Scalable Solid-State Drives (SSDs) have revolutionized the way we store and access our data across datacenters and handheld devices. Unfortunately, scaling technology can have a significant environmental impact. Across the globe, most semiconductor manufacturing use electricity that is generated from coal and natural gas. For instance, manufacturing a Gigabyte of Flash emits 0.16 Kg CO$_2$ and is a significant fraction of the total carbon emission in the system.

To better understand this concern, this paper compares the sustainability trade-offs between Hard Disk Drives (HDDs) and SSDs and recommends methodologies to estimate the embodied carbon costs of the storage system. In this paper, we outline four possible strategies to make storage systems sustainable. First, this paper recommends directions that help select the right medium of storage (SSD vs HDD). Second, this paper proposes lifetime extension techniques for SSDs. Third, this paper advocates for effective and efficient recycling and reuse of high-density multi-level cell-based SSDs. Fourth, specifically for hand-held devices, this paper recommends leveraging elasticity in cloud storage.

1 Introduction

Manufacturing, operating, transporting, and recycling computing systems, directly and indirectly, emit carbon dioxide (CO$_2$) and other greenhouse gases. As computing systems scale, their greenhouse contributions significantly impact global warming. This is highlighted by the pervasiveness of computing via hand-held devices, such as smartphones and tablets, and web services built around them. Moreover, digital data creation and consumption across the globe is snow bowling. As a result, carbon emissions due to personal devices, data centers, and networking infrastructure (known as the information and Communication Technologies (ICT) sector) are increasing rapidly. Today, about 2% of the total carbon emissions are estimated due to computing and networking devices combined [22, 23], and it is estimated to double in the next decade.

For example, the average household in the US has five to ten devices connected to the internet [30, 31]. We estimate that manufacturing and operating these devices for a year emits 2000 Kg CO$_2$ – equivalent to CO$_2$ emissions from driving a car for 5000 miles [20].

Most of the carbon emissions are because of the “conventional” electricity [6] that is used in the manufacturing and operation of computing systems [25]. For example, running and cooling the computing and networking hardware consumes significant electricity. If this electricity is generated from conventional carbon-intensive sources such as coal, natural gas, and crude oil, it will contribute to global warming. In contrast, electricity generated from renewable sources such as wind, solar, nuclear, and hydroelectric have a significantly small Global Warming Potential (GWP). Unfortunately, irrespective of whether they are hand-held devices or server nodes, manufacturing hardware and/or operating them require a significant amount of electricity – often from carbon-intensive conventional sources.

How to quantify Global Warming Potential? Typically, Global Warming Potential is the amount of CO$_2$ (or the equiv-
As a result, a significant fraction of CAPEX CO2e stems from LCA reports that quantify the CO2e of SSDs and HDDs. This paper makes the following four key contributions:

1. Analyze the overall carbon footprint of SSDs.
2. Compare overall sustainability of HDDs and SSDs.
3. Propose methodologies to estimate the embodied carbon cost of storage systems.
4. Discuss three high-level strategies for enabling sustainable storage systems.

This paper makes the following four key contributions:

1. Analyze the overall carbon footprint of SSDs.
2. Compare overall sustainability of HDDs and SSDs.
3. Propose methodologies to estimate the embodied carbon cost of storage systems.
4. Discuss three high-level strategies for enabling sustainable storage systems.

This section will first show a breakdown of embodied carbon emissions for the typical desktop system and analyze CO2e for SSDs and HDDs. Furthermore, we will discuss simple yet effective metrics to quantify the CO2e of storage systems.

2.1 System-level Breakdown of Embodied Cost

Figure 3 shows the breakdown of embodied CO2e for a Fujitsu workstation. The data is based on the Life Cycle Analysis (LCA) report published in [29]. The desktop system consists of a CPU with four cores, 8 GB DRAM, 1TB HDD, and SSD with 512GB storage capacity. For this system, the total CO2e is 706 Kg CO2e, wherein OPEX carbon emissions are 278Kg CO2 for five year lifetime, and CAPEX carbon emissions are 473 Kg CO2e. Typically, the LCA models carbon emissions from the cradle to the grave [37]. First, the LCA accounts for carbon emitted to extract the required materials during the mining, refining, and transport phases. Next, the LCA models estimate the carbon emissions for manufacturing components such as PCBs, Integrated Circuits, and Chassis. And lastly, it estimates CO2 emissions for operating, transporting, and recycling the device. Typically LCA software utilizes detailed models that account for the material models, process and packaging technology, yield, and geographic location of the manufacturing.

