Data Centers on Wheels: Emissions from Computing Onboard Autonomous Vehicles

Soumya Sudhakar, Vivienne Sze, Sertac Karaman
soumyas@mit.edu, sze@mit.edu, sertac@mit.edu
Authors are with the Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139
617.324.7352

Abstract

While much attention has been paid to data centers' greenhouse gas emissions, less attention has been paid to autonomous vehicles' (AVs) potential emissions. In this work, we introduce a framework to probabilistically model the emissions from computing onboard a global fleet of AVs and show that the emissions have the potential to make a non-negligible impact on global emissions, comparable to that of all data centers today. Based on current trends, a widespread AV adoption scenario where approximately 95% of all vehicles are autonomous requires computer power to be less than 1.2 kW for emissions from computing on AVs to be less than emissions from all data centers in 2018 in 90% of modeled scenarios. Anticipating a future scenario with high adoption of AVs, business-as-usual decarbonization, and workloads doubling every three years, hardware efficiency must double every 1.1 years for emissions in 2050 to equal 2018 data center emissions. The rate of increase in hardware efficiency needed in many scenarios to contain emissions is faster than the current rate. We discuss several avenues of future research unique to AVs to further analyze and potentially reduce the carbon footprint of AVs.

IEEE MICRO Special Issue on Environmentally Sustainable Computing

October 2022

Data Centers on Wheels: Emissions from Computing Onboard Autonomous Vehicles

Soumya Sudhakar
Massachusetts Institute of Technology
soumyas@mit.edu

Vivienne Sze
Massachusetts Institute of Technology
sze@mit.edu

Sertac Karaman
Massachusetts Institute of Technology
sertac@mit.edu

***Abstract*— While much attention has been paid to data centers' greenhouse gas emissions, less attention has been paid to autonomous vehicles' (AVs) potential emissions. In this work, we introduce a framework to probabilistically model the emissions from computing onboard a global fleet of AVs and show that the emissions have the potential to make a non-negligible impact on global emissions, comparable to that of all data centers today. Based on current trends, a widespread AV adoption scenario where approximately 95% of all vehicles are autonomous requires computer power to be less than 1.2 kW for emissions from computing on AVs to be less than emissions from all data centers in 2018 in 90% of modeled scenarios. Anticipating a future scenario with high adoption of AVs, business-as-usual decarbonization, and workloads doubling every three years, hardware efficiency must double every 1.1 years for emissions in 2050 to equal 2018 data center emissions. The rate of increase in hardware efficiency needed in many scenarios to contain emissions is faster than the current rate; we discuss several avenues of future research unique to AVs to further analyze and potentially reduce the carbon footprint of AVs.**

I. INTRODUCTION

There has been great interest in industry and academia in characterizing the emissions from data centers [1], [2], especially with respect to increased workloads expected from deep neural networks (DNNs) [3]. In 2018, data centers collectively consumed an estimated 205 TWh or 1% of the world's electricity [4] and contributed to about 0.3% of the world's emissions [1], with demand expected to grow [5]. However, less attention has been paid to the carbon footprint of computing in the emerging field of autonomous vehicles (AVs).

There is reason to expect the amount of computing will be significant onboard Level 4 or Level 5 AVs, where a human back-up driver is unnecessary [6]. For a global fleet of AVs, the overall computing workload is comparable and may even exceed current data centers' workloads if AVs are widely adopted. For example, Facebook runs trillions of DNN inferences per day across its data centers [2]; an AV that drives for an hour per day computing 10 DNN inferences at 60 Hz on each of the inputs of 10 cameras would make 21.6 million inferences per day, and one billion AVs would make 21.6 quadrillion inferences per day! Due to the computing capability onboard, AVs have even been referred to as "data centers or supercomputers on wheels" [7]. While

This work was partially supported by National Science Foundation (NSF) Cyber-Physical Systems program grant no. 1837212, NSF Real-Time Machine Learning program grant no. 1937501, and the MIT-Accenture Fellowship.

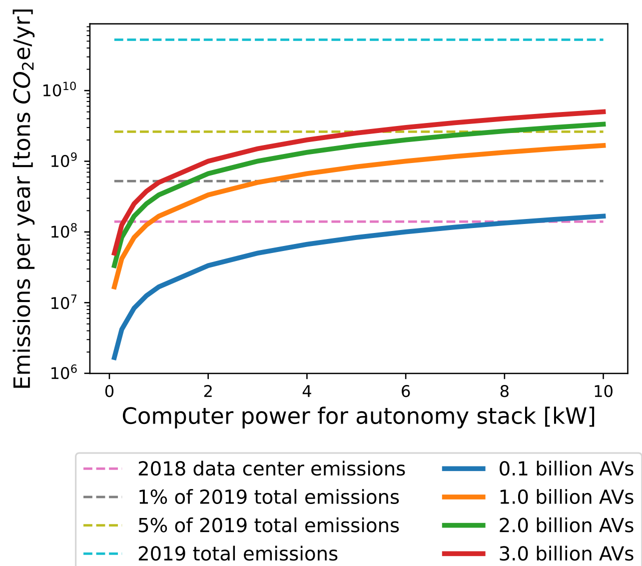


Fig. 1: Emissions from computing onboard AVs driving 1 hr/day. With one billion AVs, an avg. computer power of 0.84 kW yields emissions equal to emissions of all data centers.

previous works have explored the carbon impact of AVs due to changes in driving strategies [8], [9], [10], this work is the first to characterize the carbon emissions from the computers onboard the AV itself.

