3D-Carbon: An Analytical Carbon Modeling Tool for 3D and 2.5D Integrated Circuits

Yujie Zhao[1], Yang (Katie) Zhao[2*], Cheng Wan[1], Yingyan (Celine) Lin[1]
[1]*Georgia Institute of Technology*, [2]*University of Minnesota, Twin Cities*
{eiclab.gatech, chwan, celine.lin}@gatech.edu, yangzhao@umn.edu

Abstract

Environmental sustainability is crucial for Integrated Circuits (ICs) across their lifecycle, particularly in manufacturing and use. Meanwhile, ICs using 3D/2.5D integration technologies have emerged as promising solutions to meet the growing demands for computational power. However, there is a distinct lack of carbon modeling tools for 3D/2.5D ICs. Addressing this, we propose 3D-Carbon, an analytical carbon modeling tool designed to quantify the carbon emissions of 3D/2.5D ICs throughout their life cycle. 3D-Carbon factors in both potential savings and overheads from advanced integration technologies, considering practical deployment constraints like bandwidth. We validate 3D-Carbon's accuracy against established baselines and illustrate its utility through case studies in autonomous vehicles. We believe that 3D-Carbon lays the initial foundation for future innovations in developing environmentally sustainable 3D/2.5D ICs. Our open-source code is available at https://github.com/UMN-ZhaoLab/3D-Carbon.

ACM Reference Format:
Yujie Zhao[1], Yang (Katie) Zhao[2*], Cheng Wan[1], Yingyan (Celine) Lin[1]. 2024. 3D-Carbon: An Analytical Carbon Modeling Tool for 3D and 2.5D Integrated Circuits. In *61st ACM/IEEE Design Automation Conference (DAC '24), June 23–27, 2024, San Francisco, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3649329.3658482

1 Introduction

For decades, Moore's Law has driven Integrated Circuits (ICs), enhancing their computational power, reducing size and cost, and improving energy efficiency. These advancements have been crucial for developing new technologies like artificial intelligence (AI). However, the carbon emissions from ICs, covering their entire lifecycle from manufacturing to disposal (see Fig. 1), pose significant environmental sustainability challenges. Recent reports indicate that the carbon of Information and Communication Technology represents 2.1%~3.9% of global greenhouse gas emissions [13].

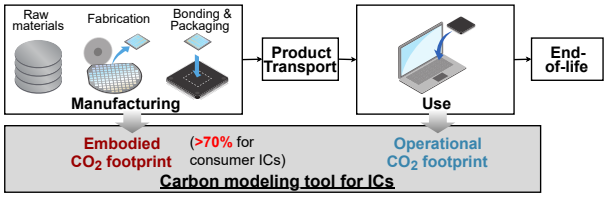

Figure 1. A carbon modeling tool tracks embodied and operational carbon emissions throughout ICs' lifecycle [17].

To ensure environmental sustainability in ICs, addressing their carbon emissions throughout their lifecycle, particularly in *manufacturing* and *use* phases, is crucial [17]. Currently, the *embodied carbon* from manufacturing activities often surpasses the *operational carbon* from energy consumption during the use of today's ICs [17]. This shift is due to extensive operational energy efficiency improvements over years. Embodied carbon can represent over 70% of the total carbon footprint for consumer ICs [11, 17].

In parallel, as the pace of Moore's Law in 2D monolithic ICs has slowed in recent years, significant advancements have been made in developing 3D/2.5D ICs to meet rising computational demands. These advanced ICs offer notable improvements over 2D ICs in power efficiency, performance, and area in various commercial products [12, 19, 23, 30].

However, currently, there's no comprehensive tool to evaluate the carbon footprint of 3D/2.5D ICs, particularly their embodied carbon, which is crucial due to their growing complexity: While additional manufacturing steps increase carbon emissions per wafer, factors like improved yield, area efficiency, use of heterogeneous technologies, and fewer metal layers could reduce the overall carbon footprint.

Current tools for estimating the embodied carbon of 3D/2.5D ICs have limitations. Industry-based databases rely on data about materials [9] or manufacturing processes [11, 17], but their practicality and accuracy are limited by the availability of up-to-date carbon emission data. First-order approaches, such as the one in [10], estimate the embodied footprint per chip based on die size. Another method, ACT+ [11], estimates 2.5D IC carbon footprint from 2D ICs based on cost comparison and simplistically treats 3D stacked dies as 2D. These

* Corresponding author.

DAC '24, June 23–27, 2024, San Francisco, CA, USA

ACM ISBN 979-8-4007-0601-1/24/06
https://doi.org/10.1145/3649329.3658482

Table 1. 3D/2.5D integration technologies summary.

