]

Embodied vs Operational Footprint

How predictive is historical data?

Rebound effect of making systems more efficient

a.k.a Jevons' paradox

more efficient system → cheaper/easier to use → increased usage and deployment → increased (embodied and operational) footprint

William Stanley Jevons (1865) first describes this rebound effect

- James Watt improved the efficiency of coal-fired steam engine
  - Each steam engine uses less coal, so coal became a more cost-effective fuel
- This led to an increased use of steam engines in a variety of industries
- The result was increased overall coal consumption

Part II

How does the global environmental footprint of computing scale?

What are the contributing factors?

[L. Eeckhout, "Kaya for Architects: Towards Sustainable Computer Systems", IEEE Micro, 2023]

Kaya Identity

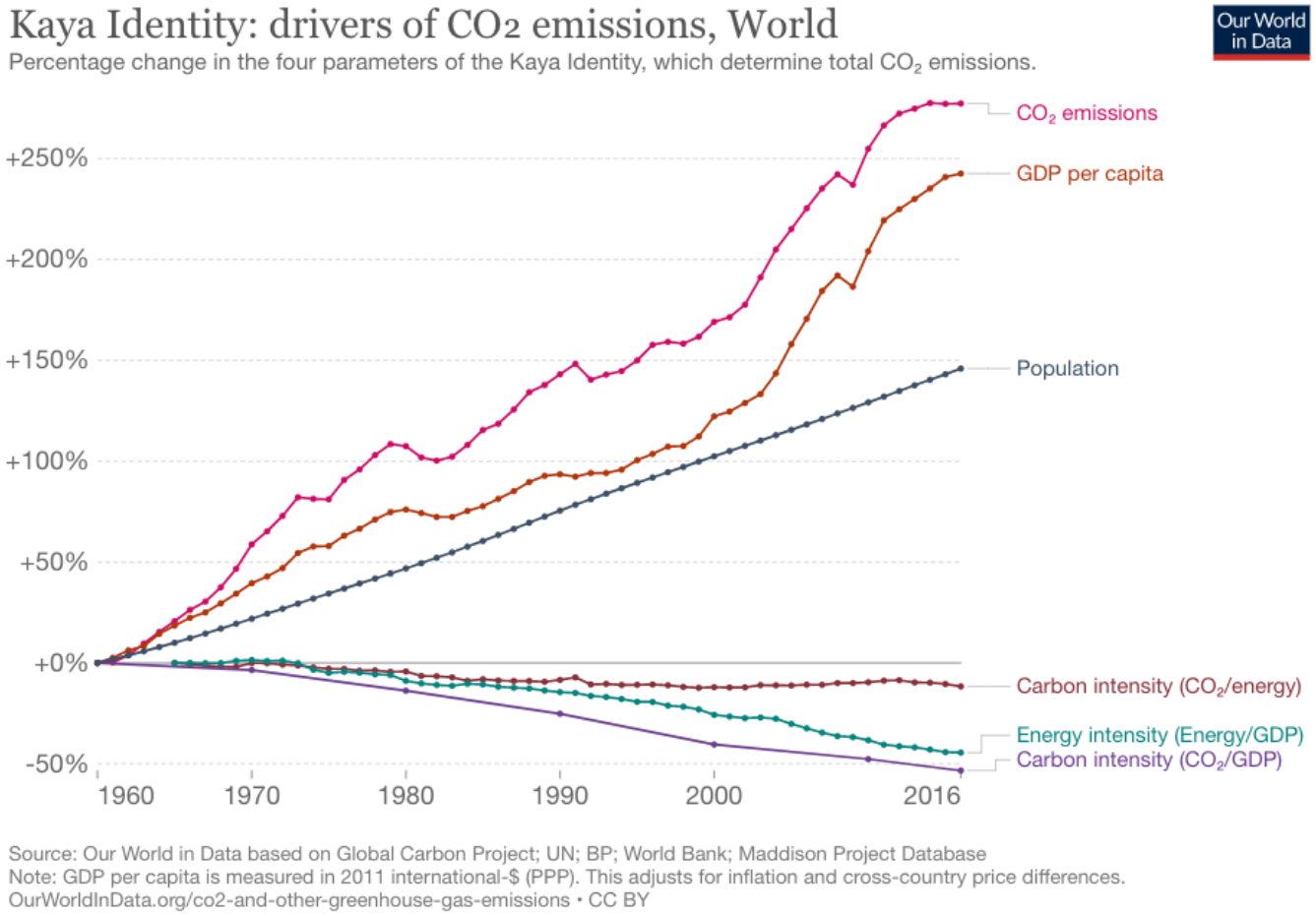
Contributing factors to carbon emissions *[by energy economist Yoichi Kaya, 1997]*

$F = P \times G/P \times E/G \times F/E$, with

- F = global CO_2 emissions
- P = global population
- G/P = GDP per capita
- E/G = energy intensity of the GDP
- F/E = carbon footprint of energy

Kaya identity is used by the Intergovernmental Panel on Climate Change (IPCC) to predict world CO_2 emission scenarios and impact on global warming

Kaya



Despite improvements in energy intensity and carbon intensity, we witness an overall increase in CO_2 emissions

How to Account for Rebound Effect in Kaya

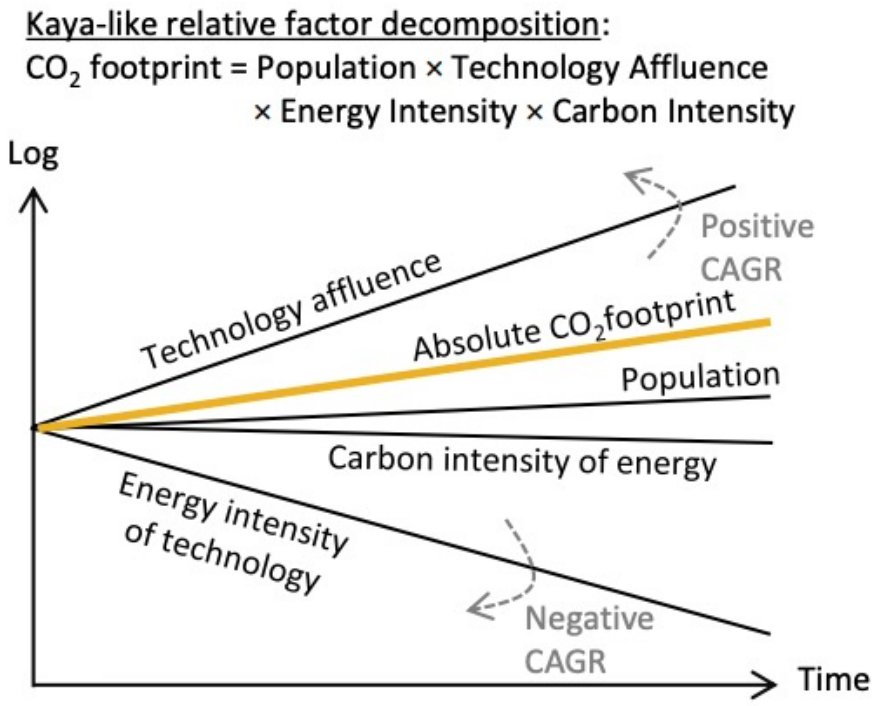
Look at trends – express each of the variables as CAGR (compound annual growth rate)

$$CAGR = \left(\frac{V_{final}}{V_{begin}}\right)^{1/t} - 1$$

V_{begin} = initial value
V_{final} = final value
t = time in years

Then, the overall CAGR equals:

$$CAGR_{overall} = \prod_{i=1}^{N}(CAGR_i + 1) - 1$$



[Bol et al., DATE 2021]

Reformulating Kaya for Architects

Can we reformulate the Kaya identity to something we, computer architects, gain insight from?

… so we can understand how to reduce environmental impact of computing?