The contributions of this paper are threefold. First, we introduce an open-source framework to probabilistically model emissions from computing onboard AVs[1]. Second, based on our literature survey to estimate parameters in the model, we find that emissions from computing onboard AVs have the potential to be comparable to that of all data centers today. We recommend targets for computer power and rate of hardware energy efficiency improvement for various scenarios. Finally, we discuss several avenues of future research unique to AVs to better characterize and reduce emissions from computing onboard AVs.

II. MODELING COMPUTING EMISSIONS

Generating the electricity needed to run the computers onboard a global fleet of AVs introduces a source of carbon emissions. The carbon dioxide equivalent (CO_2e) tons

[1] https://github.com/mit-lean/carbon-computing-avs

emitted per year from computing onboard a fleet of N AVs is given by

$$G = \alpha NPQI, \tag{1}$$

where G is the CO_2e tons emitted per year from computing, P is the average computer power for each AV, Q is the average hours per day driven by each AV, I is the average carbon intensity of the electricity used by the AVs or the grams of CO_2e emitted to produce 1 kWh of electricity, and $\alpha = 3.65 \times 10^{-7}$ is a constant that captures unit conversions. Note, Eq. 1 only considers the operational carbon emissions from the computer, and does not include operational carbon emissions from running sensors, embodied carbon emissions from manufacturing computers and sensors, or carbon emissions from prototyping algorithms and training DNNs. If each variable in Eq. 1 is known, we can directly calculate the carbon emissions from computing onboard AVs. Consider a hypothetical case where AVs operate on average for 1 hour each day running an autonomy stack that consumes 2.5 kW and drawing power from a grid with the 2020 global average carbon intensity [F3][2]. We consider the following constant baselines for emissions:

1) 0.14 Gt CO_2e, or all GHG emissions from data centers in 2018 [4], [11],
2) 0.52 Gt CO_2e, or 1% of all GHG emissions (not including land use change) in 2019 [12],

and find it would take 335 million AVs for AV computing emissions to equal 2018 data center emissions and 1.25 billion AVs for AV computing emissions to equal 1% of 2019 emissions.

Next, we sweep over the computing power for the autonomy stack and plot the emissions from computing onboard AVs in Fig. 1. With one billion AVs, less than the number of cars today [A1], the computer power must be less than 0.84 kW to have computing onboard AVs contribute less emissions than data centers. The variables in Eq. 1 are not exactly known, and there is large uncertainty with respect to future trends. We now probabilistically model each variable in Eq. 1 based on current trends (summarized in Table I) and model different scenarios based on future trends (summarized in Table II).

A. Number of AVs (N)

1) Current trends: There were an estimated 1.2 billion vehicles on the road in 2015 [A1]. Meanwhile, in 2019, there were 1,400 AVs approved for testing in the US [A2]; clearly, we are not at a point where AVs dominate the market. We model $N \sim Binomial(1.2 \times 10^9, p_n)$ for a range of adoption rates p_n.

2) Future trends: Car ownership per capita is expected to increase, with projections ranging from 1.8 to 3 billion vehicles in 2050 [G1, G2]. Projections based on aggressive cost reduction and high customer satisfaction can yield AVs capturing up to 95% of the market share by 2050 [G3]. We model two scenarios for $N_{2025-2050}$ assuming a 12 year

[2]In this section, the alphanumeric references point to references in Table I and II.

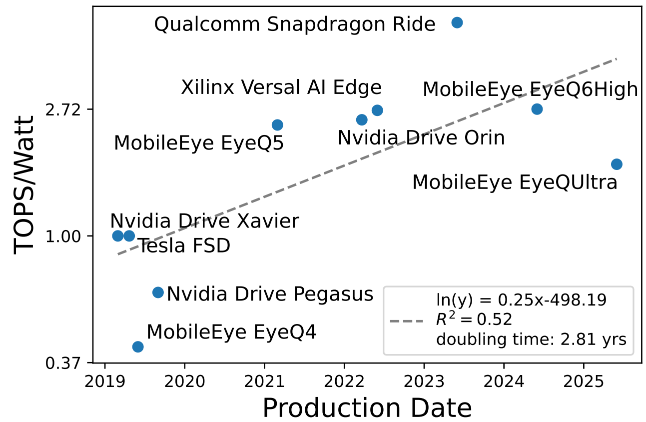


Fig. 2: TOPS (INT8) per Watt vs. production date for current and announced AV hardware [H3, H4, H5, H6, H7, H8, H9, H10, H11]. The trend shows a doubling time of 2.8 years for AV hardware energy efficiency.

lifespan for AVs, a projected 2.2% increase in vehicle sales each year [G1] with vehicle sales returning to 2019 sales by 2025 [G4], and an approximation of a high adoption projection [G3] or medium adoption projection where AVs capture 95% of the market share by 2050 or 2075 respectively.