As seen in Figure 3, semiconductor components - SSD, DRAM, HDD, and CPU constitute a majority of the embodied carbon emissions, which is a result of the highly complex and energy-intensive manufacturing of semiconductor components. Amongst all the components, SSD has the most significant footprint as it uses multiple Flash and DRAM chips, which tend to have multiple silicon dies per package. Furthermore, the Flash and DRAM fabrication centers have limited renewable electricity supply, forcing fabs to use electricity generated from carbon-intensive sources.

---

1All green gas emissions are normalized against the global warming potential of CO2. For example, Methane has 25x global warming potential compared to CO2. As a result, 1 Kg emission of Methane is equivalent to 25 Kg CO2e.
This data is organized as per the increasing capacity of SSDs.

To understand the carbon footprint of SSDs, we analyzed 94 result using four-fold cross-validation. SEF equal to 0.16 Kg-CO2e/GB on average. We validate our the storage medium. Our evaluations show that SSDs have shown in the Figure 5 is a ratio of CO2e and the capacity of Storage Embodied factor (SEF)
ture, we propose devices so that we can develop sustainable storage architec-
CO2e per GB of flash storage. Figure 5 shows the distribution of embodied cost and quantify a ballpark estimate for the CO2e that using a diverse set of LCAs can capture a trend in the CO2e increases linearly with increasing capacity. We hope However, we observe a general trend that, on average, the embodied costs as the CO2e depends on several factors. 

2.2 Embodied Carbon Cost of SSDs

To understand the carbon footprint of SSDs, we analyzed 94 LCA reports that quantify the embodied cost of SSDs. Due to limited and dated direct life cycle studies of SSDs, we compile a data set of LCA reports for Server, Workstation, Desktop, Laptop, and Chromebook products that include SSD and quantify SSD overhead specifically [11, 18, 19, 26, 32]. Figure 4 show the CO2e for five datasets that we have compiled. Each dataset consists of a diverse set of devices in terms of capacity, technology node, and device type. The majority of the LCA reports are generated by the vendor assembling the computing system, such as Dell, HP, Apple, Fujitsu, etc. This data is organized as per the increasing capacity of SSDs. SSDs with identical capacity can have significantly different embodied costs as the CO2e depends on several factors. However, we observe a general trend that, on average, the CO2e increases linearly with increasing capacity. We hope that using a diverse set of LCAs can capture a trend in the embodied cost and quantify a ballpark estimate for the CO2e per GB of flash storage. Figure 5 shows the distribution of CO2e per GB of flash storage.

To quantitatively compare the embodied cost of the storage devices so that we can develop sustainable storage architecture, we propose Storage Embodied factor (SEF), which, as shown in the Figure 5 is a ratio of CO2e and the capacity of the storage medium. Our evaluations show that SSDs have SEF equal to 0.16 Kg-CO2e/GB on average. We validate our result using four-fold cross-validation.

2.3 Embodied Carbon Cost of HDDs

Solid State Drives are significantly more energy-efficient than Hard Drive Disks (HDDs). This is because they do not have any moving parts, which dramatically reduces the idling and active power. Furthermore, SSDs offer significantly higher bandwidth and lower latency resulting in improved overall system energy efficiency.

Moreover, HDDs are bulkier, and they need a higher quantity of material for manufacturing. Thus you would expect HDDs to have higher embodied CO2e compared to SSDs. However, our analysis of previously published LCA reports suggests the contrary. We use 24 LCA reports evaluating the SFE for HDDs manufactured by four vendors, with capacities ranging from 512GB to 6TB. Figure 6 shows the distribution of SFE and the average SFE. Note that SFE measures the CO2e/GB of storage. Compared to SSDs, the embodied cost of HDDs is at least an order of magnitude lower.

Figure 4: Carbon emissions for manufacturing 94 Solid State Drives, data based on Life Cycle Analysis (LCA) reports published by eight vendors.

Key Observation: A large fraction of CO2 emissions are due to semiconductor manufacturing. This fraction is increasing with technology scaling.