We focus on carbon footprint

- But representative for other sustainability issues
- Using recently published numbers, yet to be taken with grain of salt…

Distinction between

- **Embodied emissions**: GHG emissions during manufacturing process
  - **Scope-1**: chemicals and gases emitted
  - **Scope-2**: carbon emissions from energy usage
- **Operational emissions**: GHG emissions during product lifetime

Total Carbon Footprint

Embodied Scope 2 (energy usage during production)

$$CO2e_{embodied,\ scope\text{-}2} = \#chips \times \#wafer/chips \times kWh/wafer \times CO2e/kWh$$

Embodied Scope 1 (chemicals and gases during production)

$$CO2e_{embodied,\ scope\text{-}1} = \#chips \times \#wafer/chips \times CO2e/wafer$$

Operational (energy usage during lifetime)

$$CO2e_{operational} = \#chips \times kWh/chip \times CO2e/kWh$$

[L. Eeckhout, "Kaya for Architects: Towards Sustainable Computer Systems", IEEE Micro, 2023]

Demand for Chips is Increasing

Embodied Scope 2 (energy usage during production)

$CO2e_{embodied,\ scope\text{-}2}$ = **#chips** × **wafer/chips × kWh/wafer × CO2e/kWh**

Embodied Scope 1 (chemicals and gases during production)

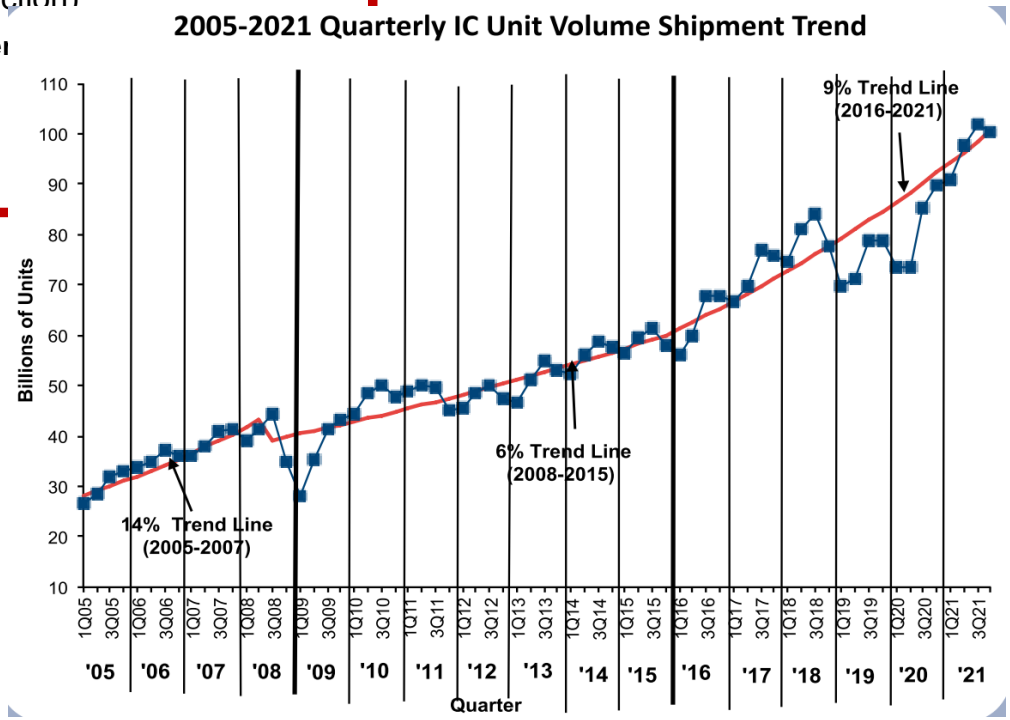
$CO2e_{embodied,\ scope\text{-}1}$ = **#chips** × **wafer/chips × CO2e/wafer**

Operational (energy usage during lifetime)

$CO2e_{operational}$ = **#chips** × **kWh/chip × CO2e/kWh**

Increasing number of chips:

CAGR = +9%



[IC Insight, 2022]

Die Size Seems to Have Stagnated

Embodied Scope 2 (energy usage during production)

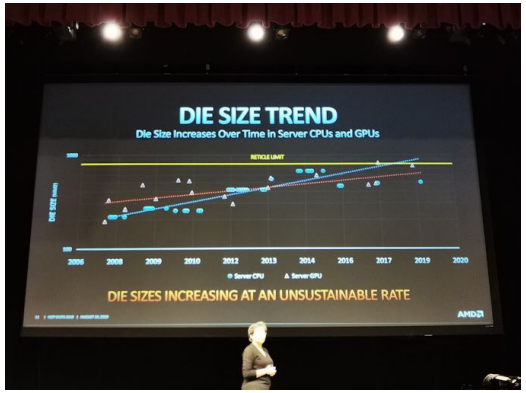
$$CO2e_{embodied,\ scope\text{-}2} = \#chips \times \#wafer/chips \times kWh/wafer \times CO2e/kWh$$

Embodied Scope 1 (chemicals and gases during production)

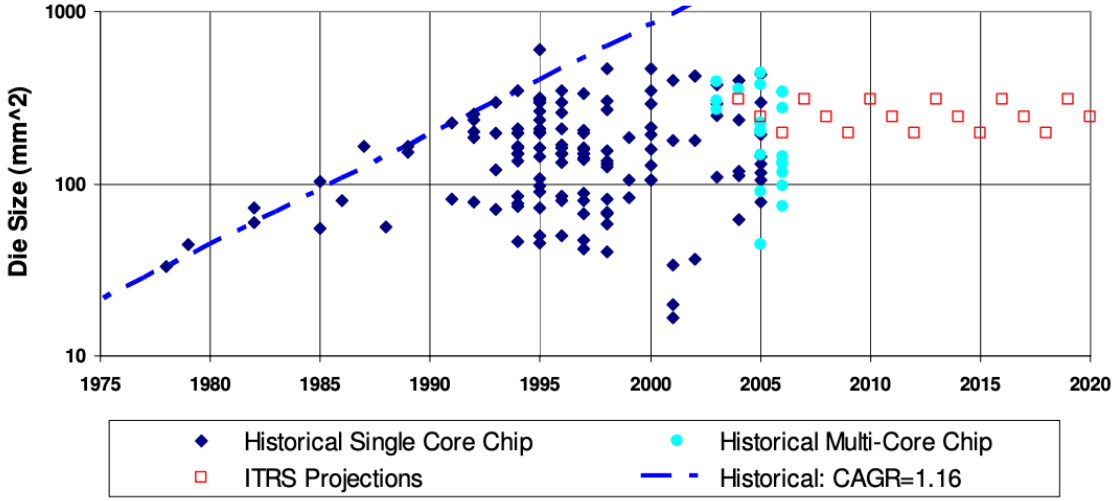
$$CO2e_{embodied,\ scope\text{-}1} = \#chips \times \#wafer/chips \times CO2e/wafer$$

Operational (energy usage during lifetime)

$$CO2e_{operational} = \#chips \times kWh/chip \times CO2e/kWh$$



Number of chips per wafer: CAGR ≃ +0%



[Kogge et al., 2008]

Increasing Energy Demand per Wafer

Embodied Scope 2 (energy usage during production)

$CO2e_{embodied,\ scope\text{-}2}$ = chips × wafer/chips × kWh/wafer × CO2e/kWh

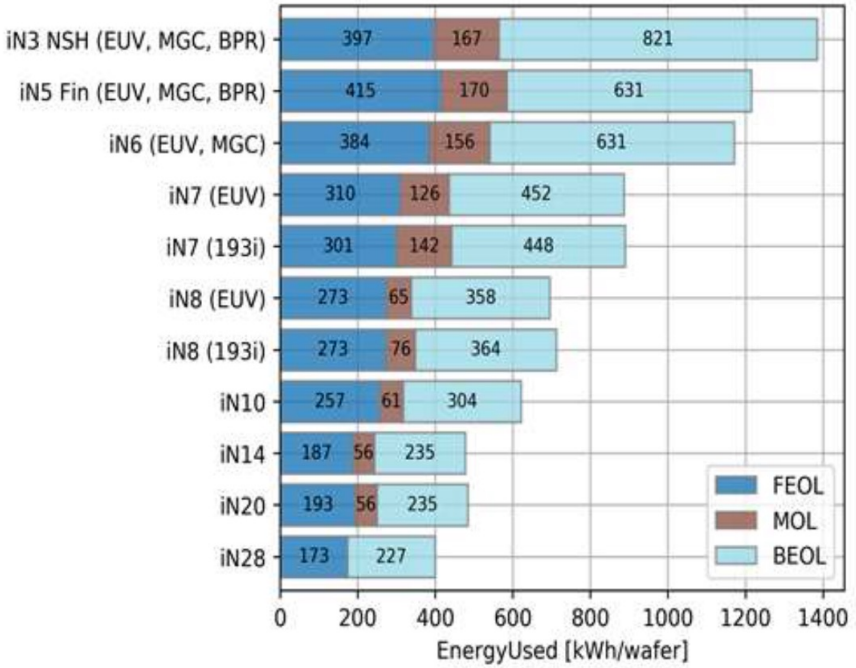
Embodied Scope 1 (chemicals and gases during production)

$CO2e_{embodied,\ scope\text{-}1}$ = chips × wafer/chips × CO2e/wafer

Operational (energy usage during lifetime)

$CO2e_{operational}$ = chips × kWh/chip × CO2e/kWh

[M. Garcia Bardon, imec, 2020]



