ACT: Designing Sustainable Computer Systems With An <u>**A**</u>rchitectural <u>**C**</u>arbon Model <u>**T**</u>ool

Udit Gupta

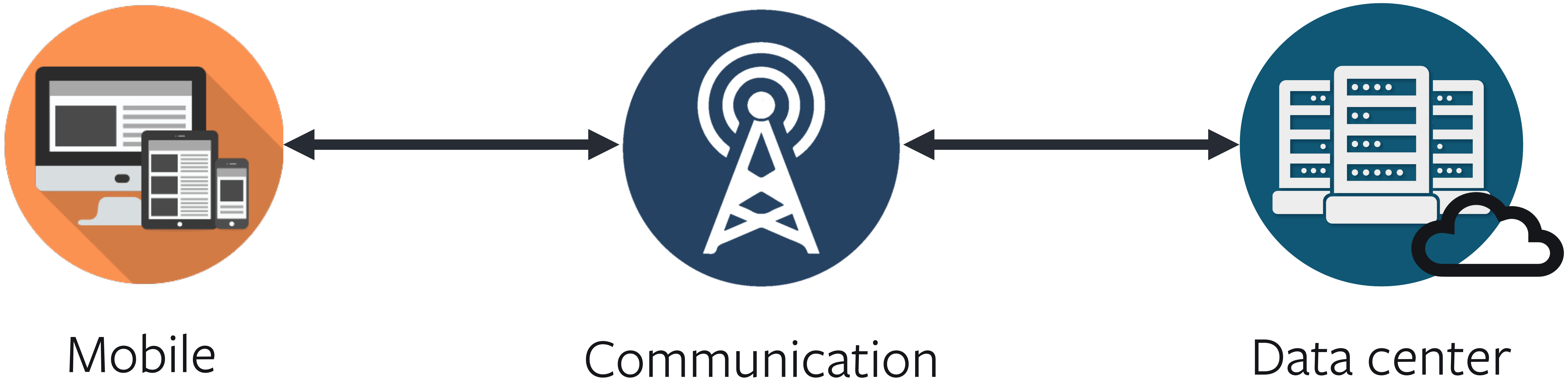
Mariam Elgamal, Gage Hills, Gu-Yeon Wei,
Hsien-Hsin S. Lee, David Brooks, Carole-Jean Wu

At International Symposium on Computer Architecture (*ISCA 2022*)

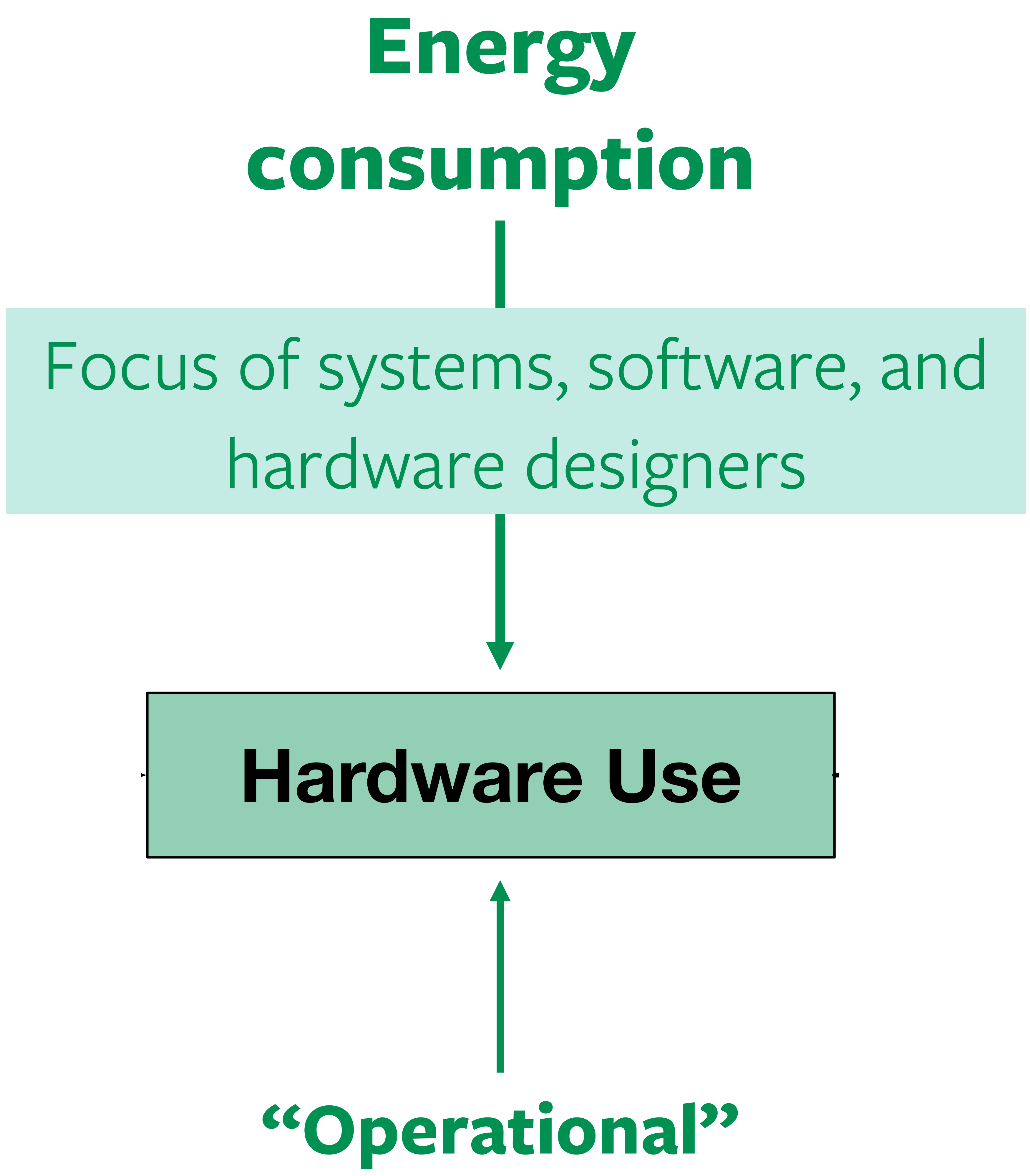
Computing incurs a growing environmental footprint

900 Million tons of CO_2

- **On par with** the aviation industry's footprint
- **2.1 - 3.9%** of worldwide emissions



Computing's emissions are rising given its growing demand!


Energy consumption
Focus of systems, software, and hardware designers
Hardware Use
“Operational”

Crucial to look at emissions across HW life cycle

Crucial to look at emissions across HW life cycle

Hardware manufacturing is a dominating source of carbon

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 of computing systems.**

***Index Terms*—Data center, mobile, energy, carbon footprint**

I. Introduction

The world has seen a dramatic advancement of information and communication technology (ICT). The rise in ICT has resulting in a proliferation of consumer devices (e.g., PCs, mobile phones, TVs, and home entertainment systems), networking technologies (e.g., wired networks and 3G/4G LTE), and data centers. Although ICT has enabled applications including cryptocurrencies, artificial intelligence (AI), e-commerce, online entertainment, social networking, and cloud storage, it has incurred tremendous environmental impacts

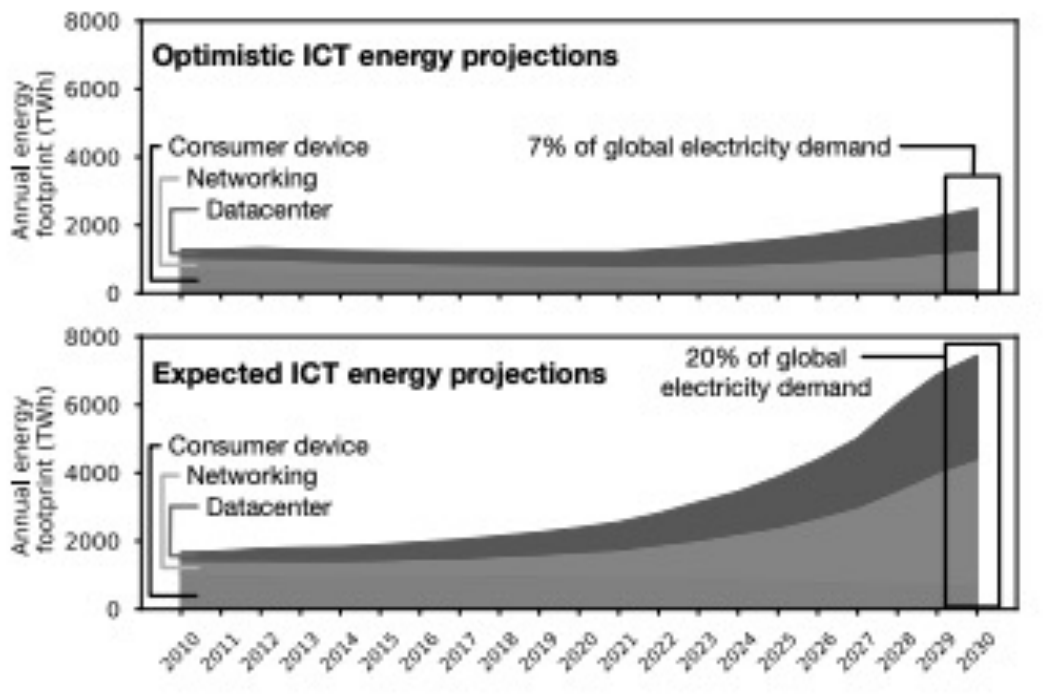
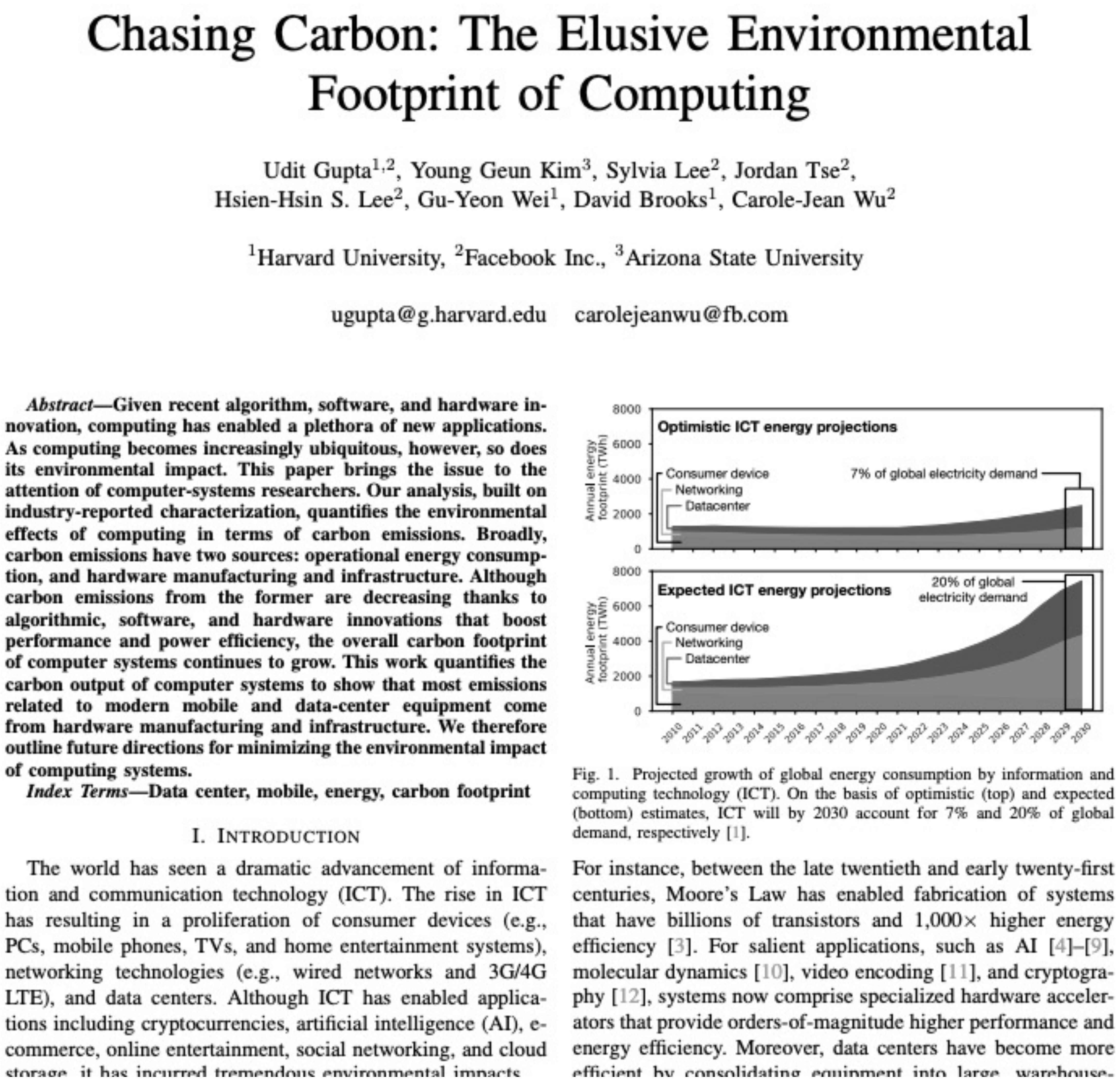


Fig. 1. Projected growth of global energy consumption by information and computing technology (ICT). On the basis of optimistic (top) and expected (bottom) estimates, ICT will by 2030 account for 7% and 20% of global demand, respectively [1].

For instance, between the late twentieth and early twenty-first centuries, Moore's Law has enabled fabrication of systems that have billions of transistors and 1,000× higher energy efficiency [3]. For salient applications, such as AI [4]–[9], molecular dynamics [10], video encoding [11], and cryptography [12], systems now comprise specialized hardware accelerators that provide orders-of-magnitude higher performance and energy efficiency. Moreover, data centers have become more efficient by consolidating equipment into large, warehouse-

Hardware manufacturing is a dominating source of carbon



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 of computing systems.**

***Index Terms*—Data center, mobile, energy, carbon footprint**

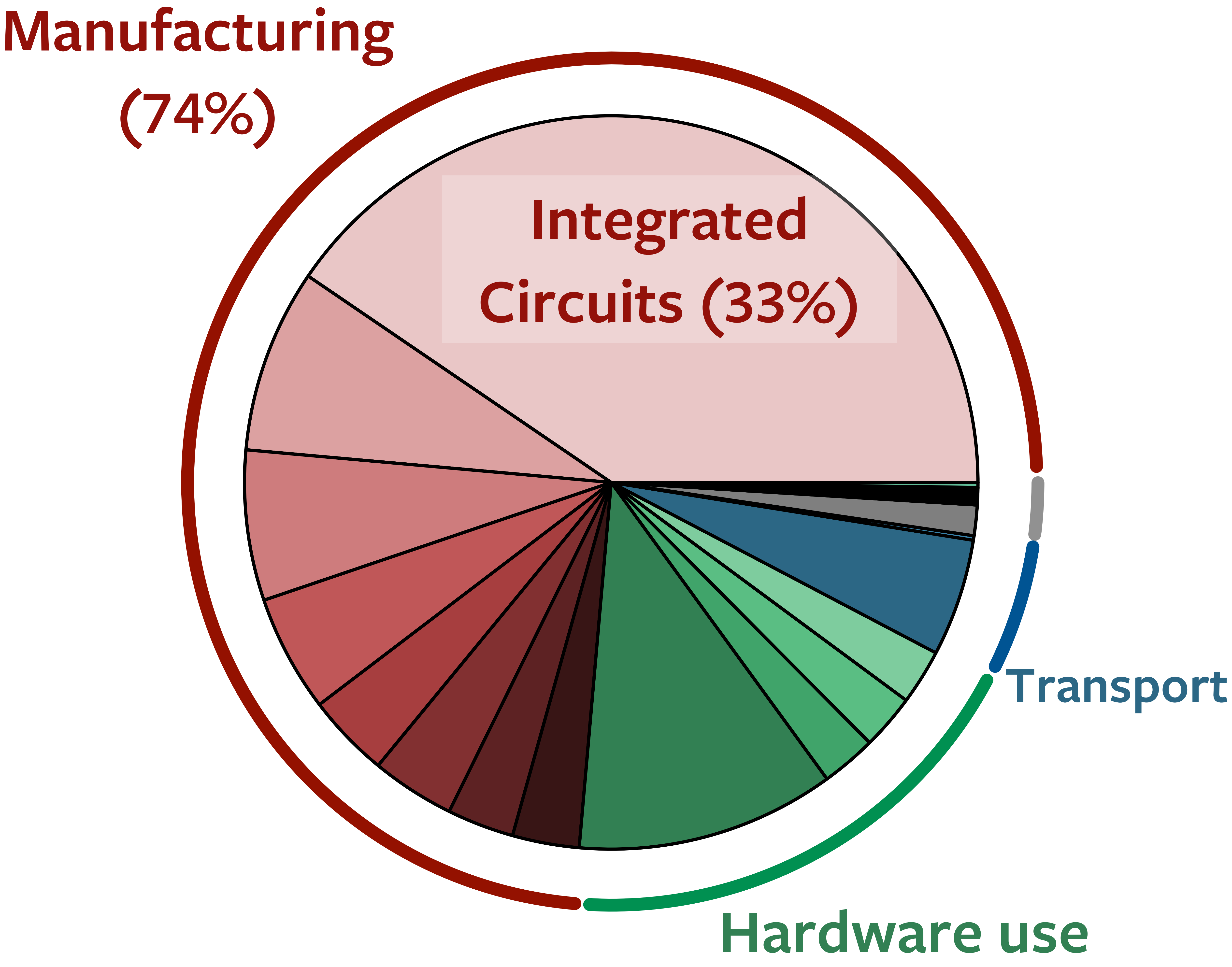
I. Introduction

The world has seen a dramatic advancement of information and communication technology (ICT). The rise in ICT has resulting in a proliferation of consumer devices (e.g., PCs, mobile phones, TVs, and home entertainment systems), networking technologies (e.g., wired networks and 3G/4G LTE), and data centers. Although ICT has enabled applications including cryptocurrencies, artificial intelligence (AI), e-commerce, online entertainment, social networking, and cloud storage, it has incurred tremendous environmental impacts

Fig. 1. Projected growth of global energy consumption by information and computing technology (ICT). On the basis of optimistic (top) and expected (bottom) estimates, ICT will by 2030 account for 7% and 20% of global demand, respectively [1].

