Understanding the Implications of Uncertainty in Embodied Carbon Models for Sustainable Computing

Anvita Bhagavathula
akb249@cornell.edu
Cornell Tech
New York, New York, USA

Leo Han
lxh4@cornell.edu
Cornell Tech
New York, New York, USA

Udit Gupta
ugupta@cornell.edu
Cornell Tech
New York, New York, USA

ABSTRACT

Quantifying the embodied carbon emissions of computer systems is a growing area of interest and several models to do so have already been established. However, these models are fundamentally deterministic and do not account for the fact that hardware and software characteristics of systems have inherent uncertainties – spatial, temporal, process-driven, and system-driven – which can cause extensive variation in carbon emission estimates. Our work aims to investigate the effect of these uncertainties on embodied carbon footprint estimates and system design choice trade-offs. We propose a novel probabilistic framework that generates distribution based outputs of embodied carbon emissions. We demonstrate the approach on modeling chip manufacturing carbon footprint by first characterizing uncertainty in the energy-per-area, gas-per-area, yield, and carbon intensity of fabrication parameters and observing how embodied carbon per area estimates change for various technology process nodes. We then apply this distribution-based framework to a case study focusing on laptop GPU chip choice, comparing the NVIDIA RTX A3000 Mobile with the Intel Arc Pro A60M from a sustainability perspective. Lastly, we investigate how the probability of making the correct carbon optimal choice changes when the carbon footprint is embodied-dominated or operational-dominated.

ACM Reference Format:
Anvita Bhagavathula, Leo Han, and Udit Gupta. 2024. Understanding the Implications of Uncertainty in Embodied Carbon Models for Sustainable Computing. In *Proceedings of 3rd Workshop on Sustainable Computer Systems (HotCarbon'24).* ACM, New York, NY, USA, 7 pages.

1 INTRODUCTION

The Information and Communication Technology (ICT) industry currently accounts for 2.1-3.9% of global carbon emissions (measured in gCO_2) [14]. This number is only projected to grow as global computing demands scale over the next decade. A key priority is to therefore collectively reduce these emissions; for example, the ICT industry has committed to reducing greenhouse gas (GHG) emissions by 45% by 2030 in accordance with the Paris Agreement [18]. An essential step to achieving this target is to accurately quantify emissions generated by computer systems. Hardware vendors currently publish life cycle assessments (LCAs) for their devices [1, 2, 21, 22, 25]. These LCAs and recent studies have shown that embodied emissions of ICT hardware — the emissions from manufacturing, raw materials, transport, and installation — can be as significant as or more significant than operational emissions — emissions from energy consumed during operation [1, 2, 17, 20–22, 32]. In addition to industry product environmental reports, researchers are actively developing new methods to enable finer-grained, component-level carbon modeling estimates, producing tools and models such as ACT [16], ECO-CHIP [24], GreenChip [19], and 3D-Carbon [33]. These tools consider hardware parameters such as process technology, silicon area, storage capacity, etc. to provide hardware embodied carbon estimates.

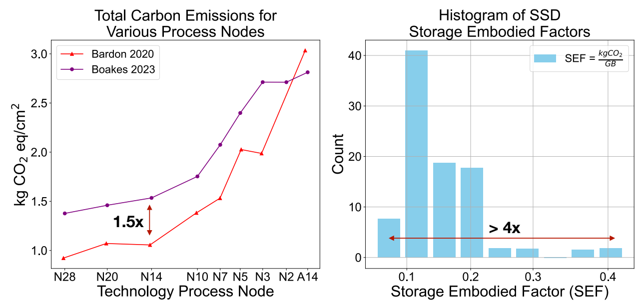

Figure 1: Benchmarking total emissions of various technology process nodes generated by two different LCA databases published in different years, illustrating how updates to embodied parameters change carbon footprint estimations (left), adapted from [6]. Distribution of estimated storage embodied factors for 94 SSDs (right), adapted from Figure 5 in [25].

However, a fundamental limitation of all of these models is that they generate singular deterministic carbon footprints. These outputs contradict the fact that hardware and software characteristics of systems have inherent uncertainties – spatial, temporal, process-driven, and system-driven – which can cause extensive variations in embodied and operational carbon emission estimates [9, 13, 25]. For example, carbon-per-area estimates between the 2020 [5] and 2023 [6] iMec assessments, shown in Figure 1 (left), vary as much as 1.5x for some process nodes. Similar variations exist for SSD embodied carbon estimates — Tannu and Nair compiled 94 LCA reports from Apple, Dell, HP, and Seagate and found that the reported embodied carbon footprint per GB can vary from less than 0.1 $kgCO_2e$ to more than 0.4 $kgCO_2e$, as shown in Figure 1 (right) [25]. Unaware of these data uncertainties, designers who use deterministic carbon models for sustainable device design may believe they are meeting their sustainability goals while actual results may deviate greatly from expectations.