B. Computer Power (P)

1) Current trends: The average computer power consumed by each AV depends on the workload of the autonomy stack and the hardware energy efficiency of the computer. The autonomy workload consists of perception, localization, planning, and control [C13]. This workload is challenging to model since solving Level 4 or Level 5 autonomy is still an active area of research [C1, C2] and is proprietary for industry. For perception tasks such as object tracking and semantic segmentation, DNNs are the leading approach in computer vision [C13, C14]. For other sensors and tasks such as planning, there is ongoing debate on whether these components will remain DNN-based, non-DNN based, or a hybrid version of both [C13]. Since recreating an entire state-of-the-art autonomy stack is beyond the scope of this paper, we choose to model only the DNN portion workload which likely plays a substantial role in the autonomy stack.

We select a multitask DNN architecture with a shared encoder and separate decoders for each task based on its popularity in academia and industry [C3, C4]. We use EfficientNet-B0 as the encoder [C5] and DeepLabV3 heads as the decoders [C6]. We consider the number of tasks T to be the number of decoders on the autoencoder and the number of cameras C to be the number of times we run the autoencoder. We measure the power $P_{meas}(T)$ and latency $L_{meas}(T)$ of the autoencoder at various values of T for 1344×1344 resolution inputs, a resolution close to that found in AV benchmarks [C15, C16]. Based on the desired throughput F of the full autonomy stack, we model multiple computers needed to achieve the desired throughput. Finally, we scale the hardware energy efficiency by multiplying by

η, the ratio of the tera operations per second (TOPS) per Watt of the measured hardware and the TOPS per Watt of the target hardware, as seen in

$$P_{target} = P_{meas}(T)L_{meas}(T)\eta FC. \quad (2)$$

While TOPS per Watt is known to not be a holistic measure [D1], we use it as an approximation since it is expensive and difficult to get access to state-of-the-art AV hardware and hardware not yet in production. We measure P_{meas} and L_{meas} on an Nvidia RTX 2080 Ti and scale the hardware energy efficiency for the Nvidia DRIVE Orin system to be the target platform, such that $\eta = 0.344$ [D2, D3]. When we substitute Eq. 2 for P in Eq. 1, we obtain

$$G = \alpha N P_{meas}(T)L_{meas}(T)\eta FCQI. \quad (3)$$

We model C, T, and F with Poisson distributions parameterized by λ_C, λ_T, λ_F and consider workloads that capture the number of cameras in commercial AV sensor suites [C7, C8, C9, C10, C11] where $\lambda_C \in \{8, 12, 16\}$, the number of tasks used in some works in academia and industry [C3, C12] where $\lambda_T \in \{10, 50, 100\}$, and the throughput based on human vision [B1] and industry [B2, B3] where $\lambda_F \in \{30, 60\}$.

2) Future trends: There is uncertainty around how autonomy workloads will change over time since Level 4 or 5 autonomy remains unsolved. If there is a paradigm shift due to a breakthrough technology, DNNs may not make up the majority of the autonomy workload. However, given significant investment in DNNs from industry and academia [I6, I7, H10, C14] and difficulty predicting breakthrough technologies, we assume DNNs will likely remain a large component of the autonomy workload.

In general, DNNs have gotten larger over time in domains such as NLP and recommendations [I1]. An exponential scaling of DNN parameters may be required for a linear gain in accuracy [3]; for such a safety-critical system, a slightly more accurate DNN may be preferable even if it is much larger. Moreover, higher resolution cameras processed at higher frame rates allow AVs to see farther and drive faster [I2, I3], and there likely will need to be uncertainty estimation and redundancy built into the hardware and algorithms [I4, I8]. On the other hand, the growth of workload size may be slowed down due to methods such as pruning and network architecture search [13] if they can maintain metrics important for safety (*e.g.,* accuracy, robustness, uncertainty quality) while decreasing latency. We model the workload increasing by multiplying L_{meas} by a factor a and we sweep over values for $a_{2025-2050}$ with doubling times equal to 3, 5, and 10 years.

To model how η will change over time, we model the rate of increase in TOPS per Watt of AV hardware. Based on historical patterns until 2009, Koomey's law states that TOPS per Watt doubles on average every 1.6 years [H1]. However, the slowdown of Dennard scaling and Moore's law has made keeping up with this rate of improvement challenging [H2]. We plot the natural log of the reported TOPS per Watt for current and announced AV hardware platforms and their production dates in Fig. 2. We fit a linear model to find the average doubling rate of TOPS per Watt for AV hardware to be 2.8 years. We model $\eta_{2025-2050}$ at various rates of hardware energy efficiency doubling including at the current pace.

C. Average Time Driven (Q)

1) Current trends: American vehicles were driven on average 0.79 hours per day according the 2017 National Household Travel Survey (NHTS) [E1]. There is uncertainty in how driving behavior may change in response to the widespread adoption of AVs. Projections for changes range from -35% to 40% [E2] due to an increase in driving due to multitasking [E3] and expansion of the transportation user-base to individuals who currently face limited mobility, or a decrease in driving due to increased car-sharing [E2, E4, E5]. We represent the range of -40% to 40% as a 95% confidence interval to obtain $Q \sim \mathcal{N}(0.79, 0.03)$.

2) Future trends: We model a scenario where the average hours driven by AVs increases by 14% due to expansion of driving to under-served populations [J1] and stays constant over time such that $Q_{2025-2050} = 0.90$.