Increasing energy demand for new tech nodes

- increasing no. processing steps
- CAGR kWh/wafer = +11.9%

According to Greenpeace East Asia:

- “Semiconductor industry electricity consumption to more than double by 2030” [April 2023]

Increasing Chemicals/Gases per Wafer

Embodied Scope 2 (energy usage during production)

$CO2e_{embodied,\ scope\text{-}2}$ = #chips × #wafer/chips × kWh/wafer × CO2e/kWh

Embodied Scope 1 (chemicals and gases during production)

$CO2e_{embodied,\ scope\text{-}1}$ = #chips × #wafer/chips × CO2e/wafer

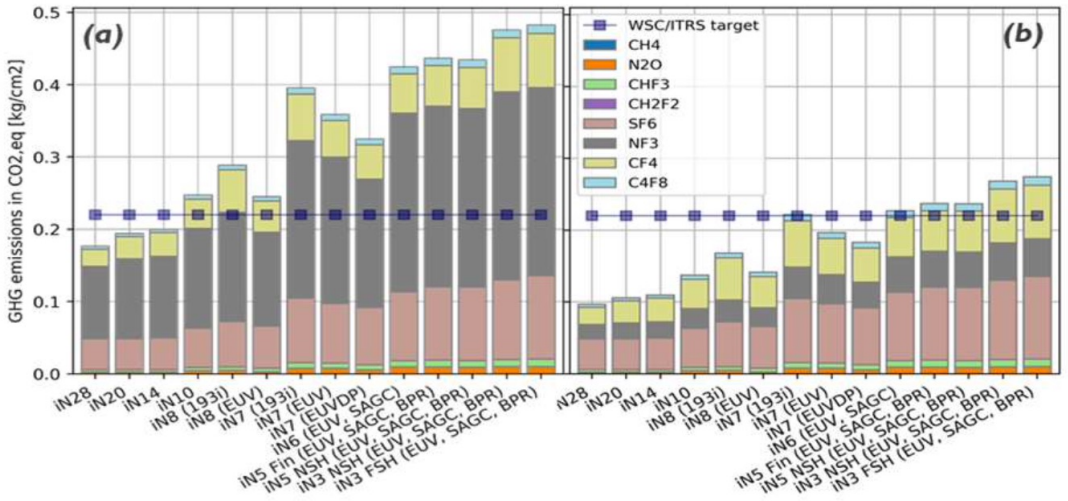
Operational (energy usage during lifetime)

$CO2e_{operational}$ = #chips × kWh/chip × CO2e/kWh

[M. Garcia Bardon, imec, 2020]

Increasing chemical/gas emissions for new tech nodes

CAGR CO2e/wafer = +9.4%



Carbon Intensity Slowly Decreasing

Embodied Scope 2 (energy usage during production)

$$CO2e_{embodied,\ scope\text{-}2} = \#chips \times \#wafer/chips \times kWh/wafer \times CO2e/kWh$$

Embodied Scope 1 (chemicals and gases during production)

$$CO2e_{embodied,\ scope\text{-}1} = \#chips \times \#wafer/chips \times CO2e/wafer$$

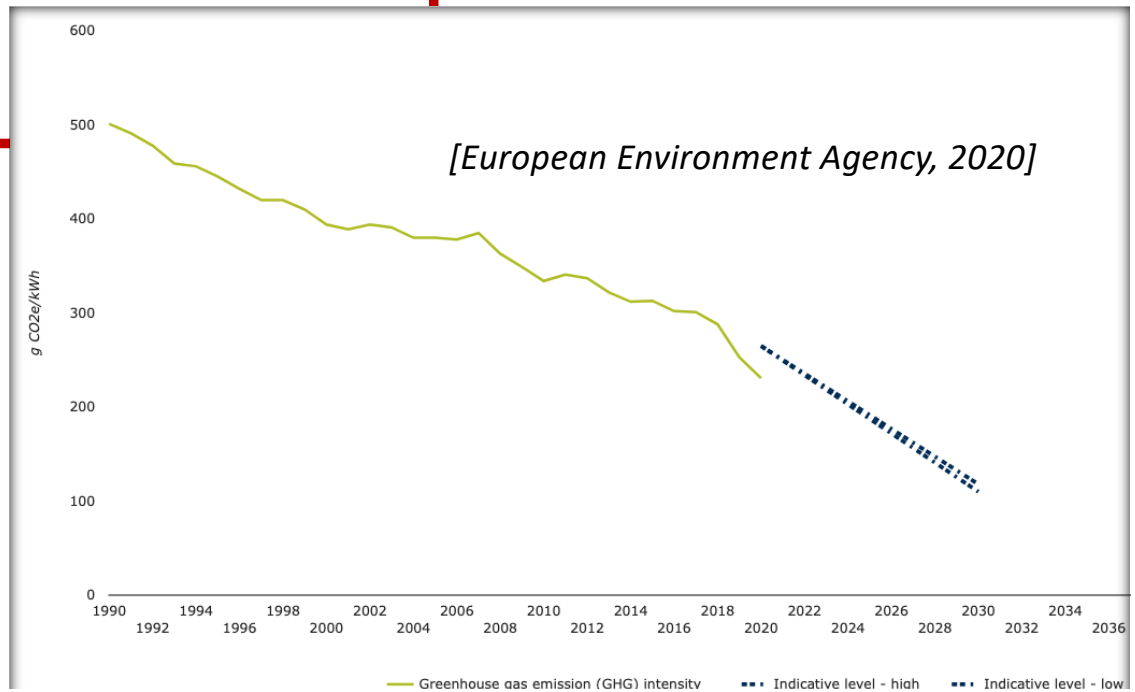
Operational (energy usage during lifetime)

$$CO2e_{operational} = \#chips \times kWh/chip \times CO2e/kWh$$

- Transition towards green energy sources

  CO2e/kWh (Europe): CAGR = -2.5%

- Much faster transition to green energy in the datacenter [Gupta et al., HPCA 2021]
  - Critical side note: green energy contracts deprive other customers from green energy



Decreasing Operational Energy

Embodied Scope 2 (energy usage during production)

$CO2e_{embodied,\ scope\text{-}2}$ = #chips × #wafer/chips × kWh/wafer × CO2e/kWh

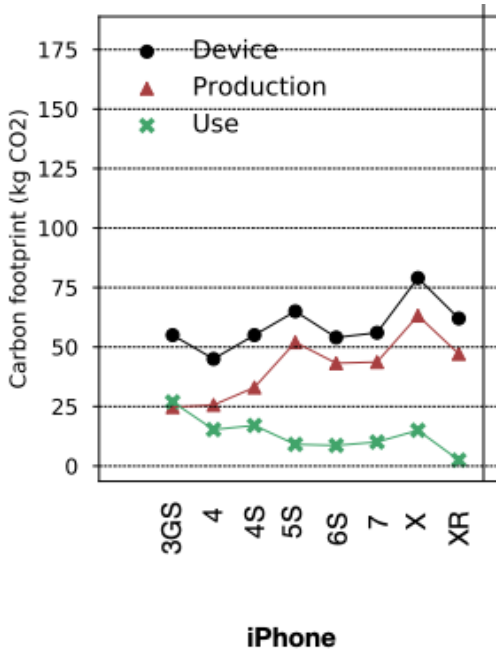
Embodied Scope 1 (chemicals and gases during production)

$CO2e_{embodied,\ scope\text{-}1}$ = #chips × #wafer/chips × CO2e/wafer

Operational (energy usage during lifetime)

$CO2e_{operational}$ = #chips × kWh/chip × CO2e/kWh

Operational energy consumption is decreasing



[U. Gupta et al. , HPCA 2021]

Total Carbon Footprint

Embodied Scope 2 (energy usage during production)

$$CO2e_{embodied,\ scope\text{-}2} = \#chips \times wafer/chips \times kWh/wafer \times CO2e/kWh$$

Embodied Scope 1 (chemicals and gases during production)

$$CO2e_{embodied,\ scope\text{-}1} = \#chips \times wafer/chips \times CO2e/wafer$$

Operational (energy usage during lifetime)

$$CO2e_{operational} = \#chips \times kWh/chip \times CO2e/kWh$$

Key take-aways:

- Demand for chips keeps increasing (Jevons' paradox?)
- GHG emissions (both scope-1 and 2) increase with new tech nodes
- Transition to green energy not moving fast enough ***and*** it doesn't impact scope-1 emissions (and other sustainability issues like raw material need, e-waste, etc.)

How to Reduce Total Carbon Emissions?