HotCarbon'24, July 9, 2024, Santa Cruz, CA

Accounting for uncertainty during hardware design has already been explored in the context of performance — past work by Cui and Sherwood proposed the notion of architectural risk which accounts for projection, process, and design uncertainties within a system's performance model [9]. In light of the large uncertainties associated with embodied carbon estimates, a similar probabilistic approach is needed to inform hardware designers of sustainability risks. This work aims to investigate the bottom-up effect of such uncertainties on embodied carbon footprint estimations to provide device designers with a probabilistic carbon modeling tool, allowing them to make risk-aware, informed sustainable design decisions. This analysis also has the potential to address complex problems such as ensuring a company's worst-case footprint remains below a certain threshold. Our contributions are as follows:

(1) We propose a novel framework that generates distribution-based outputs of embodied carbon emissions instead of singular values, allowing hardware designers to make risk-aware sustainable design decisions.
(2) We characterize uncertainty of energy-per-area (EPA), gas-per-area (GPA), yield, and carbon intensity of fabrication (CI_{fab}) parameters to estimate embodied carbon per area distributions for various technology process nodes, using semiconductor manufacturing reports, power grid data, and governmental industrial process reports.
(3) We apply the framework to a case study about optimal laptop GPU chip choice, comparing the NVIDIA A3000 Mobile with the Intel A60M from a sustainability perspective. We show that there is a 27.08% chance that using a deterministic model can lead to choosing the wrong device when optimizing for embodied carbon. We also demonstrate how the optimal GPU choice, indicated by the probability of making the correct carbon-optimal choice, changes when the carbon footprint is embodied-dominated or operational-dominated.

2 MODELING APPROACH

2.1 A Probabilistic Embodied Carbon Model

To formalize a probabilistic embodied carbon model, we transform an existing deterministic model to a distribution-based one. Specifically, let us take ACT as an example [16]. The ACT model is defined as a combination of the operational and embodied carbon footprints of a given computer system. The embodied carbon is discounted by the fraction of application run-time (T) with respect to the overall system lifetime (LT):

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

We are interested in the embodied portion of this model, which is categorized on a per-component level (E_r) for each of the application processors (SoC), memory (DRAM) and storage (SSD and HDD) elements. In this work, we characterize the uncertainty associated with several of the embodied SoC parameters:

$$\mathrm{E_{SoC}} = \mathrm{Area} \times \mathrm{CPA} = \frac{\mathrm{Area}}{\mathrm{Y}} \times (\mathrm{CI_{fab}} \times \mathrm{EPA} + \mathrm{GPA} + \mathrm{MPA}) \tag{2}$$

The embodied footprint is a combination of 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 gasses used during hardware manufacturing (GPA), emissions from procuring raw materials for fab manufacturing (MPA), and the carbon intensity of energy used during fabrication (CI_{fab}). To make this model into a probabilistic one, we transform several independent parameters into distributions as shown in Figure 2. Then, we leverage an outer sum or outer product operation if performing operations between two parameter distributions. Output matrices are then flattened and visualized as a histogram.

2.2 Model Assumptions and Distribution Construction Techniques

This work focuses on characterizing uncertainty in several of the embodied parameters of the ACT framework. Distributions for these parameters were generated either by leveraging kernel density estimation or by using a prior implied by the style of uncertainty chosen for the parameter's data. Kernel density estimation (KDE) is a technique used to estimate the probability density function of a random variable by applying a Gaussian kernel weighting technique to data. It is used when the behavior of a sparse data sample is extrapolated to resample data to a larger sample size [31]. The second distribution construction technique we used was inferring a prior based on the style of uncertainty chosen for a parameter's component data. For example, 95% confidence intervals of relative errors on emission factors indicate that the data follows a normal distribution with a standard deviation that can be obtained using the relative errors.

We augment the original ACT framework [16] by modeling EPA, GPA, yield, and CI_{fab} probabilistically and assume fixed values for MPA. The lack of available data prevents us from generating a probability distribution for MPA, but the model is parametric and can include MPA as a source of uncertainty should data become available or by using synthetic user-defined distributions. Moreover, while the number of metal layers affects EPA and GPA and can thus be a source of uncertainty, we assume that they are fixed per process node as that is the assumption made by our source on EPA and GPA data [6]; we do not have EPA and GPA data at a layer granularity.

2.3 Parameter-Level Uncertainty Characterization

2.3.1 EPA Distribution. EPA represents the energy consumed per unit area manufactured and is represented by fabrication energy numbers. Uncertainty in these values arises from temporal shifts in process energy efficiencies. To construct EPA distributions, we leveraged TSMC's annual improvement of process energy efficiency data for several technology process nodes [29]. We developed two different iterations of distributions for EPA values: one based on EPA values from the original ACT framework (ranging from 0.8 - 3.5 kWh per cm^2) and the other using updated fabrication electricity consumption data published by imec (ranging from 1.56 - 3.77 kWh per cm^2) [6, 16]. Distributions were constructed by normalizing each process node's energy efficiency numbers by the first annual value in its respective series and then dividing the corresponding EPA values by the normalized yearly improvements on the energy efficiency. A key choice we made was to assume that the discrete