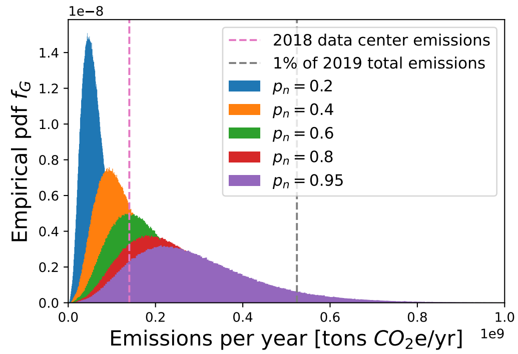
D. Carbon Intensity (I)

1) Current trends: Generating electricity to power the computers onboard the AVs generates carbon emissions, whether generated from gasoline for gasoline-powered vehicles or generated by the mix of energy sources that power the electric grid used to charge electric vehicles. We use the global average carbon intensity of electricity generation to capture the average carbon intensity across all AVs. Due to changes in economic activity due to the COVID-19 pandemic, we use the average global carbon intensity estimate for 2019, select a variance that captures differences in carbon intensity estimates between years [F4], and set $I \sim \mathcal{N}(471, 25)$.

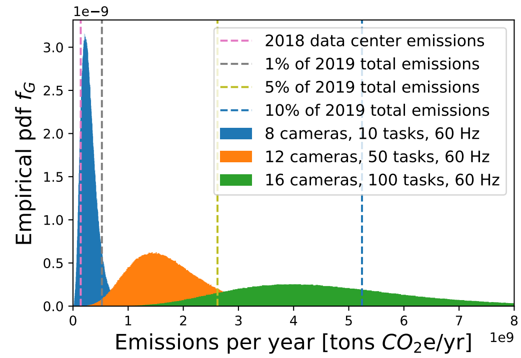
2) Future trends: We model four different scenarios for carbon intensity $I_{2025-2050}$ using different annual decarbonization rates from 471 g CO_2e/kWh: 1) 1.5% (business as usual, 2019 rate), 2) 2.5% (2020), 3) 8.1% (consistent with 2 degrees of warming), 4) 12.9% (consistent with 1.5 degrees of warming) [K1].

III. RESULTS

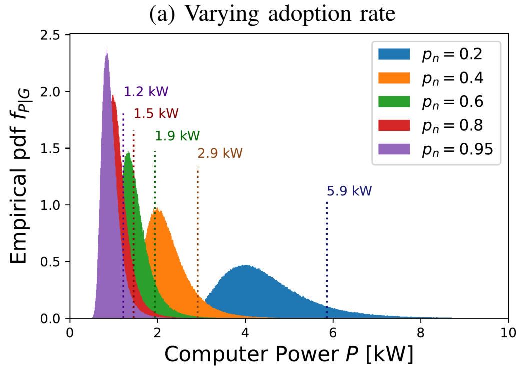
In this section, we present results from applying the carbon emissions framework using the parameters discussed in Section II. We first look at modeling emissions based on current trends and varying the adoption rate of AVs. We use a Monte Carlo simulation with one million samples to estimate the distribution of emissions G, seen in Fig. 3a for the workload with mean 8 cameras and 10 tasks at 60 Hz. The expected value of the distribution is approximately twice 2018 data center emissions when adoption rate p_n is 0.95; in 90% of the scenarios, emissions were greater than 88% of 2018 data center emissions. In Fig. 3b, we set $p_n = 0.95$ and vary the average number of cameras and tasks; the larger workloads lead to surpassing 2018 data center emissions and



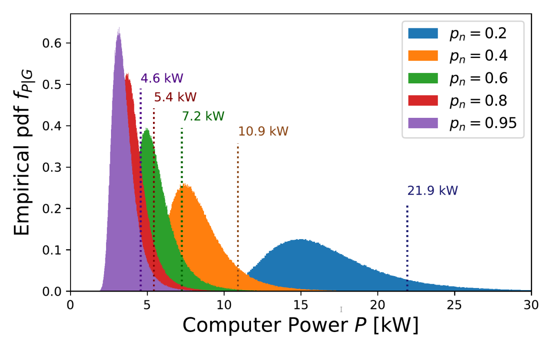
(a) Varying adoption rate



(b) Varying workloads



(c) Conditioned on G equal to data center emissions



(d) Conditioned on G equal to 1% of total emissions

Fig. 3: The top row shows probability distributions for emissions G, illustrating high probability of significant emissions as adoption rate increases and workload size increases if there is not a corresponding increase in hardware energy efficiency; the bottom row shows the probability density of computer power P given emissions reach baselines. The dotted lines show p^*, where 90% of scenarios have power less than p^*.

1% of 2019 emissions with high probability. We also estimate the probability distribution of the computer power P given emissions G is equal to the baselines, as seen in Fig. 3c and Fig. 3d. The computer power must be less than 1.2 kW in order for the emissions to be less than 2018 data center emissions for 90% of the simulated scenarios at $p_n = 0.95$.

Next, we compute the emissions over time in various scenarios with future trends. In Fig. 4, assuming a business-as-usual decarbonization rate and the workload doubling every three years, we sweep over various values for the half-life of η, or equivalently, the doubling time of the hardware energy efficiency, for the high adoption scenario (Fig. 4a) and medium adoption scenario (Fig. 4b). Maintaining the current average rate of hardware energy efficiency increase of 2.8 years in both scenarios leads to large emissions by 2050. In order to keep the emissions from computing onboard AVs in 2050 under 2018 data center emissions or under 1% of 2019 total emissions in the high adoption scenario, hardware energy efficiency must double faster than 1.1 years or 1.4 years respectively.