For instance, between the late twentieth and early twenty-first centuries, Moore's Law has enabled fabrication of systems that have billions of transistors and 1,000× higher energy efficiency [3]. For salient applications, such as AI [4]–[9], molecular dynamics [10], video encoding [11], and cryptography [12], systems now comprise specialized hardware accelerators that provide orders-of-magnitude higher performance and energy efficiency. Moreover, data centers have become more efficient by consolidating equipment into large, warehouse-

Apple's 2019 carbon emissions



Gupta et. al. Chasing Carbon: The Elusive CO_2 Footprint of Computing(HPCA 2021)

Challenge: How do we design sustainable systems by considering the footprint across lifecycles

Challenge: How do we design sustainable systems by considering the footprint across lifecycles

This work: Architectural Carbon Modeling Tools (ACT)

Challenge: How do we design sustainable systems by considering the footprint across lifecycles

This work: Architectural Carbon Modeling Tools (ACT)

Overview of ACT

Challenge: How do we design sustainable systems by considering the footprint across lifecycles

This work: Architectural Carbon Modeling Tools (ACT)

Overview of ACT

Comparing ACT to other methodologies

Challenge: How do we design sustainable systems by considering the footprint across lifecycles

This work: Architectural Carbon Modeling Tools (ACT)

Overview of ACT

Comparing ACT to other methodologies

Sustainability aware-design case studies

Challenge: How do we design sustainable systems by considering the footprint across lifecycles

This work: Architectural Carbon Modeling Tools (ACT)

Overview of ACT

Comparing ACT to other methodologies

Sustainability aware-design case studies

Architectural Carbon Model

Model	Hardware/software input

Architectural Carbon Model

Model	Hardware/software input
$Carbon = OP_{CF} + \frac{Runtime}{Lifetime} Emb_{CF}$	**Performance/power/energy** and **lifetime** of hardware

Architectural Carbon Model

Model	Hardware/software input
$Carbon = OP_{CF} + \frac{Runtime}{Lifetime} Emb_{CF}$	**Performance/power/energy** and **lifetime** of hardware
$OP_{CF} = CI_{use} \times Energy$	**Energy efficiency** and environment (**carbon intensity**)

Architectural Carbon Model

Model	Hardware/software input
$Carbon = OP_{CF} + \frac{Runtime}{Lifetime} Emb_{CF}$	**Performance/power/energy** and **lifetime** of hardware
$OP_{CF} = CI_{use} \times Energy$	**Energy efficiency** and environment (**carbon intensity**)
$Emb_{CF} = Packaging + \sum_{r}^{SoC,Memory,Storage} Emb_r$	Overhead of **hardware manufacturing**

Embodied carbon of application processors (SoC's)

$$Emb_{SoC} = Area \times \frac{\left(CI_{fab} \times Fab_{energy}\right) + Fab_{chemicals} + Fab_{materials}}{Yield}$$

Embodied carbon of application processors (SoC's)

Energy source (grid)

Device and fab characterization (industry fabs, device data)

$$Emb_{SoC} = Area \times \frac{\left(CI_{fab} \times Fab_{energy}\right) + Fab_{chemicals} + Fab_{materials}}{Yield}$$

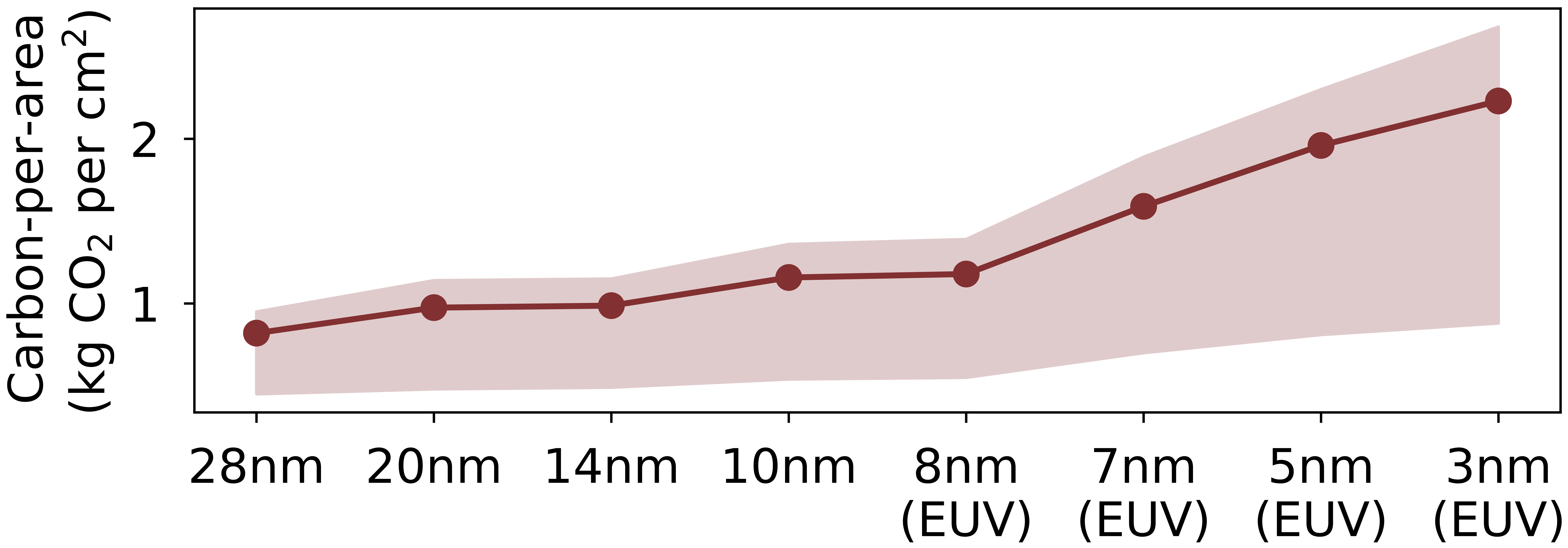
Hardware design

Embodied carbon of application processors (SoC's)

$$Emb_{SoC} = Area \times \boldsymbol{CPA}$$

Embodied carbon of application processors (SoC's)

$$Emb_{SoC} = Area \times \boldsymbol{CPA}$$

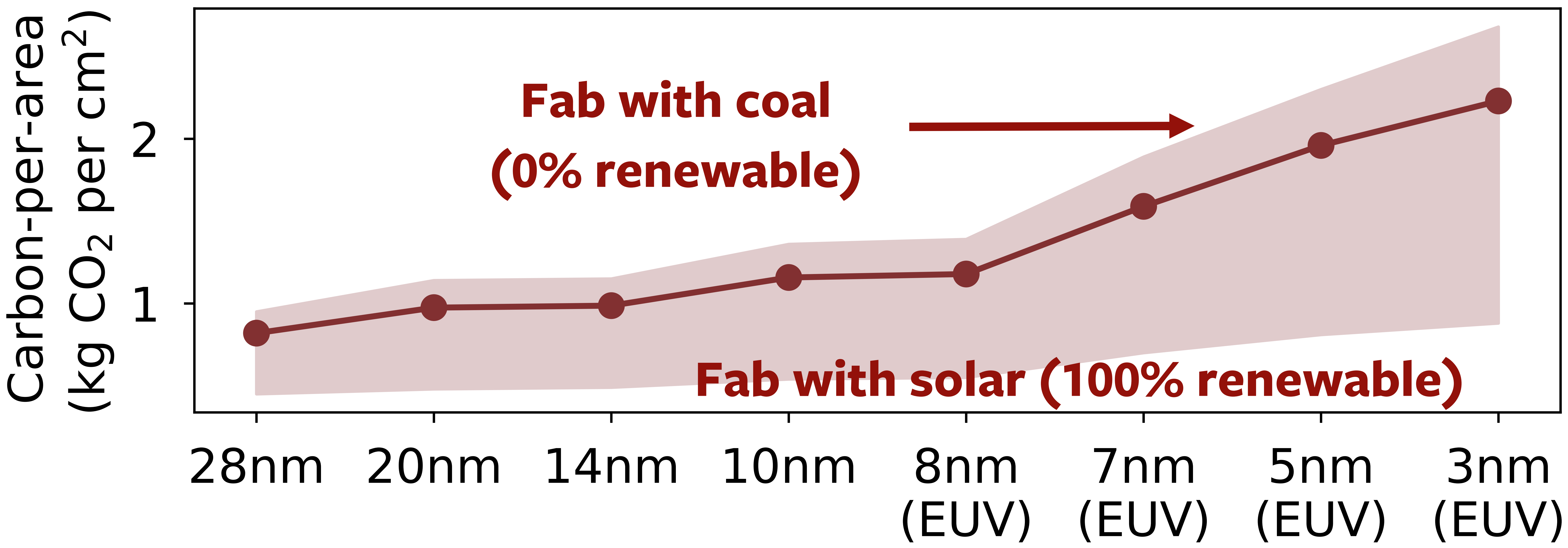


Data sources:

- *[IMEC] DTCO including Sustainability: Power-Performance-Area-Cost-Environmental score (PPACE) Analysis for Logic Technologies. Bardon et. al (IEDM 2020)*
- *[TSMC] TSMC Sustainability Reports 2018-2020*

Embodied carbon of application processors (SoC's)

$$Emb_{SoC} = Area \times \boldsymbol{CPA}$$

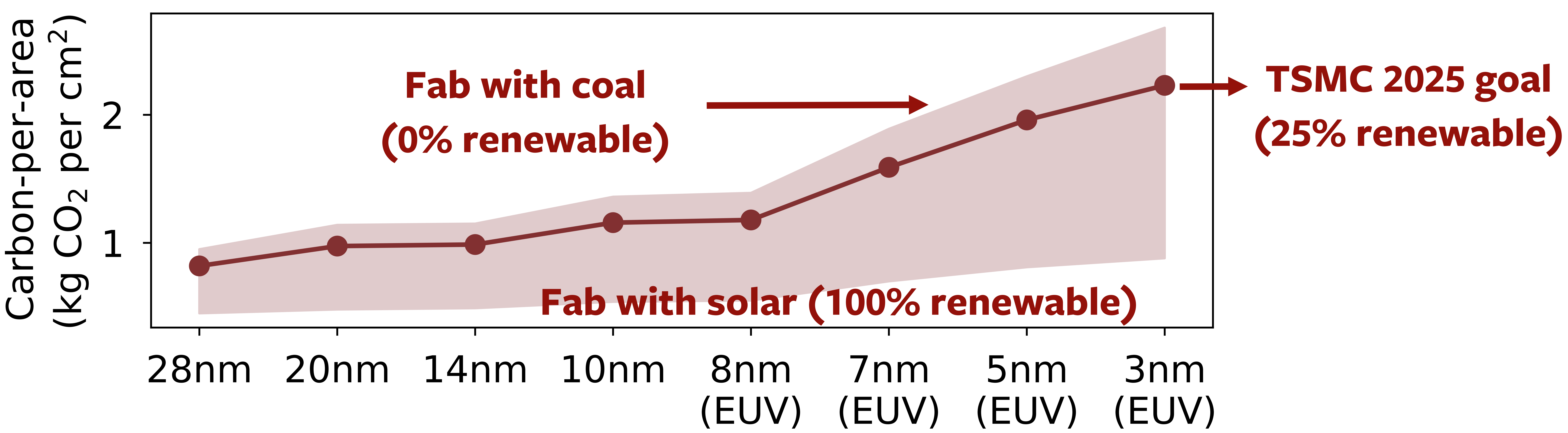


Data sources:

- *[IMEC] DTCO including Sustainability: Power-Performance-Area-Cost-Environmental score (PPACE) Analysis for Logic Technologies. Bardon et. al (IEDM 2020)*
- *[TSMC] TSMC Sustainability Reports 2018-2020*

Embodied carbon of application processors (SoC's)

$$Emb_{SoC} = Area \times \boldsymbol{CPA}$$



Data sources:

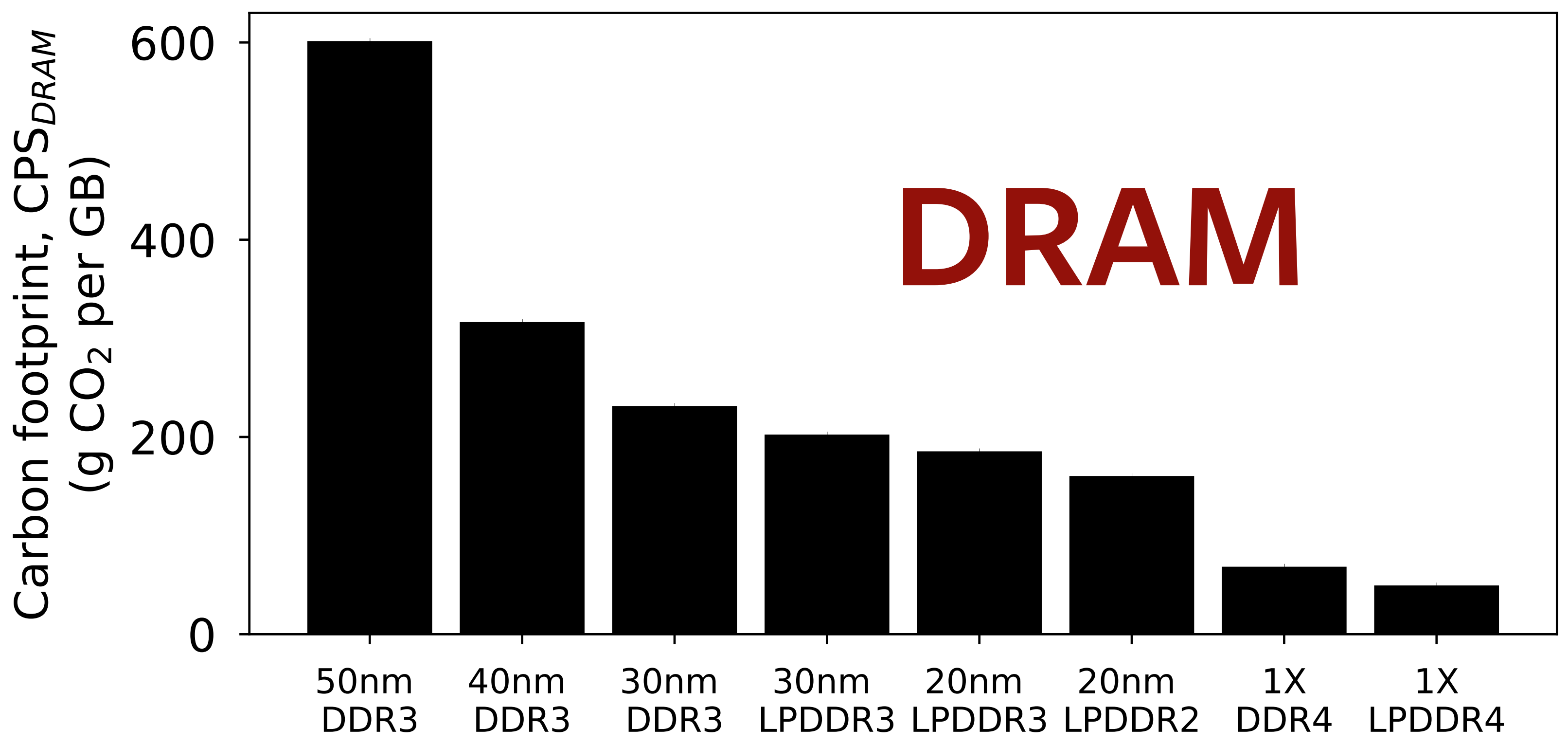
- *[IMEC] DTCO including Sustainability: Power-Performance-Area-Cost-Environmental score (PPACE) Analysis for Logic Technologies. Bardon et. al (IEDM 2020)*
- *[TSMC] TSMC Sustainability Reports 2018-2020*

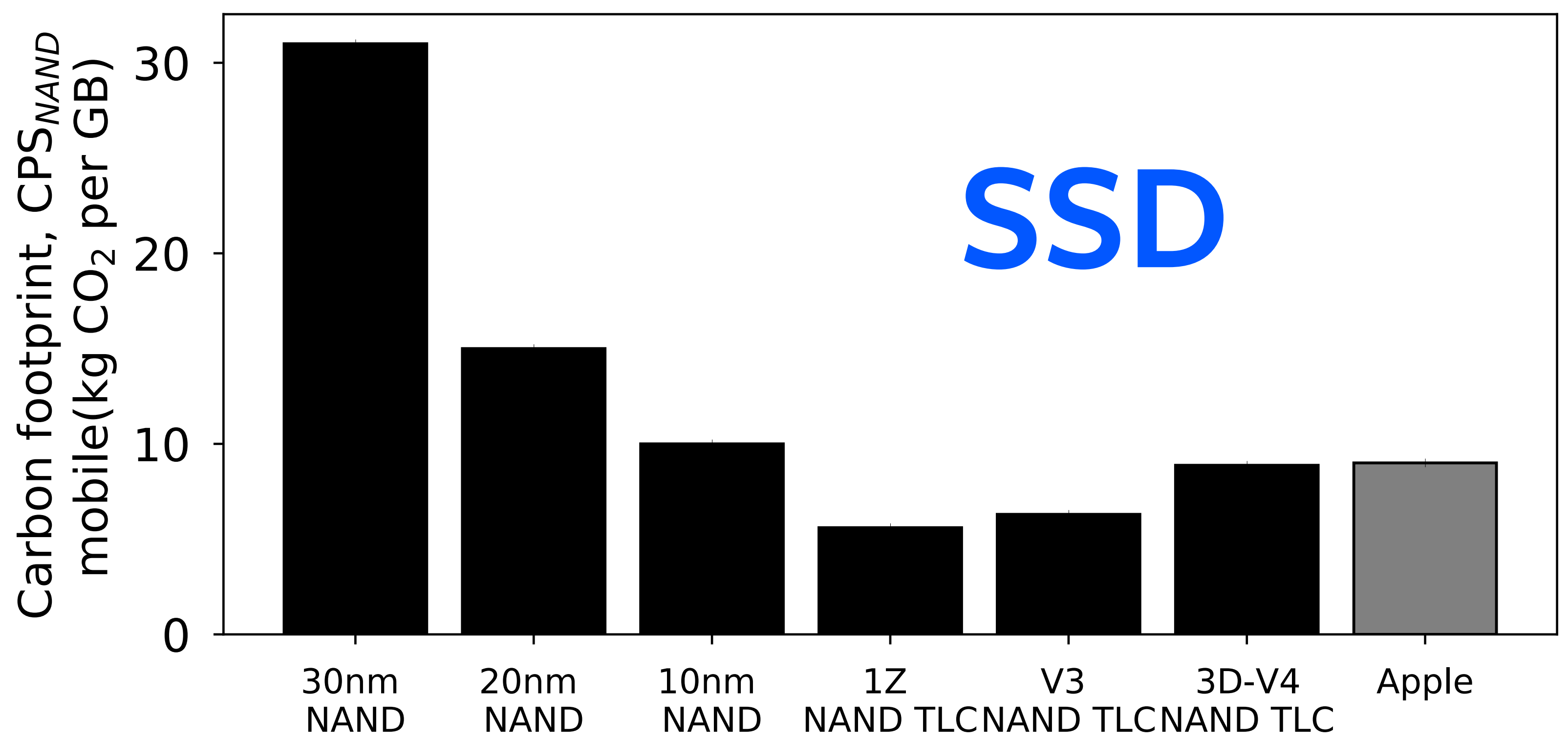
Additional details found in the paper...

Memory and storage

$$Emb_{DRAM} = DRAM_{capacity} \times CPS_{DRAM}$$

$$Emb_{SSD} = SSD_{capacity} \times CPS_{SSD}$$





Data sources: SK Hynix, Apple

ACT parameters

Parameter	Description	Range
T	App. execution time	From SW profiling
LT	HW lifetime	1-10 years
N_r	Number of ICs	From HW design
K_r	IC packaging footprint	0.15 kg CO_2
A	IC Area	From HW design (cm^2)
p	Process node	3-28 nm
MPA	Procure materials	~0.50kg CO_2 per cm^2
EPA	Fab energy	0.8-3.5 kWh per cm^2
CI_{use}	HW CO_2 intensity	30-700 g CO_2 per kWh
CI_{fab}	Fab CO_2 intensity	30-700 g CO_2 per kWh
GPA	GHG from fab	0.1-0.5 kg CO_2 per cm^2
Y	Fab yield	0-1
CPA	CO_2 from fab	0.1-0.4 kg CO_2 per cm^2
E_{DRAM}	DRAM embodied CO_2	0-0.6 kg CO_2 per GB
E_{SSD}	SSD embodied CO_2	0-0.03 kg CO_2 per GB
E_{HDD}	HDD embodied CO_2	0-0.12 kg CO_2 per GB

Challenge: How do we design sustainable systems by considering the footprint across lifecycles

This work: Architectural Carbon Modeling Tools (ACT)

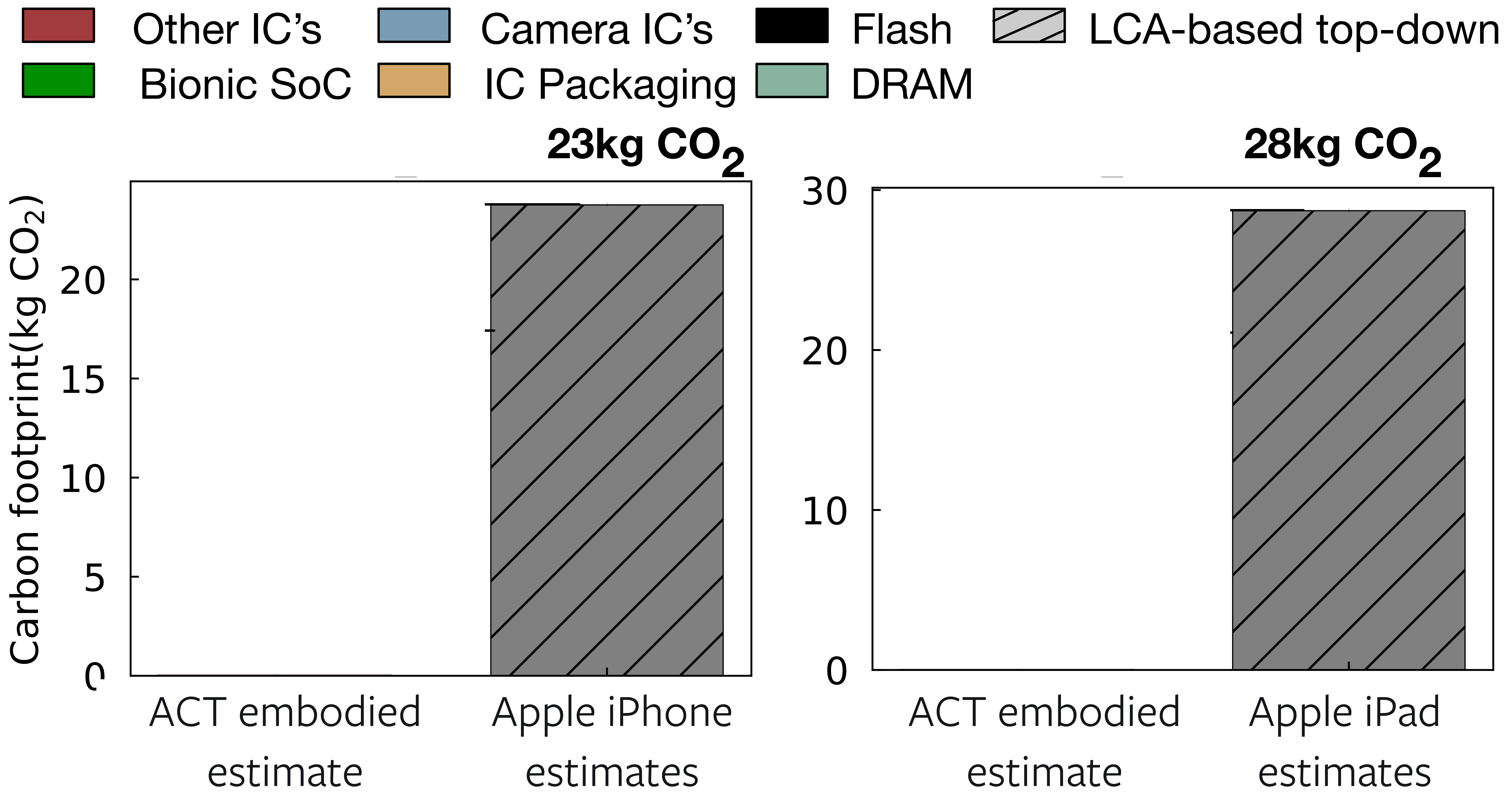
Overview of ACT model

Developed an ***extensible****, carbon model based on* ***industry data*** *for modern hardware architectures.*