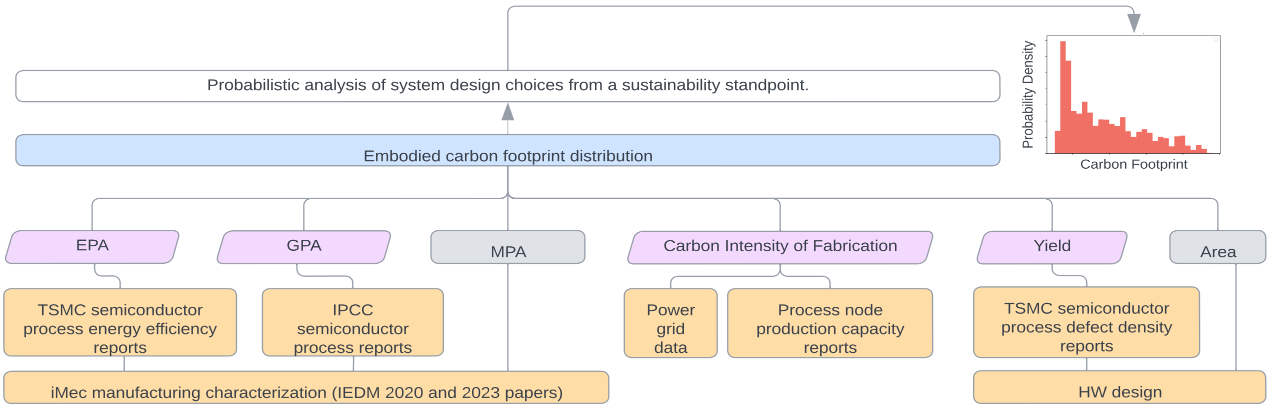

Figure 2: Diagram of our proposed framework that generates distribution-based outputs of embodied carbon emissions. The model's main parameters are hardware or manufacturing characteristics that are either single-value inputs, shown in grey, or distribution-based inputs, shown in purple. To characterize these parameters and their associated uncertainties, we use published fab characterization, hardware design attributes, power grid data, and industry environmental reports. These data and uncertainty sources are highlighted in orange. The output embodied carbon footprint distribution can be leveraged for probabilistic analysis of system design choices.

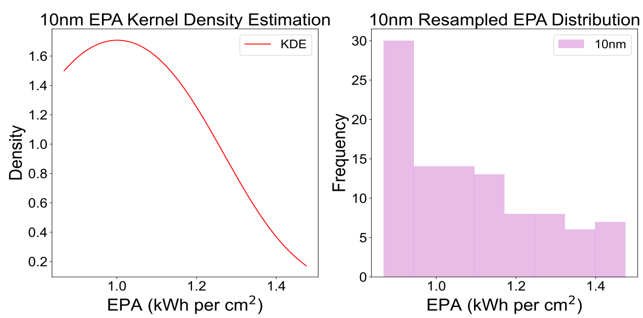

Figure 3: Kernel density estimate (KDE) of temporally distributed EPA values derived from TSMC's annual improvement of process node energy efficiency data [29] (left) and an example of a resampled distribution of EPA data using the KDE for a 10nm process node (right).

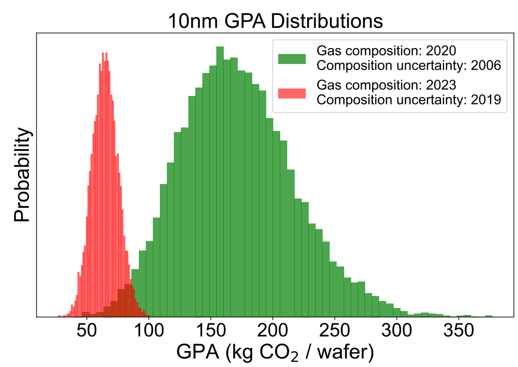

Figure 4: GPA distributions for a 10nm process node developed using two combinations of data sources. The green distribution uses 2006 iMec gas composition data in conjunction with 2020 IPCC process gas uncertainties [5, 13]. The red distribution uses 2023 iMec gas composition data in conjunction with 2019 IPCC process gas uncertainties [6, 7]. These distributions illustrate the effect of changing process-level uncertainties on the spread of GPA values.

EPA values were the ones corresponding to the energy efficiency at time of mass production. Modifying how this is instantiated can impact the numbers within the EPA distributions. These distributions were turned into histograms and their density functions were estimated using KDE. Their kernel density estimates were used to resample to larger distribution sizes as shown in Figure 3.

2.3.2 GPA Distribution. GPA represents the greenhouse gas emissions produced by fabrication facilities in semiconductor manufacturing (kgCO_2 per wafer). Uncertainty in these values is due to relative errors in process gas values during the fabrication process. To develop GPA distributions, we used 95% confidence intervals of relative errors on process gas emission factors published by the Intergovernmental Panel on Climate Change (IPCC) [7, 13]. This style of uncertainty indicates that the GPA data follows a normal distribution. Another layer of uncertainty in the GPA values also arises due to temporal change in gas composition during fabrication. We address both process-level and temporal uncertainties by developing two iterations of GPA distributions, considering changes in error values published by IPCC as well as updates made to gas compositions used in fabrication processes [5, 6].

