ACT: Designing Sustainable Computer Systems With An Architectural Carbon Modeling Tool

Udit Gupta
ugupta@g.harvard.edu
Harvard University/Meta
USA

Mariam Elgamal
mariamelgamal@g.harvard.edu
Harvard University
USA

Gage Hills
ghills@g.harvard.edu
Harvard University
USA

Gu-Yeon Wei
guyeon@seas.harvard.edu
Harvard University
USA

Hsien-Hsin S. Lee
leehs@fb.com
Meta
USA

David Brooks
dbrooks@eecs.harvard.edu
Harvard University/Meta
USA

Carole-Jean Wu
carolejeanwu@fb.com
Meta
USA

ABSTRACT

Given the performance and efficiency optimizations realized by the computer systems and architecture community over the last decades, the dominating source of computing's carbon footprint is shifting from operational emissions to embodied emissions. These embodied emissions owe to hardware manufacturing and infrastructure-related activities. Despite the rising embodied emissions, there is a distinct lack of architectural modeling tools to quantify and optimize the end-to-end carbon footprint of computing. This work proposes ACT, an architectural carbon footprint modeling framework, to enable carbon characterization and sustainability-driven early design space exploration. Using ACT we demonstrate optimizing hardware for carbon yields distinct solutions compared to optimizing for performance and efficiency. We construct use cases, based on the three tenets of sustainable design—Reduce, Reuse, Recycle—to highlight future methods that enable strong performance and efficiency scaling in an environmentally sustainable manner.

CCS CONCEPTS

• **Computer systems organization** → **Architectures**; • **Hardware** → **Integrated circuits**.

KEYWORDS

Computer Architecture, sustainable computing, mobile, energy, manufacturing

ACM Reference Format:
Udit Gupta, Mariam Elgamal, Gage Hills, Gu-Yeon Wei, Hsien-Hsin S. Lee, David Brooks, and Carole-Jean Wu. 2022. ACT: Designing Sustainable Computer Systems With An Architectural Carbon Modeling Tool. In *Proceedings of The 49th Annual International Symposium on Computer Architecture (ISCA '22).* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3470496.3527408

ISCA '22, June 18–22, 2022, New York City, NY

ACM ISBN 978-1-4503-8610-4/22/06…$15.00
https://doi.org/10.1145/3470496.3527408

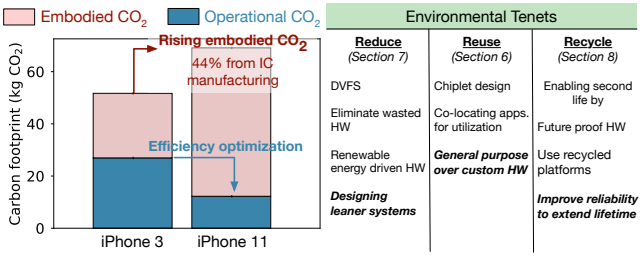

Figure 1: Given the efficiency improvements and increasing hardware complexity over the last decade, carbon footprint for mobile systems has shifted from operational to embodied emissions; a large portion of these emissions owe to manufacturing processors, memory, and storage (left). To enable sustainable computing we must consider the three tenets of environmental design: Reduce, Reuse, Recycle. This work demonstrates one example of each tenet (bolded) in mobile systems (right).

1 INTRODUCTION

Environmental sustainability is an existential threat to society that requires collective and immediate action across *all* communities and industries. Over the last two decades, the world has witnessed a dramatic rise in computing demand fueled by new applications at the edge and cloud-scale. Unfortunately, this rise incurs high energy and environmental overheads worldwide. As of 2019, estimates show the energy consumed by operating information and computing technologies (ICT) accounts for 2% of worldwide carbon emissions, half that of the aviation industry [7, 36]. If left unchecked, ICT may consume up to 8% of worldwide emissions over the next decade [7]. This work, ACT, tackles ICT's rising carbon footprint by designing a framework to quantify and optimize emissions across *end-to-end hardware lifetimes.*

Existing efforts target operational emissions. Since the early 2000s, a plethora of efficiency optimizations have been proposed and adopted to counter ICT's environmental overheads. These efficiency optimizations have primarily targeted *operational emissions* resulting from energy consumption. For example, data centers have seen a 5×, 8×, and 3× energy efficiency improvement in compute, storage, and networking, respectively [55]. At the mobile-scale, Figure 1 illustrates efficiency optimizations have reduced the operational footprint between an iPhone 3 (2009) and an iPhone 11 (2019) by 2.5×.

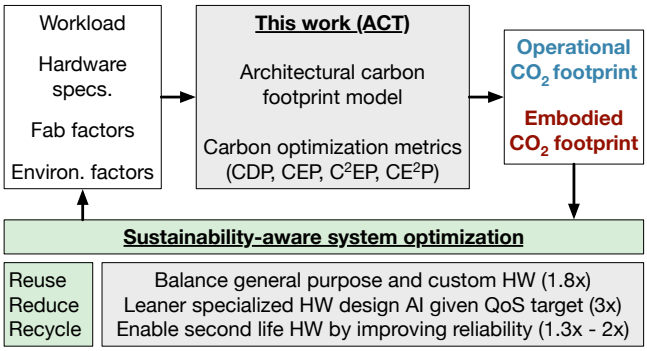

Figure 2: This work proposes ACT a high-level framework that comprises an architectural carbon footprint model and carbon-optimization metrics to characterize the end-to-end carbon footprint of systems. Based on this characterization, in this work we consider an example reuse-based, reduce-based, and recycling-based case study to demonstrate the breadth of ACT.

Dominating emissions source shifts towards manufacturing. Given the efficiency improvements and increasing prevalence of renewable energy [36, 37, 59, 103], the dominating source of ICT's carbon emissions is shifting from operational to infrastructure-related activities [36]. In 2019, Apple reported hardware manufacturing accounted for 74% of its overall footprint; in comparison, operational use of all its devices accounted for 19%. Similarly, at the individual system-level, for an iPhone 11 Figure 1 (left) shows manufacturing (red) and operational use (blue) account for 79% and 17% of the emissions, respectively; the remaining 4% owe to product transport and recycling [11]. More specifically, nearly half of Apple's emissions from hardware manufacturing come from fabricating integrated circuits (IC's) such as processors, memory, storage [8, 36]. Consequently, many technology companies (e.g., Apple, Facebook, Microsoft) have pledged to achieve carbon neutrality across their supply-chains in the coming decade [8, 28, 57]. Achieving carbon neutral supply-chains requires tackling ICT's emissions across *life cycle phases*, from *both* hardware manufacturing and use.

Pressing need for architectural carbon accounting. Tackling ICT's carbon footprint from hardware manufacturing requires us to fundamentally rethink the design and implementation of computer systems. Inspired by the tenets of environmentally sustainable design [4]—Reduce, Reuse, Recycle—Figure 1 (right) shows example hardware and software optimizations that tackle computing's rising carbon footprint. However, a pre-requisite for realizing these optimizations is understanding the salient carbon bottlenecks across hardware life cycles. Unfortunately, unlike the myriad available tools to quantify hardware energy consumption [13, 20, 25, 47, 51, 53], there is a distinct lack of methodologies to understand the carbon footprint of system hardware. Current methods to quantify emissions from system hardware require using Life Cycle Analysis tools (LCA's) [14]. While LCA tools categorize emissions across hardware manufacturing, transport, use, and recycling, LCA's do not provide detailed breakdowns of IC footprints which is necessary for carbon-aware design space exploration.

In this paper, we propose, *ACT*, an carbon modeling tool to enable carbon-aware design space exploration. As shown in Figure 2, *ACT* comprises an analytical, architectural carbon-footprint model and use-case dependent optimization metrics to estimate the carbon footprint of hardware. The proposed model estimates emissions from hardware manufacturing (i.e., embodied carbon) based on workload characteristics, hardware specifications, semiconductor fab characteristics, and environmental factors. *ACT* addresses a crucial gap in quantifying and enabling sustainability-driven hardware design space exploration, and serves as a call-to-action for computer architects to consider sustainability as a first-order citizen, alongside performance, power, and area (PPA).

Using ACT, we conduct a series of case studies to optimize carbon footprint for mobile systems. Given the wide design space, we conduct a case study inspired by each of the three tenets of sustainable design (Figure 2). First, in the context of systems, a fundamental example of *reuse* is the trade-off between programmable hardware used across applications and application-specific hardware (ASIC). We quantify the balance between PPA and carbon of programmable and specialized mobile hardware. Second, designing leaner systems *reduces* system carbon footprint. Given the efficiency benefits of ASICs, we demonstrate that minimizing hardware resources while meeting quality-of-service (QoS) targets reduces overall carbon footprint. Third, *recycling* systems and individual components can indirectly lower embodied footprints as fewer IC's are manufactured over time. We demonstrate enabling second-life through extending mobile lifetimes and improving hardware reliability can reduce overall carbon footprints.

The core contributions of this work are:

(1) We propose **ACT** an architectural carbon modeling tool that comprises an extensible model to quantify the carbon footprint of processing elements, memory, and storage. The model enables researchers to quantify emissions from hardware manufacturing and use (Section 3).
(2) We propose a set of use-case dependent carbon optimization metrics based on system and environmental factors. Using example mobile platforms, we show the metrics—carbon-delay, carbon-energy, carbon^2-energy and carbon-energy^2 product—yield distinct optimal designs from PPA optimization (Section 4).
(3) Reuse: Using ACT, we quantify the trade-off between programmable and specialized hardware. While ASICs improve efficiency, general purpose hardware incurs lower carbon emissions from manufacturing, improving overall carbon footprints by up to 1.8× (Section 6).
(4) Reduce: We demonstrate leaner accelerators can minimize carbon footprint while meeting QoS requirements. The optimal accelerator design varies based on the carbon optimization target. Compared to PPA optimized designs, carbon-aware design space exploration reduces the footprint of AI accelerators by up to 3× (Section 7).
(5) Recycle: We demonstrate extending hardware lifetime to enable second-life of systems and components can lead to lower footprint systems. Using SSD-based storage as an example, devoting additional hardware to improve reliability reduces the overall carbon footprint of devices by nearly 2× (Section 8).

Broadening access. To catalyze further investigation into sustainable system design, the underlying ACT's model and the configurable parameters are made available to the community.

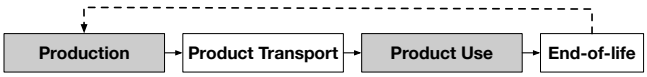

Figure 3: Enabling sustainable systems requires balancing emissions across hardware life-cycles: manufacturing, transport, use, and end-of-life processing.

2 MOTIVATION AND BACKGROUND

2.1 Sustainability-driven system design

Given the rising carbon footprint of ICT, many technology companies, including Apple, Facebook, Google, and Microsoft, have pledged to curtail their environmental footprints [8, 28, 35, 57]. Recent efforts at the cloud-scale enable renewable energy-driven warehouse-scale data centers [6, 16, 24, 66, 97]. Furthermore, application developers design tools to characterize the carbon footprint of salient workloads. Within the domain of AI, researchers have quantified the impact of large neural network training on carbon footprints [37, 59, 96, 103]. Despite the growing interest in sustainable computing, these recent efforts focus predominantly on optimizing emissions from operational use. ACT goes further to consider the impact of *both* hardware manufacturing and use over system lifetimes.

2.2 Emissions across phases of hardware life cycles

To quantify the carbon footprint of computer systems, we analyze emissions across hardware life cycles. Figure 3 illustrates hardware life cycles are split into four main phases:

- **Hardware manufacturing**: emissions from manufacturing IC's (e.g., processors, memories, storage devices) due to energy, chemicals, and gases used by fabs.
- **Hardware transport:** emissions from transporting hardware from fabs to end users.
- **Operational use:** emissions from running software on given hardware resources due to energy consumption.
- **End-of-life processing and recycling:** emissions from recycling hardware units.