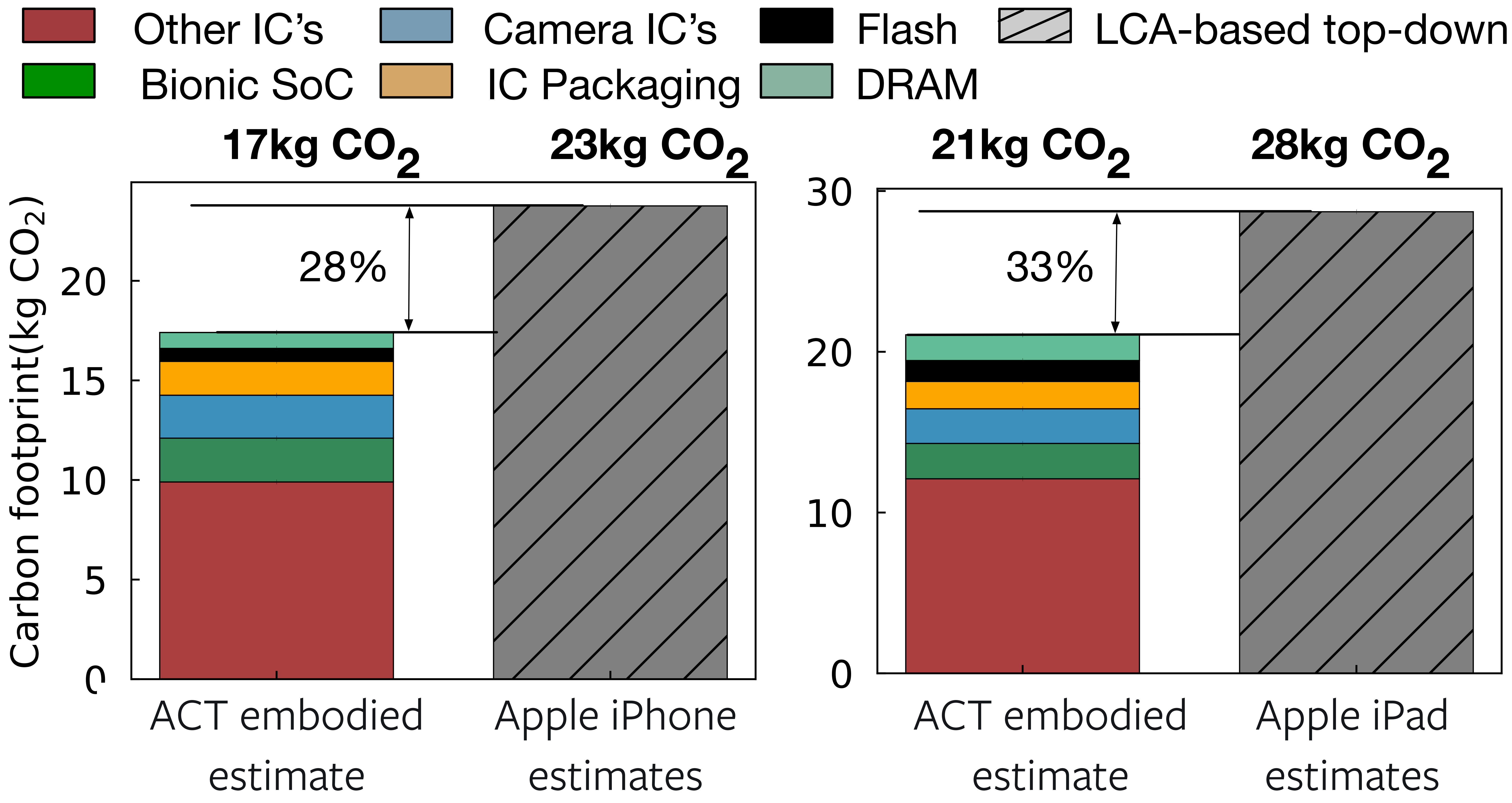
Comparing ACT to other methodologies

Sustainability aware-design case studies

Comparing ACT with Apple’s product environmental reports



Comparing ACT with Apple's product environmental reports



This work: Architectural Carbon Modeling Tools (ACT)

Overview of ACT model

*Developed an **extensible**, carbon model based on **industry data** for modern hardware architectures.*

Comparing ACT to other methodologies

*ACT provides first order carbon **estimates of modern** systems*

Sustainability aware-design case studies

Tenets of Environmental Design

Reduce

Design leaner footprint software and hardware.

Recycle

Recover discarded systems and components.

Reuse

Repurpose systems already produce.

Tenets of Environmental Design

Reduce

Design leaner footprint software and hardware.

Recycle

Recover discarded systems and components.

Reuse

Repurpose systems already produce.

Reuse: General purpose versus custom mobile HW

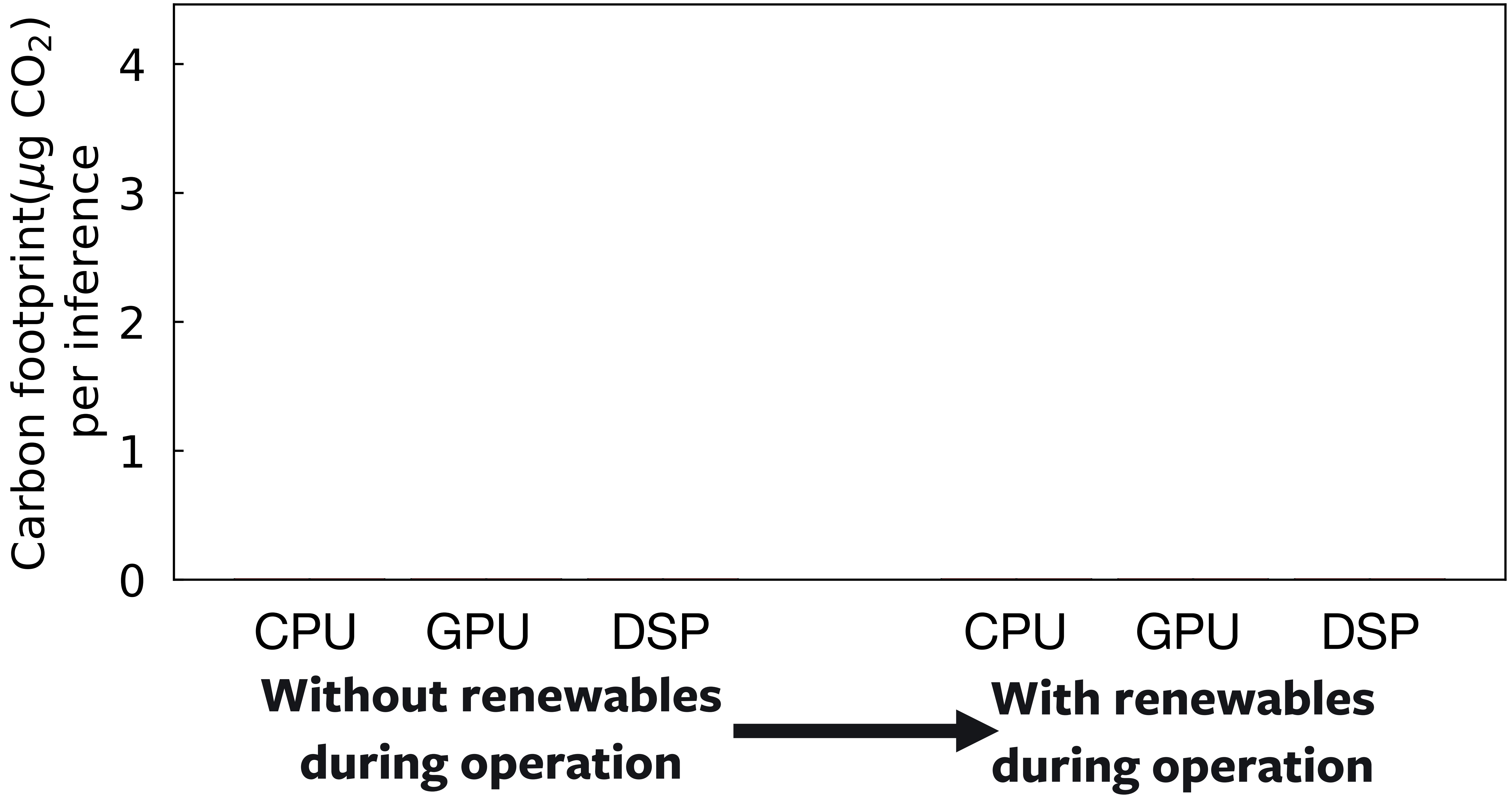
AI inference case study (MobileNet) assuming 3 year hardware lifetime, and same utilization in all cases

Commodity (mobile CPU)
Specialized (mobile GPU, DSP)
→ **Architectural Carbon Model** → Carbon per inference

Reuse: General purpose versus custom mobile HW

AI inference case study (MobileNet) assuming 3 year hardware lifetime, and same utilization in all cases

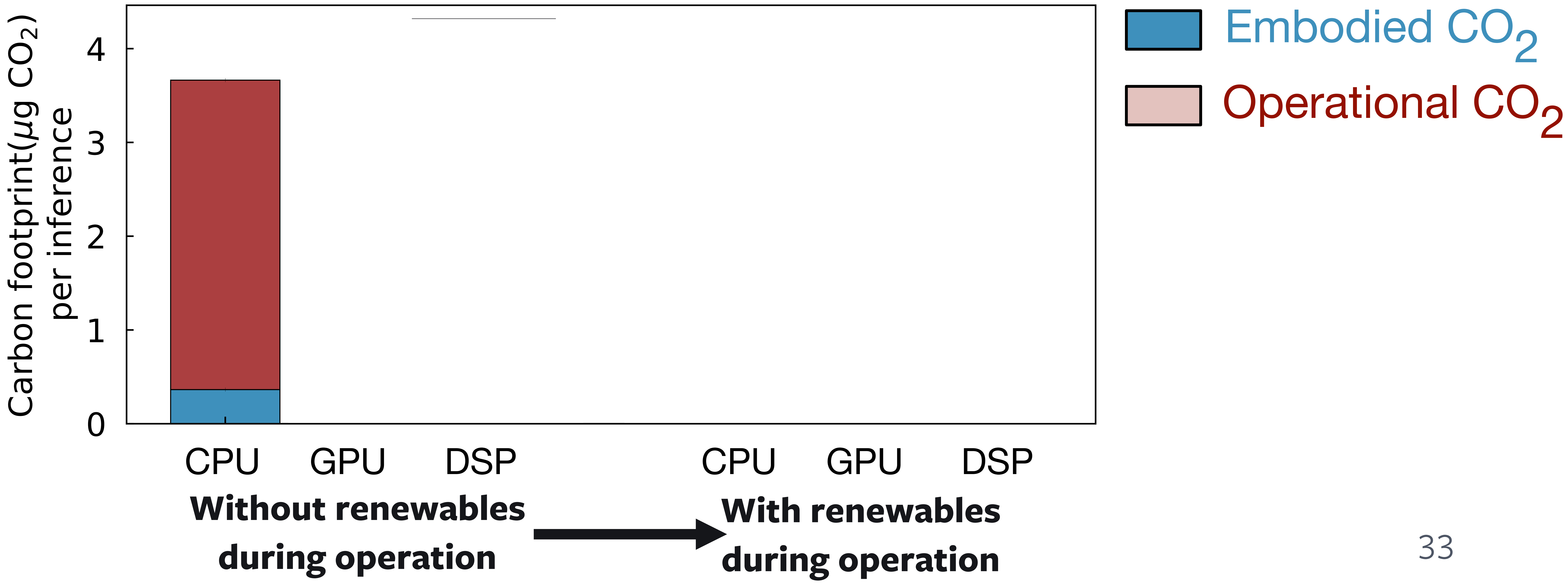
Commodity (mobile CPU)
Specialized (mobile GPU, DSP)
→ **Architectural Carbon Model** → Carbon per inference



Reuse: General purpose versus custom mobile HW

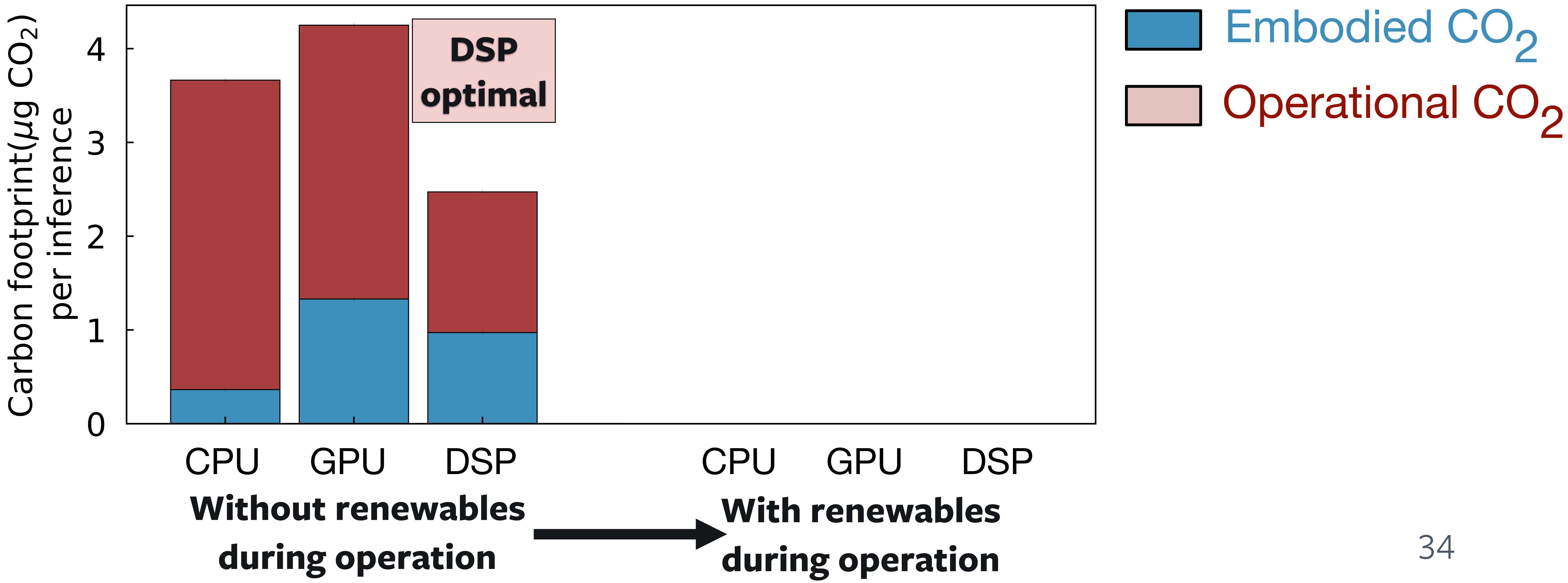
AI inference case study (MobileNet) assuming 3 year hardware lifetime, and same utilization in all cases

{ Commodity (mobile CPU)
Specialized (mobile GPU, DSP) } → **Architectural Carbon Model** → Carbon per inference



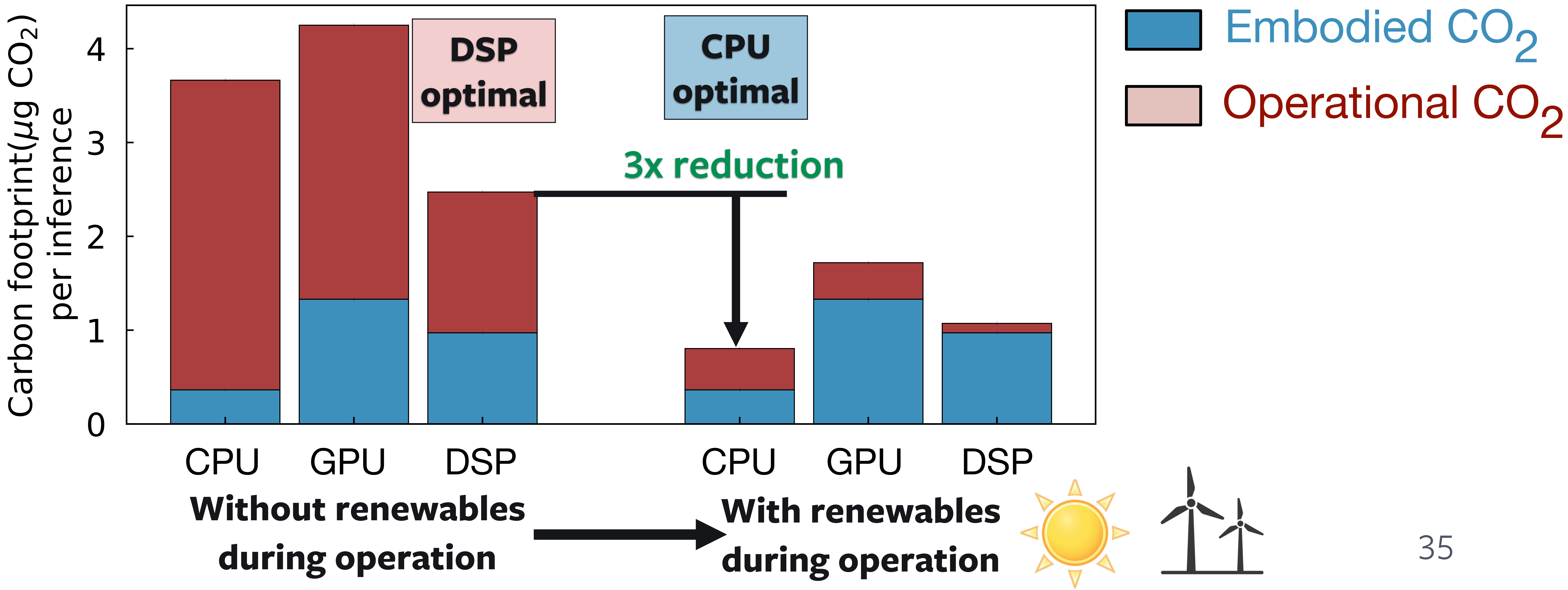
Reuse: General purpose versus custom mobile HW

AI inference case study (MobileNet) assuming 3 year hardware lifetime, and same utilization in all cases



<u>Reuse</u>: General purpose versus custom <u>mobile</u> HW

AI inference case study (MobileNet) assuming 3 year hardware lifetime, and same utilization in all cases



Tenets of Environmental Design

Reduce

Design leaner footprint software and hardware.

Recycle

Recover discarded systems and components.

Reuse

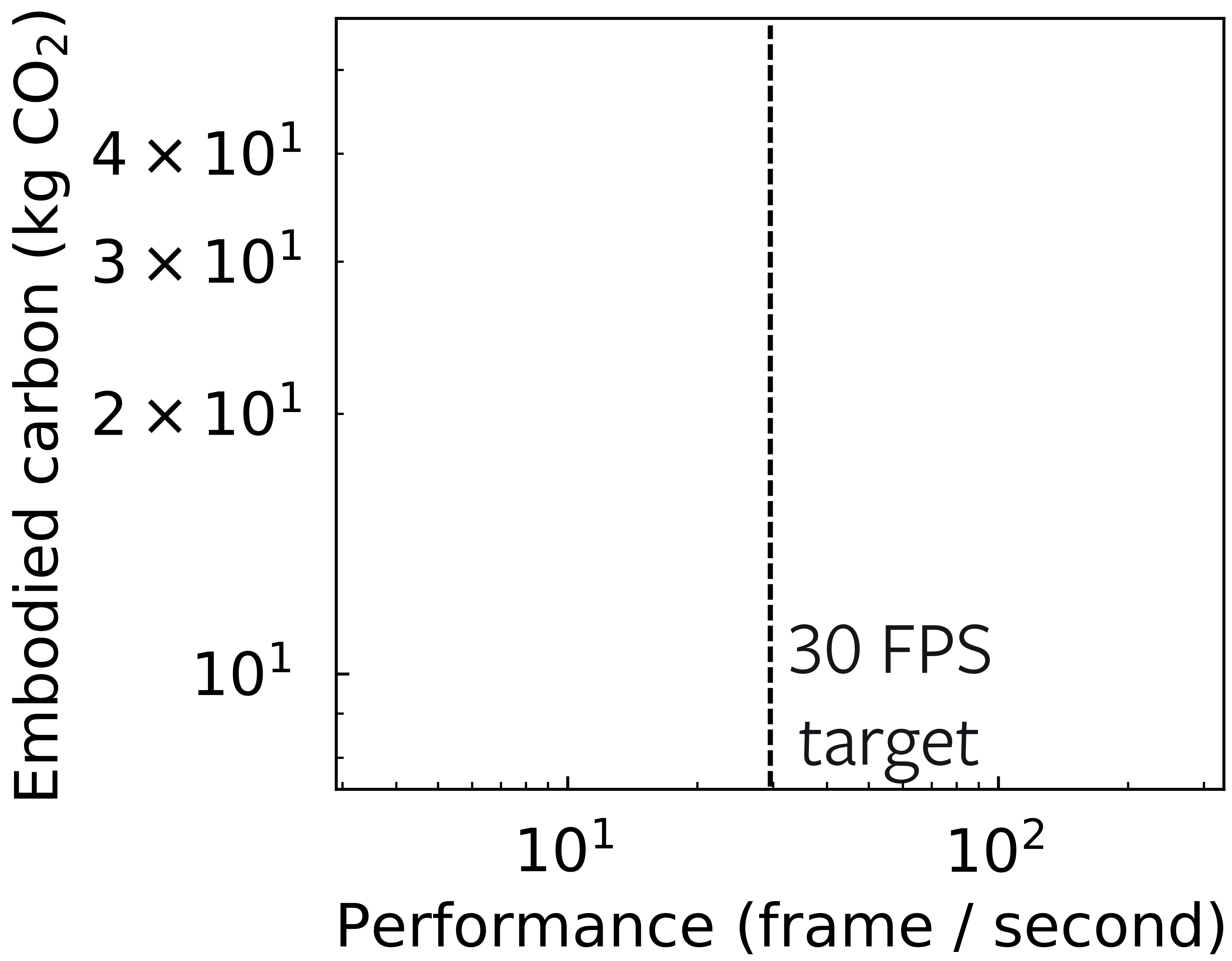
Repurpose systems already produce.