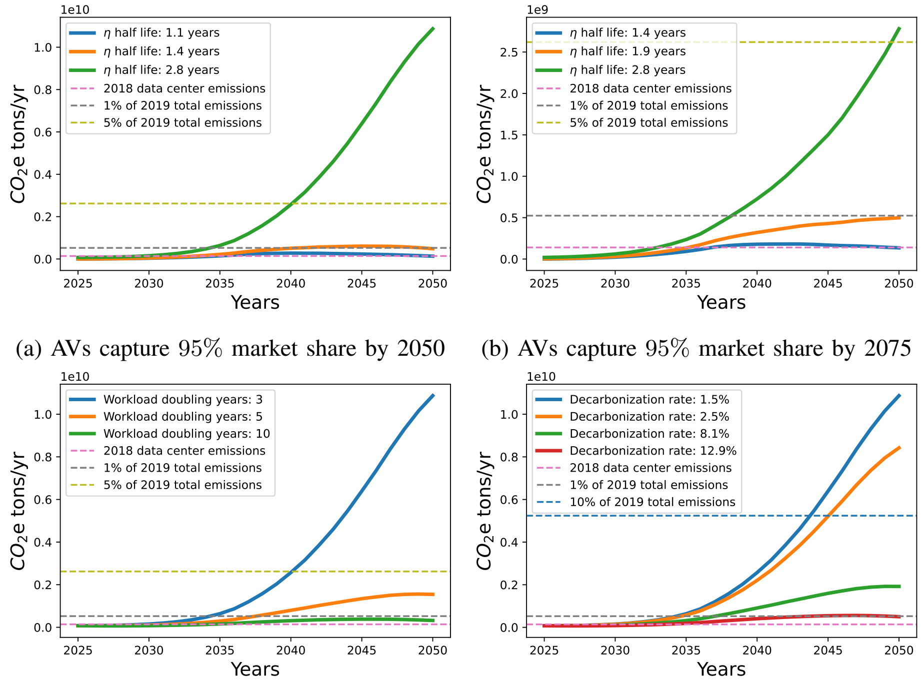
Next, we assume hardware energy efficiency doubles every 2.8 years and the high adoption scenario. In Fig. 4c, assuming a business-as-usual decarbonization rate, we sweep over the doubling time of the workload. We see a slower rate of increase of the workload yields lower emissions. Finally, in Fig. 4d, assuming the workload doubles every 3 years, we sweep over different decarbonization rates. An aggressive decarbonization rate lowers emissions, but even that scenario cannot keep AV computing emissions below that of 2018 data center emissions. A business-as-usual decarbonization rate yields high emissions in this scenario.

IV. Future Work: Reducing the Footprint

Business-as-usual trends alone are not enough to contain the operational carbon emissions from computing onboard AVs in various scenarios presented in Section III. We highlight several future research directions unique to AVs to help better characterize and potentially decrease the carbon footprint from computing onboard AVs.

1) Characterize emissions from sensing: Unlike data center servers, AVs also must sense their environment, and the power consumption of the sensors can be non-negligible for sensors such as LiDAR. Characterizing emissions for current and future trends in sensor suites will help capture the complete operational carbon footprint of AVs.

2) Characterize embodied vs. operational carbon: With the current average lifespan of a car ranging from 10 to 20 years [G1], AVs will likely have much longer lifespans than that of data center servers and mobile devices. An analysis of the embodied carbon emissions from manufacturing the computers and sensors onboard AVs would not only help cap-



(a) AVs capture 95% market share by 2050 (b) AVs capture 95% market share by 2075

(c) Constant decarbonization rate of 1.5% (d) Constant workload doubling years at 3 years

Fig. 4: Scenarios for emissions from computing onboard AVs from 2025-2050. Top row: scenarios with different hardware energy efficiency doubling times, constant workload doubling every 3 years, decarbonization rate of 1.5% for high adoption (left) and medium adoption (right) scenarios; Bottom row: scenarios with constant η half life of 2.8 years and changing workload doubling rate (left) or changing decarbonization rate (right).

ture the total carbon footprint, but also enable comparisons between operational carbon and embodied carbon over the lifespan of an AV. For example, it was found that operational carbon dominates the carbon footprint of mobile devices at device lifespans over 4 years [14]; characterization of AV embodied carbon will shed light on whether operational carbon likewise dominates AVs' total carbon footprint over its longer lifespan. In addition, understanding embodied carbon can suggest the impact of usage patterns such as car-sharing. For example, one billion privately-owned AVs driving one hour per day results in the same amount of operational carbon emissions as 100 million shared AVs driving 10 hours per day. However, the second case would result in lower embodied carbon emissions since fewer components need to be manufactured.

3) Explore trade-off between hardware specialization and generalization: Unlike data center servers, the computers onboard AVs handle constant workloads that are known ahead of time, presenting an opportunity for hardware specialization. The design of accelerators specific for autonomy tasks can deliver large reductions in energy consumption and help maintain a high rate of increase in hardware energy efficiency despite the slowdown of Dennard scaling and Moore's law [15] for both DNN and non-DNN workloads [13], [16], [17]. However, since AVs will have longer lifespans [G1], hardware will still need to maintain some ability to generalize to future workloads.