The first iteration utilizes emission errors and process gas compositions from earlier dates (2006 and 2020) [5, 13]. Normal distributions for each composite gas were generated using Tier 2a errors for both 95% and 99% abatement processes. The second iteration uses updated errors and gas compositions (2019 and 2023) [6, 7], utilizing Tier 2c abatement errors. For gases lacking specified errors, deterministic values were sampled multiple times. To generate a

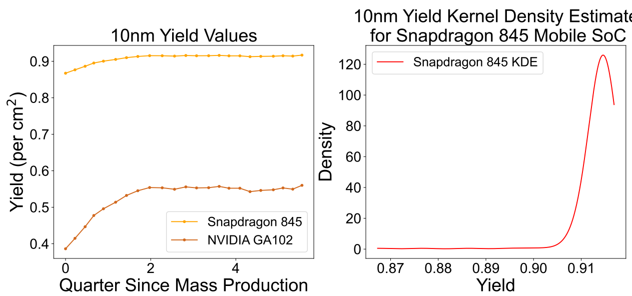

Figure 5: Yield values for the 10nm process node for two different chips (Snapdragon 845 and NVIDIA GA102) derived from TSMC defect density data [10] (left) and a kernel density estimate example for the 10nm process node in a Snapdragon 845 (right).

composite distribution, the distribution of each composite gas was randomly sampled (with replacement) and added to an array. This approach ensures that each gas in the composite distribution has an independent chance of being selected for inclusion in each sample. Figure 4 compares these two iterations of GPA distributions, illustrating how process-level and temporal uncertainties affect GPA distribution spread.

2.3.3 Yield Distribution. Yield represents the ratio of defect-free dies on a wafer relative to the total number of dies on the wafer. The uncertainty in this value arises from temporal shifts in defect density of fabrication facilities. We leverage TSMC's defect density data over time (defects/cm^2) for 3 different process nodes [10] and the Poisson yield model to generate corresponding yield numbers over time [12]. Yield data for a 10nm process node for two chips, the Snapdragon 845 mobile SoC and NVIDIA GA102 GPU, are shown in Figure 5 (left) [15, 28]. Histograms for yield data were then created and KDE was used to generate the probability density function of this data for resampling. Figure 5 illustrates this process for the Snapdragon 845 mobile SoC.

2.3.4 Carbon Intensity of Fabrication. CI_{fab} represents the energy consumed by manufacturing equivalent in carbon emissions. This value is dependent on a fab's energy source (e.g. renewables vs. non-renewables) or a fab's geographical location. The uncertainty in carbon intensity values arises from temporal shifts in carbon intensity values throughout the year due to seasonal energy demands. In the instance where we are certain about a chip's manufacturing location, each fab region will have its own distribution generated using historical CI_{fab} trends. Specifically, we leverage three years of Electricity Maps' historical carbon intensity data [3, 4] that changes over time and construct histograms using these values as shown in Figure 6 (left).

In the case where we are uncertain of a chip's manufacturing location, we can define the choice of fabrication facility region as a discrete random variable with probabilities proportional to total production capacity for relevant process nodes [8]. For instance, in the case of a 10nm process node, the global production capacity is distributed between South Korea (31% as of 2022) and Taiwan (69% as of 2022) [8]. These proportions are assumed to represent the likelihood that each geographical location is selected for fabrication. A mixture distribution for this process node was constructed

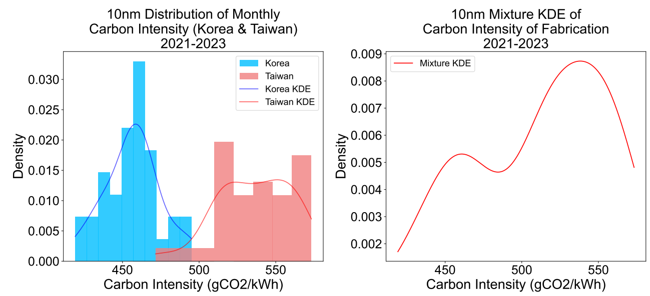

Figure 6: Histograms of monthly carbon intensity data from Korea and Taiwan between 2021-2023 with their individual kernel density estimates (KDEs) [3, 4] (left) and a weighted combination of their KDEs that form their composite mixture probability distribution function (right).

by sampling from each region. For this example, we assumed that the 10nm process node is equivalent to <10nm nodes in terms of global production capacity. Kernel density estimates (KDEs) for each region were constructed using carbon intensity data from 2021 to 2023. Lastly, a composite mixture distribution for a 10nm process node was generated by randomly sampling from each region's KDE, with the likelihood of choosing a sample from each KDE weighted by the global production capacity of that region. The composite mixture distribution is constructed by aggregating these weighted samples. Figure 6 (right) illustrates this composite distribution construction process.

2.4 Embodied Carbon Distributions

Now that we have formalized probability distributions for several embodied parameters, let us examine the outputs of our framework. We select two chips, Snapdragon 845 and NVIDIA GA102, of different sizes and generate the E_{SoC} distributions for them, focusing on their 7nm and 10nm process node emissions. We present the distribution-based carbon footprint outputs in Figure 7. We can make several observations from this data. First, the GA102 distributions exhibit higher variance than the Snapdragon 845 ones, indicating that chip area is proportional to distribution variance. This can be attributed to the fact that the embodied distributions are directly scaled by area. Second, smaller technology process nodes yield higher embodied carbon footprints, likely because smaller nodes require denser integration of components on the chip, leading to increased energy consumption during manufacturing. Additionally, advanced materials and fabrication techniques in smaller nodes may contribute to greater environmental impact throughout production (which manifests in higher EPA and GPA values). Overall, we conclude that embodied carbon footprint distributions exhibit larger variance in uncertainty for larger chip areas and smaller process nodes.