Carbon impact of life cycle phases. Gupta et al. survey myriad consumer devices (e.g., mobile phones, wearable devices, desktops, tablets, personal assistants), warehouse scale data centers (e.g., Google, Facebook), and semiconductor fabs (e.g., Intel, TSMC) [36] — the majority of emissions in computing platforms comes from hardware manufacturing. Figure 1 (left) shows that, for an iPhone 3, manufacturing and operational use account for 45% and 49% of emissions, respectively (remaining 6% owe to transport and end-of-life processing). For an iPhone 11, in comparison, manufacturing and operational use account for 79% and 17% of emissions, respectively. The shift in emissions towards hardware manufacturing owes to higher manufacturing overheads from advanced process technology and IC complexity as well as lower operational footprint from improved energy efficiency. Due to this shift, there is a pressing need to develop tools to quantify emissions across all life cycle phases.

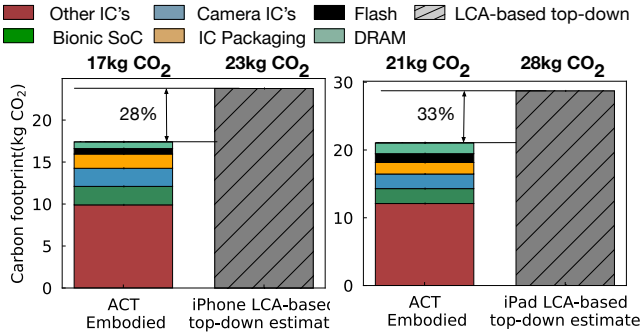

Figure 4: Embodied carbon estimates for the IC's in two example platforms: iPhone 11 (left) and iPad (right) using ACT and LCA's. Unlike the opaque LCA's, ACT provides detailed carbon footprint breakdowns for each IC enabling hardware design space exploration.

2.3 Limits of tools quantifying emissions from manufacturing

Given the importance of hardware carbon footprints, academia and industry have developed a number of tools to quantify emissions across hardware life cycles. In particular, the methodologies fall into two categories: (i) exergy-based models from the system's community and (ii) life cycle analysis tools (LCA's) and product environmental reports from environmental engineering domains. This section outlines these methods and highlights their limitations to motivate the design of ACT.

Limitations of energy balance analysis. The computer systems and architecture community has proposed *exergy*-based environmental accounting [23]. Exergy follows an energy-balance approach to quantifying the environmental impact of servers during fabrication and use. Compared to directly optimizing for carbon footprint, the energy-balance approach simplifies the design space for sustainable systems. In particular, the impact of renewable energy during manufacturing and use is not considered. Complementing exergy, this work considers the direct carbon footprint from hardware manufacturing and operational use, such as carbon and global warming potential emissions, while capturing newer process technology nodes.

Limitations of life cycle analysis tools. While LCA tools are used in industry to generate environmental product reports, they are ill-suited to guide system and hardware design space exploration. LCA tools are developed by environmental scientists to quantify the environmental footprint of products—not just computing devices—across life cycles [46, 94, 95]. Unfortunately, LCA's use coarse-grained information to estimate life cycle emissions. For instance, EIO-LCA [46] estimates carbon emissions based on the economic cost of electronics; typically, the component cost is translated to carbon based on a coarse, industry-wide scale factor. However, given the economic cost implications, the EIO-LCA's may not enable designers to conduct comparative analyses between systems or hardware components to guide design space exploration.

In addition to the economic cost-based LCA's, another class of LCA tools use database-based approaches [95]. The database-based LCA tools take as input a system's bill of materials to estimate carbon footprint. This approach can be used for characterizing the

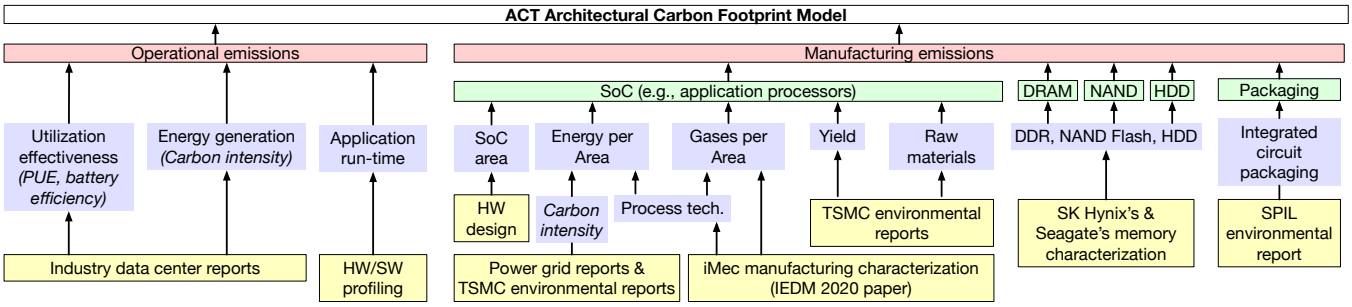

Figure 5: Proposed ACT carbon footprint model. The model accounts for emissions from both hardware manufacturing and use (red). To quantify emissions from hardware manufacturing we aggregate the footprint from SoC's, memories, storage, and packaging (green). The main modeling parameters that set the footprint of each IC are shown in blue. To instantiate the carbon model we leverage industry environmental reports and detailed fab characterization (yellow).

platform-level carbon footprint [22] but the coarse granularity of the databases is insufficient for early hardware design space exploration. Furthermore, database-based LCA tools [95] build atop emission data on older process technologies (e.g., 45nm and earlier). As such, these LCA tools are insufficient for fluid data-driven, hardware carbon accounting and design optimization for future systems.

Finally, based on the aforementioned LCA tools, product environmental reports published by industry provide only coarse-grained and opaque carbon footprint data. To highlight the need for detailed architectural carbon footprint models for early hardware design space exploration, Figure 4 estimates the embodied carbon emissions for two example platforms—iPhone (left) and iPad (right)—based on ACT and industry product environmental reports (*LCA*) [9, 10]. We compare the embodied emissions from *IC's only*. Industry product environmental reports categorize emissions into hardware manufacturing, transport, use, and recycling. Apple's sustainability reports illustrate that roughly half, 44%, the manufacturing footprint of all devices owe to IC's [8, 36]. Using this average, we estimate the IC embodied footprint to be 23kg CO_2 and 28kg CO_2 for the iPhone and iPad, respectively. While helpful in understanding platform-level footprints, these estimates lack a detailed breakdown of emissions from IC's, precluding their use for data-driven, productive, sustainability-aware hardware design.

In contrast, ACT enables hardware design space exploration by empowering hardware designers to quantify the carbon footprint for individual IC's. Central to ACT is an analytical architectural carbon model fueled primarily by publicly reported carbon and environmental footprint characterization of semiconductor fabs and hardware vendors. Publicly available LCA tools assume much older process technologies [19]. For instance, the Fairphone 3 implements 14nm LPDDR4 for DRAM and 10nm NAND Flash for storage; however, its LCA estimates both memory and storage using a 50nm DDR3 process [60]. This is not representative of state-of-the-art memory and storage manufacturing technologies. Furthermore, while the Fairphone 3 implements a 14nm CPU, the publicly available LCA estimates emissions using 32nm technology. Appendix A provides further detail. Similarly, recent work, such as the GreenChip [21, 48] framework, uses a combination of LCA tools [19], a parametric fabrication model [58], and EIO-based data to estimate the embodied carbon footprint of processors. In contrast, ACT is a predictive carbon model that builds off data directly from industry fabs and hardware vendors to provide sufficient accuracy to study carbon footprint trends from 28nm down to 3nm, allowing computer architects to estimate the footprint of future platforms in modern technologies.

To estimate the platforms' embodied emissions, ACT aggregates the footprint of each IC (i.e., *bottom-up*) including, processors (e.g., processors and SoC's), DRAM memories, and SSD-based storage. The hardware specifications for each platform are based on publicly available device-level teardowns [30, 31]. As shown in Figure 4, ACT estimates the IC's embodied footprint to be 17kg CO_2 and 21kg CO_2 for the iPhone and iPad, respectively. The detailed breakdown of emissions for each IC enable designers to quantify and optimize emissions for future sustainable systems. Section 3 details the design and implementation of the proposed carbon footprint modeling tool.

3 ACT: ARCHITECTURAL CARBON MODELING TOOL

In this section we propose ACT, an architectural carbon modeling tool to quantify the carbon footprint of systems. First, we describe the overall carbon footprint model which accounts for both operational and embodied emissions. As operational emissions are better understood by the systems and computer architecture community [16, 24, 37, 59, 97], we focus on detailing the embodied emissions aspects of the model. Second, we describe the carbon-centric optimization targets to enable sustainable system and hardware design space exploration.

3.1 Carbon Footprint Model

Figure 5 illustrates the overall carbon footprint, categorized into operational and embodied emissions, that underscores ACT. Over the last two decades, the computer architecture community has developed myriad tools and methods to quantify operational emissions by estimating and measuring utilization effectiveness [3, 15], energy consumption [20, 25, 53], and carbon intensity [1, 37]. ACT builds off these techniques to quantify operational emissions. Unlike previous tools, ACT estimates embodied carbon emissions by separately considering the footprint of application processors (e.g., SoC's), DRAM-based memories, NAND Flash based storage, HDD's, and IC packaging, as shown in green (Figure 5). Parameters that determine embodied emissions for each component are shown in blue—-for

Table 1: Input parameters in ACT architectural carbon model.

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

example, area, fab energy and gaseous intensity per area, yield, and raw materials for SoC's. Table 1 defines each of the parameters. The Appendix provides additional details on how the parameters are configured, based on data from semiconductor fabs [33] and industrial environmental reports [98, 99]. Below we present the proposed carbon model in detail.

The ACT carbon footprint model is designed to quantify the emissions of running a software application on given hardware substrate. The overall emissions, CF, are the combination of the operational (OP_{CF}) and embodied (E_{CF}) emissions. The embodied emissions are discounted based on the application run-time, T, and the overall lifetime of the system, LT. Typical lifetime for server-grade processors in datacenters are 3-5 years [15] and 2-3 years for mobile devices [8].

$$CF = OP_{CF} + \frac{T}{LT} E_{CF} \tag{1}$$

The operational emissions (OP_{CF}) are computed as the product of the energy consumed by running the workload on the target hardware and the carbon intensity during the use phase (CI_{use}). CI_{use} enables system developers to consider the implication of renewable energy driven platforms.

$$OP_{CF} = CI_{use} \times Energy \tag{2}$$

To quantify the overall embodied carbon emissions (E_{CF}), we categorize it on a per-component level (E_r) for each of the application processors (SoC), memory (DRAM) and, storage (SSD and HDD) elements. Each IC incurs additional packaging overheads (K_r). N_r represents the ICs, including application processors, memories, and storage devices, on a hardware platform. Based on industrial environmental reports, the packaging footprint is set to K_r = 0.15kg CO_2 [93]. Below we detail the embodied emissions for SoC's, memory, and storage devices.

$$E_{CF} = N_r K_r + \sum_{r}^{\text{SoC, DRAM, SSD, HDD}} E_r \tag{3}$$

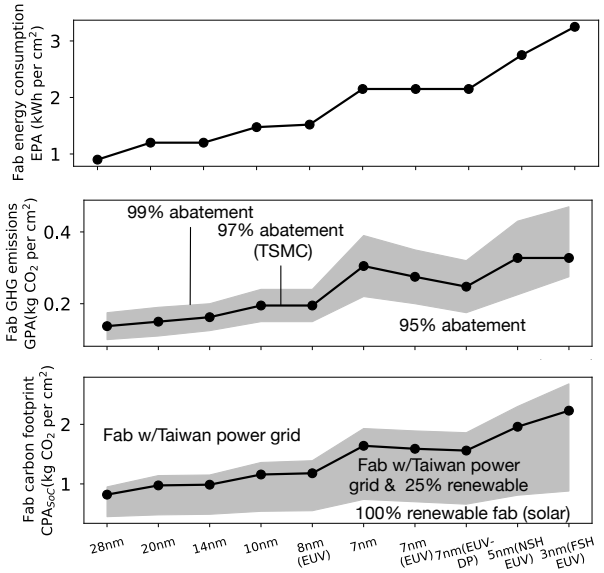

Figure 6: Embodied carbon intensities for compute hardware resources (i.e., SoC). Carbon footprint breakdown down into electricity consumed by fab (top), GHG emissions from gases and chemical emitted by fab (middle), and aggregate carbon per area (bottom). Middle and bottom show range of carbon intensities based on varying semi-conductor fab characteristics. Data is based on industry reports and detailed device-level characterization[33, 98, 99]

The embodied carbon footprint model considers the direct environmental impact of semiconductor fabrication. However, secondary overheads such as emissions from manufacturing fabs and building fabrication machines (e.g., EUV machines) are not considered. As such, the embodied carbon model can be treated as a lower bound.