2.5 or 3D	Integration Technology	F2F or F2B	Stacking	# of max 3D dies/tiers	Representative Manufacturers	Representation Products/ Prototypes
3D	Hybrid Bonding	F2F	D2W/W2W	2	TSMC's SoIC-X [29] Intel's Fevores Direct [22]	AMD 3D V-cache [30]
		F2B	D2W/W2W	≥2		AMD Ryzen 7-5800X3D [30]
	Micro-bumping	F2F	D2W/W2W	2	TSMC's SoIC-P [29] Intel's Fevores [22]	Intel Lakefield Core i5-L16G7 [15]
		F2B	D2W/W2W	≥2		HBM [23]
	M3D	F2B	N/A	2		RISC-V Core [19]
2.5D	MCM	N/A	N/A	N/A	AMD's Infinity Fabric [23]	AMD EPYC 7000 series [23]
	InFO	N/A	N/A	N/A	TSMC's InFO-2.5D CoWoS-L/R [29]	AMD Ravi 31 [5]
	EMIB	N/A	N/A	N/A	Intel's EMIB [21]	Stratix 10 [3]
	Silicon Interposer	N/A	D2W	N/A	TSMC's CoWoS-S [29]	Nvidia GPU P100 [26]

techniques provide general insights but lack the detailed breakdowns needed for effective carbon-conscious design.

To comprehensively quantify the overall carbon footprint of 3D/2.5D ICs, we propose 3D-Carbon, an analytical carbon modeling tool, in this paper. Our contributions are as follows:

- We develop 3D-Carbon, an analytical carbon modeling tool for various 3D/2.5D ICs. We believe that 3D-Carbon lays the initial foundation for future innovations in developing environmentally sustainable 3D/2.5D ICs.
- Using 3D-Carbon, we can predict the embodied carbon emissions, including the overhead associated with advanced integration technologies, and estimate the operational carbon emissions during use through surveyed parameters or third-party operational energy estimation plug-ins.
- We evaluate 3D-Carbon and demonstrate its valuable insights and broad applicability through case studies in autonomous vehicles to guide sustainable decision-making in choosing or replacing ICs.

2 Background

2.1 Commercial 3D/2.5D Integration Technologies

We examine three commercial 3D integration technologies and four 2.5D integration technologies (see Tab. 1 and Fig. 2).

2.1.1 3D Integration. Micro-bumping 3D. This method stacks multiple dies vertically using micron-level solder balls for 3D connections, with a larger pitch than other 3D technologies. **Hybrid Bonding 3D.** This technique stacks two 2D dies using bond pads through the metal layers. **Monolithic 3D (M3D).** M3D employs sequential manufacturing to create fine-pitched MIVs (typically <0.6μm) for inter-tier connections [19]. This paper emphasizes block-level M3D partitioning, where functional blocks like memory and logic macros are separated into different tiers, enabling the use of existing 2D EDA processes [4].

2.1.2 2.5D Integration. Multi-chip module (MCM). This method places multiple pre-designed dies on an organic package substrate. **Integrated fan-out (InFO).** InFO, evolving from fan-out wafer-level packaging, uses a redistribution layer (RDL) as the substrate, offering smaller line space than MCM and including chip-first and chip-last approaches. **Embedded Multi-die Interconnect Bridge (EMIB).** EMIB integrates multiple dies within a single package using a silicon interconnect bridge. **Silicon interposer.** Utilizing a silicon substrate (passive or active [30]), silicon interposers provide the finest line space but may increase carbon costs.

2.2 IC Carbon and Carbon Optimization Metrics

2.2.1 IC Total Life Cycle Carbon. Following the industry-endorsed frameworks [11, 17], we estimate an IC's total life cycle carbon footprint as follows, including both operational ($C_{operational}$) and embodied (C_{emb}) emissions:

$$C_{total} = C_{operational} + C_{emb} \quad (1)$$

2.2.2 Indifference Point and Breakeven Time Analysis. The embodied carbon (C_{emb}) of 3D/2.5D ICs includes both potential savings (e.g., fewer back-end-of-line (BEOL) layers) and overheads (e.g., advanced integration manufacturing processes). Meanwhile, their operational carbon ($C_{operational}$) benefits from shorter interconnect lengths but faces higher power needs for interfaces. Thus, 3D/2.5D ICs don't always offer clear sustainability benefits in both C_{emb} and $C_{operational}$.

