ACCSCIENCE PUBLISHING

International Journal of AI for Materials and Design

REVIEW ARTICLE

AI for chips: A systematic review of artificial intelligence in the semiconductor lifecycle

Ignacio Chechile* and **Jan Akmal**

Department of Energy and Mechanical Engineering, School of Engineering, Aalto University, Espoo, Uusimaa, Finland

Abstract

Artificial intelligence (AI) methods such as machine learning, deep learning, and large language models are increasingly used to design systems, including integrated circuits. This work provides a thorough understanding of AI adoption across the full semiconductor engineering life cycle by collecting and analyzing a corpus of peer-reviewed articles from a curated list of journals and conference proceedings. The results of this systematic review indicate active work on AI use before tapeout and depict a thinly connected lifecycle, where methods focus on specific stages and barely cross neighboring-stage boundaries. Additionally, we analyzed the geographical and industrial landscape of the global AI-for-Chips conversation and found a notable rise of China as a leader in AI research for semiconductor development, amid a general decline in contributions from other countries, with only a few exceptions. This study represents the first work bridging AI, end-to-end semiconductor engineering, and national contributions. The work benefits designers, practitioners, engineers, entrepreneurs, decision-makers, and policymakers by highlighting opportunities and challenges in using AI for chips to improve productivity and sustainability in the era of silicon dominance.

Keywords: Artificial intelligence; Semiconductors; Ontology; Systematic literature review; Challenges

***Corresponding author:**
Ignacio Chechile
(ignacio.chechile@aalto.fi)

Citation: Chechile I, Akmal J. AI for chips: A systematic review of artificial intelligence in the semiconductor lifecycle. *Int J AI Mater Design*. 2026;3(3):026340031. doi: 10.36922/IJAMD026340031

Received: August 20, 2026

Revised: September 3, 2026

Accepted: September 18, 2026

Published online: September 28, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The modern semiconductor manufacturing process sequence is the most sophisticated volume-production technology ever practiced successfully. The process consists of up to a thousand steps that must be performed nearly flawlessly to manufacture leading-edge devices from a silicon wafer.[1-3] This manufacturing process has evolved rapidly in recent years due to the adoption of modern software tools and rapid improvements in critical fabrication steps such as lithography and atomic layer deposition (ALD).[4] Among the tools used to develop semiconductors, artificial intelligence (AI) is the most recent addition.

The term AI is widely used today. Although it is often used to mean chatbots or generative coding in lay terms, AI is an umbrella term encompassing a rich set of interconnected, nested subdomains, each contributing distinct capabilities to modern engineered systems. However, none of these subdomains exists in isolation; modern AI systems require the integration of multiple techniques such as machine learning

(ML), neural networks, knowledge graphs, and computer vision. Existing surveys of AI applied to chip design focus on specific AI methods or lifecycle stages. For example, Huang *et al.*[5] cover ML broadly within electronic design automation (EDA), while Sánchez *et al.*[6] limit their analysis to graph neural networks (GNNs) in EDA. Pan *et al.*[7] review large language models (LLMs) for design automation. Ngo *et al.*[8] conducted a systematic review confined to a single fabrication task: ML-based edge-placement-error analysis and optimization. To the best of our knowledge, no prior systematic review jointly spans the full set of AI/ML method families and the complete semiconductor lifecycle from cradle to grave. The present work aims to address this gap by complementing the cited works with a method-agnostic, lifecycle-spanning analysis. Table 1 presents a summarized comparison between our work and prior review studies, factoring in lifecycle coverage, AI method coverage, protocol used, and landscape analysis.

However, to scrutinize AI use across the semiconductor lifecycle, one must first clearly define the lifecycle and its stages. A challenge quickly surfaces: Not all chips are made the same. Therefore, it is not possible to create a one-size-fits-all description of their lifecycle. For example, in digital designs, the workflow is highly automated and scripting-friendly, making it possible to go from Verilog code to tapeout with a set of scripts.[9,10] In contrast, analog design remains more manual-intensive and less amenable to abstraction and automation.[11] This means their development process will differ substantially. Nevertheless, all semiconductor devices undergo a set of common stages, from ideation to chip sign-off for fabrication and from physical realization to deployment and eventual disposal. We generalize a broader lifecycle in Figure 1 (left-hand side).

Steps 1 to 5 occur purely in the computer domain, whereas steps 6 to 10 require the chip to be physically instantiated. For added simplicity, we further simplify the stages (left side of Figure 1) into six stages: design (specification through layout and sign-off), fabrication (wafer-level), packaging (die-level and multi-die integration), transit (the chip's journey from foundry/integrator to system-level integration), in-field operation, and disposal (right side of Figure 1). We structured our review around this lifecycle discretization.

This review is organized into two parts: the first explains the methodology used for data collection and curation, using ontologies. The second part interprets the data. This includes analyzing the use of various AI subdomains at every lifecycle stage. We conclude the paper with a geographical analysis that aims to portray which actors are driving the topic on a global scale. We have designed the process to be open-source and fully reproducible, with source code, datasets, and configuration files available in the accompanying repositories.