Reduce: Designing leaner hardware systems

Nvidia DL Accelerator (NVDLA) → **Architectural Carbon Model** → Performance/energy vs. Carbon

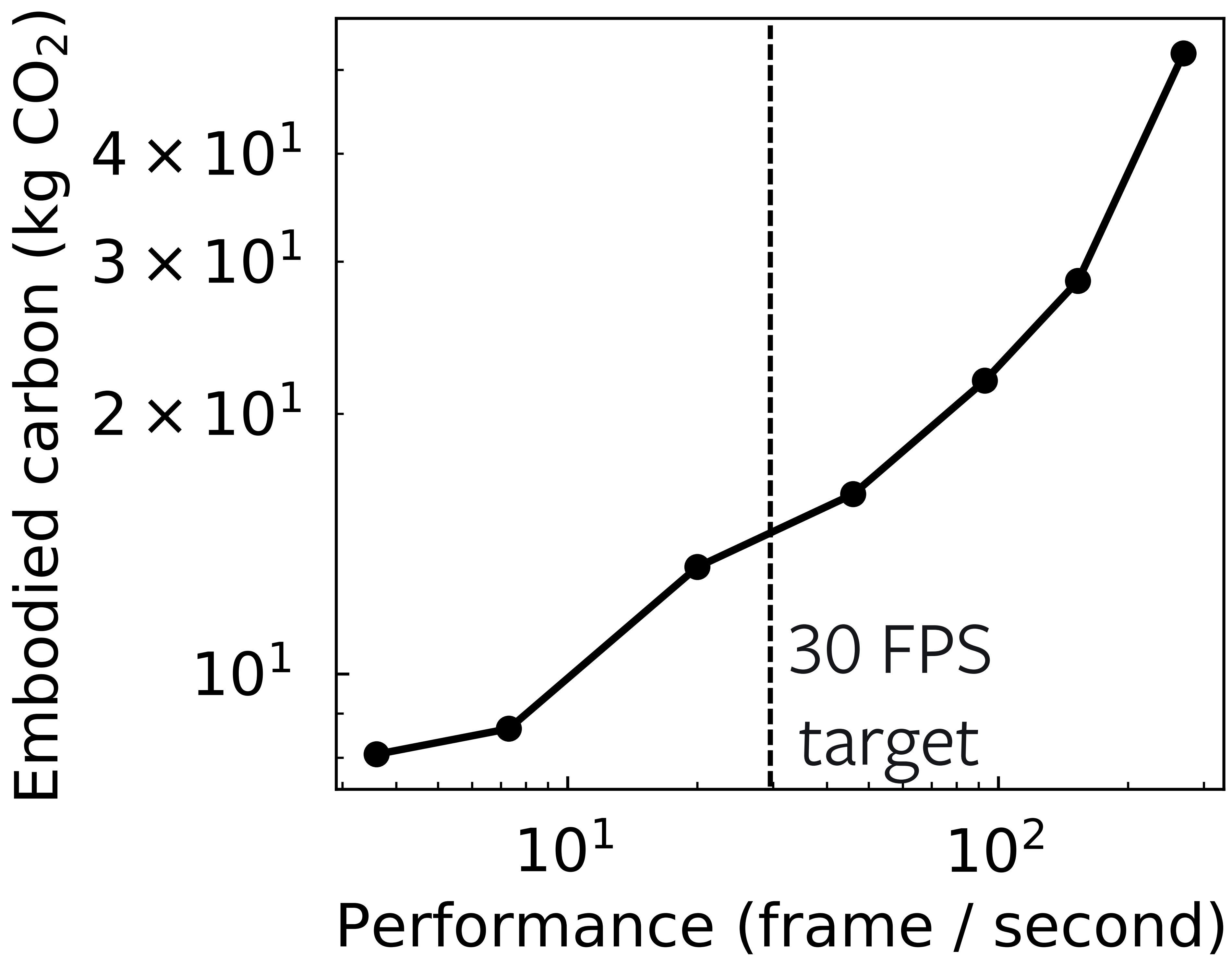
Reduce: Designing leaner hardware systems

{ Nvidia DL Accelerator (NVDLA) } → **Architectural Carbon Model** → Performance/energy vs. Carbon



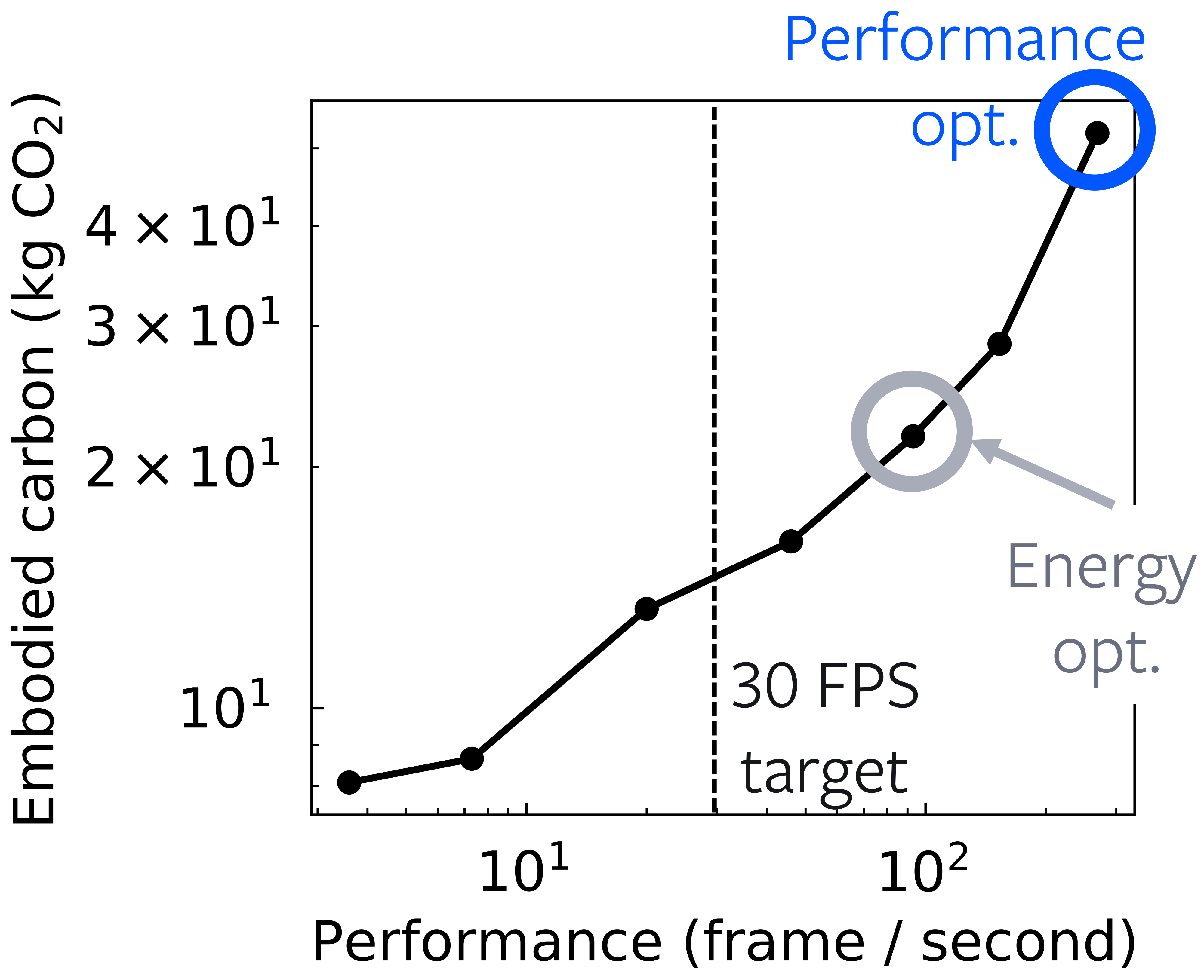
Reduce: Designing leaner hardware systems

Nvidia DL Accelerator (NVDLA) → **Architectural Carbon Model** → Performance/energy vs. Carbon

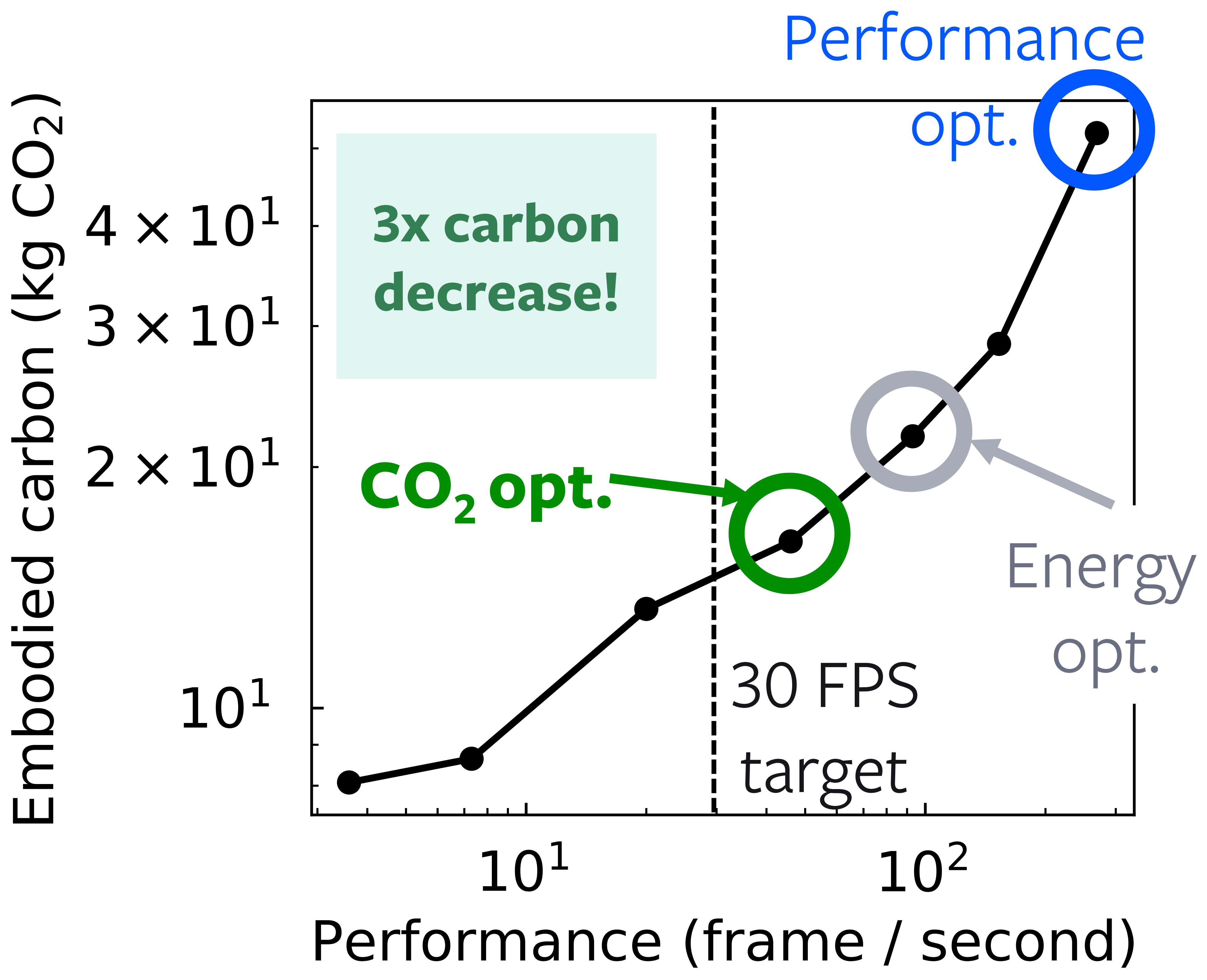


Reduce: Designing leaner hardware systems

{ Nvidia DL Accelerator (NVDLA) } → **Architectural Carbon Model** → Performance/energy vs. Carbon

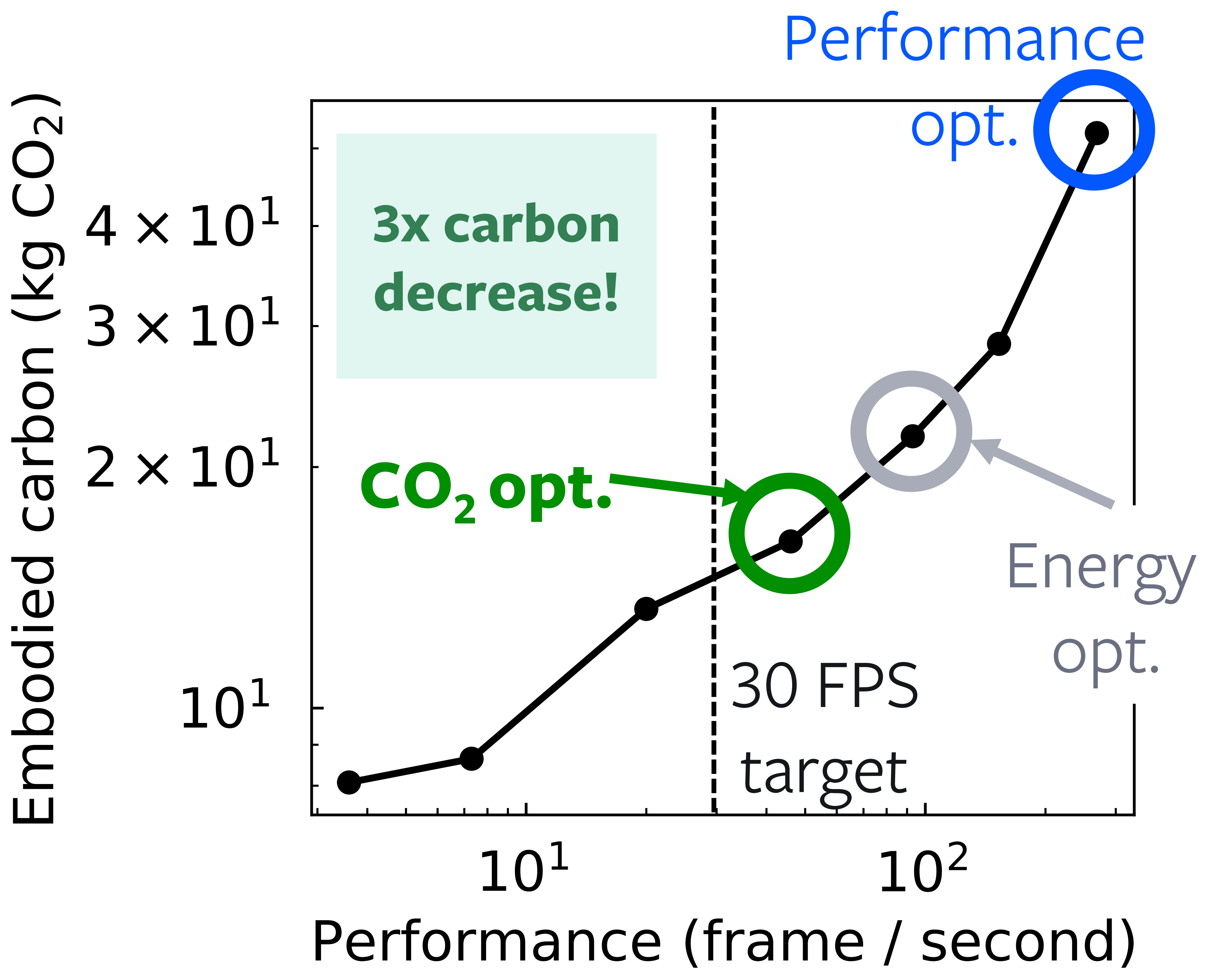


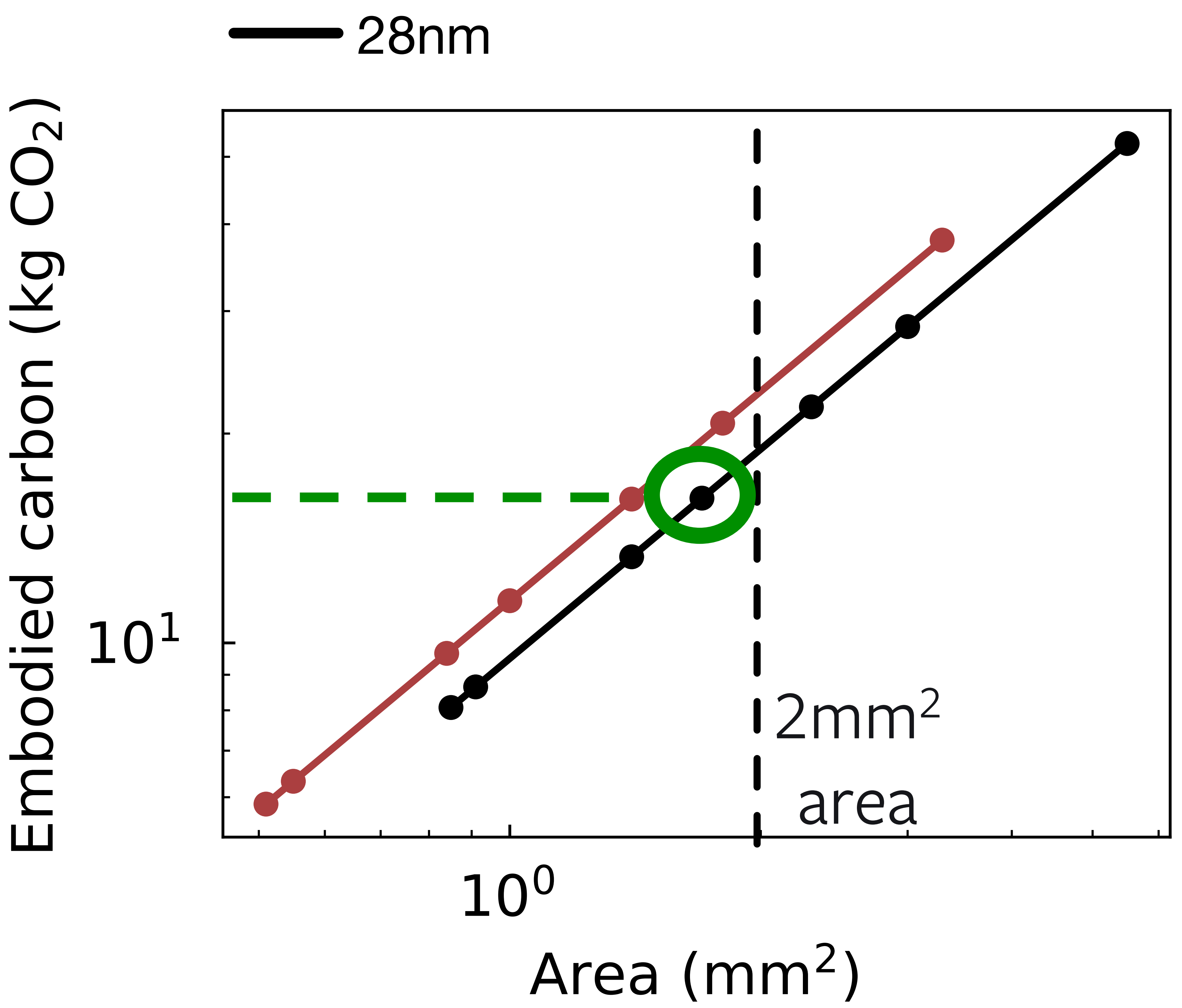
Reduce: Designing leaner hardware systems