4) Explore algorithmic efficiency improvements without sacrificing safety: AV workloads are safety-critical and cannot tolerate a decrease in performance in metrics relevant to human safety (*e.g.,* accuracy, latency, robustness, uncertainty quality). While waiting to run a workload until renewable energy is available in a data center or running a smaller DNN with lower accuracy on a mobile device may be viable strategies for those domains, they do not transfer directly to AVs due to safety concerns. Research into algorithmic changes such as compact DNN architectures [13] and efficient non-DNN algorithms that modify the algorithm to reduce the computing energy needed [18], [19] are worth exploring to understand the design space for algorithmic efficiency improvements without sacrificing safety. Moreover, Jevon's paradox may manifest when it comes to safety [20]; for example, pruning a DNN so it has half the original latency may result in AV autonomy stacks running it twice as often to increase safety.

5) Encourage an industry standard to release computer power for autonomy stack on AV hardware: Much of the difficulty in assessing the carbon impact of computing onboard AVs is due to a lack of visibility of the workloads and hardware efficiency of current AV companies' autonomy stacks. Ideally, industry would release a set of holistic metrics in order to conduct fair comparisons between different designs [D1]. However, due to concerns about intellectual property, we encourage industry to at least release computer power since that enables the community to assess the carbon impact of the autonomy stack while keeping the system details proprietary.

V. CONCLUSION

In this paper, we highlight the potential for significant emissions from computing onboard AVs, comparable to all data centers today. Our framework to estimate emissions is adaptable as the community gains more information or for industry to use based on internal numbers. We hope this work encourages research in several exciting directions unique to AVs that can help better characterize and hopefully, reduce the carbon footprint of computing onboard AVs.

VI. ACKNOWLEDGMENTS

The authors thank Jamie Koerner for insightful discussions on modeling and data visualizations.

Variable	Model	References
Number of AVs (N)	$N \sim Binomial(1.2 \times 10^9, p_n)$	A1: S. Bouton *et al.*, "Urban mobility at a tipping point," McKinsey Global Institute, Tech. Rep., 2015. A2: E. L. Chao, "Uber elevate symposium remarks," 2019. [Online]. Available: https://www. transportation.gov/briefing-room/ uber-elevate-symposium
Throughput F	$F \sim Poisson(\lambda_F)$	B1: M. F. Deering, "The limits of human vision," in *2nd International Immersive Projection Technology Workshop*, vol. 2, 1998, p. 1 B2: TeslaOwnersOnline, "Tesla full self driving explained by Andrej Karpathy," August 2021. [Online]. Available: https://www.youtube.com/watch?v=3SypMvnQT_s&t=403s B3: ARM, "ARM mali-c78ae product brief," 2021. [Online]. Available: https://www.arm.com/ products/silicon-ip-multimedia/ image-signal-processor/mali-c78ae
Architecture	C multitask DNNs with EfficientNet-B0 shared encoder and T DeepLabV3 decoders, $C \sim Poisson(\lambda_C)$, $T \sim Poisson(\lambda_T)$	C1: M. A. Khan *et al.*, "Level-5 autonomous driving—are we there yet? a review of research literature," *ACM Computing Surveys (CSUR)*, vol. 55, no. 2, pp. 1–38, 2022. C2: M. Anderson, "The road ahead for self-driving cars: The av industry has had to reset expectations, as it shifts its focus to level 4 autonomy-[news]," *IEEE Spectrum*, vol. 57, no. 5, pp. 8–9, 2020 C3: S. Vandenhende et al., "Multi-task learning for dense prediction tasks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2021. C4: See B2. C5: M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International con- ference on machine learning*. PMLR, 2019, pp. 6105– 6114. C6: L.-C. Chen *et al.*, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017. C7: Designing the Waymo driver," March 2020. [Online]. Available: https://www.youtube.com/watch?v= o8rCOKSDMcg C8: Autopilot." [Online]. Available: https://www. tesla.com/autopilot C9: Presenting the mobileye drive™ self-driving system," April 2021. [Online]. Available: https://www.mobileye.com/blog/ mobileye-drive-self-driving-system/ C10: Introducing nvidia drive hyperion 9: Next- generation plat-form for software-defined autonomous vehicle fleets," March 2022. [Online]. Available: https://blogs.nvidia.com/blog/2022/03/ 22/drive-hyperion-9-atlan/ C11: "Zoox: 1-hour fully autonomous drive in San Francisco with commentary," April 2020. [Online]. Available: https://www.youtube.com/watch?v= 6r7vDhPXmiM C12: Matroid, "Andrej Karpathy - AI for full-self driving at tesla," April 2020. [Online]. Available: https://www. youtube.com/watch?v=hx7BXih7zx8 C13: S. Liu et al., "Autonomous vehicles powered by FPGAs," in Robotic Computing on FPGAs. Springer, 2021, pp. 133–148. C14: H. Chen and Z. Deng, "Bibliometric analysis of the application of convolutional neural network in computer vision," IEEE Access, vol. 8, pp. 155 417–155 428, 2020. C15: P. Sun et al., "Scalability in perception for autonomous driving: Waymo open dataset," in Proceedings of the IEEE/CVF con-ference on computer vision and pattern recognition, 2020, pp. 2446–2454. C16: M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.
Ratio of TOPS/Watt η	$\eta = 0.344$	D1: V. Sze *et al.*, "How to evaluate deep neural network processors: Tops/w (alone) considered harmful," *IEEE Solid-State Circuits Magazine*, vol. 12, no. 3, pp. 28–41, 2020. D2: Nvidia Turing GPU Architecture," 2018. D3: "Nvidia drive gets amped: Scalable plat- form moves to nvidia ampere architecture," May 2020. [Online]. Available: https:// blogs.nvidia.com/blog/2020/05/14/ drive-platform-nvidia-ampere-architecture/