To aid in sustainable decision-making for choosing and replacing 3D/2.5D ICs over 2D ICs for fixed workload applications, we use the indifference point metric (T_c) and the breakeven time metric (T_r) as defined in [20], which are based on the embodied carbon costs ($C_{emb}^{2D}/C_{emb}^{3D/2.5D}$), the use-phase carbon intensity (CI_{ues}), and the power consumptions ($P_{app}^{2D}/P_{app}^{3D/2.5D}$, see Eq. (17)):

$$T_c = \frac{C_{emb}^{3D/2.5D} - C_{emb}^{2D}}{CI_{use}(P_{app}^{2D} - P_{app}^{3D/2.5D})}, T_r = \frac{C_{emb}^{3D/2.5D}}{CI_{use}(P_{app}^{2D} - P_{app}^{3D/2.5D})} \quad (2)$$

In scenarios of "choosing", T_c indicates when the saved C_{emb} of 3D/2.5D ICs is offset by increased $C_{operational}$. In scenarios of "replacing", T_r shows the breakeven time when the increased C_{emb} of 3D/2.5D ICs is compensated by reduced $C_{operational}$ with the assumption that the embodied carbon cost has already been invested for the 2D ICs. We compare T_c/T_r to the IC's remaining lifetime (T_{life}) to guide sustainable decision-making in choosing or replacing ICs.

3 3D-Carbon Modeling Tool

3.1 3D-Carbon: High-Level Overview

Fig. 3 shows an overview of 3D-Carbon, concentrating on both the embodied and operational carbon emissions of 3D/2.5D ICs. The relevant parameters have been obtained from industry environmental reports, as listed in Tab. 2.

For estimating *embodied carbon emission*, 3D-Carbon uses a hardware design description consisting of 3D/2.5D configurations and IC area details, a technology information description of the technology node and the maximum number of BEOL layers, and the manufacturing location.

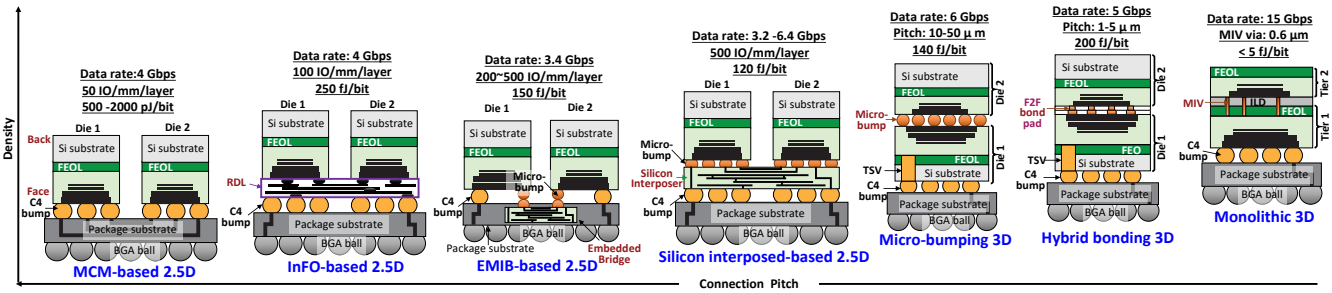

Figure 2. The vertical stack diagram of 3D and 2.5D integration options studied in this paper.

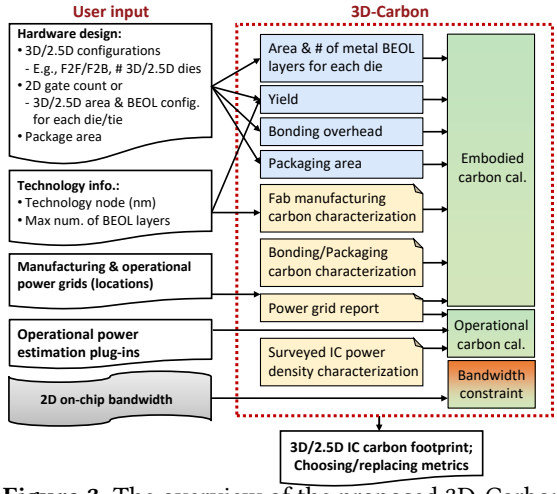

Figure 3. The overview of the proposed 3D-Carbon.

For calculating *operational carbon emission*, 3D-Carbon integrates with operational power consumption plugins like [2, 16, 18] and utilizes the use-location's carbon intensity.

Additionally, 3D/2.5D ICs employ off-die I/O interfaces for data movements, which utilize on-chip wire resources in 2D ICs. Thus, 3D-Carbon introduces an I/O bandwidth constraint to assess the viability of 3D/2.5D ICs compared to their 2D counterparts in terms of data movements.