Reduce: Designing leaner hardware systems

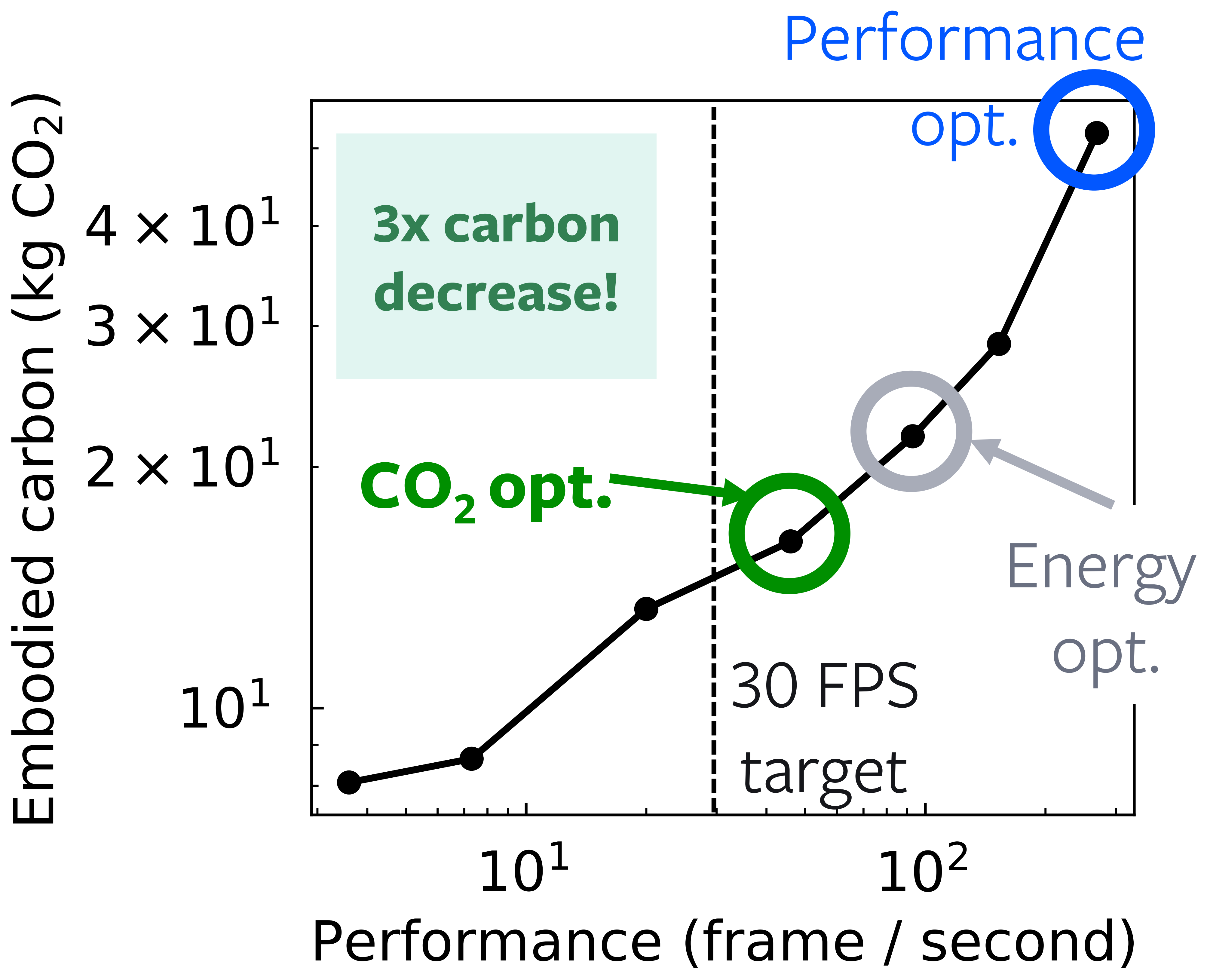
{ Nvidia DL Accelerator (NVDLA) } → **Architectural Carbon Model** → Performance/energy vs. Carbon

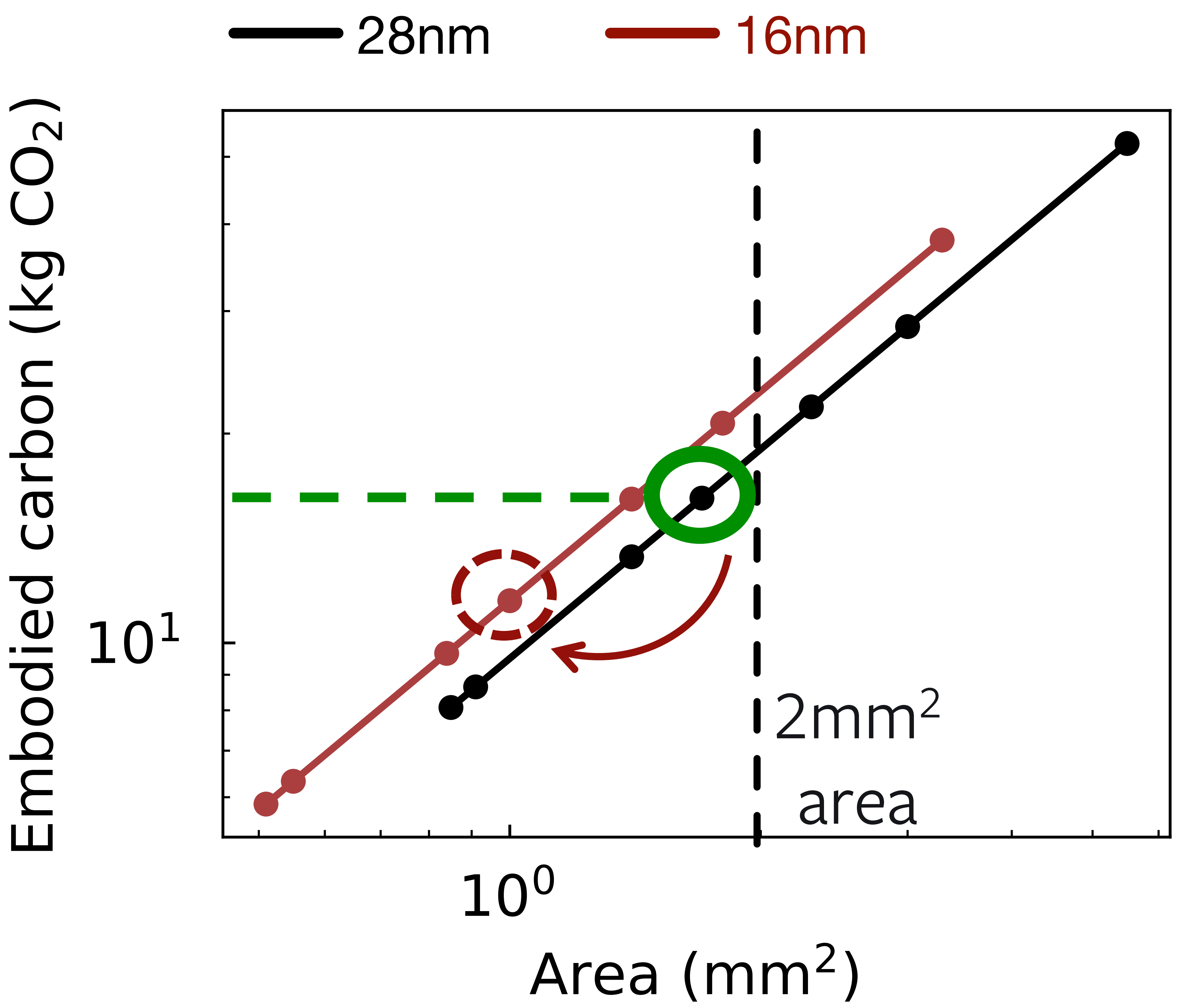




Reduce: Designing leaner hardware systems

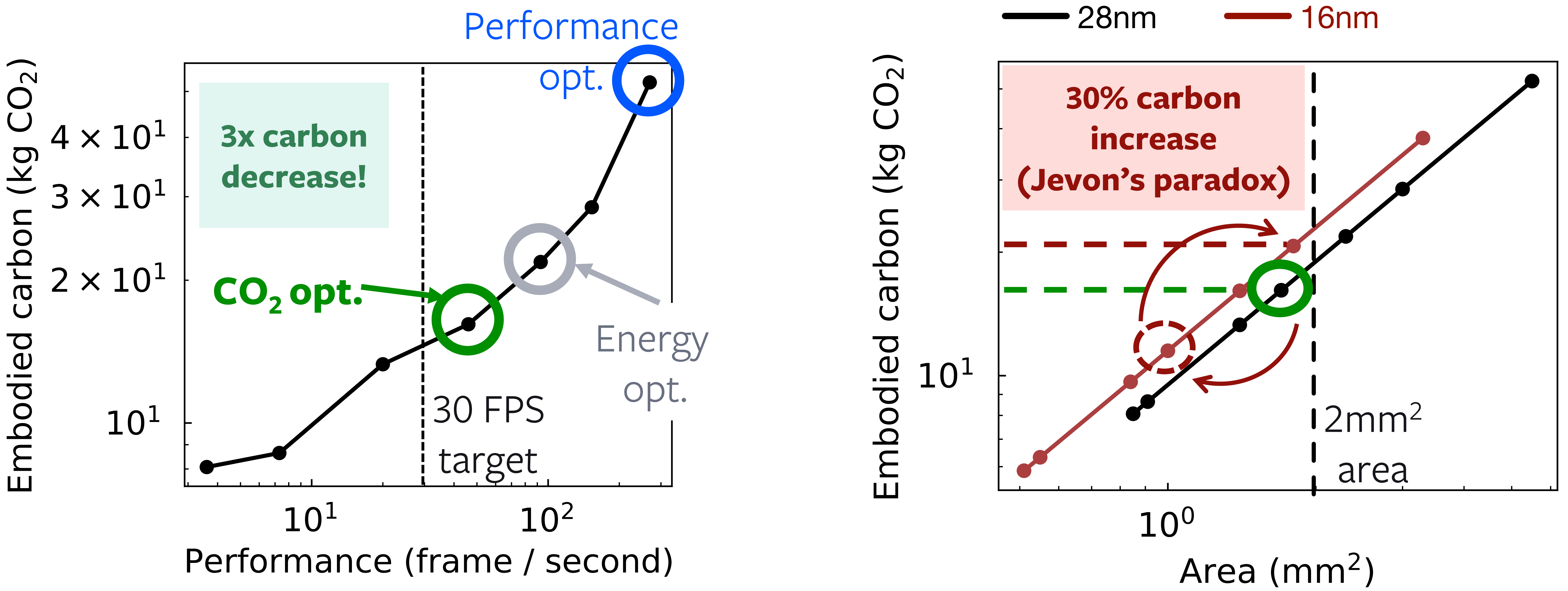
{ Nvidia DL Accelerator (NVDLA) } → **Architectural Carbon Model** → Performance/energy vs. Carbon





Reduce: Designing leaner hardware systems

{ Nvidia DL Accelerator (NVDLA) } → **Architectural Carbon Model** → Performance/energy vs. Carbon



Challenge: How do we design sustainable systems by considering the footprint across lifecycles

This work: Architectural Carbon Modeling Tools (ACT)

Overview of ACT model

*Developed an **extensible**, carbon model based on **industry data** for modern hardware architectures.*

Comparing ACT to other methodologies

*ACT provides first order carbon **estimates of modern** systems*

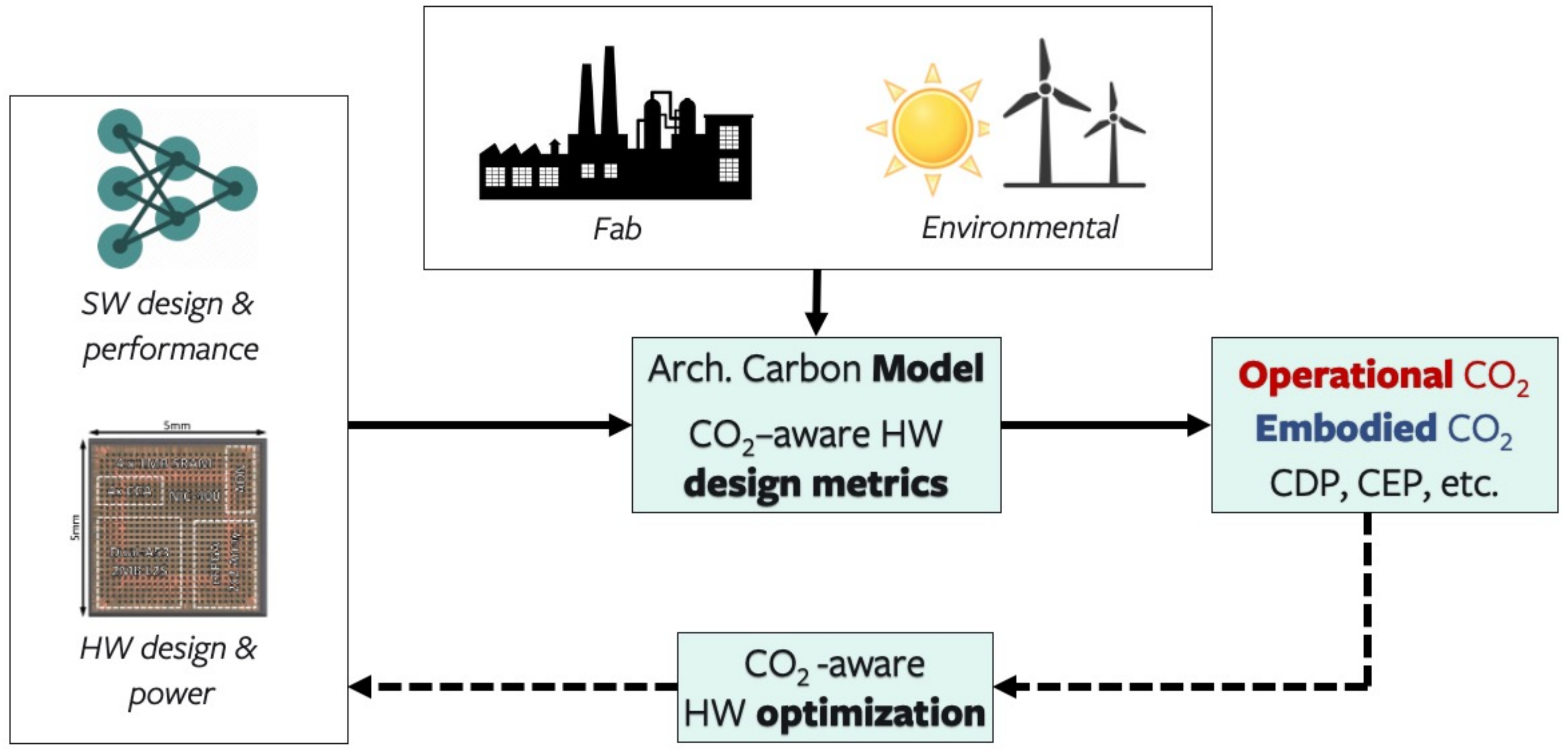
Sustainability aware-design case studies

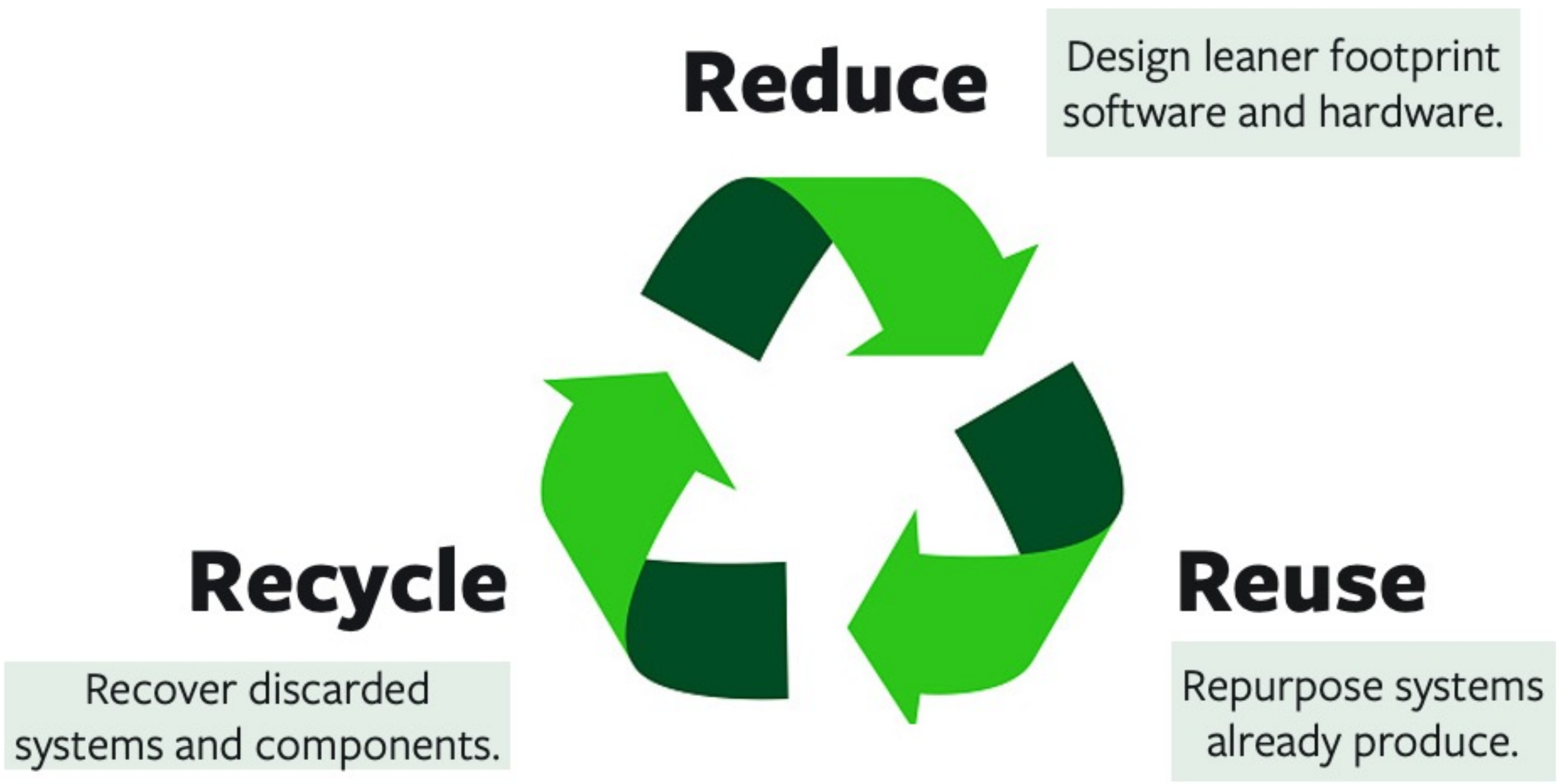
*Can eliminate carbon footprint by up **3x** (Reduce) and **2x** (Reuse) and **2x** (Recycle)*