3 CASE STUDY

Now that we have established this distribution-based framework, we apply it to a system design choice case study. In our case study scenario, we are building a sustainable notebook computer and choosing from two notebook GPUs of the same performance class:

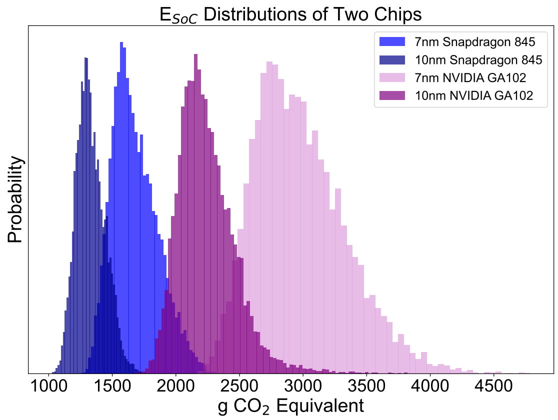

Figure 7: Embodied carbon footprint distributions for 7nm and 10nm process nodes for two different chips: Snapdragon 845 and NVIDIA GA102. These distributions show that larger chip areas and smaller process nodes have greater variance in their embodied carbon footprint estimates.

the Intel A60M and the NVIDIA A3000 Mobile. The two GPUs have different chip areas and are made with different processes: the Intel A60M is 269 mm^2 using TSMC's N6 process where as the NVIDIA A3000 Mobile is 392 mm^2 using Samsung's 8N process [26, 27]. TSMC's N6 process is part of its 7nm class of processes and we use public 7nm manufacturing data for it [30]; Samsung's 8N process is part of its 10nm class of processes and we use public 10nm manufacturing data for it [23]. We assume the probability distributions of the Intel A60M and NVIDIA A3000 Mobile are independent as they are from different vendors, built with different processes at different fabs, and released in different years (2021 and 2023).

From a power-efficiency standpoint, NVIDIA A3000 Mobile is more efficient, with a thermal design power (TDP) consumption of 70 W as compared to the Intel A60M's 95 W TDP [26, 27]. However, the Intel A60M is more optimal in terms of area. We generate the E_{SoC} distributions for both hardware components as depicted in Figure 8 by first generating a set of carbon-per-area (CPA) samples per our methods and data in Section 2.3.

Initially, our instinct is to use the central tendencies (e.g. mean, median, mode) of these distributions to inform our decision-making process. Based on this approach, we observe that the embodied carbon footprint of the Intel A60M's distribution is lower than that of the NVIDIA A3000 Mobile, suggesting that the Intel A60M may be the better choice from a sustainability standpoint. However, there is significant overlap in the distributions of the NVIDIA A3000 Mobile and the Intel A60M, suggesting there is a substantial possibility than the Intel A60M carbon footprint can actually be higher. With the distributions generated by our method, as shown in Figure 8, we calculate there to be a 27.08% chance that the embodied carbon footprint of the Intel A60M is actually higher than that of NVIDIA A3000 Mobile.

Embodied carbon alone does not represent the entirety of a device's carbon footprint; we must also consider operational carbon emissions which are driven by operational power. We leverage the Normalized Carbon Footprint (NCF) metric first proposed as part of the FOCAL framework [11]. This metric can be used to compare the carbon footprint of two different designs and is a weighted sum of the normalized embodied and operational footprint of the two designs X and Y. The weight, defined by the parameter α, is determined by an anticipated use case (whether a workload is foreseen as more operational or embodied intensive).

$$\mathrm{NCF}_{\mathrm{ft},\alpha}(\mathrm{X},\mathrm{Y}) = \alpha\frac{\mathrm{E_X}}{\mathrm{E_Y}} + (1-\alpha)\frac{P_X}{P_Y} \tag{3}$$

where the normalized embodied footprint is computed as the ratio of the embodied distributions, E_X and E_Y, and the normalized operational footprint is the ratio of the power consumption, P_X and P_Y, respectively. An NCF value <1 implies that X has a lower footprint than Y whereas an NCF value >1 implies that X has a higher footprint than Y. We want to ensure that we are making the correct decision for a given X and Y configuration; this happens when we choose X with an NCF < 1. In cases where X and Y are not perfect substitutes and we would prefer X for another reason such as cost or performance, we can define a threshold NCF value higher than 1, $\mathrm{NCF}_{Threshold}$, to represent the sustainability trade-off we are willing to make. On the other hand, if X is less preferred than Y given the same carbon footprint, we can set $\mathrm{NCF}_{Threshold}$ lower than 1 to represent the amount of sustainability gains required to shift the preference in favor of X. This allows us to define the probability of choosing the correct carbon-optimal device as the area under an NCF distribution curve that falls below $\mathrm{NCF}_{Threshold}$.