3.2 3D-Carbon: Embodied Carbon Emission

Unlike 2D ICs, 3D/2.5D ICs integrate N 2D dies with additional bonding emissions. For interposer-based ICs, an extra RDL/silicon interposer is manufactured. We calculate the overall embodied emission ($C_{emb}^{3D/2.5D}$) by summing the emissions from die manufacturing ($C_{die}^{3D/2.5D}$), bonding ($C_{bonding}^{3D/2.5D}$), packaging ($C_{packaging}^{3D/2.5D}$), and the 2.5D interposer ($C_{int}^{2.5D}$):

$$C_{emb}^{3D/2.5D} = C_{die}^{3D/2.5D} + C_{bonding}^{3D/2.5D} + C_{packaging}^{3D/2.5D} + C_{int}^{2.5D} \quad (3)$$

3.2.1 $C_{die}^{3D/2.5D}$ Model. To calculate the embodied carbon from die manufacturing in 3D/2.5D ICs, we begin from the carbon emission per wafer ($C_{wafer_{die_i}}$) for each die (denoted as die_i), the corresponding die-per-wafer count (DPW_{die_i}), and the yield (Y_{die_i}) (as detailed in Sec. 3.2.5):

$$C_{die}^{3D/2.5D} = \sum_{i=1}^{N} \frac{C_{wafer_{die_i}}}{DPW_{die_i}} \cdot \frac{1}{Y_{die_i}} \quad (4)$$

Table 2. 3D/2.5D IC embodied carbon model parameters.

Parameter	Range	Source
Hardware design related parameters		
N_g^{2D}	User input (optional)	Input
N	≥ 2	Input
A_{die_i}	User input (optional)	Input/Eq. (7)
N_{BEOL_i}	User input (optional)	Input/Eq. (10)
Foundary related parameters		
Process	$3 \sim 28\ nm$	Input
$A_{wafer_{die_i}}$	$31{,}415.93 \sim 159{,}043.13\ mm^2$	Input
$Y_{IO}^{micro\_3D/2.5D}$	0~1	[12]
D_{TSV}	$0.3 \sim 25\ \mu m$	[19]
GPA_i, MPA_i	$0.1 \sim 0.5\ kg\ CO_2/cm^2$	[7, 17]
EPA_i	$0.4 \sim 0.1\ kWh/cm^2$	[7]
N_{fan}	$1 \sim 5$	[27]
p	$0.6 \sim 0.8$	[27]
β	$450 \sim 850\ M$	[27]
λ	$3 \sim 28\ nm$	[27]
ω	3.6λ	[27]
Bonding related parameters		
$CPA_{RDL}^{InFO_{2.5D}}$	RDL characterization	[7][24]
$EPA_{D2W/W2W}^{micro/hybrid/C4}$	$0.9 \sim 2.75\ kwh/cm^2$	[1]
$y_{W2W}^{micro/hybrid}$	0~ 1	[1][23]
Substrate related parameters		
$s_{RDL/EMIB/Si\_int}$	≥ 1	[12]
D_{gap}	$0.5 \sim 2\ mm$	[12]
Packaging related parameters		
$s_{package}^{3D/2.5D}$	≥ 1	[24][12]
$CPA_{packaging}$	Package characterization	[24]
Carbon intensity		
CI_{emb}, CI_{use}	$30 \sim 700\ g\ CO_2 kWh$	[17]

We calculate the die-per-wafer count (DPW_{die_i}) by dividing wafer area ($A_{wafer_{die_i}}$) by die area (A_{die_i}) [27], similarly applied to interposer-per-wafer count (DPW_{int}):

$$DPW_{die_i/int} = \frac{\pi \cdot \left(A_{wafer_{die_i}}/2\right)^2}{A_{die_i/int}} - \frac{\pi \cdot A_{wafer_{die_i}}}{\sqrt{2 \cdot A_{die_i/int}}} \quad (5)$$

Similar to work [17], 3D-Carbon formulates the wafer carbon footprint ($C_{wafer_{die_i}}$) for die_i as follows:

$$C_{wafer_{die_i}} = (CI_{emb} \cdot EPA_i + GPA_i + MPA_i) \cdot A_{wafer_{die_i}} \quad (6)$$

where CI_{emb} is the carbon intensity of the fab's electrical grid (location), $EPA/GPA/MPA$ is fab energy/gas emissions/raw materials carbon foot print per unit 2D die area.

Area Estimation: Additional areas for Through-Silicon Vias (TSVs) ($A_{TSV_i}^{3D}$) and interface I/O drivers ($A_{IO_i}^{2.5D/Micro\_3D}$) are

required in 3D/2.5D ICs for die-to-die transmission. The total die area for die_i includes gate area (A_{gate_i}), TSVs, and I/Os:

$$A_{die_i}^{3D/2.5D} = A_{gate_i} + A_{TSV_i}^{3D} + A_{IO_i}^{2.5D/Micro\_3D} \tag{7}$$

Gate area (A_{gate_i}) is calculated from gate count ($N_{g_i}^{2D}$), feature size (λ), and scaling term (β):

$$A_{gate_i} = N_{g_i}^{2D} \cdot \beta \cdot \lambda^2 \tag{8}$$

$A_{TSV_i}^{3D}$ relates to the size of each TSV (D_{TSV}) and the TSV count (X_{TSV_i}). D_{TSV} corresponds to each technode. Meanwhile, X_{TSV_i} varies depending on the die stacking method (i.e., F2F/F2B). For F2B, the TSV count is calculated using Rent's rule as [27]. F2F TSV count equals the IO number.