This work: ACT

Develop the model

Case studies

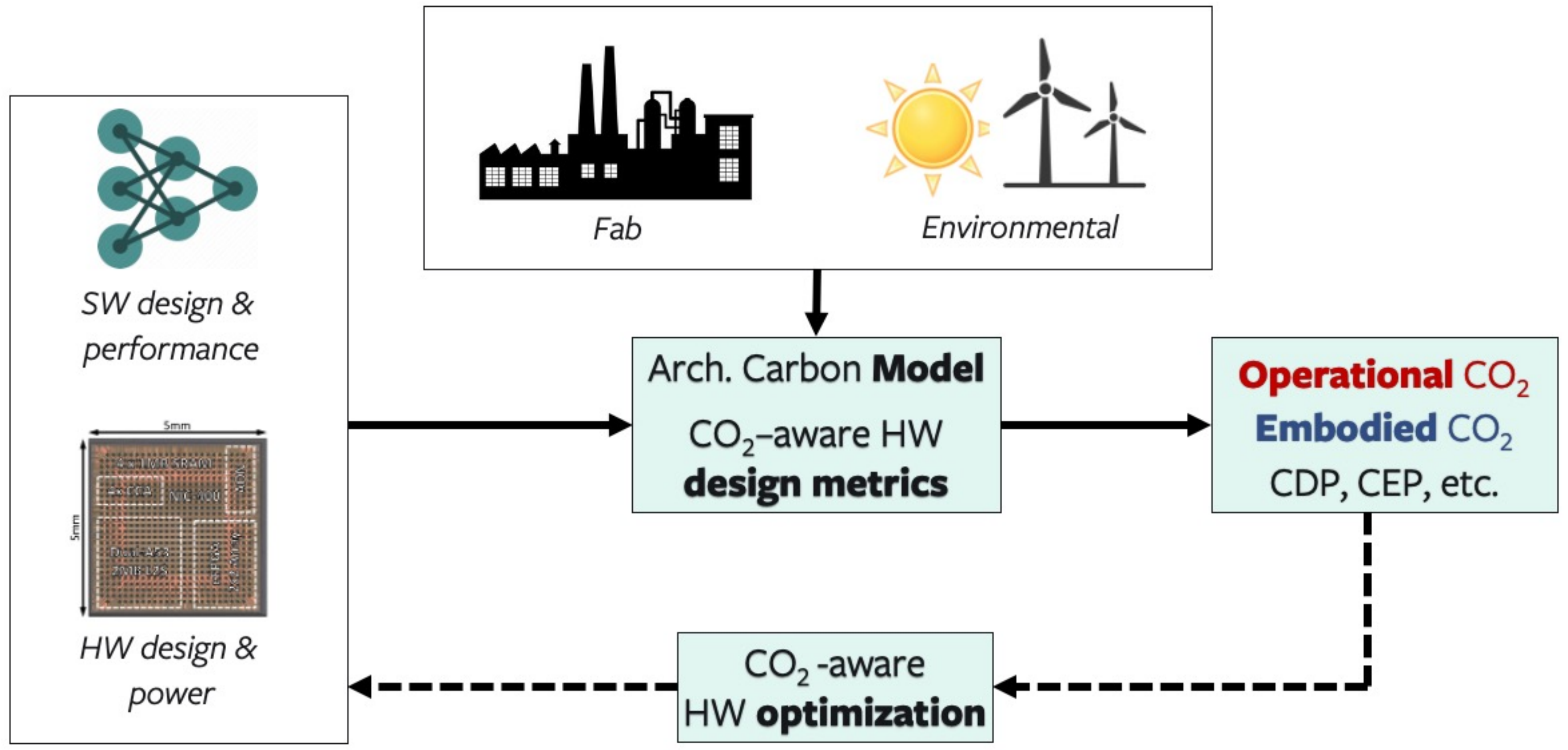


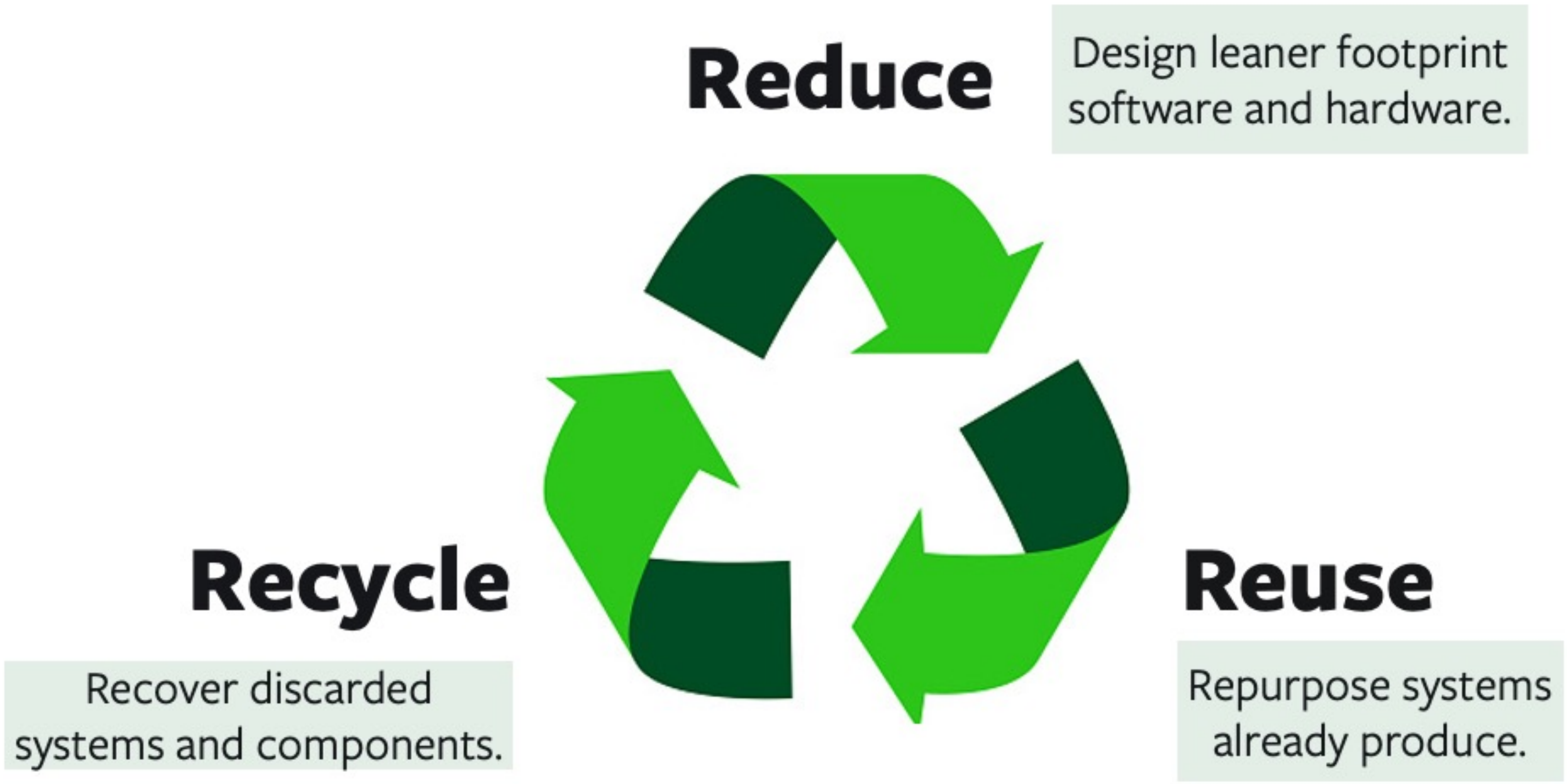


This work: ACT

Develop the model

Case studies



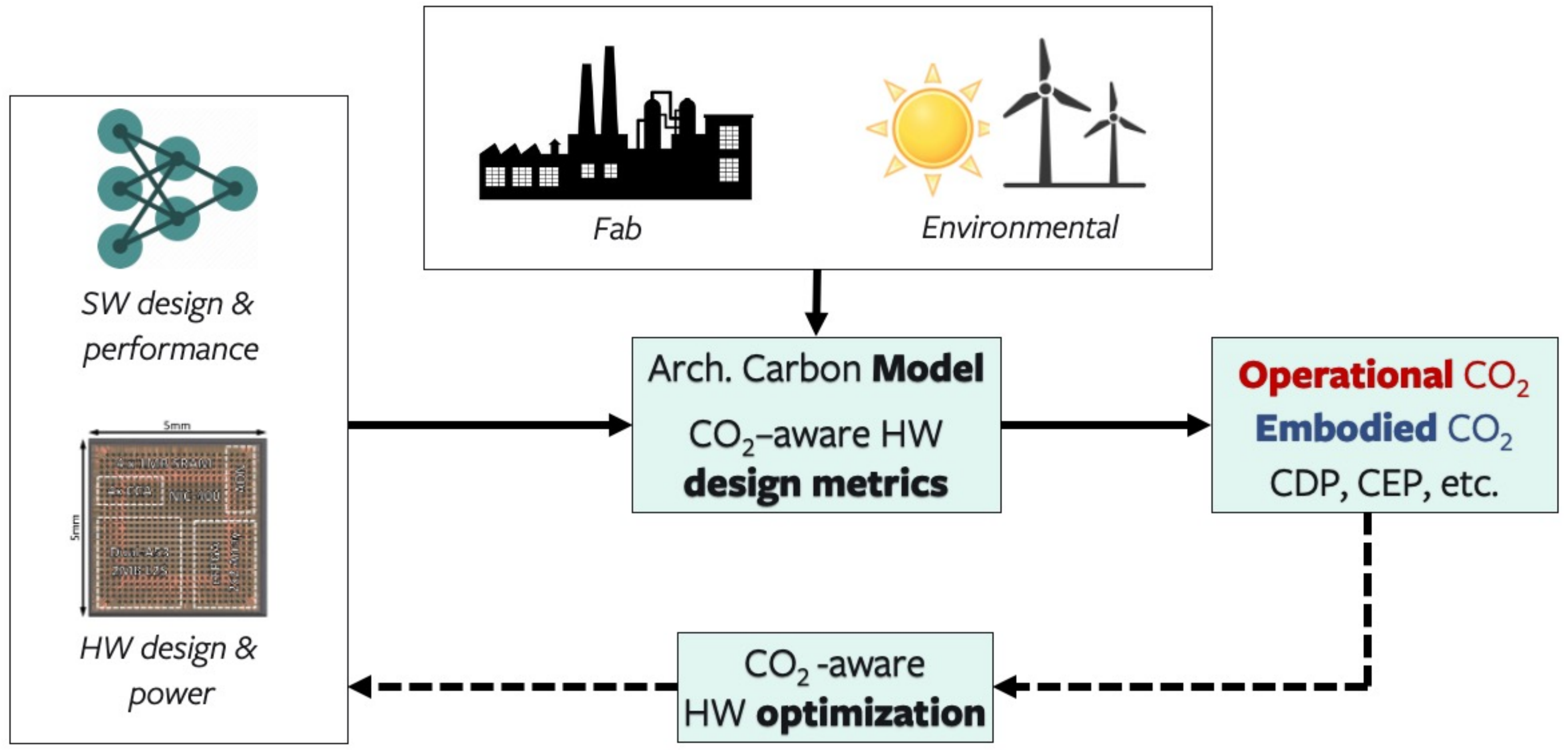


More details in the paper!

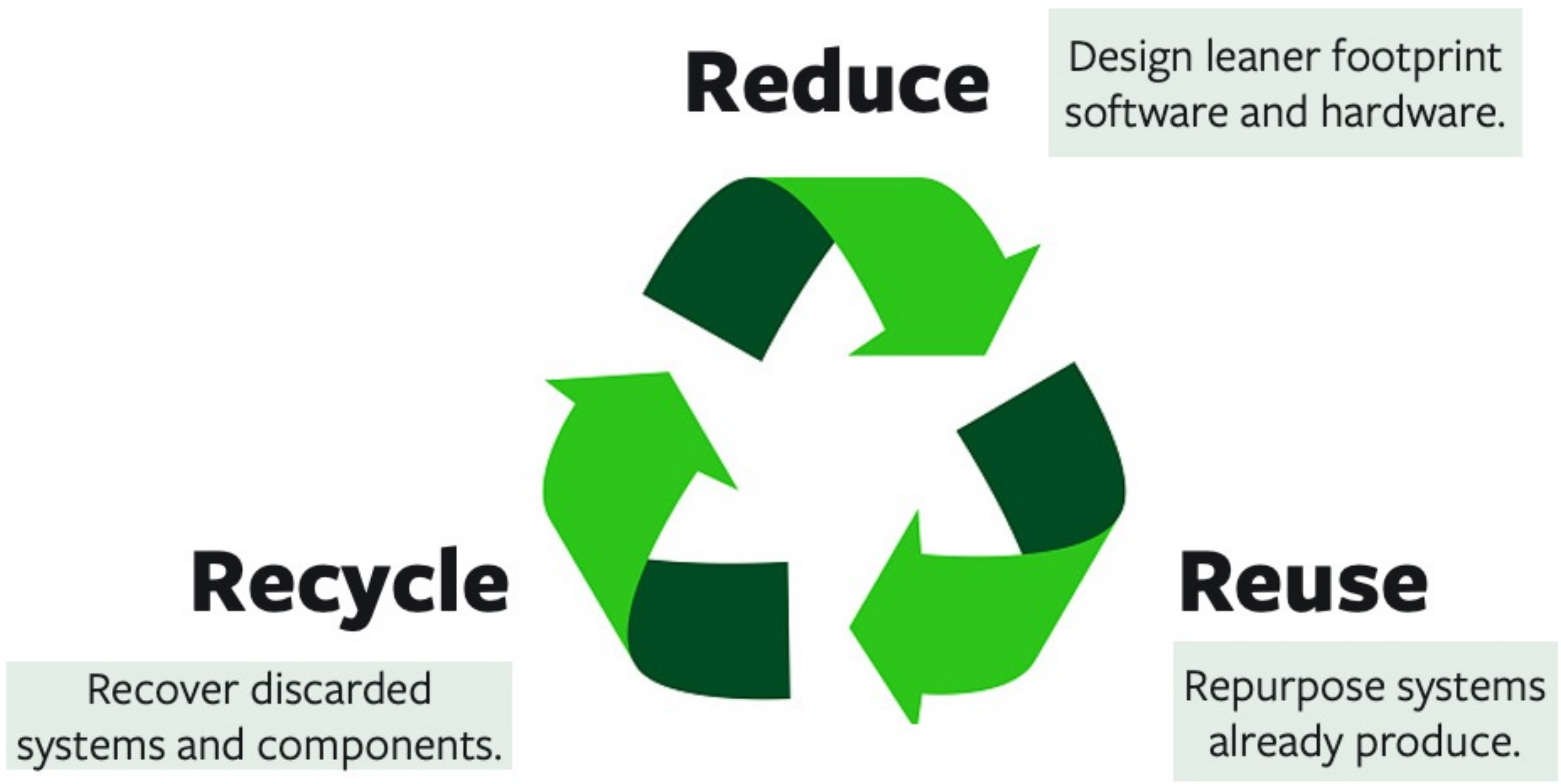
- **Modeling parameters** and industry **sources** for data
- **Carbon-aware metrics** for early DSE (e.g., EDP, CDP, CEP)
- Detailed **comparison** against industry LCA's
- Reuse case study: impact of **reconfigurable accelerators** (FPGA's)
- Recycle case study: Enabling **second life** & SSD provisioning

This work: ACT

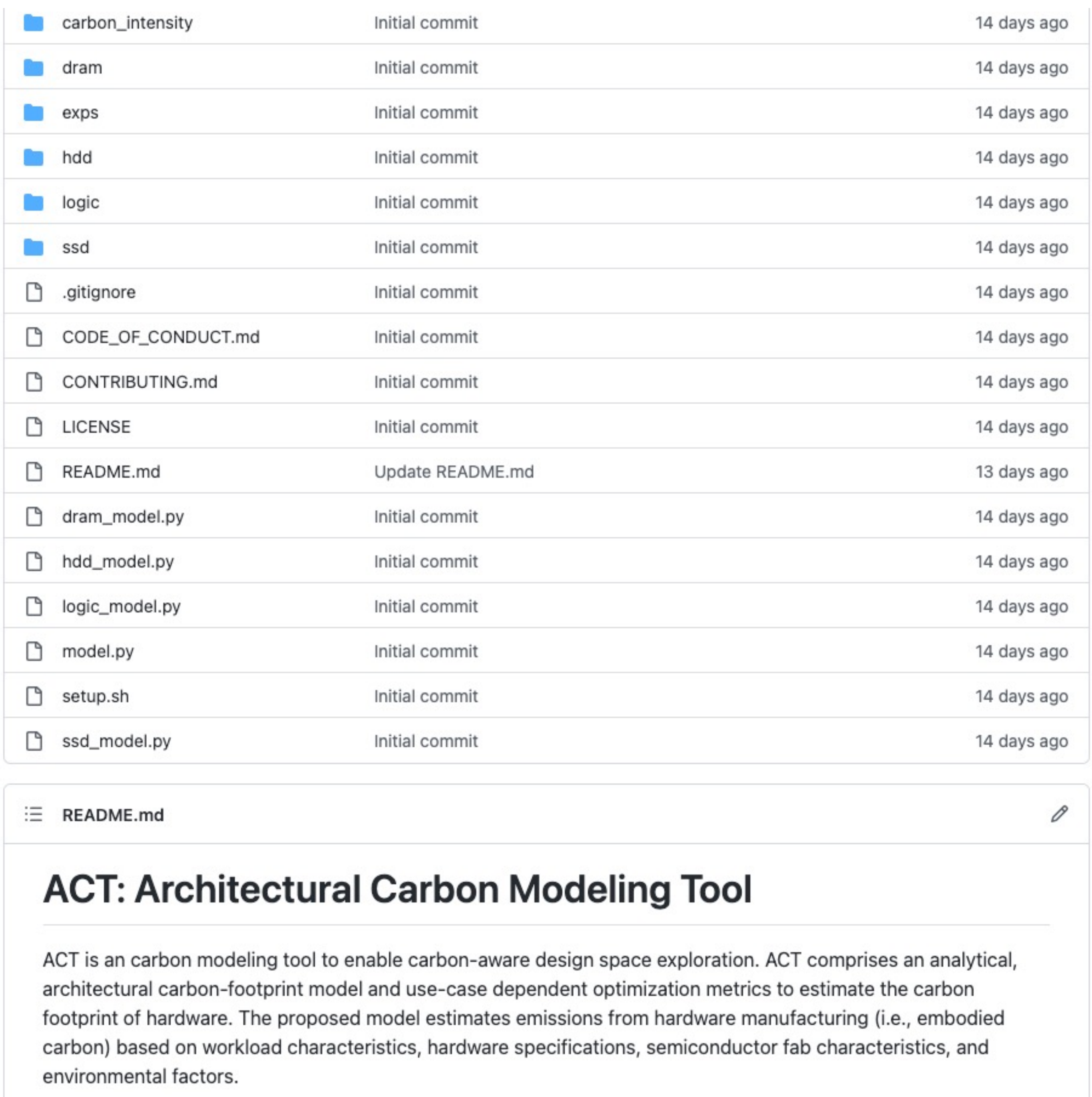
Develop the model



Case studies



Open-source!



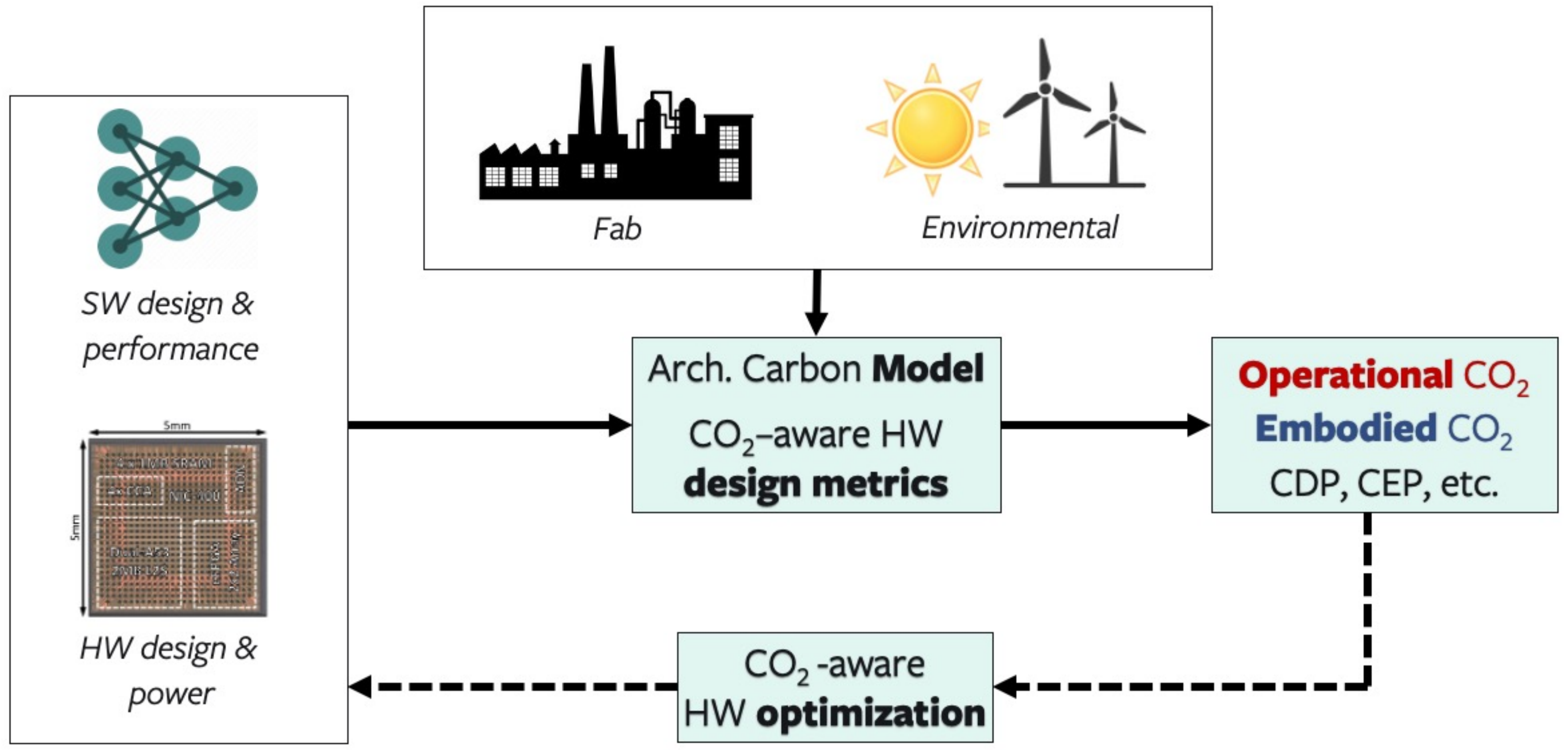
More details in the paper!

- **Modeling parameters** and industry **sources** for data
- **Carbon-aware metrics** for early DSE (e.g., EDP, CDP, CEP)
- Detailed **comparison** against industry LCA's
- Reuse case study: impact of **reconfigurable accelerators** (FPGA's)
- Recycle case study: Enabling **second life** & SSD provisioning

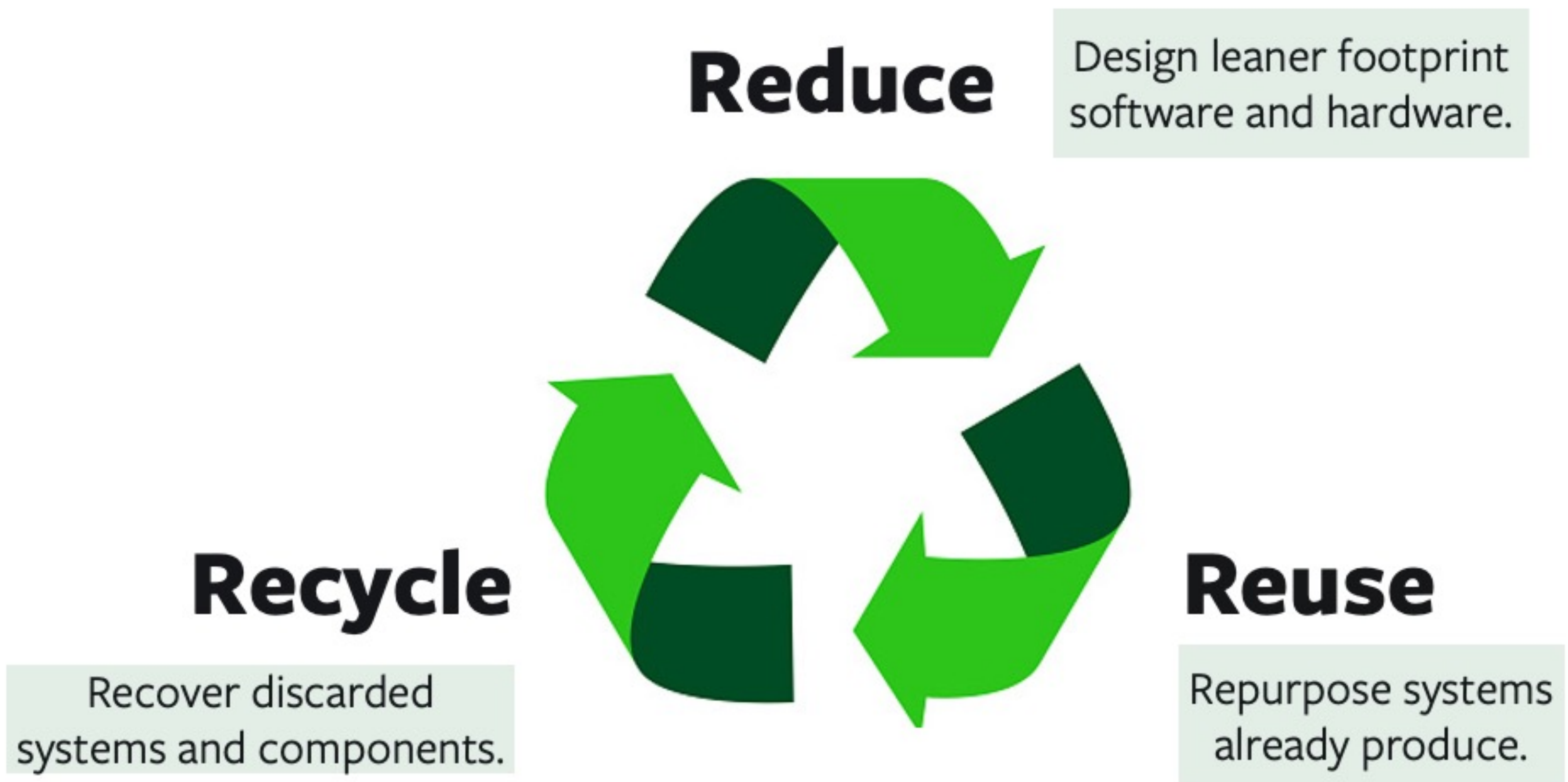
Thank you!