In the default scenario where X and Y are perfect substitutes otherwise, the $\mathrm{NCF}_{Threshold}$ is 1. The carbon-optimal choice varies across operational intensive and embodied intensive regimes which are characterized by varying α values — when α is large, the embodied carbon footprint dominates overall carbon footprint, while operational carbon footprint dominates when alpha is small. For example, when a CPU is powered off or idle for most of its use life, then its lifetime operational carbon footprint is relatively small compared to its embodied carbon footprint. This case is in the embodied-dominant regime and modeled with a large α. On the other end, a CPU that consistently runs at high utilization levels will have its lifetime carbon emissions dominated by operational carbon emissions and is modeled by a small α.

To investigate this, we sweep 100 α values to generate $P(NCF < 1)$values for two configurations: (X=NVIDIA A3000 Mobile, Y=Intel A60M) and the inverse (X=Intel A60M, Y=NVIDIA A3000 Mobile), denoted as configuration 1 and 2 respectively. These results are presented in Figure 9. We define high certainty decision making regions as ones where $P(NCF < 1) > 90\%$ or $P(NCF < 1) < 10\%$. When $P(NCF < 1) > 90\%$ for a given X, Y configuration, we pick X as the optimal design with high certainty. In contrast, when $P(NCF < 1) <$ 10%, we choose Y as the optimal design with high certainty. Our data reveals that for an $\mathrm{NCF}_{Threshold}$ of 1, the NVIDIA A3000 Mobile is the optimal choice for operational-intensive workloads ($0 \leq \alpha \leq 0.5$) with high certainty. The NVIDIA A3000 Mobile is also the optimal choice with relative certainty when ($0.5 < \alpha \leq 0.8$), whereas the Intel A60M is the optimal choice with relative certainty for the highly embodied-dominant regime ($0.8 < \alpha \leq 1$). ($0.5 \leq \alpha \leq 1$) define the region of where we have a <90% chance of choosing correctly. Through this analysis, we demonstrate that

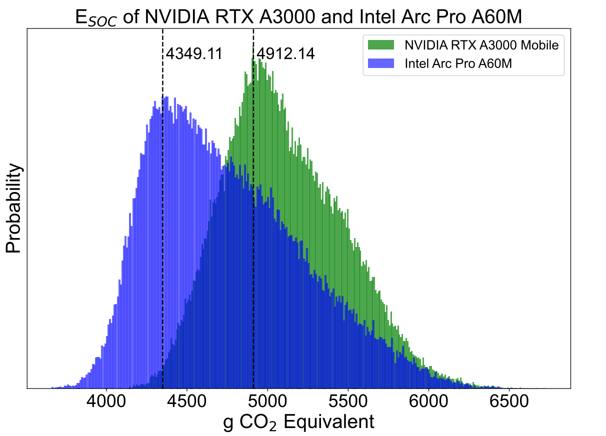

Figure 8: Embodied carbon distributions of two laptop GPU chips: the NVIDIA A3000 Mobile and the Intel A60M. While the distribution peaks indicate that the Intel A60M incurs less embodied carbon, there is a region of uncertainty in which both distributions intersect, leading to a 27.08% chance that the Intel A60M actually incurs more carbon than the NVIDIA A3000 Mobile.

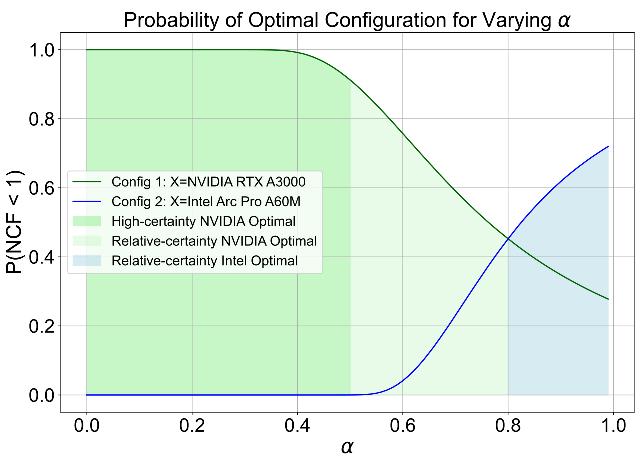

Figure 9: Probability of choosing correctly (P(NCF < NCF$_{Threshold}$)) when choosing between the NVIDIA A3000 Mobile and the Intel A60M across embodied and operational dominant regimes. We identify distinct regions where the optimal decision can either be made with high or relative certainty. The relative-certainty regions imply there is a significant probability that we will select a non-optimal GPU.

an optimal design choice from a sustainability perspective cannot be made based on a single deterministic carbon footprint. Instead, workload characteristics (embodied-intensive versus operational intensive) and hardware parameter uncertainty distributions must also be considered.

We can further extend this analysis by varying the desired NCF$_{Threshold}$, or the factor of improvement we want X to exhibit with respect to Y. We vary NCF$_{Threshold}$ between 0.7 - 1.5 and plot the P(NCF < NCF$_{Threshold}$) values as shown in Figure 10. These

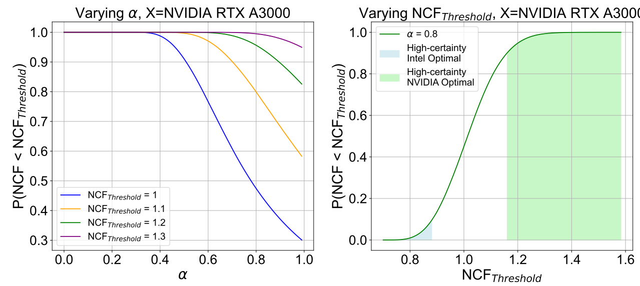

Figure 10: Probability of choosing correctly when choosing between the NVIDIA A3000 Mobile and the Intel A60M for various NCF$_{Threshold}$ values (left) and for a fixed embodied-dominant regime where we show high-certainty decision making regions defined by the choice of NCF$_{Threshold}$ (right).