Hinges on relative importance of embodied vs operational emissions

Varies across devices *[Gupta et al., HPCA 2021]*

- Battery-powered devices: embodied emissions dominate
- Always-connected devices: operational emissions dominate

Two scenarios:

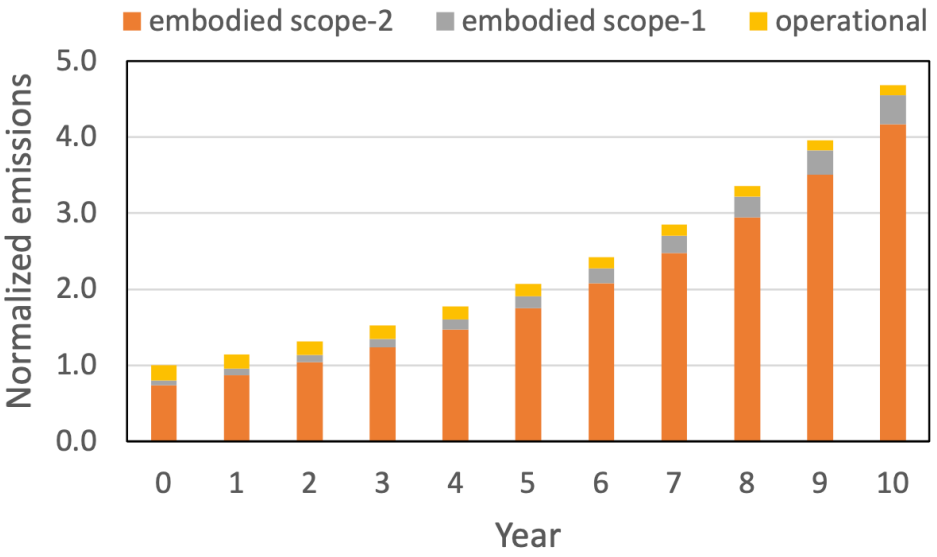
(a) Embodied emissions dominate initially: 80% embodied vs 20% operational

(b) Operational emissions dominate initially: 20% embodied vs 80% operational

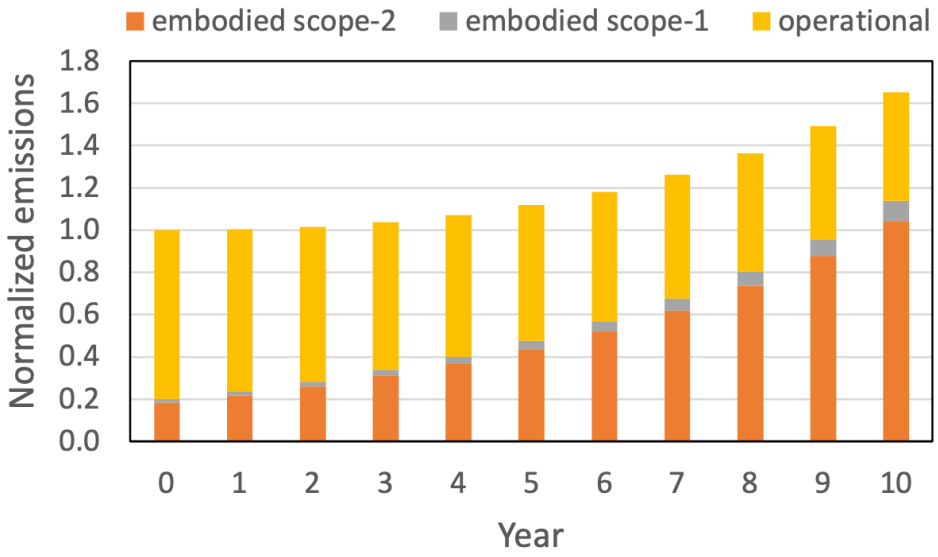
Other assumptions:

- Constant die size (CAGR = 0%)
- Lifetime operational energy consumption decreases (CAGR = -10%)

Current Trends for Total Carbon Emissions



(a) initially dominating embodied emissions



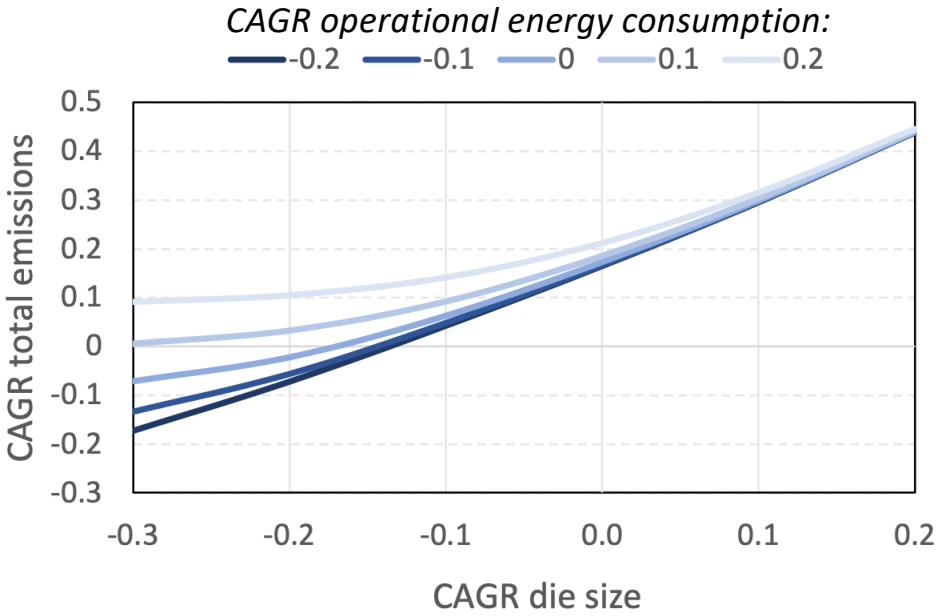
(b) initially dominating operational emissions

Total carbon footprint continues to grow

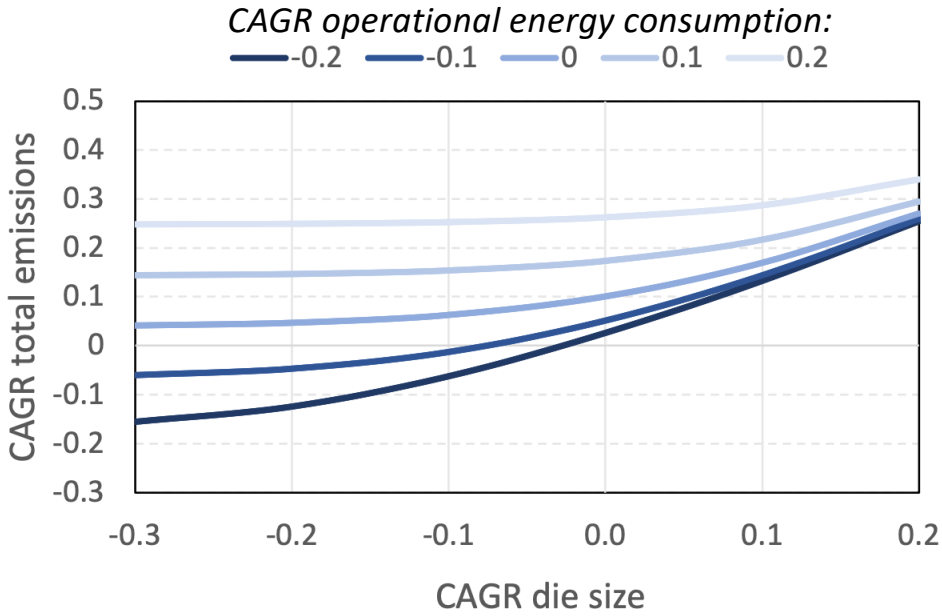
Embodied emissions grow in importance and (will) dominate, in both scenarios

Reason: increasing number of chips *and* growing energy intensity of chip manufacturing