This work: ACT

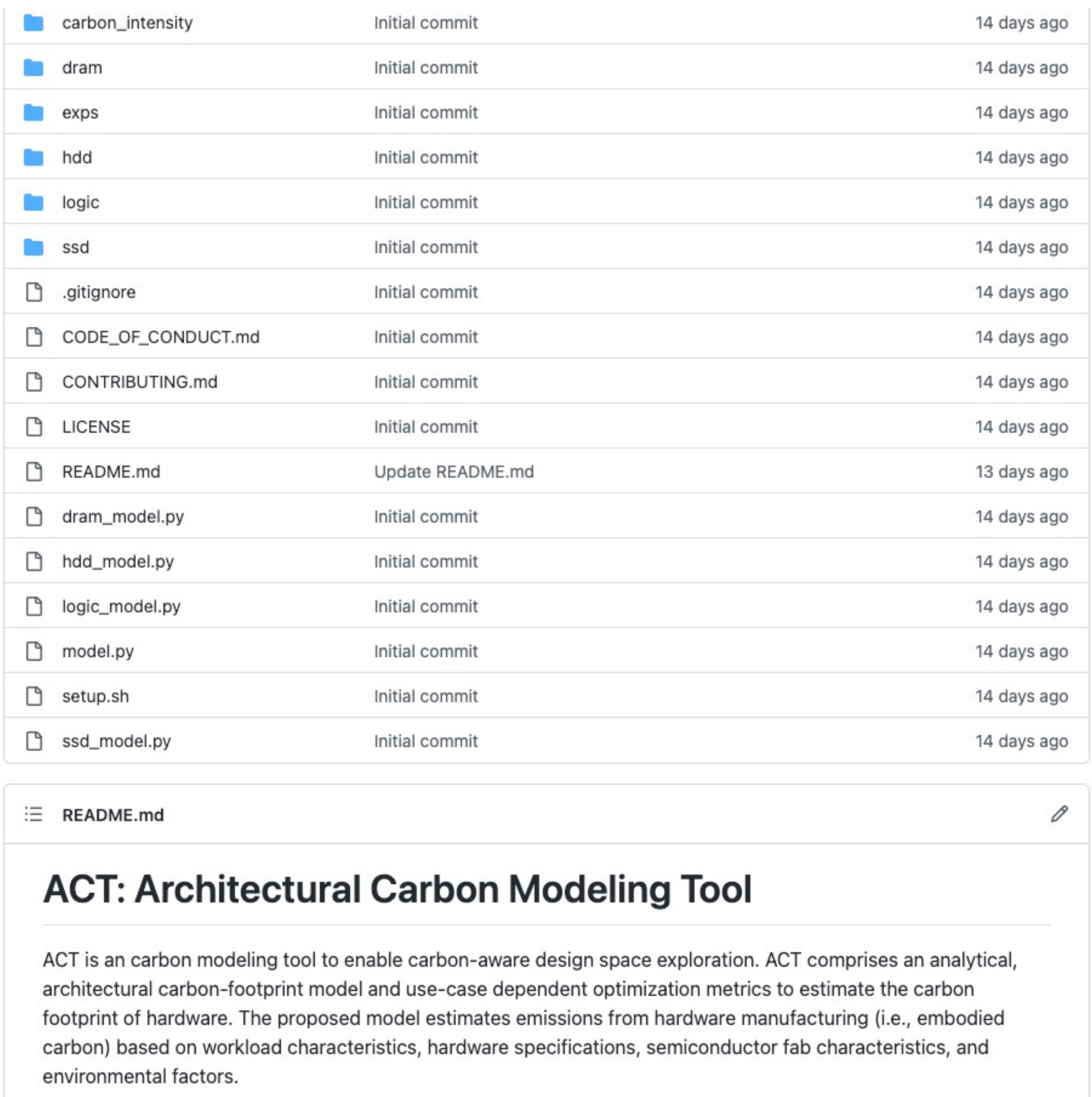
Develop the model



Case studies



Open-source!

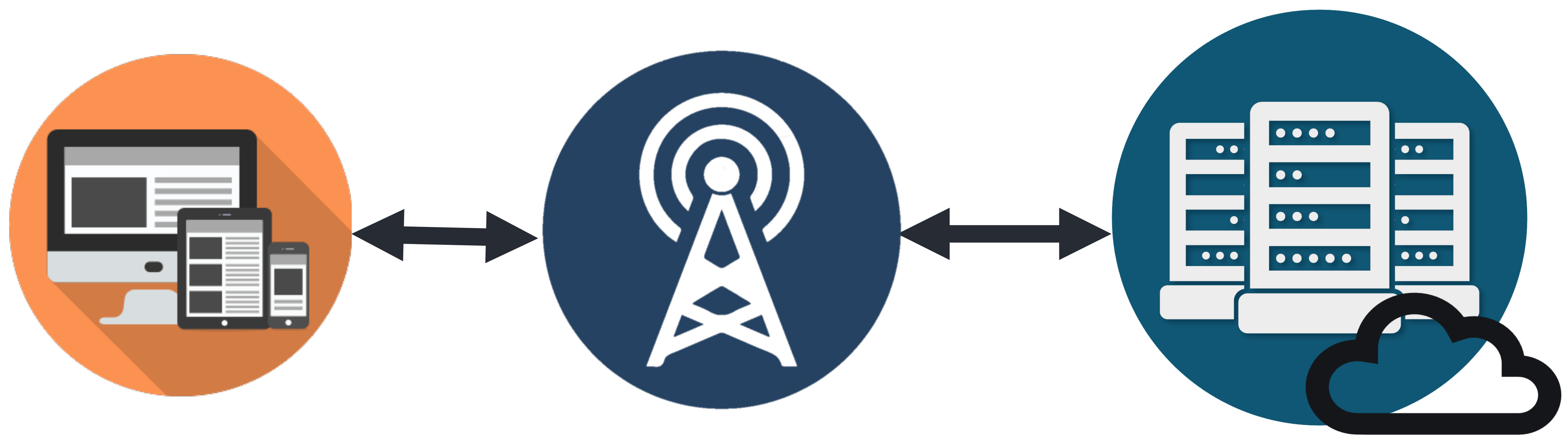


More details in the paper!

- **Modeling parameters** and industry **sources** for data
- **Carbon-aware metrics** for early DSE (e.g., EDP, CDP, CEP)
- Detailed **comparison** against industry LCA's
- Reuse case study: impact of **reconfigurable accelerators** (FPGA's)
- Recycle case study: Enabling **second life** & SSD provisioning

700 Million tons of CO_2

- Half the aviation industry's footprint
- 2.1 - 3.9% of worldwide emissions

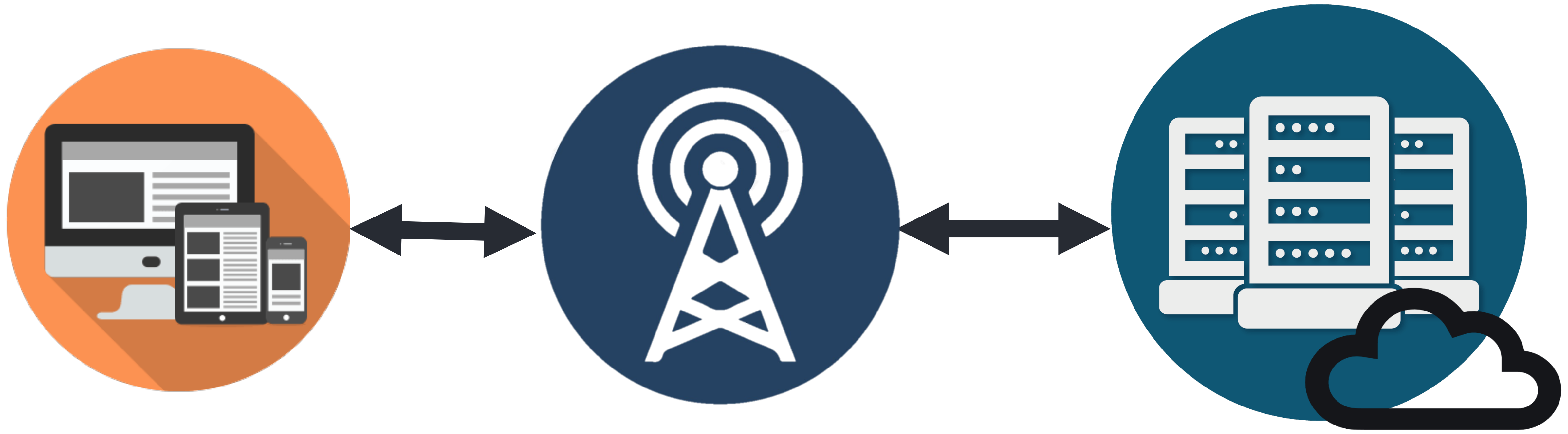


Mobile **Communication** **Data center**

Computing incurs a growing environmental footprint

700 Million tons of CO_2 → **Doubling over the next decade!**

- Half the aviation industry's footprint
- 2.1 - 3.9% of worldwide emissions



Mobile **Communication** **Data center**

59% of world online

Future improvements challenging (e.g., slowing Moore's Law, PUE)

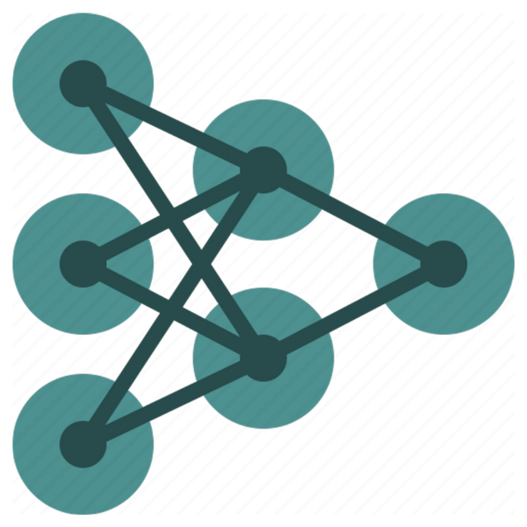
Emerging applications demanding higher compute