Considering the large size of micro-bumps and 2.5D connectors, additional I/O driver area ($A_{IO_i}^{micro\_3D/2.5D}$) is required, calculated using a ratio ($\gamma_{IO}^{micro\_3D/2.5D}$) of the gate area [12]:

$$A_{IO_i}^{micro\_3D/2.5D} = \gamma_{IO}^{micro\_3D/2.5D} \cdot A_{gate_i} \tag{9}$$

BEOL Configuration: Reducing metal layers in the BEOL offers a more environmentally-friendly approach. The number of BEOL layers (N_{BEOL_i}) is estimated as follows [27]:

$$N_{BEOL_i} = \frac{N_{fan} \cdot \omega \cdot N_{g_i} \cdot \bar{L}_i}{\eta \cdot A_{die_i}} \tag{10}$$

3.2.2 $C_{bonding}^{3D/2.5D}$ Model. 3D/2.5D ICs involve wafer-to-wafer (W2W) or die-to-wafer (D2W) integration, stacking multiple dies or wafers [1]. The bounding energy per unit area $EPA_{D2W/W2W}^{micro/hybrid/C_4}$ and yield $Y_{bonding_i}^{micro/hybrid/C_4}$ depends on the choice of D2W or W2W, and on the bonding method (C4-bump, micro-bumping, or hybrid bonding).

$$C_{bonding}^{3D/2.5D} = \sum_{i=1}^{N-1} \frac{CI_{emb} \cdot EPA_{D2W/W2W}^{micro/hybrid/C4} \cdot A_{die_i}}{Y_{bonding_i}^{micro/hybrid/C4}} \tag{11}$$

3.2.3 $C_{packaging}^{3D/2.5D}$ Model. In 3D-Carbon, the packaging carbon footprint is estimated using packaging carbon emissions per area ($CPA_{packaging}$):

$$C_{packaging}^{3D/2.5D} = CPA_{packaging} \cdot A_{package}^{3D/2.5D} \tag{12}$$

The package area $A_{package}^{3D/2.5D}$ is calculated using a linear empirical equation from [12], applying a scaling factor $s_{package}^{3D/2.5D}$, which is based on the largest die area in 3D ICs and the total die area for 2.5D ICs.

3.2.4 $C_{int}^{2.5D}$ Model. The carbon footprint for RDL, EMIB, and silicon interposer substrates is modeled similarly to die carbon footprint. The interposer area calculation differs for each integration method:

$$A_{Si\_int} = s_{Si\_int} \cdot \sum_{i=1}^{N} A_{die_i} \tag{13}$$

$$A_{RDL/EMIB} = s_{RDL/EMIB} \cdot D_{gap} \cdot \sum_{i=1}^{N} l_{adjacent_i} \tag{14}$$

Table 3. Stacking Yields

3D Type	$Y_{die_i}^{micro/hybrid}$	$Y_{bonding_i}^{micro/hybrid}$
D2W	$y_{die_i} \cdot (y_{D2W}^{micro/hybrid})^{N-i}$	$(y_{D2W}^{micro/hybrid})^{N-i}$
W2W	$\prod_{j=1}^{N} y_{die_j} \cdot (y_{W2W}^{micro/hybrid})^{N-1}$	$\prod_{j=1}^{N} y_{die_j} \cdot (y_{W2W}^{micro/hybrid})^{N-1}$

2.5D Type	$Y_{die_i}^{2.5D}$	$Y_{substrate}^{2.5D}$	$Y_{bonding_i}^{2.5D}$
Chip-first	$y_{die_i} \cdot y_{substrate}^{2.5D}$	$y_{substrate}^{2.5D}$	1
Chip-last	$y_{die_i} \cdot \prod_{j=1}^{N} y_{bonding_j}^{2.5D}$	$y_{substrate}^{2.5D} \cdot \prod_{j=1}^{N} y_{bonding_j}^{2.5D}$	$\prod_{j=1}^{N} y_{bonding_j}^{2.5D}$

where s_{Si_int} and $s_{RDL/EMIB}$ are a scaling factors, $l_{adjacent_i}$ is the total length of adjacent sides for all dies, and D_{gap} is the gap between two adjacent dies.

3.2.5 Yield Model. We estimate yields for different process technologies using data from [24] for bonding and packaging yields, and a yield distribution model from [12] for die and substrate yields.

$$y_{die_i} = (1 + \frac{A_{die_i}(d,p) \times D_0}{\alpha})^{-\alpha} \tag{15}$$

where D_0 is the defect density and α is a parameter determined by process complexity which both given by [12].