We Need to Reduce Both Embodied and Operational Emissions



(a) initially dominating embodied emissions

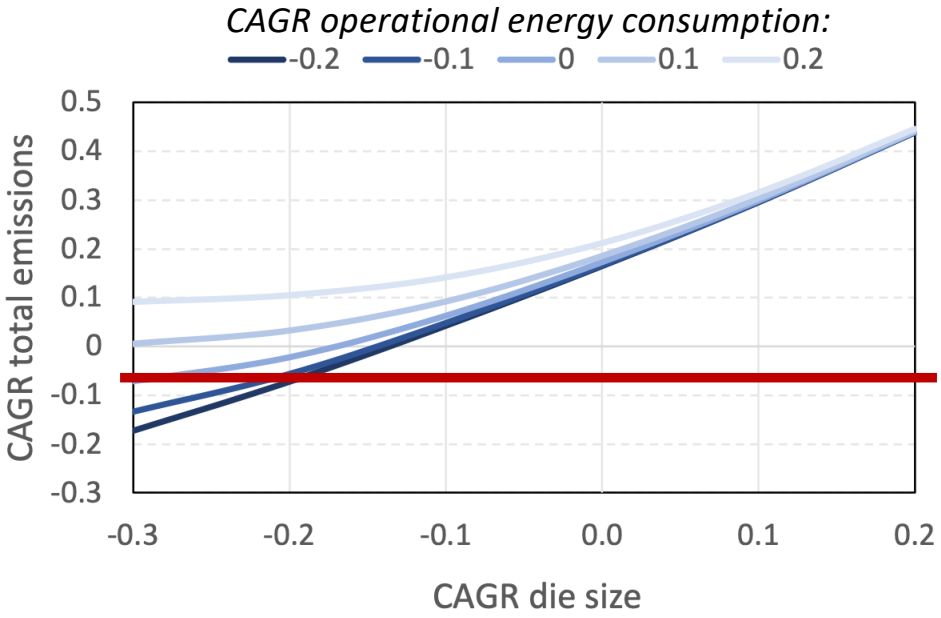


(b) initially dominating operational emissions

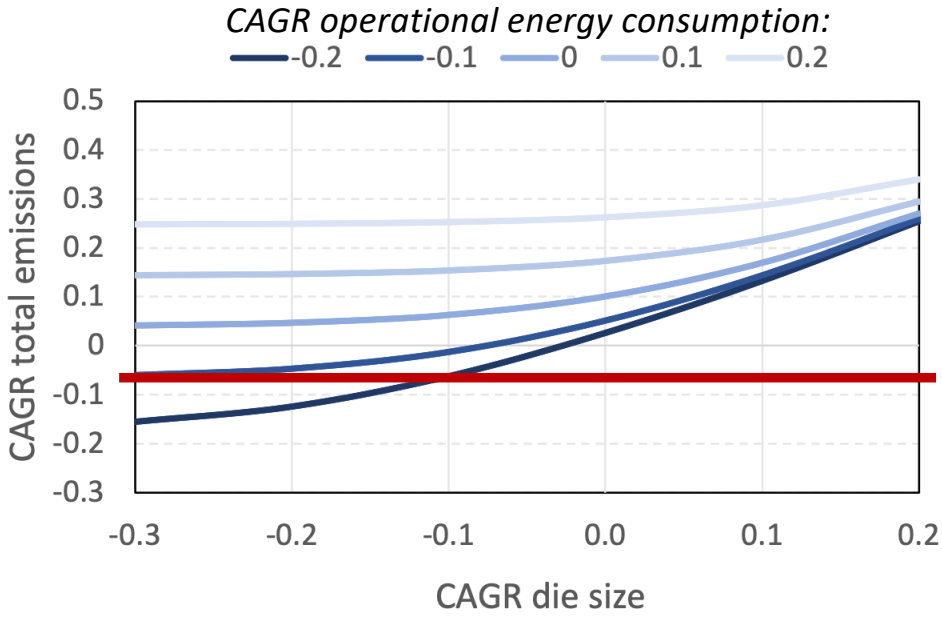
- We need to significantly reduce die size
- Reducing operational energy intensity is less critical, but not unimportant

- We need to reduce die size, AND
- We need to reduce operational energy intensity

We Need to Significantly Reduce Both Die Size and Operational Energy Consumption



(a) initially dominating embodied emissions



(b) initially dominating operational emissions

Paris agreement says we should decrease total CO2e by 7.6% per annum *[United Nations]*

- Aim to limit global warming to less than 2℃, and try to limit the increase to 1.5℃ relative to pre-industrial level (1850-1900)

Key Take-Aways

Total carbon emissions continue to grow under current scaling trends

Embodied emissions are most dominant contributor to the total carbon footprint

Computer architects can (and should) **reduce total carbon footprint by reducing die size *and* operational emissions**

- Reducing die size is more important
- Leverage Moore's Law in a sober way: limit the number additional transistors used for providing new functionality

Transitioning to green chip manufacturing alone is not the solution

- It does not affect Scope-1 emissions, nor does it diminish the need for raw materials and ultra-pure water during production
- This reinforces key observation that we need to design smaller chips (and continue to reduce operational energy intensity)

Part III

What can we, as computer architects, do to reduce the environmental impact of computing?

How to reason about sustainable computer system design in light of inherent data uncertainty?

Embrace Inherent Data Uncertainty

Develop a first-order model that is

- ***deliberately simple,***
- ***built upon first principles, and***
- ***provides insight***

Key idea:

- ***use proxies for embodied and operational footprint that we, computer architects, have control over, and***
- ***parameterize relative importance of embodied versus operational footprint***

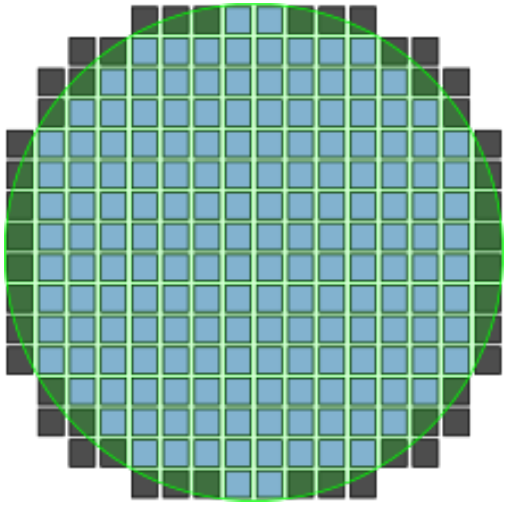
Embodied Footprint

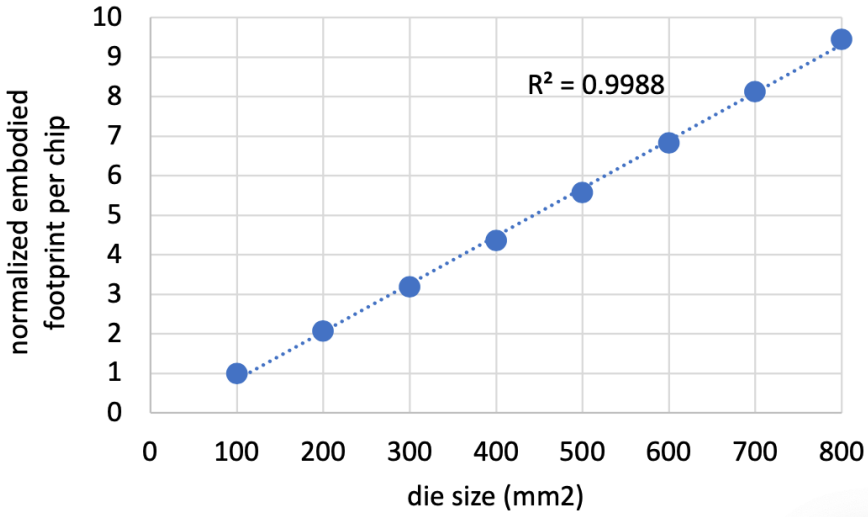
Wafer = production unit in semiconductor fab

- Environmental impact for producing a wafer: energy consumed, chemicals and gases emitted, ultra pure water used, materials used

The bigger the chip, the higher the embodied footprint per chip

Pr





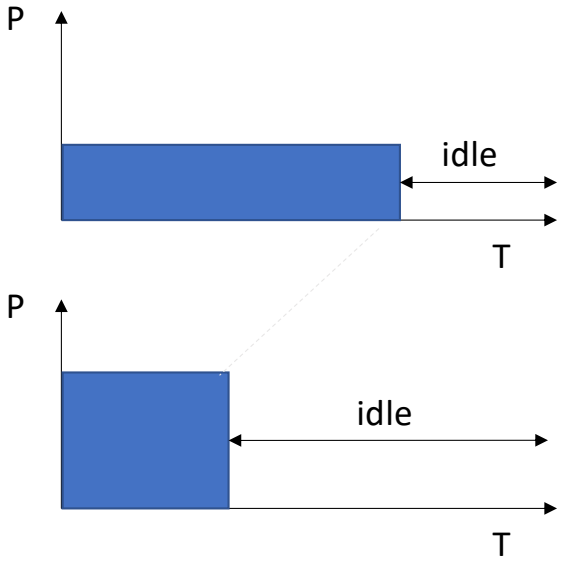
Operational Footprint (1/2)

(1) Fixed-work scenario

- Assumption: a device performs fixed amount of work over its entire lifetime

The higher the energy consumption, the higher the operational footprint

Proxy = energy consumption (E)