Figure 5: Distribution of estimated storage embodied factor for 94 Solid State Disks (SSDs).

Figure 6: Distribution of estimated storage embodied factor for 24 Hard Drive Disks (HDDs).
2.4 Impact of Technology Scaling on CO2e

![Figure 7: Carbon emissions for manufacturing cm² of silicon die for different technology nodes for two type of electricity generation schemes.]

By cramming more transistors within a fixed chip area, we can increase the Flash storage density. To that end, chip manufacturers use transistor scaling and 3D die stacking. By reducing transistor size, we can increase the number of transistors within the chip, and by stacking multiple wafer dies on top of each other, we add more capacity. Unfortunately, both of the strategies require a complex and energy-intensive manufacturing process. Moreover, chip manufacturing is becoming more and more complicated with the scaling of transistors. As a result, the carbon emissions for fabricating silicon die are increasing rapidly as transistor sizes are shrinking [12].

Figure 7 shows the carbon emissions from fabricating a logic die using different technology nodes. The number of fabrications steps and their energy intensity increase with reducing transistor feature sizes, resulting in higher carbon emissions. One of the key reasons for the high carbon emissions is the lack of renewable energy sources in the geographic locations where the majority of the semiconductor manufacturing happens. Thereby, most semiconductor fabrication plants rely on the electricity generated by coal and natural gas plants. The data in Figure 7 is derived from a recent study [12], which provides a detailed analysis of the impact of using carbon-intensive sources for chip manufacturing. Currently, a small fraction of electricity used in semiconductor manufacturing is renewable [33]. And although semiconductor manufacturers are committed to achieving carbon neutrality, transitioning from coal to renewable will be challenging and slow.

3 Strategies to Reduce Storage CO2e

We recommend three key strategies to reduce the embodied cost of carbon for storage devices. These strategies rely on selecting the appropriate storage medium, improved resource efficiency and endurance frameworks, recycling and reuse, and leveraging elasticity within cloud storage.

Table 1: Emissions of SSD and HDD.

<table>
<thead>
<tr>
<th>Storage</th>
<th>Energy (KWh)</th>
<th>OPEX CO2e (Kg)</th>
<th>CAPEX CO2e (Kg)</th>
<th>Total CO2e (Kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. Life</td>
<td>5yr</td>
<td>10yr</td>
<td>5yr</td>
<td>10yr</td>
</tr>
<tr>
<td>HDD (1TB)</td>
<td>183.9</td>
<td>367.9</td>
<td>79.6</td>
<td>159</td>
</tr>
<tr>
<td>SSD (1TB)</td>
<td>56.9</td>
<td>113.8</td>
<td>24.6</td>
<td>49.2</td>
</tr>
</tbody>
</table>

3.1 Selecting the Right Storage Medium

Flash-based storage is energy proportional as opposed to HDDs, which consume significant idle power [24]. Therefore, conventional wisdom suggests that replacing HDDs by SSDs would reduce energy consumption and make storage environmentally sustainable [28]. However, we argue that basing all the decisions only on the OPEX CO2e can be misleading and one must account for the embodied cost when deciding on the storage architecture. To illustrate this, we evaluate the CAPEX and OPEX CO2e for 1TB of SSD and HDD devices using the average Storage Embodied Factor (SEF) data and average power consumption of HDD and SDD devices.

For instance, considering a workload with 20% active and 80% idle cycles, Table 1 estimates the energy consumed by 1TB SSD and HDD for a five and ten-year time span. In our model, we assume the average HDD power consumption to be 4.2 W, whereas SSD consumes 1.3 W power. We evaluate the OPEX CO2e from total energy consumption using an emission factor specified by EPA². For calculating the CAPEX, we use the average SEF for both HDD and SSD. Furthermore, we assume both SSD and HDD have five year lifetime; thus, for storing the data for ten years, we account for an additional CAPEX upgrade cost. Table 1 shows that the overall CO2e for HDD is significantly lower compared to the SSDs. This is due to a relatively lower embodied cost. However, this first order analysis does not consider the impact of SSDs on the overall performance, power, and energy consumption.

3.2 Extending the Lifetime of SSDs

The embodied carbon cost can also be amortized by extending the lifetime of SSD devices. To this end, we advocate using four approaches.