We also consider the impact of each process on the overall yield. Denoting individual process yield as y and overall yield as Y, the yield of a process like W2W bonding ($Y_{bonding_i}$) is affected by its own yield (y_b) and the die yield (y_{die}), as defective dies cannot be separated before bonding. The results for different Y values are listed in Tab. 3.

3.3 3D-Carbon: Operational Carbon Emission

We focus on the fixed-throughput approach, widely adopted in applications like autonomous vehicles (AVs) [28]. Given the varied energy consumption patterns that different power benchmarks can produce, the operational carbon footprint ($C_{operational}$) of ICs is determined by the diverse run-time (T_{app_k}), the use carbon intensity (CI_{use}), and power consumption (P_{app_k}) of each application.

$$C_{opreational} = \sum_{k} CI_{use} \cdot P_{app_k} \cdot T_{app_k} \tag{16}$$

The power (P_{app_k}) is calculated as:

$$\begin{aligned} P_{app_k} &= \sum_{i=1}^{N} (\frac{Th_{app_k}}{Eff_{die_i}} + P_{IO_i}) \\ P_{IO_i} &= P_{per\_pitch_i} \cdot N_{pitch_i} \\ N_{pitch_i} &= L_{edge_i} \cdot D_{pitch_i} \cdot N_{BEOL_i} \end{aligned} \tag{17}$$

where Th is the fixed throughput, Eff_{die_i} is the energy efficiency of die_i, and P_{IO_i} is each die's interface I/O driver power. In the absence of specific input for Eff_{die_i}, we utilize surveyed parameters (e.g., as in [19]) to estimate Eff_{die_i}. For 2.5D ICs and Micro-bumping 3D ICs, the I/O power (P_{IO_i}) should be included. We presume P_{IO_i}, using the energy cost per pitch ($P_{per\_pitch_i}$), the length of die_i's edge (L_{edge_i}), the surveyed pitch density (D_{pitch_i}), and the number of BEOL layers (N_{BEOL_i}).

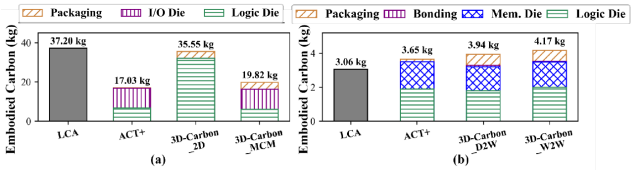

Figure 4. Validation of 3D-Carbon against (a) 2.5D EPYC 7452 [8] and (b) 3D Lakefield [15].

3.4 3D-Carbon: Bandwidth Constraint

A key constraint for 3D/2.5D ICs is ensuring sufficient I/O interface bandwidth for applications. We assume that 3D ICs' I/O bandwidth matches the on-chip bandwidth of their 2D counterparts [6]. The 2.5D ICs' I/O bandwidth is:

$$BW_{die_i} = N_{pitch_i} \cdot BW_{per\_pitch_i} \tag{18}$$

With deep neural networks being the primary workload for AVs [28], we establish the bandwidth constraint that the performance (i.e., throughput) degradation of 2.5D ICs exceeds 20% if the I/O bandwidth is reduced by half than the 2D on-chip bandwidth [6]. Based on this, if the 2.5D ICs' interface causes performance to fall below the throughput requirement, we categorize these instances as "invalid".

4 Validation of 3D-Carbon

We validate 3D-Carbon by comparing its predicted embodied carbon emissions with those obtained from Life Cycle Assessment (LCA) reports [14] and the latest ACT+ tool [17].

4.1 Validation against One 2.5D IC: EPYC 7452

We validate our 3D-Carbon model against one MCM 2.5D IC, EPYC 7452 [8]. Inputs for both 3D-Carbon and ACT+ are based on the EPYC 7452's specifications: 7nm technology for four CPU dies and 14nm for one I/O die.

As shown in Fig. 4(a), the LCA [14], designed for 2D monolithic ICs, reports higher embodied emissions than 3D-Carbon and ACT+. When we adjust 3D-Carbon for a 2D IC, the discrepancy in embodied emissions between LCA and 3D-Carbon is about 4.4%. Unlike ACT+, our model includes manufacturing complexity: it considers BEOL configurations, adjusting carbon footprint for CPU dies with fewer BEOL layers, and estimates packaging carbon emissions based on actual packaging area, resulting in higher packaging carbon emission (3.47 kg) compared to ACT+'s fixed 0.15 kg carbon.

4.2 Validation against One 3D IC: Lakefield

We validate our model using Intel's Lakefield 3D IC [15], which features heterogeneous integration with a 7nm top logic die and a 14nm base memory die, and show the results in Fig. 4(b). Both 3D-Carbon and ACT+ use the Lakefield's technology specifications.

We reference the GaBi LCA database [14] for LCA validation baseline. Since GaBi doesn't cover the 7 nm process, it assume 14nm for both dies, leading to an underestimation

Table 4. NVIDIA GPU DRIVE series specifications [25].