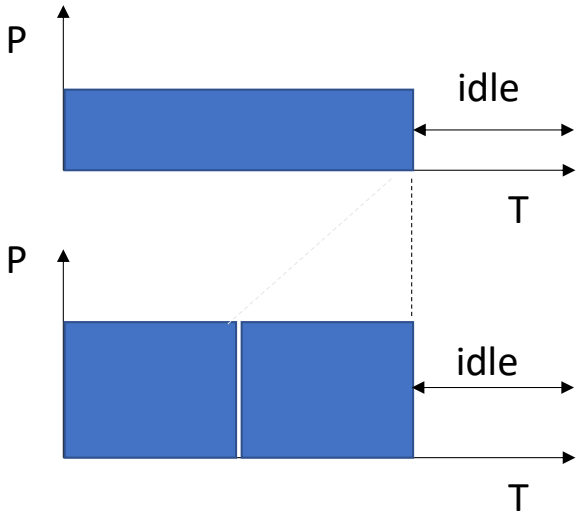
Operational Footprint (2/2)

(2) Fixed-time scenario – *more realistic scenario(?)*

- Assumption: we use the device for the same amount of time
- Or we use the new device even more frequently, cf. Jevons' paradox

The higher the power consumption, the higher the operational footprint

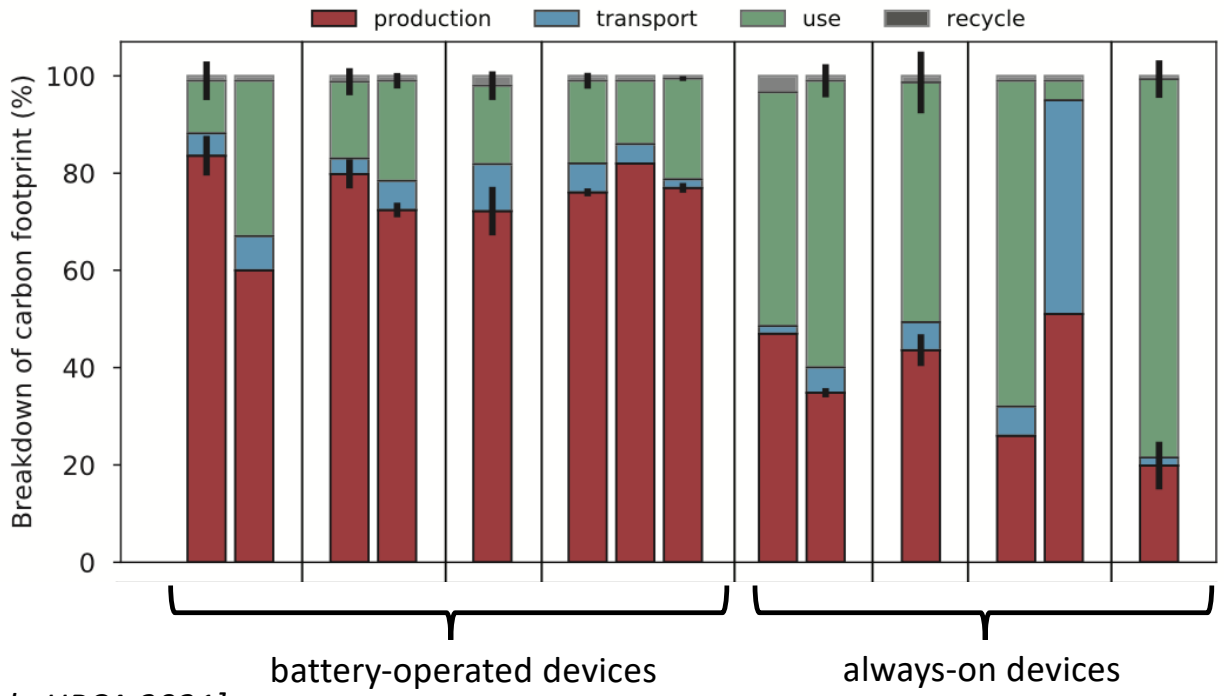
Proxy = power consumption (P)



Embodied versus Operational Footprint

Ratio of embodied vs operational footprint depends on

- **Device type**



[Gupta et al., HPCA 2021]

Embodied versus Operational Footprint

Ratio of embodied vs operational footprint depends on

- **Device type**
  - **Battery-operated vs always-on devices**
- **Lifetime**
  - **The longer the lifetime, the higher the relative weight of operational footprint**
- **Energy mix**
  - **The greener the energy mix, the higher the relative weight of embodied footprint**

Solution: parameterize the embodied-to-operational footprint ratio

First-Order Sustainability Model

Total footprint = embodied footprint + operational footprint

= α × A + (1-α) × E *// assuming fixed-work scenario*

= α × A + (1-α) × P *// assuming fixed-time scenario*

with A = chip area, E = energy, P = power

α parameter is a function of device type/usage, lifetime of device, rebound effect, energy source during manufacturing vs lifetime

Parameterization allows for considering different scenarios:

- Embodied emissions dominate (assume α = 80% or α = 70-90%)
- Operational emissions dominate (assume α = 20% or α = 10-30%)

[L. Eeckhout, "A First-Order Model to Assess Computer Architecture Sustainability", Best of CAL 2022]

Comparison to ACT*

ACT*

- Relies on detailed carbon emission numbers
- Implicitly assumes fixed-work scenario
- Derives the embodied-to-operational ratio from the numbers

First-order model

- Deliberately simple to gain insight and explore sensitivity
- Includes fixed-time scenario and accounts for Jevons' paradox
- Useful complement to ACT

**[Gupta et al., ISCA 2022]*

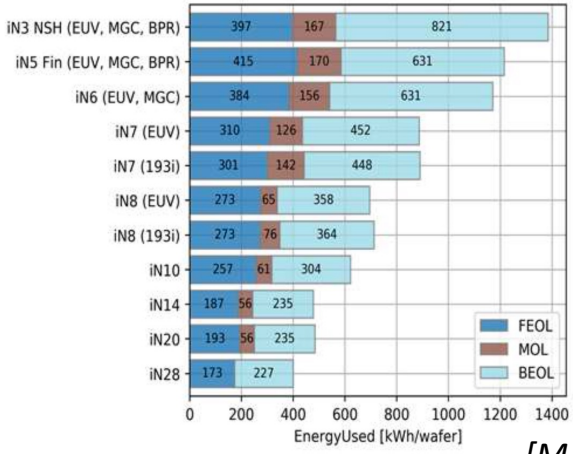
What insight can we gain from this simple model?

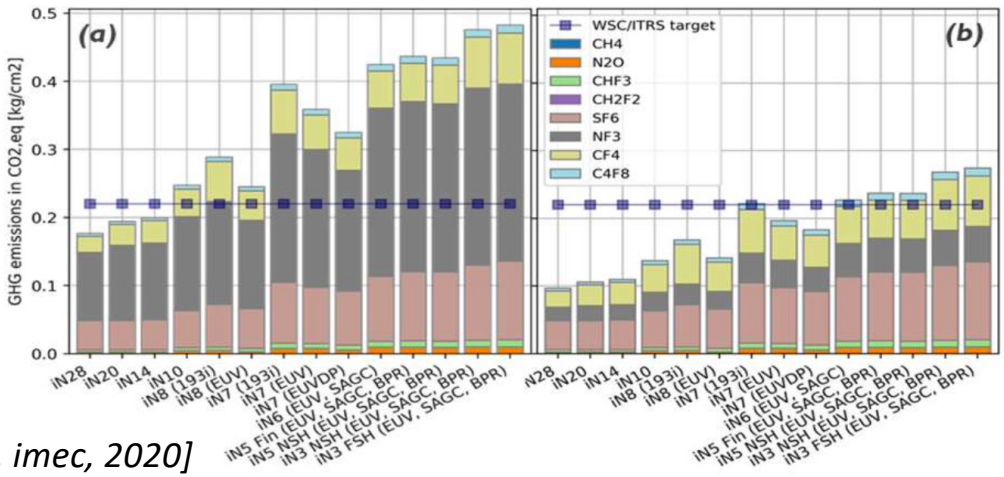
Case Study #1: Die Shrink

Implement an existing microarchitecture in a new tech node

Embodied emissions: net decrease

- Reduction of chip area by 50%
- This offsets the increase in energy consumption during manufacturing (+25%) and increase in chemicals/gases emitted (+19.5%)





[M. Garcia Bardon, imec, 2020]

Case Study #1: Die Shrink

Implement an existing microarchitecture in a new tech node

Embodied emissions: net decrease

- Reduction of chip area by 50%
- This offsets the increase in energy consumption during manufacturing (+25%) and increase in chemicals/gases emitted (+19%) *[Imec, 2020]*