---

²0.7Kg/KWh is a typical emission factor based on the U.S. national weighted average CO2 marginal emission rate based on the year 2019 data [21]
3.2.1 Inter-Node Wear Leveling

The storage cells in modern Flash-based SSDs have low endurance and can only be written 10K to 100K times [2]. To improve endurance, modern SSD devices use table-based wear-leveling [3, 15]. In this technique, a table is used to dynamically redirect updated data into locations that have lower intensity of writes. This ensures that all cells in the SSD, on average, wear-out uniformly. To reduce wear-out, it would be useful to explore wear-leveling strategies by looking across several storage nodes. Essentially, this will explore trade-offs in increasing access latencies while broadly extending the lifetime of flash devices in the system.

3.2.2 Intelligent Data Placement: SLC versus MLC

The endurance of a flash device depends on whether the device uses single-level cells (SLC) or multi-level cells (MLC) [14]. MLCs store 2+ bits in each cell and thereby increase the information capacity of flash [13]. However, writing an MLC cell is an iterative process. Thus, the endurance of MLC is significantly lower than SLC. This affects the lifetime and the density of flash devices. Going into the future, one can rethink how user data can be distributed across systems that have SLC and MLC flash devices to maximize their effective lifetime. Moreover, techniques such as Zoned Namespaces (ZNS), designed to improve performance isolation, can be tuned to improve SSDs’ lifetime and sustainability.

3.2.3 Recycling and Reusing Flash Devices

We can try to recycle and reuse the MLC devices as low-capacity SLC devices. This is primarily because, while MLC devices wear out quickly, they can be morphed into lower capacity SLC or a hybrid storage medium that uses both SLC and MLC devices. Unfortunately, from the perspective of a datacenter node or a hand-held device, SLC devices tend to offer significantly lower capacity benefits. We argue that given the CAPEX embodied carbon costs of SSD devices, it would be useful to re-employ (or recycle) these morphed SLC SSDs into other storage nodes – such as those that employ data logging, store dormant container images, etc.

3.2.4 Efficient Error Correction Codes (ECC)

We can also better-utilize the existing SSDs by employing strong and efficient error correcting codes (ECCs). Modern SSDs already employ Low-Density Parity-Check Codes (LPDC). These codes are capable of correcting several faulty bits within a block of data [38]. Depending on the usage pattern, it would be beneficial to design new ECC codes that help increase the lifetime of flash devices while incurring capacity overheads. Complex ECC can also have longer encoding and decoding latencies [16] thereby increasing the access time of the SSD [27] – presenting an interesting trade-off to explore.

3.3 Leverage Elasticity in Cloud Storage

Unlike datacenters, fabricating composable (or modular) hand-held systems is a major challenge [36]. Recently, there has been a renewed interest in repairing and recycling hand-held electronics [1, 5]. While these strategies help reduce the effective CO2e overheads for the aggregate device lifetime, they also introduce life-style changes for the end users. Furthermore, interconnecting components that span across different technology generations may not be very efficient. To overcome these challenges, we can rethink data management across the devices and the cloud.

Key advantages of cloud storage is that it is scalable, secure, composable, and can have higher durability by using effectively using data redundancy while providing elastic capacities [7, 8, 17, 34, 35]. Thus, manufactures and service providers can determine which data is stored locally on the hand-held device and propose a cloud storage pricing model that takes the costs of embodied carbon, cloud latency, and network throughput into account.

4 Summary

Aggressive technology scaling has enabled flash-based SSDs to provide high-density storage from datacenters to hand-held devices. However, before we settle on the choice of SSD-based systems, we must understand their environmental impacts. In this paper, we highlight the embodied carbon costs of SSDs during their manufacturing and operation. We show that, in general, SSDs consume a significant fraction of the total embodied carbon costs. To better understand this concern, this paper compares the sustainability trade-offs between HDDs and SSDs and recommends methodologies to estimate the embodied carbon costs of the storage system. To tackle this concern, this paper presents strategies, namely selecting the storage medium appropriately, resilience and endurance optimizations, recycling and reusing flash devices, and elastically using cloud storage for hand-held devices. We contend that the recommended changes to storage architectures will help guide long-term storage system design, impact the design of billions of devices, and reduce their embodied carbon costs in the coming decades.

Acknowledgment

We thank the anonymous reviewers for their insightful suggestions and feedback.

References