Application processor embodied emissions. The embodied carbon footprint of hardware begins at the procurement of raw materials and continues to the energy consumption and GHG emissions released during wafer processing in the fab. For an application processor, the embodied carbon emissions, E_{SoC}, are based on the die area (Area) and carbon emitted per unit area manufactured (CPA).

CPA depends on various semiconductor fab parameters including the fab yield ($0 \leq Y \leq 1$), the energy consumed per unit area manufactured (EPA), emissions per unit area from chemicals burned (e.g., gases, PFCs) during hardware manufacturing (GPA), and emissions from procuring raw materials for fab manufacturing (MPA). We convert the energy consumed by manufacturing into carbon emissions using the carbon intensity of fabs (CI_{fab}), which is dependent on the fab's energy source (i.e. renewables vs. non-renewables).

$$\begin{aligned} E_{SoC} &= \text{Area} \times \text{CPA} \\ &= \frac{1}{Y}((CI_{fab} \times \text{EPA} + \text{GPA} + \text{MPA}) \times \text{Area}) \end{aligned} \tag{4}$$

$$\text{CPA} = \frac{1}{Y} \times (CI_{fab} \times \text{EPA} + \text{GPA} + \text{MPA}) \tag{5}$$

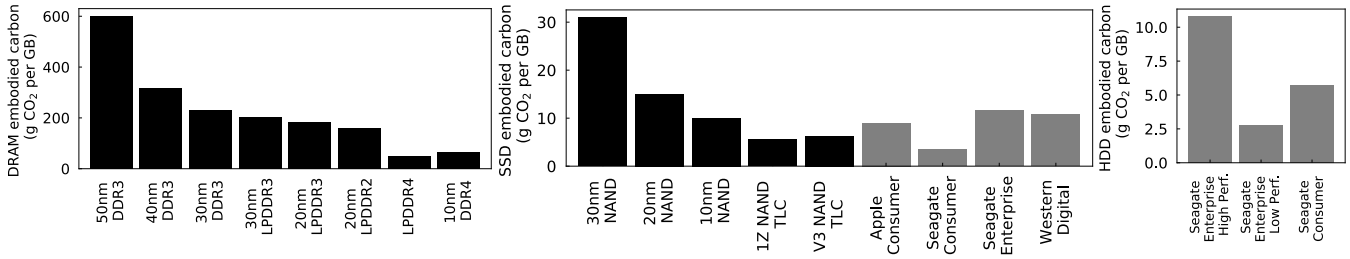

Figure 7: Embodied carbon emissions for varying DRAM-based memories (left), NAND Flash-based storage devices (center), and HDD's (right) used in ACT's carbon footprint model. Bars in black represent characterization from device-level carbon characterization while bars in grey represent characterization from component-level analyses. All data points are based on industry environmental reports[42–45].

Figure 6 illustrates the carbon footprint of manufacturing application processors across technology nodes from 28nm down to 3nm (x-axis). On the top, we show rising energy consumed per unit area (EPA) based on device-level characterization [33]. The rising energy consumption owes to more sophisticated lithographic processes (e.g., extreme ultraviolet lithography). Similarly, Figure 6 (middle) illustrates the rising carbon emissions from chemicals and gases (GPA) [33]. The shaded regions illustrate the lower-bound and upper-bound of emissions based on gaseous abatement strategies used by fabs to improve fab efficiency. Aggregating the parameters together, Figure 6 (bottom) illustrates overall CPA across nodes. The shaded region illustrates the upper bound, which assumes the average Taiwan power grid, and the lower bound, which assumes renewable energy powered fabs. Unless specified, the remainder of this paper instantiates CPA with the solid line, which assumes a fab powered by 25% renewable energy based on industry fab characteristics [98, 99].

Memory and storage embodied emissions. ACT quantifies the embodied emissions from memory and storage systems based on the capacity of memory ($\text{Capacity}_{\text{DRAM}}$), SSD ($\text{Capacity}_{\text{SSD}}$), and HDD ($\text{Capacity}_{\text{HDD}}$). The capacity is translated to carbon footprint based on a carbon-per-size factor for DRAM (CPS_{DRAM}), HDD (CPS_{HDD}), and SSD (CPS_{SSD}).

$$\text{E}_{\text{DRAM}} = \text{CPS}_{\text{DRAM}} \times \text{Capacity}_{\text{DRAM}} \quad (6)$$

$$\text{E}_{\text{HDD}} = \text{CPS}_{\text{HDD}} \times \text{Capacity}_{\text{HDD}} \quad (7)$$

$$\text{E}_{\text{SSD}} = \text{CPS}_{\text{SSD}} \times \text{Capacity}_{\text{SSD}} \quad (8)$$

Figure 7 illustrates the carbon-per-size for various memory and storage devices across process technologies. The DRAM (E_{DRAM}), SSD (E_{SSD}), and HDD (E_{HDD}) embodied carbon emissions are shown on the left, middle, and right, respectively. In black we show embodied carbon characterization from semiconductor manufacturers (e.g., SK Hynix [42–45]); in grey we show component-level analysis from mobile (e.g., Apple [8]) and SSD vendors (e.g., Seagate [71–91], Western Digital [26, 27]). At commensurate technology nodes, the carbon intensity of DRAM is higher than that of SSD and HDD. Across process technologies for DRAM and SSD, newer nodes exhibit lower carbon footprint per unit capacity.

Instantiating ACT's model parameters. The core contribution of ACT is the design of a comprehensive and extensible architectural carbon footprint model shown in Figure 5. We instantiate the model parameters (Table 1 based on state-of-the-art publicly 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.

Table 2: ACT's use-case dependent sustainability optimization metrics. C represents embodied carbon, E represents energy, D represents delay, and A represents area.

Metric	Use case
EDP	Energy optimization (e.g., mobile)
EDAP	Energy and cost optimization (e.g., mobile)
CDP	Balance $CO_2$and perf. (e.g., sustainable data center)
CEP	Balance $CO_2$and energy (e.g., sustainable mobile device)
C^2EP	Sustainable device dominated by embodied footprint
CE^2P	Sustainable device dominated by operational footprint

3.2 Carbon Optimization Metrics

Designing sustainable systems requires not only models to quantify emissions but also standardized optimization metrics. Thus, along with the proposed carbon footprint model, ACT comprises use-case dependent optimization metrics, shown in Table 2. Where it is infeasible to quantify the lifetime carbon footprint of a hardware platform early in the design cycle, these metrics can aid designers to incorporate sustainability into hardware design space exploration. To gain intuition we start with two frequently used metrics by computer architects to design hardware systems: energy-delay product (EDP) and energy-delay-area (EDAP) product [53]. EDP optimizes systems for operational energy consumption while EDAP balances both operational overheads and capital manufacturing overheads [53]. Extending these metrics ACT introduces four new carbon optimization metrics: carbon-delay product (CDP), carbon-energy product (CEP), carbon^2-energy product (C^2EP), and carbon-energy^2 product (CE^2P). Here, carbon (C) represents the embodied carbon emissions.

- **CDP** balances embodied carbon emissions and performance. CDP is relevant when designing high performance sustainable systems such as data center hardware [28, 57].
- **CEP** balances embodied carbon emissions and energy consumption. Considering both capital and operational characteristics, CEP targets sustainable mobile devices.
- **C^2EP** prioritizes optimizing embodied emissions over energy. C^2EP is especially relevant when designing end-to-end sustainable mobile systems dominated by embodied emissions (e.g., powered by renewable or carbon-free energy).
- **CE^2P** prioritizes optimizing energy over embodied emissions. CE^2P is especially relevant for systems dominated by operational emissions (e.g., powered by "brown" energy).

Based on these carobn-aware optimization metrics, Section 4 demonstrates how designing systems for sustainability yields new design spaces.

4 CARBON OPTIMIZATION OPENS NEW DESIGN SPACES

Based on the ACT carbon footprint model and carbon-centric optimization targets, this section characterizes the sustainability design space of commodity mobile platforms. We demonstrate the optimal system in terms of environmental sustainability varies compared to typical hardware optimization approaches focused on performance, power and energy, and area (PPA) [20, 53]. Consequently, future hardware systems must consider environmental sustainability as a first-order design metric alongside PPA. To highlight the distinct sustainability design space exposed by ACT, this section comprises two main discussions. First, we compare trends in embodied emissions and area. Second, we compare ACT's carbon-based optimization metrics to performance and energy-centric optimization strategies.

4.1 Embodied carbon emissions versus hardware area

While hardware area is a component of embodied emissions, there are fundamental differences between embodied carbon and area. For instance, fab characteristics such as renewable energy procurement (CI_{fab}) and yield impact embodied emissions at constant area. Furthermore, the end-to-end carbon footprint depends not only on embodied emissions but also operational footprint. Operational characteristics such as the lifetime of the system, utilization, power consumption, and availability of renewable energy during use all play a crucial role in determining the overall carbon footprint (see Section 3). As such, to realize sustainable systems, it is insufficient for designers to optimize for area alone; we must go beyond to consider end-to-end environmental impacts.

4.2 Performance and energy versus carbon optimization

In addition to area, we highlight the distinct hardware design spaces available when optimizing for sustainability versus performance and energy. Figure 8 illustrates the performance, energy, carbon, and environmental footprint of three mobile SoC families, i.e., Exynos [67–70], Snapdragon [61–65], and Kirin [38–41].

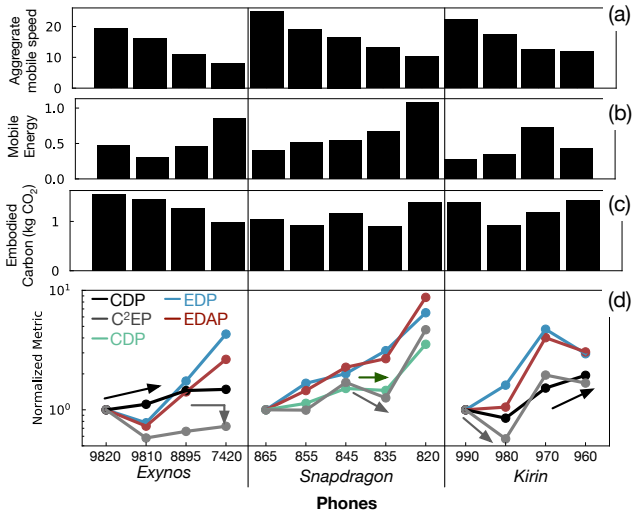

Figure 8: Carbon-optimization design space for commodity mobile systems across Exynos, Snapdragon, and Kirin SoC families. Using Geekbench-based mobile workloads, we show the performance (a), energy (b), and embodied carbon (c) characteristics for each platform. Using the carbon-optimization metrics we show the optimal performance, efficiency, and sustainability-aware design vary based on use case (d).