Operational emissions: net decrease or neutral

- Classical scaling: power reduces by 2x, performance increases by 1.41x, and energy reduces by 2.82x
- Post-Dennard scaling: power remains the same, energy reduces by 1.41x

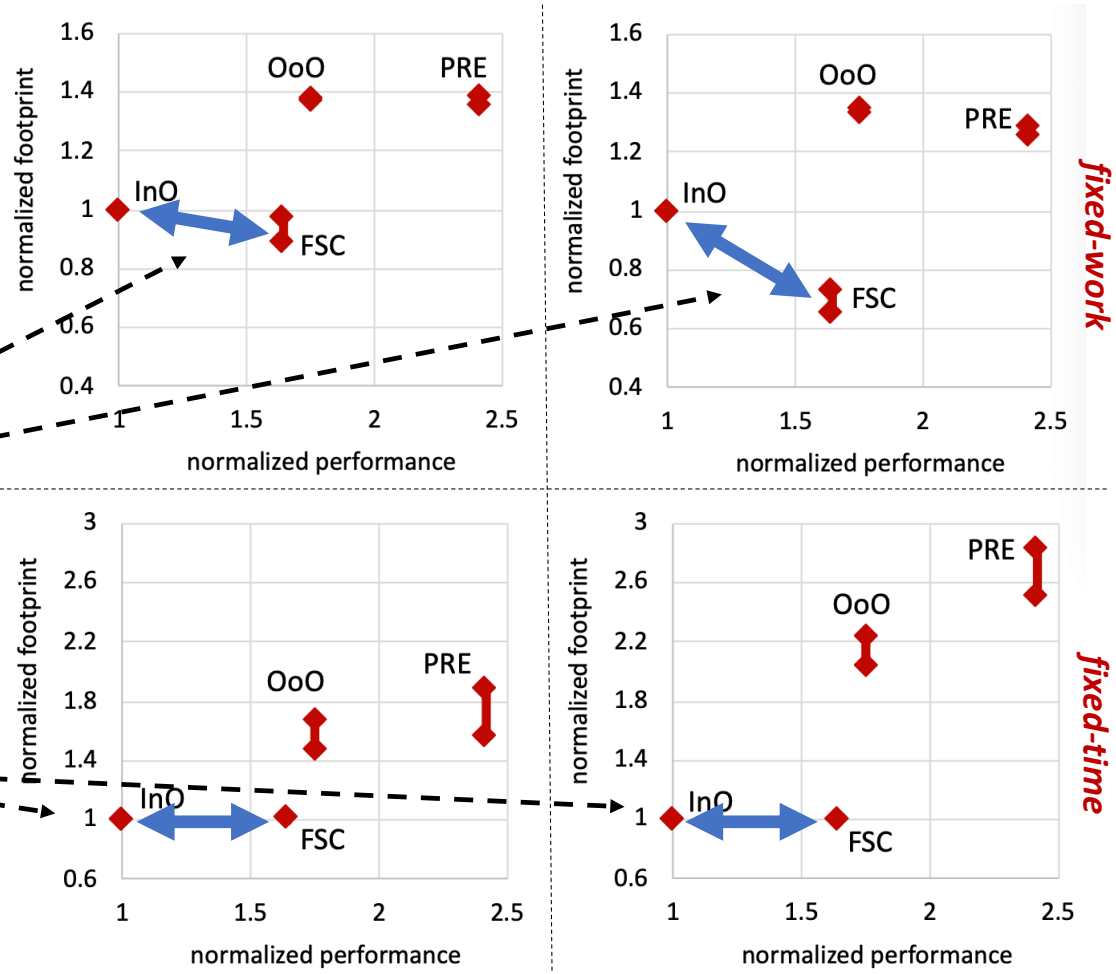
Overall conclusion: net reduction in environmental footprint

Case Study #2: Microarchitecture

In-order vs FSC*

FSC: lower footprint due to lower energy consumption and higher performance

FSC: similar footprint due to similar chip area and similar power consumption



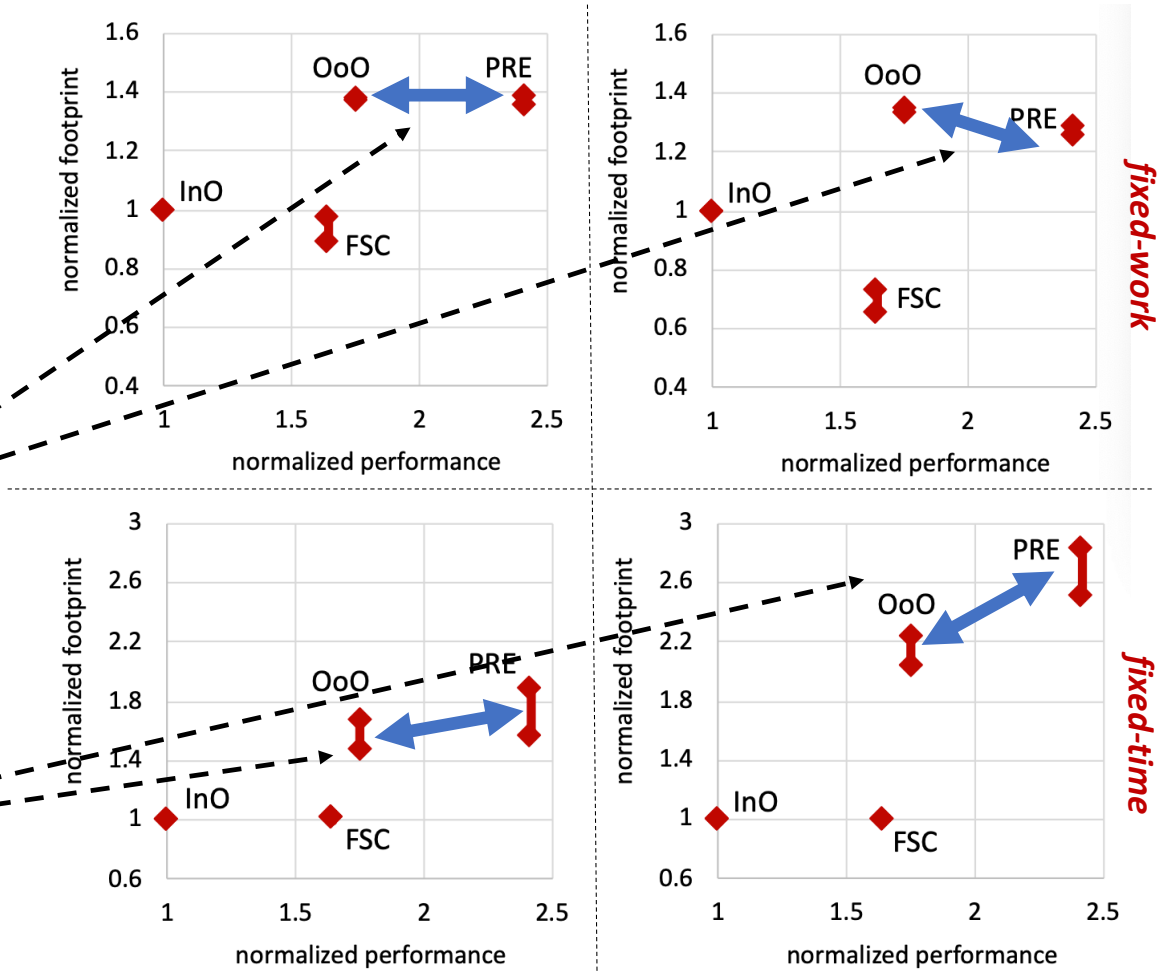
*Forward Slice Core [PACT 2020]

Case Study #2: Microarchitecture

Out-of-order vs PRE*

PRE: similar/lower footprint due to lower energy consumption, higher performance and small chip area overhead

PRE: higher footprint due to higher power consumption



Precise Runahead Execution [HPCA 2020]

embodied dominated (α = 70% - 90%)

operational dominated (α = 10% - 30%)

Case Study #3: Accelerators

H.264 accelerator* versus general-purpose CPU implementation

- Accelerator consumes 500x less energy
- Accelerator is 15x smaller
- Similar performance