Overview of ACT

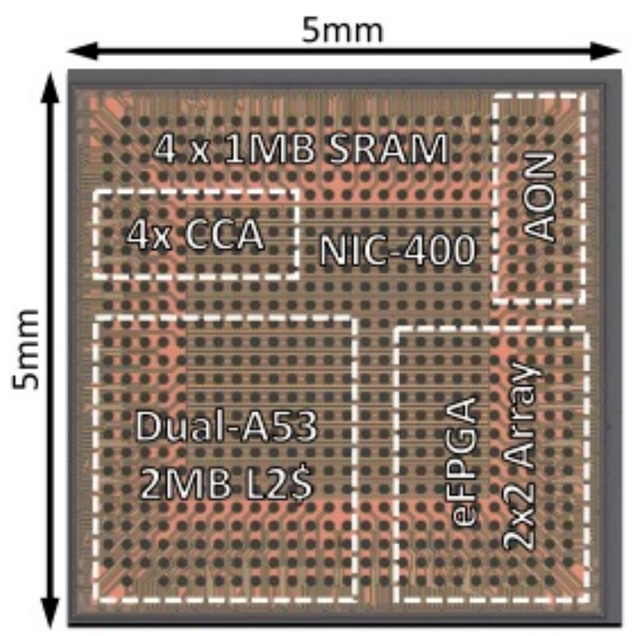
Arch. Carbon **Model**

Carbon-aware HW **design metrics**

Overview of ACT

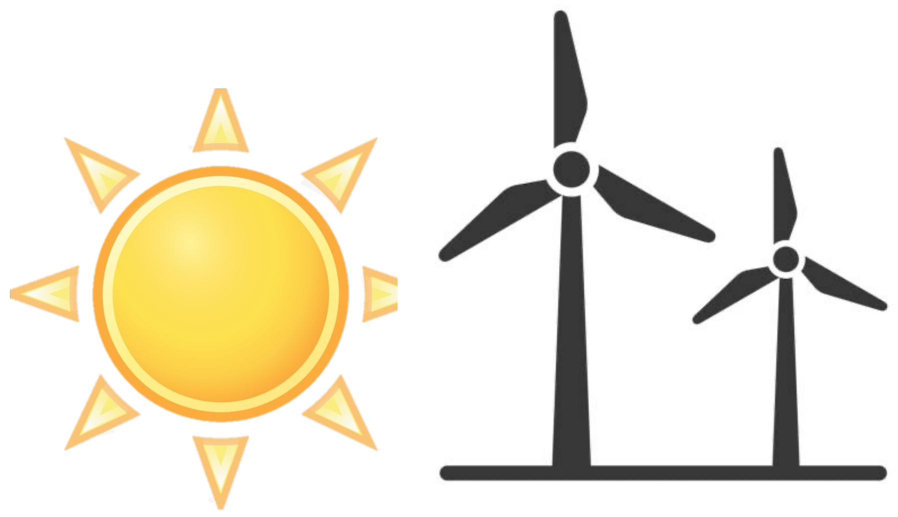


SW design & performance



HW design & power

Fab

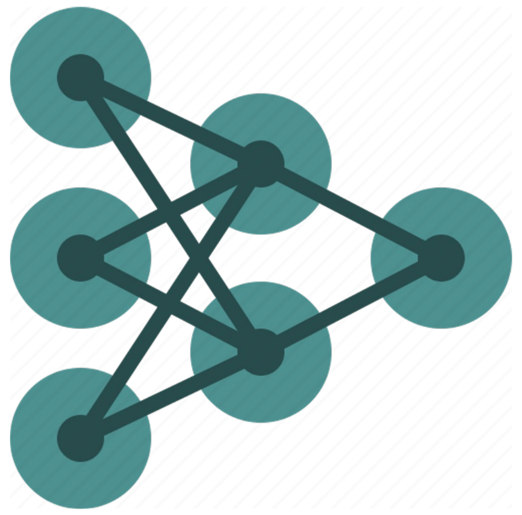


Environmental

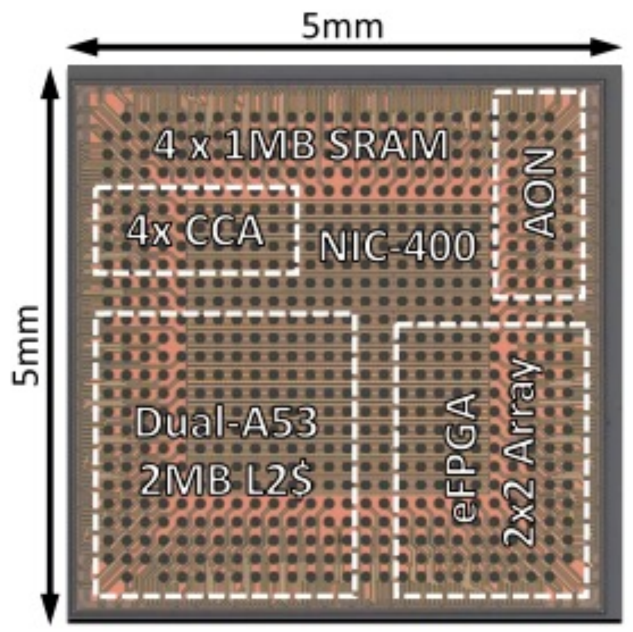
Arch. Carbon **Model**

Carbon-aware HW **design metrics**

Overview of ACT

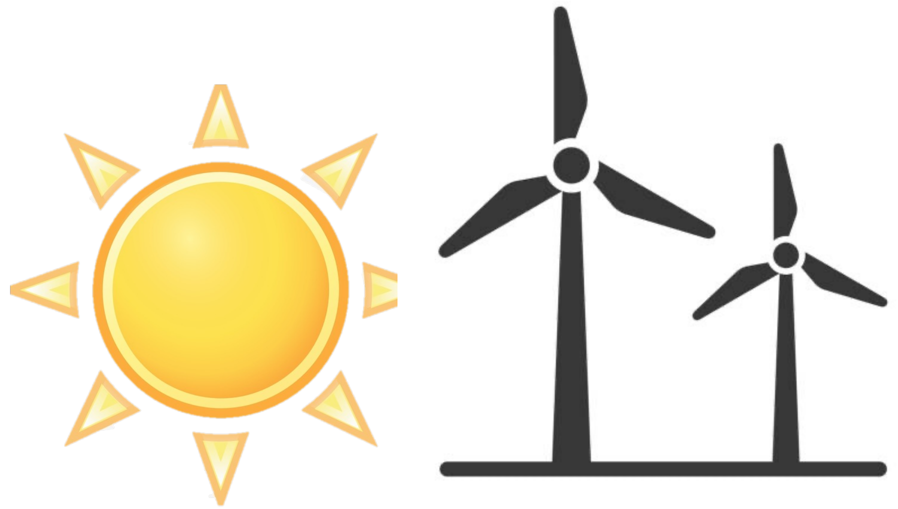


SW design & performance



HW design & power

Fab



Environmental

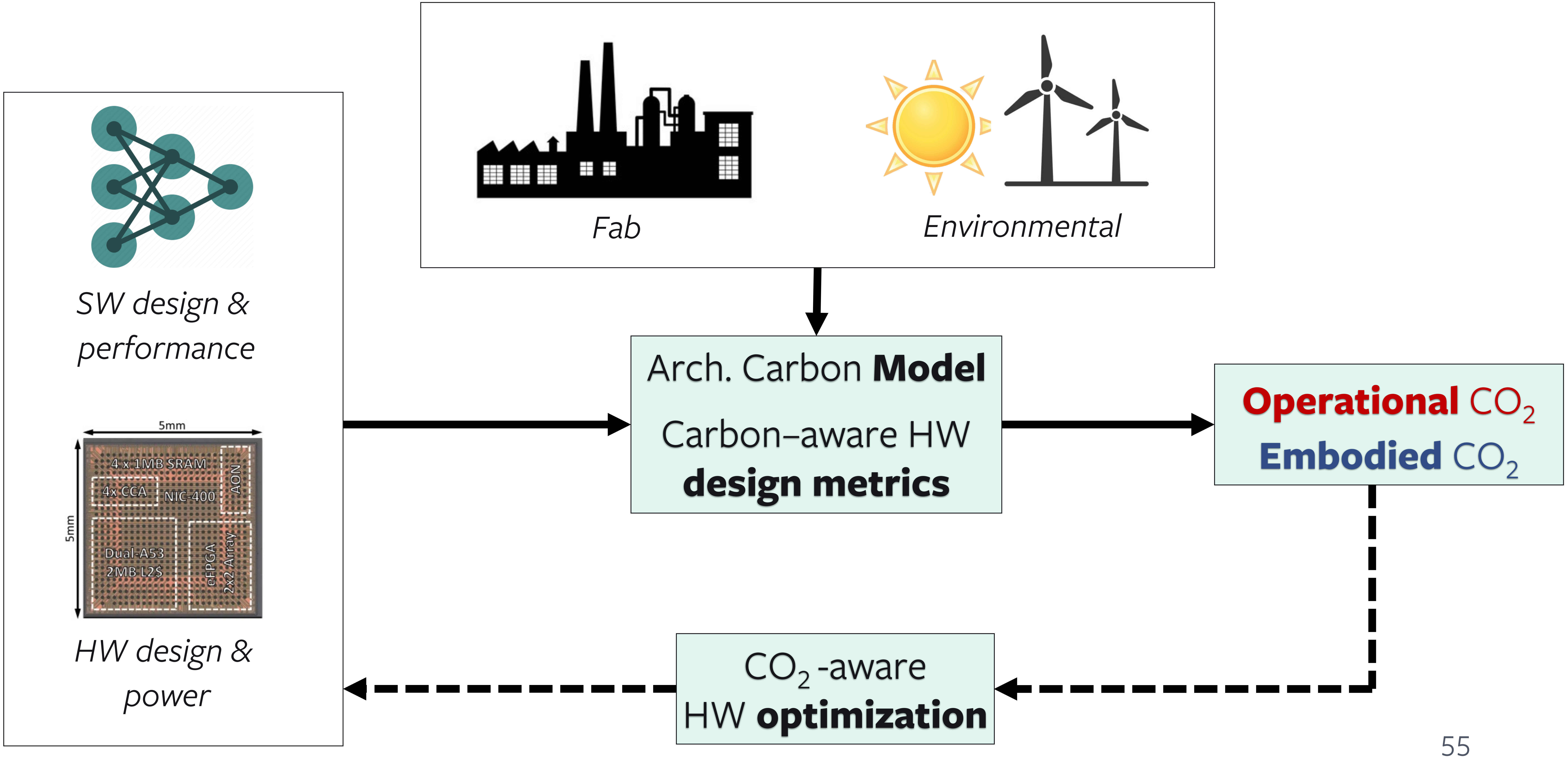
Arch. Carbon **Model**

Carbon–aware HW **design metrics**

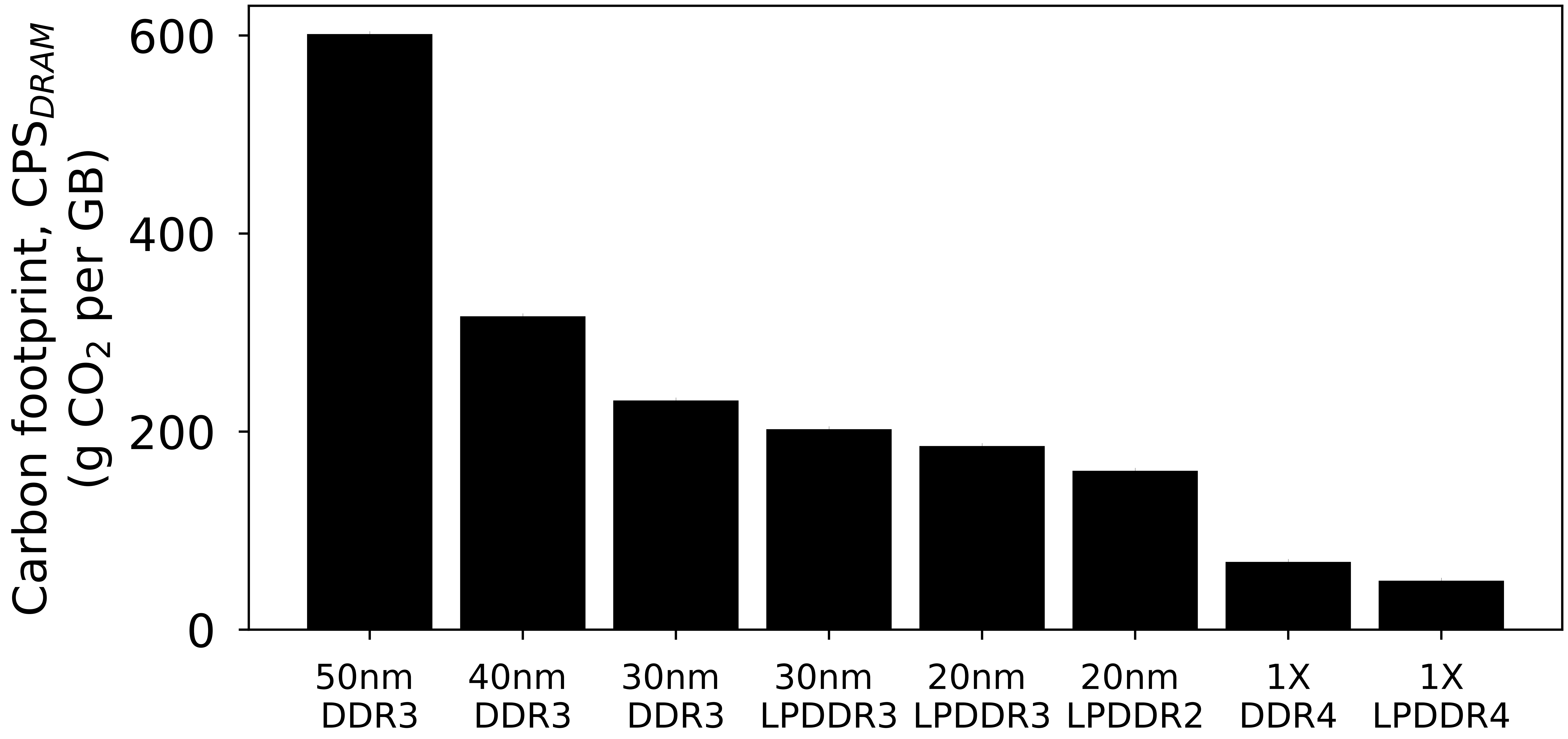
Operational CO_2

Embodied CO_2

Overview of ACT

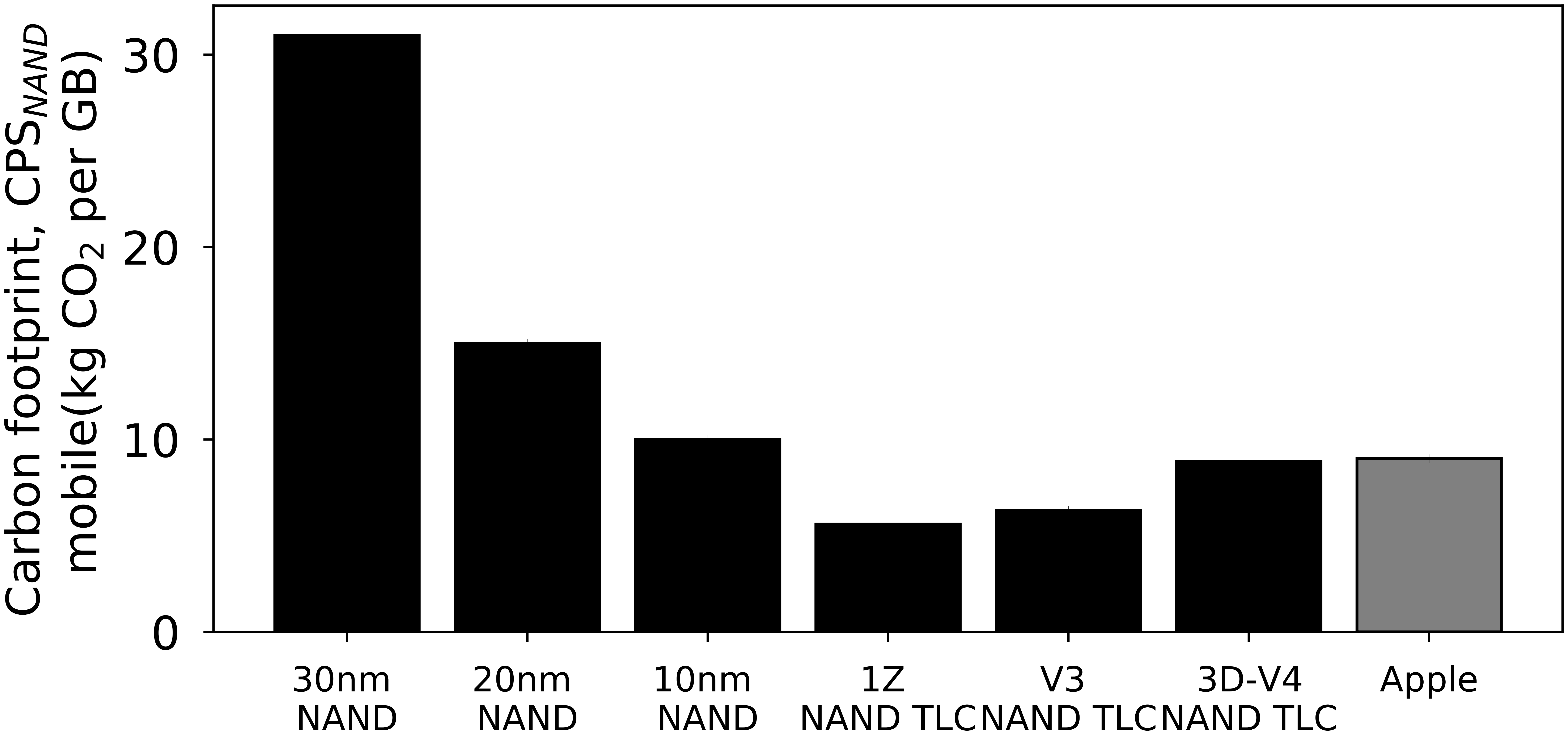


DRAM carbon per storage model



Data sources: SK Hynix

SSD carbon per storage model



Data sources: SK Hynix, Apple

More comparisons (ACT vs. LCA's) in the paper…

	ACT vs. Dell R740 server LCA	ACT vs. Fairphone 3 mobile device LCA

More comparisons (ACT vs. LCA's) in the paper...

<table>
<tr><th>IC component</th><th>ACT vs. Dell R740 server LCA</th><th>ACT vs. Fairphone 3 mobile device LCA</th></tr>
<tr><td>Compute (processors, SoC's)</td><td>Within 2.2x</td><td>Within 1.18x</td></tr>
<tr><td>Memory</td><td>Within 1.62x</td><td rowspan="2">Within 2.1x</td></tr>
<tr><td>Storage</td><td>Within 1.05-2.2x</td></tr>
</table>

Takeaways

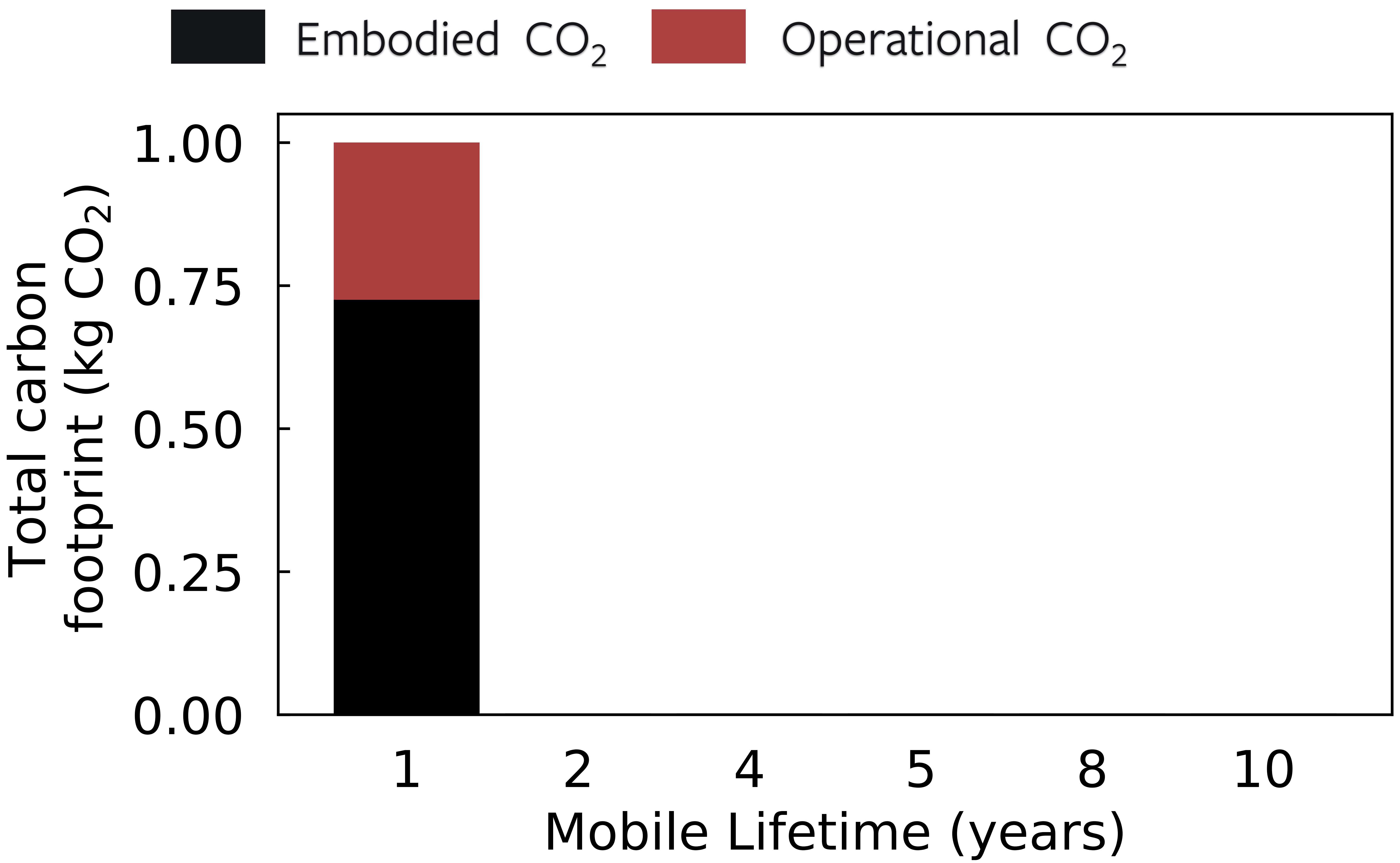
(1) ACT provides first-order approximate of LCA's that use old technology nodes (45nm NAND, 32nm CPU)

(2) ACT enables architects to study new technology nodes

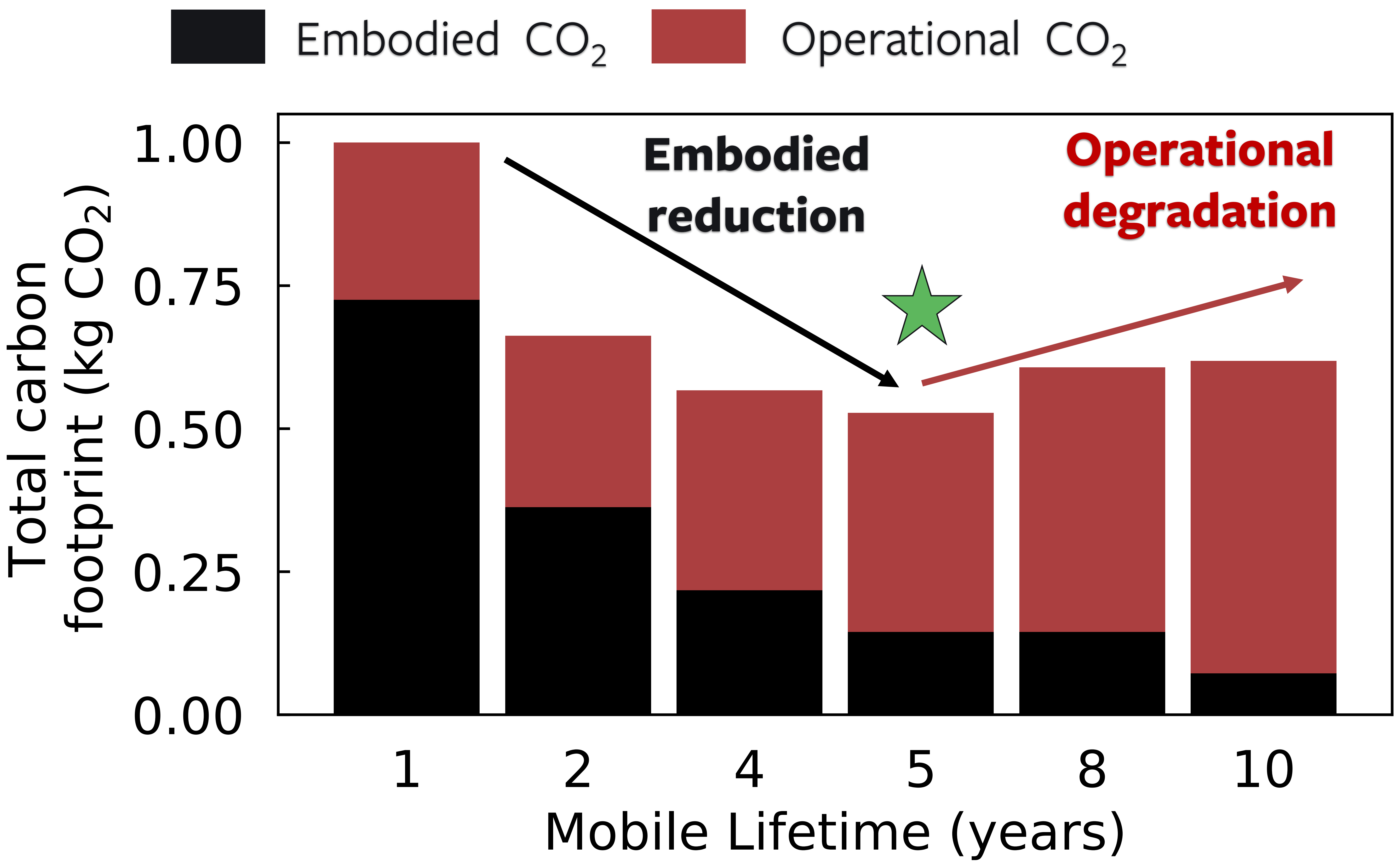
Recycle: Extending hardware lifetime

Mobile SoC's
Geekbench characterization → **Architectural Carbon Model** → Operational vs. Embodied carbon

Recycle: Extending hardware lifetime

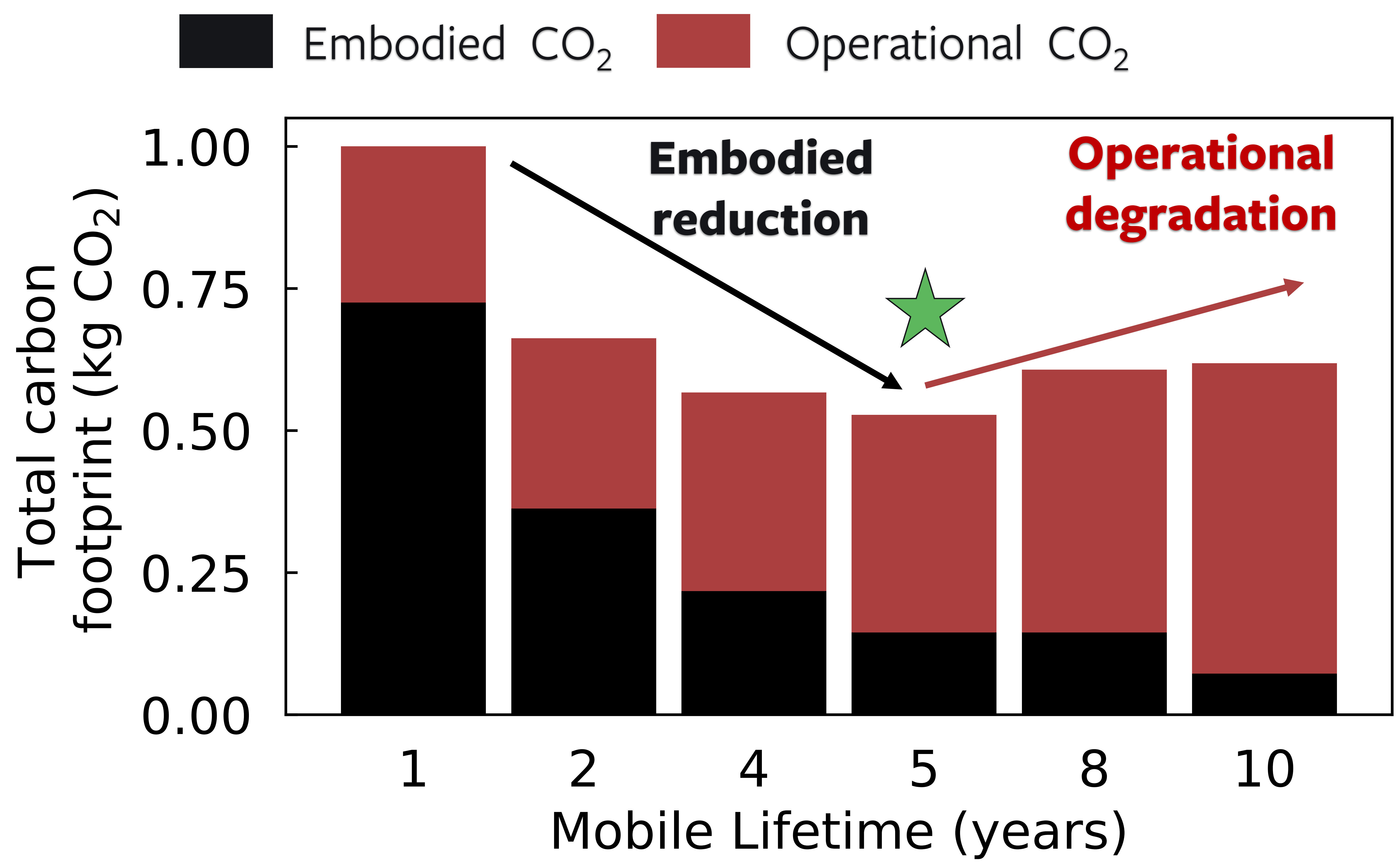


Recycle: Extending hardware lifetime



Recycle: Extending hardware lifetime

{ Mobile SoC's Geekbench characterization } → **Architectural Carbon Model** → Operational vs. Embodied carbon



Enabling **2nd life** requires enhancing **HW reliability**

See paper for case study on storage reliability using **SSD overprovisioning**