Across the three families we consider multiple generations of mobile chipsets. Performance is measured based on the geometric mean of seven mobile Geekbench 5 workloads[34]: HTML 5 rendering, AES encryption, text compression, image compression, face detection, speech recognition, and AI-based image classification. For representative characterization we average performance for each workload across 10 mobile chipsets in the wild [34]. Power for the different mobile SoC's is based on their TDP.

Figure 8(a) and (b) illustrate the mobile SoC performance and energy respectively. Following intuition, newer architectures within each SoC family have higher performance. Energy efficiency does not monotonically increase or decrease given the interplay between performance and power. Compared to performance and energy, embodied carbon, as shown in Figure 8(c) follow a distinct trend given the impact of varying process technology, SoC die area, and DRAM capacity. In particular, the embodied carbon footprint of the Snapdragon and Kirin processors fluctuates across SoC versions. The varying trends highlight the distinct hardware design space between performance, energy, and embodied carbon footprint.

Figure 8(d) quantifies the mobile SoC's in terms of the sustainability optimization metrics discussed in Section 3.2. The y-axis is normalized to the newest processor within each family (i.e., Exynos 9820, Snapdragon 865, Kirin 980). Across the processor families we see varying trends between energy-centric metrics (i.e., EDP, EDAP) and carbon-centric metrics (i.e., CDP, CEP, C^2EP). These differences are highlighted by the arrows shown in Figure 8(d). Furthermore, across all the SoC's the optimal platform varies based on the optimization metric. The optimal hardware in terms of EDP, EDAP, embodied carbon, CEP, and C^2EP are the Kirin 990, Snapdragon 865, Snapdragon 835, Kirin 980, and Kirin 980, respectively.

Table 3: Tenets of environmental design (e.g., reuse, reduce, recycle) motivating the example case studies showing the sustainable hardware optimization and design space. For each case study we highlight the main ACT parameters optimized.

Sustainability Tenet	Definition	Example case study	ACT model parameters
Reuse	Reuse systems to amortize embodied CO_2	Balance general purpose hardware with application specific hardware (**Section 6**)	CI_{use}, Energy, T, CI_{fab}
Reduce	Eliminate embodied carbon by designing leaner systems	Designing learner accelerators for salient application like AI (**Section 7**)	Energy, T, p, SoC_{area}
Recycle	Enable second life of platforms and components to amortize embodied CO_2	Extending hardware lifetime by improving reliability (**Section 8**)	OP_{CF}, Emb_{CF}, LT, $SSD_{capacity}$

The following sections show how ACT enables sustainable system design based on the tenets of environmental design.

5 TENETS OF ENVIRONMENTAL DESIGN: METHOD

This section explores distinct case studies to highlight the extensibility and applicability of ACT to design sustainable systems. Table 3 summarizes the main case studies shown in this paper. For completeness we motivate each of the case studies based on the three **R**'s of an environmentally-sustainable process: **R**euse, **R**educe, **R**ecycle [4]. For each case study we list the main ACT modeling parameters to co-optimize performance, energy, and end-to-end carbon footprint.

Reuse: From a sustainability perspective, it is prudent to reuse products. In the context of systems design, one example is *reusing* hardware components for myriad applications. In this work, we consider the trade-off between general-purpose hardware and ASIC's. The analysis explores the trade-off between hardware/software performance, energy efficiency, carbon, and the implication of renewable energy during system hardware manufacturing and operational use (see Section 6).

Reduce: The best way to reduce carbon footprint is to eliminate emissions production. Given the performance and efficiency benefits of ASICs, one example of *reducing* emissions is designing leaner hardware accelerators to balance performance, power, energy, and carbon. In this work we consider a particular application domain—mobile AI inference. The analysis explores the trade-off between performance, energy, process technology, and accelerator area (see Section 7).

Recycle: Finally, recycling systems and individual hardware components into future systems further eliminates wasted carbon in discarded hardware. *Recycling* introduces new challenges and opportunities including the design of modular and repairable computing platforms [12] and enabling second-life of systems and components. Enabling second-life requires future-proofing hardware to provide sufficient performance on emerging applications, and improving hardware reliability. The analysis explores the trade-off between operational and embodied emissions by enabling second-life and extending hardware lifetimes. We focus on storage through improved reliability lifetimes of mobile systems (see Section 8).

6 REUSE CASE STUDY: BALANCING GENERAL PURPOSE AND SPECIALIZED HARDWARE

In the context of sustainable systems, **reuse** introduces new, fundamental trade-offs between programmable and application-specific compute substrates. In a post-Moore's law era, computer systems and architecture researchers have turned to specialization to continue scaling performance and efficiency for salient applications. Despite their performance and energy benefits, ASICs require manufacturing additional hardware which exacerbates embodied emissions. On the other hand, general purpose hardware can be *reused* across applications, amortizing embodied emissions at lower performance and efficiency. This section explores this trade-off between general purpose and specialized hardware. Using mobile AI inference as a driving example, we consider the impact of provisioning mobile systems with programmable CPUs versus GPU-based and DSP-based co-processors. The optimal system design varies based on the availability of renewable energy during manufacturing and use — "green" fabs favor specialization while renewable energy during operation favors general purpose compute. Going further, we show designing more programmable accelerators, such as FPGAs, offers a promising balance between performance, efficiency, and sustainability.

6.1 Mobile system provisioning

Table 4 illustrates the inference latency, power, operational carbon footprint per inference, and overall embodied emissions for a Snapdragon 845 mobile SoC [29, 63]. The analysis considers three cases: SoC based on programmable CPUs, augmenting the CPU with a GPU, and augmenting the CPU with a DSP. The estimation for operational and embodied emissions assumes the average carbon intensity of the United States (e.g., 300g CO_2 per kWh) [1] and average fab characteristics based on the ACT model (refer to Section 3).

The optimal hardware configuration depends optimization target and use case. As expected, the GPU and DSP achieve 1.08× and 2.2× lower energy per inference than the CPU, yielding similar operational carbon footprint reductions. Unfortunately, the efficiency improvements come at the expense of embodied emissions. Compared to CPUs, the GPU's and DSP's additional silicon area increases the embodied footprint by 1.9× and 1.8×, respectively. Given the energy efficiency improvements of GPUs and DSPs, offsetting the additional embodied footprint requires an average lifetime utilization (i.e., *reuse* frequency) higher than 5% and 1%, respectively. These reuse frequencies linearly increase in the presence of renewable energy during operation—for example, with solar the GPU requires an average 35% utilization, well beyond the typical usage behavior of mobile platforms [32].

Carbon-optimization trade-offs. We use ACT's carbon-aware optimization targets to summarize the use-case dependent hardware design strategies. Figure 9 illustrates the design space between

Table 4: Mobile AI inference latency, power, operational footprint, and embodied footprint of a CPU, GPU, and DSP. CPU's are carbon optimal while co-processors are efficiency optimal.

Hardware	Latency	Power	OP_{CF}	E_{CF}
CPU	6.0 ms	6.6W	3.3 μg CO_2	253 g CO_2
DSP(+CPU)	12.1 ms	2.9W	3.1μg CO_2	205 (+253)g CO_2
GPU(+CPU)	9.2 ms	2.0W	1.5μg CO_2	189 (+253)g CO_2

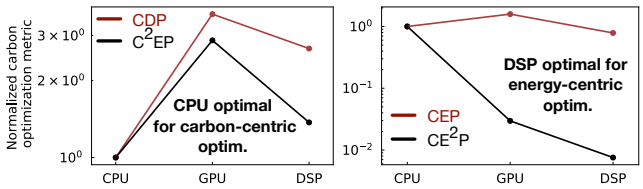

Figure 9: Using the proposed ACT carbon metrics, the optimal hardware design point varies. For embodied carbon-centric optimization targets, the CPU-based SoC is optimal due to lower manufacturing overheads (left). For operational carbon-centric optimization targets, the DSP-based SoC is optimal given the energy efficiency benefits of the co-processor.

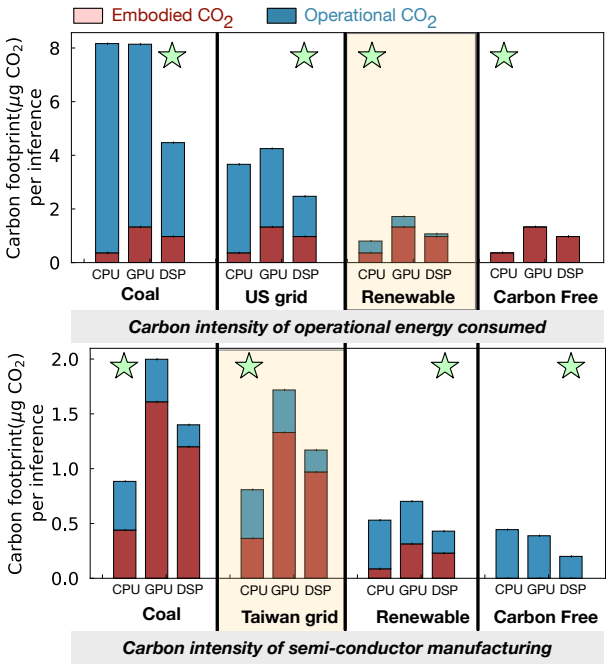

Figure 10: Varying carbon intensities during manufacturing (CI_{fab}) and use (CI_{use}). With increasing renewable energy during operation, the optimal hardware design point shifts from specialized (DSP) to general purpose (CPU) hardware given the reduced impact of energy efficiency optimization on overall carbon (top). With increasing renewable energy during manufacturing, the optimal design shifts from general purpose to specialized hardware given the lower embodied overhead of implementing co-processors (bottom).

the mobile CPU, GPU, and DSP based on the proposed carbon-optimization targets; results are normalized to the CPU-only design. The CPU-based SoC incurs the lowest manufacturing overheads given the lower die area; as such, the general purpose system is optimal for embodied carbon-centric optimization metrics, CDP and C^2EP. On the other hand, the DSP-based SoC is optimal for operational carbon-centric optimization targets, CEP and CE^2P, given the energy efficiency improvements. Recall, the choice of optimization target may depend on the availability of renewable energy.

Impact of renewable energy during fabrication and operational use. Figure 10 illustrates the impact of varying carbon intensities and renewable energy during hardware manufacturing and operational use on the overall carbon footprint of the CPU, GPU, and DSP-based SoC's. To survey the space, we separately explore the impact of renewable energy during use (top) and manufacturing (bottom). The top assumes fixed hardware manufacturing overheads based on the average power grid in Taiwan. The bottom assumes fixed operational overheads based on a renewable energy grid. The y-axis illustrates the embodied and operational emissions per-AI inference, assuming a 3-year lifetime.