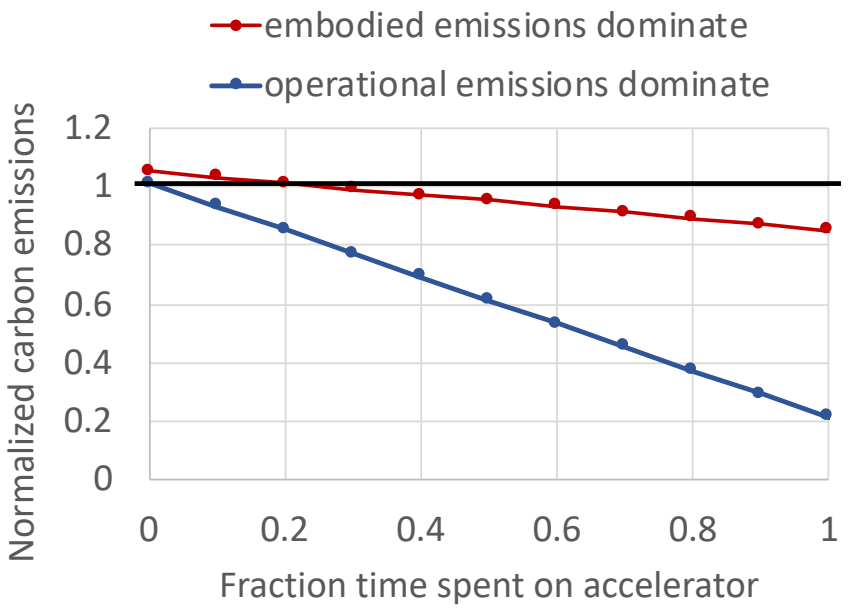
Embodied emissions = CPU + accelerator

- Accelerator = 6.5% extra chip area over CPU

Operational emissions = (1-f) x E_{CPU} + f x $E_{accelerator}$

- Depends on fraction f spent on accelerator

If embodied emissions dominate, it is critical that the accelerator is used for a sufficient fraction of time to be more sustainable than a CPU implementation



*[Hameed et al., ISCA 2010]

Case Study #3: Dark Silicon

Same assumptions as in previous slide

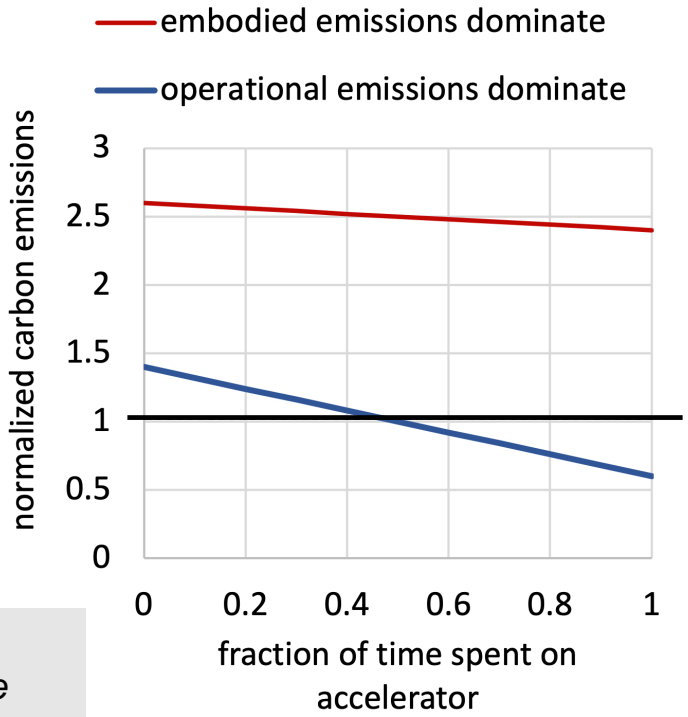
- Accelerator consumes 500x less energy
- Similar performance

BUT: Accelerator occupies 2/3 of the total chip area

If embodied emissions dominate (likely case today, esp. for mobile devices), dark silicon leads to a substantial increase in overall footprint

If operational emissions dominate, it is critical to use the accelerator(s) to amortize the embodied footprint to reduce overall footprint

Research question: can we design reconfigurable accelerators to drastically reduce embodied footprintenvironmental footprint while still achieving high performance and low energy?

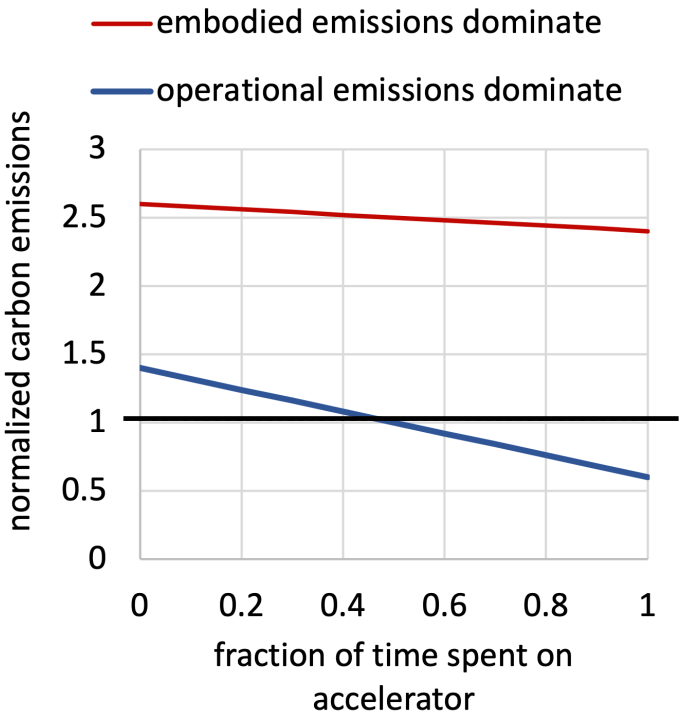


Dark Silicon Considered Harmful

If embodied emissions dominate (likely case today, esp. for mobile devices), dark silicon leads to a substantial increase in overall footprint

If operational emissions dominate, it is critical to use the accelerator(s) to amortize the embodied footprint to reduce overall footprint

Research question for this community: *Can we design reconfigurable accelerators to drastically reduce the total environmental footprint while still achieving high performance and low energy for many applications?*



What Can We, Computer Architects, Do?

1. **Temper the demand for new chips**
   - Integrate more functionality per chip, happening today → nevertheless, demand for new chips continues to grow… (rebound effect, Jevons' paradox?)
   - Engage fault-tolerance, reconfiguration or graceful degradation techniques to improve effective yield and/or extend product lifetime
   - Transition from product-based business model to service-oriented model? Design for repairability?
2. **Design smaller chips** to reduce embodied emissions
   - Keep chip area constant (current practice): CAGR total CO2e = +18.9%
   - Keep transistor count constant (die shrink): CAGR wafer/chips = -29.3% → CAGR total CO2e = -16.8%
   - Is there a middle ground?
     - Reduce die size by 15% per year: carbon-neutral
     - Reduce die size between 15% and 29% per year → reduction in CO2e
   - Leverage Moore's law in a sober way → reduce die size *and* add new functionality in new tech nodes
   - Seems to contradict current trends…

[L. Eeckhout, "Kaya for Architects: Towards Sustainable Computer Systems", IEEE Micro, 2023]

What Can We, Computer Architects, Do?

1. **Temper the demand for new chips**
   - Integrate more functionality per chip, happening today → nevertheless, demand for new chips continues to grow… (rebound effect, Jevons' paradox?)
   - Engage fault-tolerance, reconfiguration or graceful degradation techniques to improve effective yield and/or extend product lifetime
   - Transition from product-based business model to service-oriented model? Design for repairability?
2. **Design smaller chips** to reduce embodied emissions
   - Doesn't mean we should stop adding new functionality!
   - Leverage Moore's law in a sober way → reduce die size *and* add new functionality in new tech nodes
   - Seems to contradict current trends…
3. **Improve energy and power efficiency** to reduce operational emissions
   - Reduce energy consumption when work is to be done
   - Reduce power consumption when idle
   - Is happening today, although Jevons paradox is looming around the corner
     - Improving power/energy efficiency may trigger increased deployment, increasing the embodied emissions

[L. Eeckhout, "Kaya for Architects: Towards Sustainable Computer Systems", IEEE Micro, 2023]

Summary

Assessing computer architecture sustainability is challenging

- Multi-faceted problem, inherent data uncertainty

Total carbon emissions continue to grow under current scaling trends

Embodied emissions are most dominant contributor to the total carbon footprint

Computer architects can (and should) **reduce total carbon footprint by reducing die size** (primarily) ***and*** **operational emissions** (secondarily)

First-order model using proxies for embodied/operational footprint and parameterized embodied/operational ratio

- Deliberately simple, yet accounts for Jevons' paradox
- Provides insight and intuition
- Framework to reason about computer architecture sustainability trade-offs

Sustainable Computer Systems

Lieven Eeckhout
Ghent University, Belgium

[L. Eeckhout, "Kaya for Architects: Towards Sustainable Computer Systems", IEEE Micro, 2023]

[L. Eeckhout, "A First-Order Model to Assess Computer Architecture Sustainability", Best of CAL 2022]