curves demonstrate that as the NCF$_{Threshold}$ increases, the range of α values for which X, the NVIDIA A3000 Mobile in this case, is the optimal solution widens, indicating that the choice of this threshold value can also impact which design choice is optimal for a given workload. We fix $\alpha = 0.8$, which is a scenario where the embodied footprint dominates, and identify NCF$_{Threshold}$ values for which we can make a optimal decision with high certainty. In Figure 10, we use configuration 1 and identify high-certainty decision making regions as being where one of the choices has a >90% chance of being carbon-optimal, i.e. P(NCF < NCF$_{Threshold}$) > 90% or < 10%. We find that the NVIDIA A3000 Mobile is optimal with high certainty at $\alpha = 0.8$ for NCF$_{Threshold}$ > 1.16 and the Intel A60M is optimal with high certainty at the same α for NCF$_{Threshold}$ < 0.87.

4 SUMMARY AND FUTURE

This work, as far as we know, is the first attempt at integrating uncertainty into embodied carbon footprint estimations for computer systems. We proposed a novel probabilistic framework to generate distribution-based outputs of embodied carbon. Next, we characterized the temporal and process-driven uncertainties present in EPA, GPA, yield, and CI$_{fab}$ parameters. We then applied our distribution-based framework to a system design choice case study and demonstrated that our probabilistic carbon footprint estimates can help designers make more risk-aware sustainable design choices. Future avenues of research include mapping carbon footprint probabilities to a cost function (e.g., financial cost, carbon cost) to incentivize more robust configuration practices. In the long term, we also aim to integrate this style of probabilistic analysis into the hardware design loop itself to guide sustainable hardware design.

ACKNOWLEDGMENTS

We sincerely thank Michael Shen and the anonymous reviewers for their valuable feedback. We would also like to acknowledge the CloudLab infrastructure support (NSF 1419199) for providing resources to conduct these experiments. This work is also in part supported by NSF Grant CCF-2324860.

REFERENCES

[1] Apple. 2023. *Product environmental report: iPhone 15 Pro and iPhone 15 Pro Max.* Technical Report.

[2] Apple. 2023. *Product environmental report: iPhone 15 Pro and iPhone 15 Pro Max.* Technical Report.

[3] Electricity Maps ApS. 2024. Electricity Maps - South Korea - 2021 to 2023. Website. Retrieved June 29, 2024 from https://app.electricitymaps.com/zone/US-CAR-SCEG

[4] Electricity Maps ApS. 2024. Electricity Maps - Taiwan - 2021 to 2023. Website. Retrieved June 29, 2024 from https://app.electricitymaps.com/zone/US-CAR-SCEG

[5] Marie Garcia Bardon and Bertrand Parvais. 2020. The environmental footprint of logic CMOS technologies: A DTCO-based analysis. Website. Retrieved May 6, 2024 from https://www.imec-int.com/en/articles/environmental-footprint-logic-cmos-technologies

[6] Lizzie Boakes, M. Garcia Bardon, V. Schellekens, I.-Y. Liu, B. Vanhouche, G. Mirabelli, F. Sebaai, L. Van Winckel, E. Gallagher, Cedric Rolin, and L.-Å. Ragnarsson. 2023. Cradle-to-gate Life Cycle Assessment of CMOS Logic Technologies. *2023 International Electron Devices Meeting (IEDM)* (2023), 1–4. https://api.semanticscholar.org/CorpusID:267525112

[7] Eduardo Calvo Buendia, Kiyoto Tanabe, Andrej Kranjc, Baasansuren Jamsranjav, Maya Fukuda, Sekai Ngarize, Akira Osako, Yurii Pyrozhenko, Pavel Shermanau, and Sandro Federici. 2019. *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Volume 3 Industrial Processes and Product Use.* Intergovernmental Panel on Climate Change (IPCC), Switzerland, Chapter 6.

[8] Matthew Connatser. 2024. America Will Make 28% of Advanced Chips in 2032. Website. Retrieved June 29, 2024 from https://www.theregister.com/2024/05/08/us_china_sia/

[9] Weilong Cui and Timothy Sherwood. 2017. Estimating and understanding architectural risk. In *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture* (Cambridge, Massachusetts) *(MICRO-50 '17)*. Association for Computing Machinery, New York, NY, USA, 651–664. https://doi.org/10.1145/3123939.3124541

[10] Ian Cutress. 2020. 'Better yield on 5nm than 7nm': TSMC update on defect rates for N5. Website. Retrieved May 6, 2024 from https://www.anandtech.com/show/16028/better-yield-on-5nm-than-7nm-tsmc-update-on-defect-rates-for-n5

[11] Lieven Eeckhout. 2024. FOCAL: A First-Order Carbon Model to Assess Processor Sustainability. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM. https://doi.org/10.1145/3620665.3640415

[12] EESemi. 2005. Test yield models. Website. Retrieved May 6, 2024 from https://www.eesemi.com/test-yield-models.htm

[13] Simon Eggleston, Leandro Buendia, Kyoko Miwa, Todd Ngara, and Kiyoto Tanabe. 2006. *2006 IPCC Guidelines for National Greenhouse Gas Inventories: Volume 3 Industrial Processes and Product Use.* Institute for Global Environmental Strategies (IGES) on behalf of the Intergovernmental Panel on Climate Change (IPCC), Hayama, Japan, Chapter 6.