	PX 2	XAVIER	ORIN	THOR
Technology node (nm)	16	12	7	5
Num. of transistors (Billion)	15.3	21	17	77
Energy efficiency (TOPS/W)	0.75	1	2.74	12.5
Announced (year)	2016	2017	2019	2022

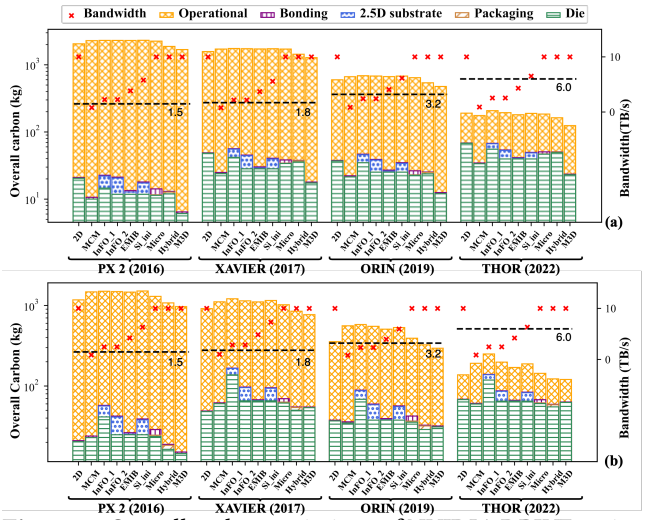

Figure 5. Overall carbon emissions of NVIDIA DRIVE series: 2-die 3D/2.5D ICs with the (a) homogeneous and (b) heterogeneous approaches [6]. Note that InFO_1/InFO_2 represent chip-first/chip-last approaches, respectively.

than 3D-Carbon and ACT+ as the 7nm node is more complex and carbon-intensive. Compared to ACT+, our model accounts for manufacturing complexities and differences between D2W and W2W stacking methods. D2W, involving advanced bonding technology, results in lower yield for the bonding process but allows pre-stacking die availability checks, leading to higher individual die yields (Y_{die}). Specifically, the logic die yield in D2W is 89.3%, the memory die is 88.4%, whereas in W2W, both dies have a yield of 79.7%.

5 Case study: Sustainable Decision-Making for NVIDIA Autonomous Vehicle GPUs

We conduct sustainable decision-making case studies using 3D-Carbon on the NVIDIA GPU DRIVE series for AVs as detailed in Tab. 4. These compare carbon emissions of original 2D designs with hypothetical 3D/2.5D designs. The hypothetical designs involve two die division approaches: homogeneous (splitting the 2D IC into two similar dies) and heterogeneous (isolating the memory and IOs from the main logic die and implementing them separately in an older 28nm node). For 3D ICs, we consider F2F with D2W stacking.

5.1 Overall Carbon Footprint

Fig. 5(a) and (b) show NVIDIA DRIVE series' carbon emissions for homogeneous and heterogeneous 3D/2.5D approaches, respectively. The black line indicates the required I/O bandwidth, and the red "x" marks achieved bandwidth. InFO and

Table 5. Case studies for choosing/replacing the NVIDIA DRIVE ORIN 2D IC with the 3D/2.5D ICs.

3D/2.5D ICs	EMIB	Si_int	Micro	Hybrid	M3D
Embodied carbon save ratio	23.69%	-9.59%	25.88%	35.64%	65.53%
Overall carbon save ratio	6.5%	-9.86%	7.63%	21.71%	41.03%
Choosing metric T_c (years)	<22	∞	<25	>0	>0
Replacing metric T_r (years)	∞	∞	∞	> 75	> 19

silicon-interposer 2.5D ICs increase embodied carbons due to large substrate areas and low substrate yields. Other 3D/2.5D designs constantly reduce/maintain the embodied carbons, particularly in the homogeneous approach (see Fig. 5(a)), while the heterogeneous approach (see Fig. 5(b)) introduces lesser saving due to smaller memory die areas and limited benefits from the older technology. With the exponential growth of energy efficiency over time, the operational carbon emissions decrease, as detailed in Tab. 4. Operational carbon emissions are higher for 2.5D ICs than 2D/3D ICs, due to the performance degradation (i.e., throughput) from 3D-Carbon's bandwidth constraint (see Sec. 3.4). For THOR, none of the four 2.5D ICs meet the necessary bandwidth, rendering them "invalid".