Overall, the optimal hardware design, between programmable and *reusable* versus specialized hardware, varies based on the availability of renewable energy during manufacturing and use. With increasing renewable energy during operation, going from coal to carbon-free, the optimal provisioning decision changes from more specialized DSP's to generalized CPU's with 1.8× reduction (Figure 10 (top)). Intuitively, the co-processors optimize operational energy efficiency but benefits are mitigated in the presence of operational renewable energy since use-phase emissions account for a smaller fraction of the total footprint. On the other hand, Figure 10 (bottom) shows that, with increasing renewable energy, going from coal to carbon-free fabs, the optimal provisioning decision changes from general purpose CPU's to DSP's. Evidently, the additional manufacturing footprint of co-processors is reduced with "green" semiconductor fabs. As such, we imagine one way to design sustainable platforms for target markets more broadly, is for mobile vendors to offer different chipsets to users based on the fab characteristics and availability of renewable energy in the field.

6.2 Re-configurable hardware accelerators

Apart from provisioning systems to balance general purpose and specialized accelerators, designing re-configurable accelerators can maximize *reuse* across applications. Therefore, we build off of an example case study designing hardware for mobile SoC's [50, 102] to understand the performance, efficiency, and sustainability trade-offs across general purpose compute, specialized ASICs, and re-configurable accelerators. Figure 11 illustrates the performance and efficiency of three applications—FIR, AES encryption, and AI inference—on dual core ARM A53 CPUs ("CPU"), specialized AI ASIC ("Accel"), and an embedded FPGA ("FPGA") [104].

The optimal hardware design—CPU versus ASIC versus FPGA—varies based on the optimization target. In terms of performance, Figure 11 (top) shows ASIC achieves 26× higher AI performance than CPU but relies on the general purpose host processor to run remaining applications. Reconfigurable FPGA achieves 50×, 80×, and 24× higher performance than CPU, a geometric mean improvement of 45×. On the other hand, Figure 11 (bottom left) shows in

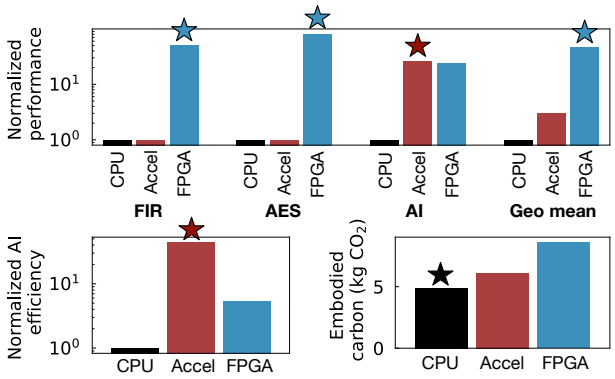

Figure 11: Based on the efficiency and carbon requirements the optimal hardware varies in terms of programmable CPU's, specialized ASIC's (Accel), and reconfigurable ASIC's (e.g., *FPGA*). Compared to CPU and *Accel*, an *FPGA* offers significant performance benefit across various applications such as FIR, AES, AI inference (top). *Accel* provides the highest energy efficiency for AI (bottom left). CPUs incur the lowest embodied footprint (bottom right).

terms of AI efficiency, ASIC is optimal providing 44× and 5× energy reduction per inference compared to CPU and FPGA, respectively. Finally, Figure 11 (bottom right) shows that in terms of embodied emissions, CPU incurs 1.3× and 1.8× lower footprint compared to ASIC and FPGA-based designs. Putting it all together, when designing SoC's for a variety of workloads, reconfigurable FPGA's offer a promising substrate providing both specialization and *reuse* to balance performance, efficiency, and carbon. In fact, across CDP, CEP, CE^2P, C^2EP, FPGA outperforms CPU and ASIC-based designs. However, when designing domain-specific SoC's for salient applications, such as AI, specialized ASICs provide higher performance and efficiency at lower carbon footprint.

7 REDUCE CASE STUDY: DESIGNING APPLICATION-SPECIFIC HARDWARE

Designing leaner specialized accelerators can **reduce** the footprint of hardware systems, where we minimize embodied carbon emissions while meeting the QoS targets set by applications. Given the benefits of specialized hardware for AI we saw earlier, this section studies the performance, efficiency, and sustainability design space of neural-processing units (NPU) based on an example NVIDIA Deep Learning Accelerator (NVDLA) [2]. First, we demonstrate that the optimal hardware design varies based on the optimization target (e.g., performance, efficiency, or carbon). Next, we show designing AI accelerators to meet a performance-driven QoS target of 30 FPS for image processing while minimizing embodied carbon footprints, yields up to 3× lower footprints than the performance and energy optimal configurations. Finally, we show hardware designers must be wary of Jevons paradox—the benefits of efficiency improvements are often overshadowed by increasing application demands. We find while AI accelerators in newer process nodes achieve higher efficiency, they also incur higher embodied emissions. Therefore, to curb and *reduce* the growing footprint from hardware advancement, designers must design systems under strict resource-budget.

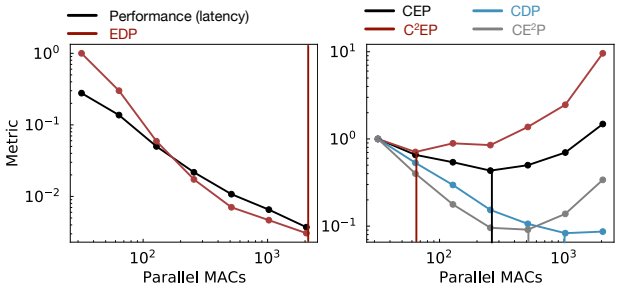

Figure 12: Design space for specialized AI accelerator based on NVDLA [2]. Varying the compute intensity (x-axis) we show the optimal design point for performance and EDP (left) varies compared to carbon-aware optimization metrics (right). Compared to the performance optimal design (2048 MACs), optimizing directly for sustainability reduces the carbon targets by up to 10×.

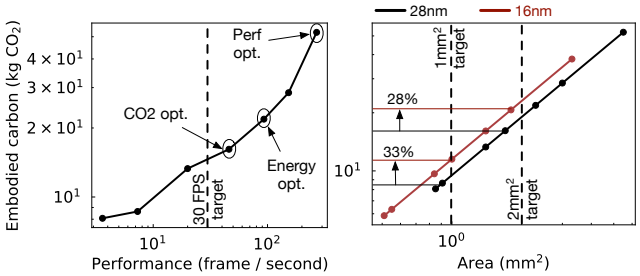

Figure 13: To reduce hardware footprint we must design leaner systems to minimize emissions while meeting QoS targets. (Left) Setting a QoS-target of 30 FPS, we find the optimal hardware configuration in terms of performance incurs a 3× higher footprint compared to the carbon-optimal design. (Right) Setting an area-based resource constraint we find newer technology nodes (28nm to 16nm) incur a roughly 30% higher footprint.

Carbon-aware AI ASIC optimization. Following the reuse analysis we find the optimal design point for AI accelerators varies across optimization targets. Figure 12 illustrates the impact of varying compute resources (i.e., parallelism) on performance and EDP (left) as well as carbon-aware optimization metrics (right) of a 16nm NVDLA based NPU [2]. Intuitively, the most parallel and compute-intensive design (2048 MACs) achieves the optimal performance and EDP. However, the optimal configuration for CDP, CE^2P, CEP, C^2EP are 1024, 512, 256, 128 MACs, respectively. This is due to the successively increasing embodied carbon overhead in more parallel hardware designs. Compared to the most parallel configuration, designing the accelerator based on the sustainability target *reduces* the carbon-aware optimization target by up to an order of magnitude. Accordingly, even when designing ASIC's for performance and efficiency, computer architects must consider sustainability as a first-order target.

QoS-driven sustainability design. To *reduce* embodied footprints we must design leaner systems to meet application-driven performance targets while minimizing hardware resources. Figure 13 (left) illustrates the trade-off between inference throughput versus embodied carbon footprint across a range of NVDLA designs

comprising 64 to 2048 MAC's in powers of 2. To achieve a QoS target of 30 FPS, we find the hardware design incurring the minimum embodied-carbon footprint (16g CO_2) comprises 256 MACs. On the other hand, the performance and energy optimal configurations incur 3.3× and 1.4× higher embodied emissions. This overhead owes to over-provisioning hardware as the performance and energy optimal points achieve 9× and 3× higher throughput than the QoS target. Therefore, designing leaner systems based on strict application-requirement is crucial to sustainable computing.

Resource constrained sustainability design. To *reduce* the rising embodied footprints overheads, in addition to QoS-constrained carbon optimization, we must also design systems within strict resource budgets to protect against rising overheads from Jevons paradox. Figure 13 (right) illustrates the trade-off between area (x-axis) and embodied carbon footprint for a 28nm and 16nm NVDLA-based NPU across a designs comprising 64 to 2048 MAC's in powers of 2. With two example resource constraints, 1mm^2 and 2mm^2 we show the optimal hardware configurations and their corresponding embodied carbon footprints. Within these resource constraints, going from 28nm to the newer 16nm technology node increases the embodied carbon footprint of the accelerator by 33% at 1mm^2 and 28% at 2mm^2. The increase in carbon footprint demonstrates the impact of Jevons paradox [5]; while improving technology nodes should lower the overall footprint, in practice the reductions are overshadowed by deploying more advanced systems with higher efficiencies.

8 RECYCLE CASE STUDY: EXTENDING HARDWARE LIFETIMES

The final tenet is **recycling** systems and components to enable second-life and eliminate wasted embodied emissions from unused resources. Recycling systems is a broad challenge that requires designing modular devices and extending hardware lifetimes. This section considers the sustainability implications of extending hardware lifetimes. Generally, extending hardware lifetimes can minimize embodied emissions as fewer systems are produced over time. Unfortunately, longer lifetimes may also degrade overall carbon footprint. Older hardware components exhibit lower performance and energy efficiency compared to state-of-the-art platforms, leading to higher operational emissions. Furthermore, extending lifetimes also requires improving hardware reliability characteristics which may incur higher operational and embodied emissions. Hardware design processes must consider the target lifetime of systems to balance operational and embodied emissions.

Extending mobile lifetime to balance life-cycle emissions. The design space of extending hardware lifetime and its impact on environmental footprint is complex given evolving factors that impact the computing landscape; such as, the evolution of workload demands and characteristics, varying availability of renewable energy over time, varying user behaviors interacting with their devices, and varying performance hardware performance across generations. To constrain the design space for the purposes of this *recycling* study, we assume fixed workloads, renewable energy, and user behaviors. The study then focuses on the trade-off between operational efficiency improvements and embodied carbon overheads of mobile devices.

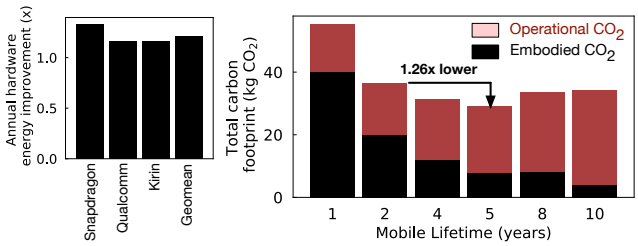

Figure 14: Hardware advancements yield a 1.21× annual energy efficiency improvement across a host of workloads on mobile SoC's (left). Extending mobile lifetimes reduces embodied carbon emissions over time but degrades operational emissions, as annual energy efficiencies are sacrificed (right). For example, over a 10 year period we find the optimal lifetime for mobile SoC's to be around 5 years, lowering the overall footprint by 1.26× compared to current average lifetimes of 2-3 years.

Figure 14 (left) illustrates the annual energy efficiency improvement for mobile SoC's. The analysis considers generations of Snapdragon [61–65], Exynos [67–70], and Kirin [38–41] SoC's and Geekbench-based mobile workloads (see Section 3.2). The average annual energy efficiency improvement is 21%.