[14] Mariam Elgamal, Doug Carmean, Elnaz Ansari, Okay Zed, Ramesh Peri, Srilatha Manne, Udit Gupta, Gu-Yeon Wei, David Brooks, Gage Hills, and Carole-Jean Wu. 2023. Carbon-Efficient Design Optimization for Computing Systems. In *Proceedings of the 2nd Workshop on Sustainable Computer Systems* (Boston, MA, USA) *(HotCarbon '23)*. Association for Computing Machinery, New York, NY, USA, Article 16, 7 pages. https://doi.org/10.1145/3604930.3605712

[15] Andrei Frumusanu. 2018. The samsung galaxy S9 and S9+ review: Exynos and Snapdragon at 960fps. https://www.anandtech.com/show/12520/the-galaxy-s9-review/2

[16] 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* (New York, New York) *(ISCA '22)*. Association for Computing Machinery, New York, NY, USA, 784–799. https://doi.org/10.1145/3470496.3527408

[17] 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.

[18] ITU. 2020. ICT industry to reduce greenhouse gas emissions by 45 per cent by 2030. Website. Retrieved May 2, 2024 from https://www.itu.int/en/mediacentre/Pages/PR04-2020-ICT-industry-to-reduce-greenhouse-gas-emissions-by-45-percent-by-2030.aspx

[19] Donald Kline, Nikolas Parshook, Xiaoyu Ge, Erik Brunvand, Rami Melhem, Panos K. Chrysanthis, and Alex K. Jones. 2019. GreenChip: A tool for evaluating holistic sustainability of modern computing systems. *Sustainable Computing: Informatics and Systems* 22 (2019), 322–332. https://doi.org/10.1016/j.suscom.2017.10.001

[20] Paul Lin, Robert Bunger, and Victor Avelar. 2023. *Quantifying Data Center Scope 3 GHG Emissions to Prioritize Reduction Efforts.* Technical Report.

[21] Microsoft. 2024. *Surface Pro (11th edition) Ecoprofile.* Technical Report.

[22] Thinkstep on behalf of Dell Technologies. 2019. *Life Cycle Assessment of Dell R740.* Technical Report.

[23] David Schor. 2018. VLSI 2018: Samsung's 8nm 8LPP, a 10nm extension. Website. Retrieved July 2, 2024 from https://fuse.wikichip.org/news/1443/vlsi-2018-samsungs-8nm-8lpp-a-10nm-extension/

[24] Chetan Choppali Sudarshan, Nikhil Matkar, Sarma B. K. Vrudhula, Sachin S. Sapatnekar, and Vidya A. Chhabria. 2024. ECO-CHIP: Estimation of Carbon Footprint of Chiplet-based Architectures for Sustainable VLSI. In *IEEE International Symposium on High-Performance Computer Architecture, HPCA 2024, Edinburgh, United Kingdom, March 2-6, 2024.* IEEE, 671–685. https://doi.org/10.1109/HPCA57654.2024.00058

[25] Swamit Tannu and Prashant J. Nair. 2023. The Dirty Secret of SSDs: Embodied Carbon. *SIGENERGY Energy Inform. Rev.* 3, 3 (oct 2023), 4–9. https://doi.org/10.1145/3630614.3630616

[26] TechPowerUp. 2021. NVIDIA RTX A3000 Mobile Specs | TechPowerUp GPU Database. Website. Retrieved July 2, 2024 from https://www.techpowerup.com/gpu-specs/rtx-a3000-mobile.c3806

[27] TechPowerUp. 2023. Intel Arc Pro A60M Specs | TechPowerUp GPU Database. Website. Retrieved July 2, 2024 from https://www.techpowerup.com/gpu-specs/arc-pro-a60m.c4161

[28] TechPowerUp. 2024. NVIDIA GA102 GPU Specs | TechPowerUp GPU Database. https://www.techpowerup.com/gpu-specs/nvidia-ga102.g930

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

[30] TSMC. 2024. N7/N6. Website. Retrieved July 2, 2024 from https://www.tsmc.com/english/dedicatedFoundry/technology/platform_DCE_N7_N6

[31] Jake Vanderplas. 2024. 2.8. Density Estimation — scikit-learn.org. Website. Retrieved May 6, 2024 from https://scikit-learn.org/stable/modules/density.html

[32] 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. arXiv:2111.00364 [cs.LG] https://arxiv.org/abs/2111.00364

[33] Yujie Zhao, Yang Zhao, Cheng Wan, and Yingyan Lin. 2024. 3D-Carbon: An Analytical Carbon Modeling Tool for 3D and 2.5D Integrated Circuits. arXiv:2307.08060 [cs.AR] https://arxiv.org/abs/2307.08060