Average time driven Q	$Q \sim \mathcal{N}(0.79, 0.03)$	E1: 2017 National Household Travel Survey," 2017, ac- cessed: 2022-04-14. [Online]. Available: https:// nhts.ornl.gov E2: C. D. Harper *et al.*, "Estimating potential increases in travel with autonomous vehicles for the non-driving, el- derly and people with travel-restrictive medical condi- tions," *Transportation research part C: emerging technolo- gies*, vol. 72, pp. 1–9, 2016 E3: N. Thomopoulos and M. Givoni, "The autonomous car—a blessing or a curse for the future of low carbon mobility? an exploration of likely vs. desirable outcomes," *European Journal of Futures Research*, vol. 3, no. 1, pp. 1–14, 2015 E4: S. Kaplan *et al.*, "The future of autonomous vehicles: Lessons from the literature on technology adoption," *Applied Economic Perspectives and Policy*, vol. 41, no. 4, pp. 583–597, 2019. E5: S. Pan et al., "Shared use of electric autonomous vehicles: Air quality and health impacts of future mobility in the united states," Renewable and Sustainable Energy Reviews, vol. 149, p. 111380, 2021.
Carbon intensity I	$I \sim \mathcal{N}(471, 25)$	F1: N. Scarlat et al., "Quantification of the carbon intensity of electricity produced and used in Europe," Applied Energy, vol. 305, p. 117901, 2022. F2: "Frequently asked questions (FAQs)," 2021. [Online]. Available: https://www.eia.gov/tools/faqs/faq.php?id=74&t=11 F3: IEA, "Tracking power 2021," Nov 2021. [Online]. Available: https://www.iea.org/reports/tracking-power-2021 F4: IEA, "Tracking power 2020," June 2020. [Online]. Available: https://www.iea.org/reports/tracking-power-2020

TABLE I: Modeling current trends for variables to compute carbon emissions of computing onboard AVs

Variable	Model	References
Number of AVs ($N_{2025-2050}$)	$N_{2025-2050}$ reaches 95% market share in 2050 (high adoption) or in 2075 (medium adoption)	G1: H. Hao *et al.*, "Carbon footprint of global passenger cars: Scenarios through 2050," *Energy*, vol. 101, pp. 121–131, 2016. G2: P. Gadonneix *et al.*, "Global transport scenarios 2050," *World Energy Council*, vol. 456, 2011 G3: A. Talebian and S. Mishra, "Predicting the adoption of connected autonomous vehicles: A new approach based on the theory of diffusion of innovations," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 363– 380, 2018 G4: "Global car sales by key markets, 2005-2020," May 2020. [Online]. Available: https://www. iea.org/data-and-statistics/charts/global-car-sales-by-key-markets-2005-2020
Ratio of TOPS/Watt $\eta_{2025-2050}$	$\eta_{2025-2050}$ doubles every 2.8 years	H1: J. Koomey *et al.*, "Implications of historical trends in the electrical efficiency of computing," *IEEE Annals of the History of Computing*, vol. 33, no. 3, pp. 46–54, 2010. H2: J. Hennessy and D. Patterson, "A new golden age for computer architecture: Domain-specific hardware/software co-design, enhanced," in *ACM/IEEE 45th Annual Interna- tional Symposium on Computer Architecture (ISCA)*, 2018. H3: L. Liu *et al.*, "Computing systems for autonomous driving: State of the art and challenges," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6469–6486, 2020. H4: "CES 2022 under the hood an hour with Amnon," Jan 2022. [Online]. Available: https://www.youtube. com/watch?v=1mXy0oi8d60 H5: Watch the first mobileye investor summit," Nov 2019. [Online]. Available: https://www.youtube.com/ watch?v=9JWvzuOlAKs H6: "Qualcomm announces expansion of scalable snapdragon ride platform portfolio," Jan 2022. [Online]. Available: qualcomm.com/news/releases/2021/01/26/ qualcomm-announces-expansion-scalable-snapdragon-ride-platform-portfolio H7: E. Talpes *et al.*, "Compute solution for tesla's full self- driving computer," IEEE Micro, vol. 40, no. 2, pp. 25–35, 2020. H8: K. Freund, "Xilinx readies versal AI edge for 2022 avail- ability," Jun 2021. [Online]. Available: https://www. forbes.com/sites/karlfreund/2021/06/09/ xilinx-readies-versal-ai-edge-for-2022-availability/ ?sh=348e0b8066c5 H9: "Tesla autonomy day," April 2019. [Online]. Available: https://www.youtube.com/watch?v= Ucp0TTmvqOE H10: "Nvidia drive gets amped: Scalable plat- form moves to nvidia ampere architecture," May 2020. [Online]. Available: https:// blogs.nvidia.com/blog/2020/05/14/ drive-platform-nvidia-ampere-architecture/ H11: "Nvidia Drive PX." [Online]. Available: https: //www.nvidia.com/content/nvidiaGDC/sg/ en_SG/self-driving-cars/drive-px/