Figure 14 (right) illustrates the impact of varying mobile lifetimes on their overall carbon—embodied (black) and operational (red) footprints of mobile IC's—over an example 10 year period. On the horizontal axis we vary the hardware lifetime; one year represents frequent hardware replacement while ten years represents infrequent replacement. Distinctly, extending hardware lifetime leads to lower embodied emissions as users acquire fewer mobile devices. However, extending lifetime leads to a degradation in operational emissions given the lower energy efficiency of older systems compared to state-of-the-art platforms available in the market. Hence, we analyze the interplay between lowering embodied emissions and raising operational emissions, which yields an optimal trade-off point around 5 years for the example SoC's. Compared to current lifetimes of 2-3 years, enabling second-life and extending hardware lifetimes reduce overall carbon footprint by up to 1.26×. Similarly, Sharhad et al. analyzed the total cost of ownership benefit of deploying decommissioned mobile devices as cheap and energy-efficient servers versus conventional servers [92]. Enabling longer lifetimes requires designers to carefully consider architectural characteristics, such as future proofing hardware to provide sufficient QoS and ensuring hardware reliability.

Improving SSD reliability to extend hardware lifetime. As an example of the impact of improving hardware reliability on *recycling* and system carbon footprint, we consider SSD-based storage systems. Based on SSD failures in the field, recent work [56] models the lifetime of storage systems as:

$$\text{Lifetime (years)} = \frac{\text{PEC} \times (1 + \text{PF})}{365 \times \text{DWPD} \times \text{WA} \times \text{R}_{\text{compress}}}$$

. Where PEC represents the the program-erase-cycles, PF represents the over provisioning factor, DWPD represents the number of full physical disk writes per day, WA represents the write amplification factor, and $\text{R}_{\text{compress}}$ represents the storage compression rate. Based on prior work, we fix PEC, DWPD, and $\text{R}_{\text{compress}}$ [56]. Figure 15

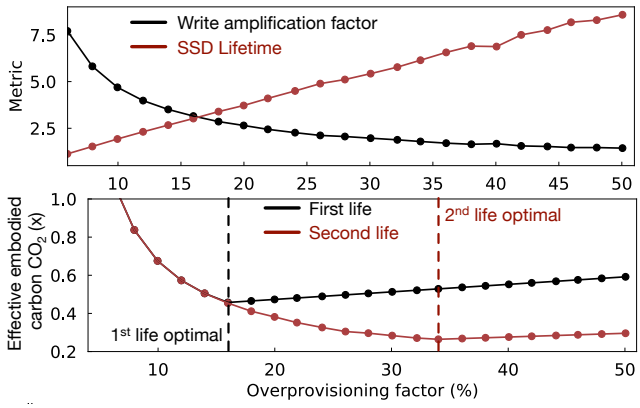

Figure 15: Extending hardware lifetimes to enable second-life (e.g., recycling) requires improving hardware reliability. We consider the impact of over provisioning SSD's to improve reliability and lifetime. The top shows the impact of increasing over provisioning on device write-amplification, which together determine the device lifetime. The bottom varies over provisioning ratios to extend lifetime, decreasing SSD's embodied footprints. Enabling second-life requires increasing the over provisioning ratio from 16% to 34%.

(top) illustrates the impact of varying PF (x-axis) on the required WA factor (black) for realistic SSD systems. Together the two determine the SSD lifetime (red). Intuitively, increasing the degree of over-provisioning protects storage systems from long term failures and extends hardware lifetime. Unfortunately, over provisioning also incurs higher embodied carbon footprint given the higher storage capacity.

Figure 15 (right) illustrates the impact of hardware reliability (i.e., over-provisioning SSD's) on the overall embodied footprint based on the ACT carbon model for storage devices (see Section 3). The y-axis illustrates the embodied footprint normalized to a baseline PF of 4%. We find for single mobile lifetime of about 2 years, the optimal over-provisioning factor is 16%; additional reliability guards incur higher capacity and exacerbate embodied emissions. Extending hardware lifetime to a second life reduces the embodied footprint by 1.8×; however, this requires increasing the over-provisioning factor to 34%. Improving reliability support and hardware lifetimes is crucial to allow systems and components to be *recycled* and re-purposed in future platforms.

9 CONCLUSION AND FUTURE WORK

With the rising ubiquity of computing, there is a pressing need to focus on innovating more environmentally sustainable computing technologies. Optimal system design and optimization strategies not only depend on workload and hardware characteristics but also semiconductor fabrication and environmental factors. Therefore, the computer system and architecture community must look beyond the efficiency advances from the last two decades towards the environmental impact of *end-to-end hardware life cycles*, including *both* manufacturing and operational use. To push the frontiers of sustainable system design, we propose our architectural carbon modelling tool ACT. Through three-principled case studies, we demonstrate designing systems and hardware for PPA versus sustainability yields unique solutions.

We summarize the main takeaways using ACT:

- Future system and hardware design cycles must consider sustainability as a first-order design metric alongside performance, power, area, and energy.
- Reducing the carbon footprint of systems requires architects to design leaner systems based on QoS requirements and identify the growing emissions with Jevon's paradox.
- Reusing hardware components to amortize emissions requires balancing general purpose and specialized architectures and improving overall hardware utilization.
- Recycling systems requires future proofing systems to preserve operational efficiency as workloads evolve and improving hardware reliability to extend hardware lifetimes.

We hope this work lays the foundation for future investment from the systems and architecture communities to design environmentally-sustainable systems.

ACKNOWLEDGMENTS

We would like to thank Frances Amatruda, Jordan Tse, Bobbie Manne, Doug Carmean, Jeff Glick, and the anonymous reviewers for their feedback on this paper.

A APPENDIX

To enable further investigation into carbon-aware system design, we describe, ACT, the proposed architectural carbon model in detail and the configurable parameters within it. We summarize the carbon intensity of energy sources (e.g., coal, solar, wind) and of geographic locations used in this paper. Next, we enumerate the embodied carbon intensities for memory, storage (e.g., SSD, HDD), and application processors. We provide the operational and embodied carbon intensities corresponding to ACT's modeling parameters in Table 1. Finally, we compare ACT's embodied carbon estimates to mobile and server scale LCA's.

A.1 Operational carbon intensities

Table 5 summarizes the average carbon intensity of energy generation based on various energy sources including coal, gas, solar, hydro-power, nuclear, and wind. Table 6 summarizes the average carbon intensity of energy generation based on geographic locations across the world. While these are average values, carbon intensity can fluctuate over time. Depending on the use case, the carbon intensity values represent the CI_{use} and CI_{fab} parameters of Table 1.

A.2 Embodied carbon intensities

Table 7 summarizes the embodied carbon parameters, energy per area (EPA) and gas per area (GPA), for application processor manufacturing, from 28nm to 3nm process nodes. Table 8 summarizes the materials per area (MPA) based on characterization from life-cycle analyses [19]. Table 9, Table 10, and Table 11 summarize the embodied carbon footprint for DRAM, SSD and HDD storage technologies, respectively.

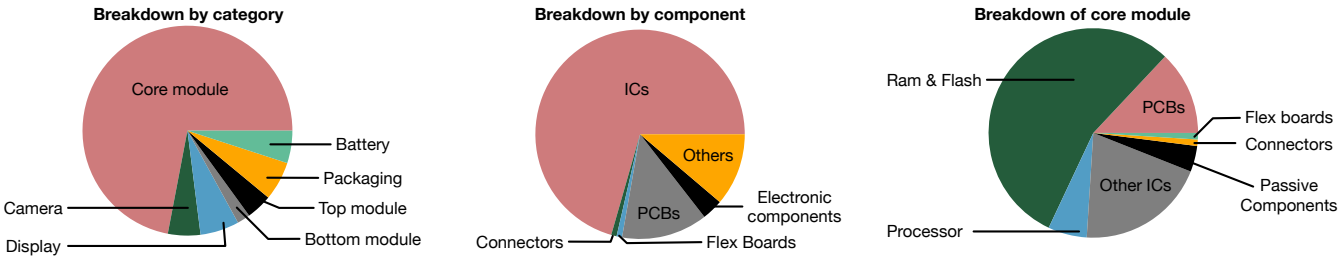

Figure 16: Breakdown of Fairphone 3 LCA by (a) category of module, (b) component-type, and (c) of core module.

Table 5: Carbon efficiency of various energy sources.

Source	Carbon intensity (g CO_2/kWh)	Energy-payback time (months)
Coal	820	2 [101]
Gas	490	1 [101]
Biomass	230	~12 [54]
Solar	41	~36 [49]
Geothermal	38	72 [52]
Hydropower	24	~12–36 [100, 101]
Nuclear	12	2 [101]
Wind	11	≤12 [18]

Table 6: Global carbon efficiency to produce energy [1, 17, 37].

Geographic average	Carbon intensity (g CO_2 / kWh)	Dominant energy source
World	301	–
India	725	Coal/gas
Australia	597	Coal
Taiwan	583	Coal/gas
Singapore	495	Gas
United States	380	Coal/gas
Europe	295	–
Brazil	82	Wind/hydropower
Iceland	28	Hydropower

Table 7: Embodied carbon parameters, EPA and GPA, for application processor manufacturing [33].

Process node	Energy per area (kWh/cm^2)	Gas per area (g CO_2/cm^2) (95% abated)	(99% abated)
28	0.90	175	100
20	1.2	190	110
14	1.2	200	125
10	1.475	240	150
7	1.52	350	200
7-EUV	2.15	350	200
7-EUV-DP	2.15	350	200
5nm	2.75	430	225
3nm	2.75	470	275

Table 8: Embodied carbon of raw material procurement.

Source	Embodied Carbon (g CO_2/ cm^2)
LCA [19]	500

Table 9: Embodied carbon of DRAM storage.

Technology node	Embodied Carbon (g CO_2/GB)
50nm DDR3	600
40nm DDR3	315
30nm DDR3	230
30nm LPDDR3	201
20nm LPDDR3	184
20nm LPDDR2	159
LPDDR4	48
10nm DDR4	65

Table 10: Embodied carbon of SSD storage [26, 27, 83, 84, 91]

Technology	Embodied Carbon (g CO_2/GB)
30nm NAND	30
20nm NAND	15
10nm NAND	10
1z NAND TLC	5.6
V3 NAND TLC	6.3
Western Digital 2016	24.4
Western Digital 2017	17.9
Western Digital 2018	12.5
Western Digital 2019	10.7
Seagate Nytro 1551	3.95
Seagate Nytro 3530	6.21
Seagate Nytro 3331	16.92

Table 11: Embodied carbon of Seagate's HDD storage [71–73, 78, 79, 82, 85–88]

Technology	Type	Embodied Carbon (g CO_2/GB)
BarraCuda	Consumer	4.57
BarraCuda2	Consumer	10.32
BarraCuda Pro	Consumer	2.35
FireCuda	Consumer	5.1
FireCuda 2	Consumer	9.1
Exos2x14	Enterprise	1.65
Exosx12	Enterprise	1.14
Exosx16	Enterprise	1.33
Exos15e900	Enterprise	20.5
Exos10e2400	Enterprise	10.3

A.3 ACT comparison

We compare ACT's embodied carbon estimates with the LCA-based results of existing systems, i.e., Dell R740 [22], Apple iPhone [8],

Table 12: IC LCA and ACT comparison.