5.2 Sustainable Decision-Making

We chose the five valid 3D/2.5D ICs with the homogeneous approach for NVIDIA DRIVE ORIN. Tab. 5 presents the embodied carbon savings, overall carbon savings, and metrics for choosing and replacing relative to the original 2D IC. These 3D/2.5D ICs can save up to 65.53% embodied carbon emission and up to 41.03% overall carbon emission. Given the average 10-year lifetime of AV devices, the EMIB 2.5D IC and all three types of 3D ICs can save carbon emissions compared to the 2D IC, as the 10-year lifetime falls within their choosing metric (T_c) ranges. For the decision-making of replacing the 2D IC with 3D/2.5D ICs, we advise against replacing the original 2D IC due to the significant embodied carbon emissions in the new 3D/2.5D ICs, which cannot be compensated by operational carbon savings over their lifetime (see their replacing metric (T_r) ranges).

6 Conclusion

This work introduces 3D-Carbon, an analytical carbon modeling tool designed to understand the carbon emissions of commercial-grade 3D/2.5D ICs at the early design stage. Addressing the need for such tools amidst the growing use of advanced integration technologies, 3D-Carbon aims to pave the way for future developments in environmentally sustainable 3D and 2.5D ICs.

Acknowledgement

This work was supported in part by CoCoSys, one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

References

[1] 2023. EVG. https://www.evgroup.com/products/bonding/.
[2] 2023. GPU: Power vs Performance. http://uu.diva-portal.org/smash/get/diva2:767619/FULLTEXT01.pdf
[3] 2024. Intel® Stratix® 10 FPGAs Overview - High Performance Intel® FPGA.
[4] Akgun, , et al. 2019. Network-on-chip design guidelines for monolithic 3-D integration. IEEE Micro (2019).
[5] AMD. 2023. AMD Navi 31.
[6] Arunkumar et al. 2017. MCM-GPU: Multi-chip-module GPUs for continued performance scalability. In ISCA.
[7] Bardon et al. 2020. DTCO including sustainability: Power-performance-area-cost-environmental score (PPACE) analysis for logic technologies. In IEDM. IEEE.
[8] Cristian Constantinescu. 2021. AMD EPYC™ 7002 Series – A Processor with Improved Soft Error Resilience. In DSN-S.
[9] Dell. ,2021. LCA of Dell Servers .
[10] Eeckhout et al. 2022. A First-Order Model to Assess Computer Architecture Sustainability. IEEE Computer Architecture Letters (2022).
[11] Elgamal et al. 2023. Design Space Exploration and Optimization for Carbon-Efficient Extended Reality Systems. arXiv preprint arXiv:2305.01831 (2023).
[12] Feng et al. 2022. Chiplet actuary: a quantitative cost model and multi-chiplet architecture exploration. In DAC.
[13] Freitag et al. 2021. The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. Patterns (2021).
[14] GaBi. 2022. GaBi database.
[15] Gomes et al. 2020. 8.1 Lakefield and Mobility Compute: A 3D Stacked 10nm and 22FFL Hybrid Processor System in 12×12mm2, 1mm Package-on-Package. In ISSCC.
[16] Guler et al. 2020. McPAT-Monolithic: An area/power/timing architecture modeling framework for 3-D hybrid monolithic multicore systems. VLSI (2020).
[17] Gupta et al. 2022. ACT: Designing sustainable computer systems with an architectural carbon modeling tool. In ISCA.
[18] Huang et al. 2010. Interaction of scaling trends in processor architecture and cooling. 198 – 204.
[19] Kim et al. 2021. Micro-bumping, hybrid bonding, or monolithic? A PPA study for heterogeneous 3D IC options. In DAC. IEEE.
[20] Kline et al. 2019. GreenChip: A tool for evaluating holistic sustainability of modern computing systems. Sustainable Computing: Informatics and Systems 22 (2019), 322–332.
[21] Mahajan et al. 2016. Embedded Multi-die Interconnect Bridge (EMIB) – A High Density, High Bandwidth Packaging Interconnect. In ECTC.
[22] Mahajan et al. 2021. Advanced packaging technologies for heterogeneous integration. In IEEE Hot Chip Conference. 22–24.
[23] Naffziger et al. 2021. Pioneering Chiplet Technology and Design for the AMD EPYC™ and Ryzen™ Processor Families : Industrial Product. In ISCA.
[24] Nagapurkar et al. 2022. Economic and embodied energy analysis of integrated circuit manufacturing processes. Sustainable Computing: Informatics and Systems (2022).
[25] Nvidia. 2022. Nvidia architecture.
[26] NVIDIA. 2023. NVIDIA Tesla P100.
[27] Stow et al. 2016. Cost and thermal analysis of high-performance 2.5D and 3D integrated circuit design space. In ISVLSI. IEEE.
[28] Sudhakar et al. 2023. Data Centers on Wheels: Emissions From Computing Onboard Autonomous Vehicles. IEEE Micro (2023).
[29] TSMC. 2023. CoWoS.
[30] Wuu et al. 2022. 3D V-Cache: the Implementation of a Hybrid-Bonded 64MB Stacked Cache for a 7nm x86-64 CPU. In ISSCC.