Workload factor $a_{2025-2050}$	$a_{2025-2050}$ doubles every 3,5, or 10 years	I1: C.-J. Wu *et al.*, "Sustainable AI: Environmental impli- cations, challenges and opportunities," arXiv preprint arXiv:2111.00364, 2021. I2: A. Kelly, "Adaptive perception for autonomous vehicles," Carnegie Mellon University, Tech. Rep., 1994. I3: B. Schoettle, "Sensor fusion: A comparison of sensing capa- bilities of human drivers and highly automated ve- hicles," *University of Michiga*n, 2017. I4: P. Koopman and M. Wagner, "Autonomous vehicle safety: An interdisciplinary challenge," *IEEE Intelligent Trans- portation Systems Magazine*, vol. 9, no. 1, pp. 90–96, 2017. I5: J. Huang *et al.*, "Speed/accuracy trade-offs for modern convo- lutional object detectors," in *Proceedings of the IEEE confer- ence on computer vision and pattern recogni- tion*, 2017, pp. 7310–7311, I6: "How to accelerate the journey to au- tonomous vehicles," 2020. [Online]. Available: https://armkeil.blob.core.windows.net/developer/ Files/pdf/research/arm-automotive-deep-dive-2020.pdf I7: Han et al., "Convolutional neural network with int4 optimization on Xilinx device," Xilinx, Tech. Rep., 2020. I8: B. Lakshminarayanan et al., "Simple and scalable predictive uncer- tainty estimation using deep ensembles," Advances in neural informa- tion processing systems, vol. 30, 2017.
Average time driven $Q_{2025-2050}$	$Q_{2025-2050} = 0.9$	J1: See E2.
Carbon intensity $I_{2025-2050}$	$I_{2025-2050}$ decreases 1.5%, 2) 2.5%, 3) 8.1%, 4) 12.9%	K1: I. P. Milborrow *et al.*, "Net zero economy index 2021 code red to go green," PwC, Tech. Rep., 2021.

TABLE II: Modeling future trends for variables to compute carbon emissions of computing onboard AVs

REFERENCES

[1] N. Jones, "How to stop data centres from gobbling up the world's electricity," *Nature*, vol. 561, no. 7722, pp. 163–167, 2018.

[2] C.-J. Wu *et al.*, "Sustainable ai: Environmental implications, challenges and opportunities," *arXiv preprint arXiv:2111.00364*, 2021.

[3] R. Schwartz *et al.*, "Green ai," *Communications of the ACM*, vol. 63, no. 12, pp. 54–63, 2020.

[4] E. Masanet *et al.*, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020.

[5] A. S. Andrae and T. Edler, "On global electricity usage of communication technology: trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.

[6] "Automated vehicles for safety," 2022. [Online]. Available: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety#resources

[7] K. Johnson, "Why not use self-driving cars as supercomputers?" *WIRED*, Jul 2021. [Online]. Available: https://www.wired.com/story/use-self-driving-cars-supercomputers/

[8] V. Jayawardana and C. Wu, "Learning eco-driving strategies at signalized intersections," *arXiv preprint arXiv:2204.12561*, 2022.

[9] M. A. Figliozzi, "Carbon emissions reductions in last mile and grocery deliveries utilizing air and ground autonomous vehicles," *Transportation Research Part D: Transport and Environment*, vol. 85, p. 102443, 2020.

[10] Ó. Silva *et al.*, "Environmental impacts of autonomous vehicles: A review of the scientific literature," *Science of The Total Environment*, p. 154615, 2022.

[11] "Greenhouse gas emissions from energy: Overview," 2021. [Online]. Available: https://www.iea.org/reports/greenhouse-gas-emissions-from-energy-overview

[12] U. UNEP, "Emissions gap report 2020," *UN Environment Programme*, 2020.

[13] V. Sze *et al.*, "Efficient processing of deep neural networks," *Synthesis Lectures on Computer Architecture*, vol. 15, no. 2, pp. 1–341, 2020.

[14] U. Gupta *et al.*, "Act: designing sustainable computer systems with an architectural carbon modeling tool," in *Proceedings of the 49th Annual International Symposium on Computer Architecture*, 2022, pp. 784–799.

[15] J. Hennessy and D. Patterson, "A new golden age for computer architecture: Domain-specific hardware/software co-design, enhanced," in *ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*, 2018.

[16] S. Murray *et al.*, "Robot motion planning on a chip." in *Robotics: Science and Systems*, vol. 6, 2016.

[17] K. Gupta *et al.*, "Efficient computation of map-scale continuous mutual information on chip in real time," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 6464–6470.

[18] S. Sudhakar *et al.*, "Balancing actuation and computing energy in motion planning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4259–4265.

[19] P. Z. X. Li *et al.*, "Memory-efficient gaussian fitting for depth images in real time," *2022 IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[20] B. Alcott, "Jevons' paradox," *Ecological economics*, vol. 54, no. 1, pp. 9–21, 2005.