IC	Device	Actual HW node	LCA node	LCA CO_2	ACT node 1	ACT CO_2	ACT node 2	ACT CO_2
RAM	Dell R740	10nm DDR4	50nm DDR3	533 kg	50nm DDR3	329 kg	10nm DDR4	64 kg
	Fairphone 3	14nm LPDDR4	50nm DDR3	see Flash + RAM	50nm DDR3	2.9 kg	1Xnm DDR4	0.5 kg
Flash	Apple iPhone	iPhone 11 - 10nm CPU	-	0.56 kg	10nm NAND	0.6 kg	V3 TLC	0.48 kg
	Dell R740 31TB	10nm NAND 10nm DDR4	45nm NAND 50nm RAM	3373 kg	30nm NAND 50nm DDR3	1440 kg	V3 TLC	583 kg
	Dell R740 400GB	10nm NAND 10nm DDR4	45nm NAND 50nm RAM	67 kg	30nm NAND 50nm DDR3	63 kg	V3 TLC	14 kg
	Fairphone 3	10nm NAND	50nm	see Flash + RAM	30nm NAND	2.3 kg	3V3 TLC 1Xnm LPDDR4	0.9 kg
Flash + RAM	Fairphone 3	10nm NAND 14nm LPDDR4	50nm NAND 50nm RAM	11 kg	30nm NAND 50nm RAM	5.2 kg	V3 TLC 1Xnm LPDDR4	0.9 kg
CPU	Dell R740	14nm	32nm	47 kg	28nm	22 kg	14nm	27 kg
	Fairphone 3	14nm	32nm	1.07 kg	28nm	0.9 kg	14nm	1.1 kg
Other ICs	Fairphone 3	14nm	32nm	5.3 kg	28nm	5.6 kg	14nm	6.2 kg

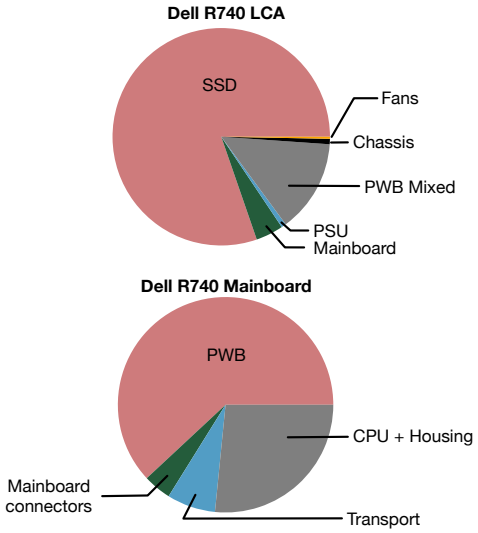

Figure 17: Breakdown of Dell R740 LCA.

and Fairphone 3 [60]. While LCA tools are widely-used in industry to quantify the environmental footprint of products, the primary purpose of the LCA is to generate environmental product reports but *not* directly for system and hardware design space exploration. It is, however, useful to see the comparison between ACT's estimates and the LCA-based results for the ICs — memory, storage, CPU, and other ICs — in Table 12. We present the actual hardware parameters used by the Dell R740 system, the Fairphone 3, and the Apple iPhone in the **Actual HW node** column whereas the columns of **LCA node**, **ACT node 1**, and **ACT node 2** present the hardware nodes used for the LCA and for the ACT model, respectively. W e attempt to model the hardware process used for the LCA using **ACT node 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 difference to come from the lack of up-to-date carbon emission data for the latest compute, memory, and storage technologies. The best-available, comprehensive environmental footprint study for semiconductors is Life-Cycle Assessment of Semiconductors [19]. The database and analysis is, however, dated, covering process technologies of CMOS (350nm to 32nm), DRAM (150nm to 57nm), and 2D NAND (150nm to 45nm). In comparison, the Dell R740 system features Intel 14nm Xeon CPUs, 1z-nm DDR4 (the third generation 10nm-class manufacturing technology) and 10nm NAND-based SSD storage. Furthermore, key semiconductor fab characteristics, such as carbon intensity of power grids used by hardware manufacturing, yield and efficiency, are evolving over time and vary by manufacturer, facility, and product line.

Figure 16 illustrates the carbon footprint breakdown over the system components for Fairphone 3 whereas Figure 17 illustrates the carbon footprint breakdown for the Dell R740 system. While IC's (e.g., processors, memory, storage) account for the majority of embodied emissions—roughly 70% for Fairphone 3 and 80% for Dell R740—other components have a non-negligible impact. As ACT enables research to characterize and optimize the architectural carbon emissions from IC's, designers should be careful in using ACT in reporting emissions for end-to-end mobile and server-class computing platforms.

REFERENCES

[1] 2020. electricityMap. https://www.electricitymap.org/map
[2] 2021. NVIDIA Deep Learning Accelerator. http://nvdla.org/
[3] 2021. Smart Cap Divider Charger: RT9759. https://www.richtek.com/Products/Battery%20Management/Single%20Cell%20Li-Ion%20Battery%20Charger/RT9759?sc_lang=en&specid=RT9759
[4] United States Environmental Protection Agency. 2021. Reduce, Reuse, Recycle. https://www.epa.gov/recycle
[5] Blake Alcott. 2005. Jevons' paradox. *Ecological economics* 54, 1 (2005), 9–21.
[6] Pradeep Ambati, Noman Bashir, David Irwin, and Prashant Shenoy. 2021. Modeling and Analyzing Waiting Policies for Cloud-Enabled Schedulers. *IEEE Transactions on Parallel and Distributed Systems* (2021).
[7] Anders SG Andrae and Tomas Edler. 2015. On global electricity usage of communication technology: trends to 2030. *Challenges* 6, 1 (2015), 117–157.
[8] Apple. 2019. Apple Environmental Responsibility Report. https://www.apple.com/environment/pdf/Apple_Environmental_Responsibility_Report_2019.pdf
[9] Apple. 2019. Product Environmental Report iPad. https://www.apple.com/environment/pdf/products/ipad/iPad_PER_sept2019.pdf
[10] Apple. 2019. Product Environmental Report iPhone 11. https://www.apple.com/environment/pdf/products/iphone/iPhone_11_PER_sept2019.pdf
[11] Apple. 2019. Product Environmental Report iPhone 11 Pro Max. https://www.apple.com/environment/pdf/products/iphone/iPhone_11_Pro_Max_PER_sept2019.pdf
[12] Apple. 2021. Apple announces Self Service Repair. https://www.apple.com/newsroom/2021/11/apple-announces-self-service-repair/
[13] Akhil Arunkumar, Evgeny Bolotin, David Nellans, and Carole-Jean Wu. 2019. Understanding the Future of Energy Efficiency in Multi-Module GPUs. In *2019 IEEE International Symposium on High Performance Computer Architecture (HPCA)*.
[14] Robert U Ayres. 1995. Life cycle analysis: A critique. *Resources, conservation and recycling* 14, 3-4 (1995), 199–223.
[15] Luiz André Barroso, Urs Hölzle, and Parthasarathy Ranganathan. 2018. The datacenter as a computer: Designing warehouse-scale machines. *Synthesis Lectures on Computer Architecture* 13, 3 (2018), i–189.
[16] Noman Bashir, Tian Guo, Mohammad Hajiesmaili, David Irwin, Prashant Shenoy, Ramesh Sitaraman, Abel Souza, and Adam Wierman. 2021. Enabling Sustainable Clouds: The Case for Virtualizing the Energy System. *arXiv preprint arXiv:2106.08872* (2021).
[17] Sewa Bhawan and RK Puram. 2018. CO2 Baseline Database for the Indian Power Sector.
[18] Alexandra Bonou, Alexis Laurent, and Stig I Olsen. 2016. Life cycle assessment of onshore and offshore wind energy-from theory to application. *Applied Energy* 180 (2016), 327–337.
[19] Sarah B Boyd. 2011. *Life-cycle assessment of semiconductors*. Springer Science & Business Media.
[20] David Brooks, Vivek Tiwari, and Margaret Martonosi. 2000. Wattch: A framework for architectural-level power analysis and optimizations. *ACM SIGARCH Computer Architecture News* 28, 2 (2000), 83–94.
[21] Erik Brunvand, Donald Kline, and Alex K. Jones. 2018. Dark Silicon Considered Harmful: A Case for Truly Green Computing. In *2018 Ninth International Green and Sustainable Computing Conference (IGSC)*. 1–8.
[22] Andreas Busa, Malcolm Hegeman, Jeff Vickers, Natalia Duque-Ciceri, and Constantin Herrmann. 2019. Life Cycle Assessment of Dell R740. https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/Full_LCA_Dell_R740.pdf
[23] Jichuan Chang, Justin Meza, Parthasarathy Ranganathan, Amip Shah, Rocky Shih, and Cullen Bash. 2012. Totally green: evaluating and designing servers for lifecycle environmental impact. *ACM SIGPLAN Notices* 47, 4 (2012), 25–36.
[24] Andrew A Chien, Richard Wolski, and Fan Yang. 2015. The zero-carbon cloud: High-value, dispatchable demand for renewable power generators. *The Electricity Journal* 28, 8 (2015), 110–118.
[25] Howard David, Eugene Gorbatov, Ulf R Hanebutte, Rahul Khanna, and Christian Le. 2010. RAPL: Memory power estimation and capping. In *2010 ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED)*. IEEE, 189–194.
[26] Western Digital. 2020. Sustainability Report 2020. https://documents.westerndigital.com/content/dam/doc-library/en_us/assets/public/western-digital/collateral/company/western-digital-2020-sustainability-report.pdf
[27] Western Digital. 2021. Life Cycle Assessment: Western Digital WD GreenTM SATA Solid State Drive (SSD). https://documents.westerndigital.com/content/dam/doc-library/en_us/assets/public/western-digital/collateral/analyst-report/life-cycle-assesment-wd-green-sata-ssd-western-digital.pdf
[28] Facebook. 2020. Facebook Sustainability Data 2019. https://sustainability.fb.com/wp-content/uploads/2020/05/2019-Sustainability-Data-Disclosure_Final-1.pdf
[29] Andrei Frumusanu. 2018. The Snapdragon 835 - A Quick Recap. https://www.anandtech.com/show/12520/the-galaxy-s9-review/2
[30] Andrei Frumusanu. 2019. The Apple iPhone 11, 11 Pro & 11 Pro Max Review: Performance, Battery, & Camera Elevated. https://www.anandtech.com/show/14892/the-apple-iphone-11-pro-and-max-review
[31] Andrei Frumusanu. 2019. The Apple iPhone 11, 11 Pro & 11 Pro Max Review: Performance, Battery, & Camera Elevated, The Apple A13 SoC: Lightning & Thunder. https://www.anandtech.com/show/14892/the-apple-iphone-11-pro-and-max-review/2
[32] Cao Gao, Anthony Gutierrez, Madhav Rajan, Ronald G. Dreslinski, Trevor Mudge, and Carole-Jean Wu. 2015. A study of mobile device utilization. In *2015 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*.
[33] M. Garcia Bardon, P. Wuytens, L.-A. Ragnarsson, G. Mirabelli, D. Jang, G. Willems, A. Mallik, A. Spessot, J. Ryckaert, and B. Parvais. 2020. DTCO including Sustainability: Power-Performance-Area-Cost-Environmental score (PPACE) Analysis for Logic Technologies. In *2020 IEEE International Electron Devices Meeting (IEDM)*. 41.4.1–41.4.4. https://doi.org/10.1109/IEDM13553.2020.9372004
[34] Geekbench. 2019. Geekbench 5. https://www.geekbench.com/doc/geekbench5-cpu-workloads.pdf
[35] Google. 2021. Google Environmental Report 2020. https://www.gstatic.com/gumdrop/sustainability/google-2020-environmental-report.pdf
[36] Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin S Lee, Gu-Yeon Wei, David Brooks, and Carole-Jean Wu. 2021. Chasing Carbon: The Elusive Environmental Footprint of Computing. In *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. IEEE, 854–867.
[37] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. 2020. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *arXiv preprint arXiv:2002.05651* (2020).
[38] HiSilicon. 2021. Kirin 960. https://www.hisilicon.com/en/products/Kirin/Kirin-flagship-chips/Kirin-960/
[39] HiSilicon. 2021. Kirin 970. https://www.hisilicon.com/en/products/Kirin/Kirin-flagship-chips/Kirin-970/
[40] HiSilicon. 2021. Kirin 980. https://www.hisilicon.com/en/products/Kirin/Kirin-flagship-chips/Kirin-980/
[41] HiSilicon. 2021. Kirin 990 5G. https://www.hisilicon.com/en/products/Kirin/Kirin-flagship-chips/Kirin-990-5G/
[42] SK hynix. 2018. SK hynix Sustainability Report 2018: DBL, SKHynix's New Growth Strategy. https://www.skhynix.com/sustainability/UI-FR-SA1601
[43] SK hynix. 2019. SK hynix Sustainability Report 2019. https://www.skhynix.com/sustainability/UI-FR-SA1601
[44] SK hynix. 2020. SK hynix Sustainability Report 2020. https://www.skhynix.com/sustainability/UI-FR-SA1601
[45] SK hynix. 2021. SK hynix Sustainability Report 2021. https://www.skhynix.com/sustainability/UI-FR-SA1601
[46] Carnegie Mellon University Green Design Institute. 2008. Economic Input-Output Life Cycle Assessment (EIO-LCA), US 1997 Industry Benchmark model. http://www.eiolca.net
[47] Vijay Kandiah, Scott Peverelle, Mahmoud Khairy, Junrui Pan, Amogh Manjunath, Timothy G. Rogers, Tor M. Aamodt, and Nikos Hardavellas. 2021. AccelWattch: A Power Modeling Framework for Modern GPUs. In *Proceedings of the 54th Annual IEEE/ACM International Symposium on Microarchitecture*.
[48] Donald Kline, Nikolas Parshook, Xiaoyu Ge, Erik Brunvand, Rami Melhem, Panos K. Chrysanthis, and Alex K. Jones. 2016. Holistically evaluating the environmental impacts in modern computing systems. In *2016 Seventh International Green and Sustainable Computing Conference (IGSC)*. 1–8.
[49] The US National Renewable Energy Laboratory. 2004. Photovoltalics FAQs.
[50] Sae Kyu Lee, Paul N Whatmough, Marco Donato, Glenn G Ko, David Brooks, and Gu-Yeon Wei. 2021. SMIV: A 16-nm 25-mm^2 SoC for IoT With Arm Cortex-A53, eFPGA, and Coherent Accelerators. *IEEE Journal of Solid-State Circuits* (2021).
[51] Jingwen Leng, Tayler Hetherington, Ahmed ElTantawy, Syed Gilani, Nam Sung Kim, Tor M. Aamodt, and Vijay Janapa Reddi. 2013. GPUWattch: Enabling Energy Optimizations in GPGPUs. In *Proceedings of the 40th Annual International Symposium on Computer Architecture*.
[52] Kewen Li, Huiyuan Bian, Changwei Liu, Danfeng Zhang, and Yanan Yang. 2015. Comparison of geothermal with solar and wind power generation systems. *Renewable and Sustainable Energy Reviews* 42 (2015), 1464–1474.
[53] Sheng Li, Jung Ho Ahn, Richard D Strong, Jay B Brockman, Dean M Tullsen, and Norman P Jouppi. 2009. McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures. In *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*. 469–480.
[54] Kristian Madsen and Niclas Scott Bentsen. 2018. Carbon debt payback time for a biomass fired CHP plant—A case study from Northern Europe. *Energies* 11, 4 (2018), 807.
[55] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. 2020. Recalibrating global data center energy-use estimates. *Science* 367, 6481 (2020), 984–986.

[56] Justin Meza, Qiang Wu, Sanjev Kumar, and Onur Mutlu. 2015. A large-scale study of flash memory failures in the field. *ACM SIGMETRICS Performance Evaluation Review* 43, 1 (2015), 177–190.

[57] Microsoft. 2021. 2020 Environmental Sustainability Report. https://www.microsoft.com/en-us/corporate-responsibility/sustainability/report

[58] Cynthia F. Murphy, G. Kenig, David T. Allen, J. Laurent, and D. E. Dyer. 2003. Development of parametric material, energy, and emission inventories for wafer fabrication in the semiconductor industry. *Environmental science & technology* 37 23 (2003).

[59] David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350* (2021).

[60] Marina Proske, David Sánchez, Christian Clemm, and Sarah-Jane Baur. 2020. LIFE CYCLE ASSESSMENT OF THE FAIRPHONE 3. https://www.fairphone.com/wp-content/uploads/2020/07/Fairphone_3_LCA.pdf

[61] Qualcomm. 2021. Snapdragon 820 Mobile Platform. https://www.qualcomm.com/products/snapdragon-820-mobile-platform

[62] Qualcomm. 2021. Snapdragon 835 Mobile Platform. https://www.qualcomm.com/products/snapdragon-835-mobile-platform

[63] Qualcomm. 2021. Snapdragon 845 Mobile Platform. https://www.qualcomm.com/products/snapdragon-845-mobile-platform

[64] Qualcomm. 2021. Snapdragon 855 Mobile Platform. https://www.qualcomm.com/products/snapdragon-855-mobile-platform

[65] Qualcomm. 2021. Snapdragon 865 Mobile Platform. https://www.qualcomm.com/products/snapdragon-865-mobile-platform

[66] Ana Radovanovic, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, et al. 2021. Carbon-Aware Computing for Datacenters. *arXiv preprint arXiv:2106.11750* (2021).

[67] Samsung. 2021. Mobile Processor Exynos 7 Octa (7420). https://www.samsung.com/semiconductor/minisite/exynos/products/mobileprocessor/exynos-7-octa-7420/

[68] Samsung. 2021. Mobile Processor Exynos 8895. https://www.samsung.com/semiconductor/minisite/exynos/products/mobileprocessor/exynos-9-series-8895/

[69] Samsung. 2021. Mobile Processor Exynos 9810. https://www.samsung.com/semiconductor/minisite/exynos/products/mobileprocessor/exynos-9-series-9810/

[70] Samsung. 2021. Mobile Processor Exynos 9820. https://www.samsung.com/semiconductor/minisite/exynos/products/mobileprocessor/exynos-9-series-9820/

[71] Seagate. 2017. Seagate EXOS 15E900 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/exos-15e900-report/

[72] Seagate. 2017. Seagate Exos X12 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/exos-x12-sustainability-report/

[73] Seagate. 2017. Seagate FireCuda Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/firecuda-sustainability-report/

[74] Seagate. 2017. Seagate Skyhawk 3TB Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/skyhawk-3tb-sustainability-report/

[75] Seagate. 2017. Seagate Skyhawk 6TB Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/skyhawk-6tb-sustainability-report/

[76] Seagate. 2017. Seagate Skyhawk Surveillance HDD Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/skyhawk-surveillance-hdd-report/

[77] Seagate. 2017. Seagate Video 3.5 HDD Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/video-3.5-hdd-report/

[78] Seagate. 2019. Seagate BarraCuda Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/barracuda-sustainability-report/

[79] Seagate. 2019. Seagate Exos 10E2400 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/exos-10e2400-sustainability-report/

[80] Seagate. 2019. Seagate Exos 5E8 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/exos-5e8-sustainability-report/

[81] Seagate. 2019. Seagate Exos 7E8 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/exos-7e8-sustainability-report/

[82] Seagate. 2019. Seagate Exos X16 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/exos-x16-sustainability-report/

[83] Seagate. 2019. Seagate Nytro 1551 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/nytro-1551-sustainability-report/

[84] Seagate. 2019. Seagate Nytro 3530 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/nytro-3530-sustainability-report/

[85] Seagate. 2020. Seagate BarraCuda Pro Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/barracuda-pro-sustainability-report/

[86] Seagate. 2020. Seagate BarraCuda Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/barracuda-2.5-sustainability-report/

[87] Seagate. 2020. Seagate Exos 2X14 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/exos-2x14-sustainability-report/

[88] Seagate. 2020. Seagate FireCuda Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/firecuda-3.5-sustainability-report/

[89] Seagate. 2020. Seagate IronWolf Pro Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/ironwolf-pro-report/

[90] Seagate. 2020. Seagate IronWolf Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/ironwolf-sustainability-report/

[91] Seagate. 2020. Seagate Nytro 3331 Sustainability Report. https://www.seagate.com/global-citizenship/product-sustainability/nytro-3331-sustainability-report/

[92] Mohammad Shahrad and David Wentzlaff. 2017. Towards Deploying Decommissioned Mobile Devices as Cheap Energy-Efficient Compute Nodes.. In *9th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 17)*.

[93] Ltd Siliconware Precision Industries Co. 2019. SPIL 2019 Corporate Social Responsibility. http://csr.spil.com.tw/uploads/2018-SPIL-CSR-EN.pdf

[94] SimaPro. 2021. LCA software to help you drive change. https://simapro.com/

[95] Sphera. 2021. Life Cycle Assessment Software. https://gabi.sphera.com/america/index/

[96] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243* (2019).

[97] Jennifer Switzer, Rob McGuinness, Pat Pannuto, George Porter, Aaron Schulman, and Barath Raghavan. 2021. TerraWatt: Sustaining Sustainable Computing of Containers in Containers. *arXiv preprint arXiv:2102.06614* (2021).

[98] TSMC. 2018. TSMC Corporate Social Responsibility Report 2018. https://esg.tsmc.com/download/csr/2018_tsmc_csr_report_published_May_2019/english/pdf/e_all.pdf

[99] TSMC. 2019. TSMC Corporate Social Responsibility Report 2019. https://esg.tsmc.com/download/csr/2019-csr-report/english/pdf/e-all.pdf

[100] Varun, Ravi Prakash, and I. K. Bhat. 2010. Life Cycle Energy and GHG Analysis of Hydroelectric Power Development in India. *International Journal of Green Energy* 7, 4 (2010), 361–375. https://doi.org/10.1080/15435075.2010.493803 arXiv:https://doi.org/10.1080/15435075.2010.493803

[101] Daniel Weißbach, G Ruprecht, A Huke, K Czerski, S Gottlieb, and A Hussein. 2013. Energy intensities, EROIs (energy returned on invested), and energy payback times of electricity generating power plants. *Energy* 52 (2013), 210–221.

[102] Paul N Whatmough, Sae Kyu Lee, Marco Donato, Hsea-Ching Hsueh, Sam Xi, Udit Gupta, Lillian Pentecost, Glenn G Ko, David Brooks, and Gu-Yeon Wei. 2019. A 16nm 25mm 2 SoC with a 54.5 x flexibility-efficiency range from dual-core Arm Cortex-A53 to eFPGA and cache-coherent accelerators. In *2019 Symposium on VLSI Circuits*. IEEE, C34–C35.

[103] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga Behram, James Huang, Charles Bai, Michael Gschwind, Anurag Gupta, Myle Ott, Anastasia Melnikov, Salvatore Candido, David Brooks, Geeta Chauhan, Benjamin Lee, Hsien-Hsin S. Lee, Bugra Akyildiz, Maximilian Balandat, Joe Spisak, Ravi Jain, Mike Rabbat, and Kim Hazelwood. 2022. Sustainable AI: Environmental Implications, Challenges and Opportunities. In *Proceedings of the Conference on Machine Learning and Systems*.

[104] Fang-Li Yuan, Cheng C Wang, Tsung-Han Yu, and Dejan Marković. 2014. A multi-granularity FPGA with hierarchical interconnects for efficient and flexible mobile computing. *IEEE Journal of Solid-State Circuits* 50, 1 (2014), 137–149.