2. Methodology

This work follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement[12], adapted to the engineering domain where the

Table 1. Comparison of this review with prior reviews across four dimensions: lifecycle coverage, AI method coverage, screening protocol, and landscape analysis

Review	Lifecycle stages covered	AI method families covered	Ontology-based semantic screening	Geographic & patent landscape
Huang *et al.*[5] (2021; ML for EDA)	Design only (EDA flow, including mask synthesis and test)	Broad: classical ML, deep learning, reinforcement learning	No; narrative survey, expert-curated inclusion	No
Sánchez *et al.*[6] (2023; EDA and GNNs)	Design only	Single family: GNNs	No; narrative survey	No
Pan *et al.*[7] (2025; LLMs for EDA)	Design only	Single family: LLMs	No; narrative survey	No
Ngo *et al.*[8] (2023; edge placement error)	Fabrication only, and a single task within it	ML methods for one task	No; systematic protocol, but no semantic layer	No
This review	Six phases: design, fabrication, packaging, transit, in-field, disposal	12 families, including symbolic and evolutionary methods	Yes; PRISMA 2020 with ontology-based automated screening of 14,551 records	Yes; geographic analysis of 660 papers and patent landscape analysis of 233 patent families

Abbreviations: AI: Artificial intelligence; EDA: Electronic design automation; GNN: Graph neural network; LLM: Large language model; ML: Machine learning; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

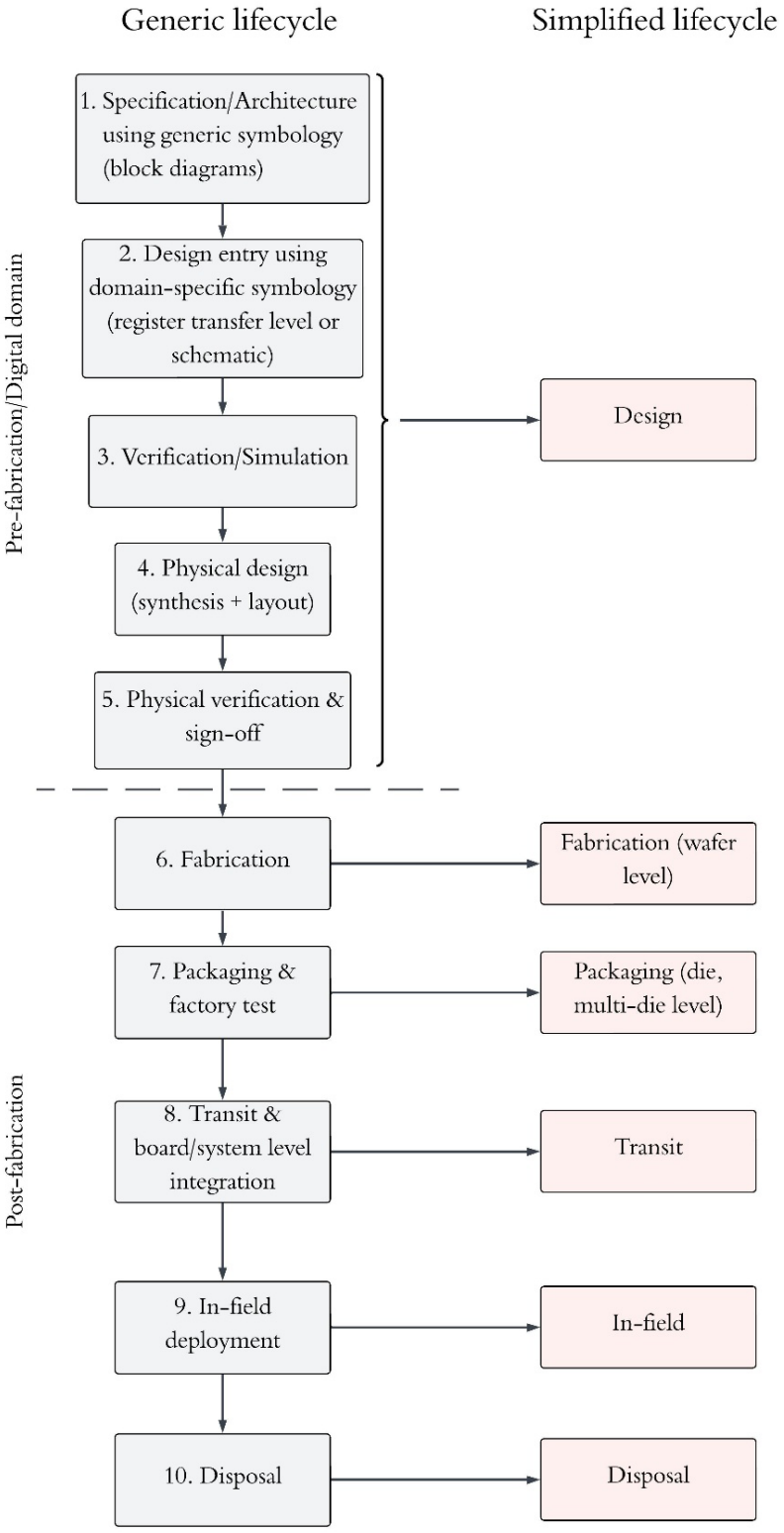


Figure 1. Generic semiconductor lifecycle stages (left: full lifecycle; right: simplified)

risk-of-bias assessment and certainty-of-evidence scoring are replaced by (i) a peer-reviewed journal and conference filter at retrieval and (ii) a high-confidence ontology-match threshold at classification. The Scopus query strings (one per lifecycle phase) are built programmatically in the repository code, based on the ontology defined in slr_ontology.py. The retrieval window is 2015–2026 inclusive, against a curated venue allow-list (venues_eda.txt). Duplication removal, ontology-based classification, high-confidence filtering, false-positive correction, and manual curation are each implemented as separate reproducible pipeline stages. Figure 2 presents the PRISMA 2020 flow diagram for this review.

Our data pipeline returned 19,055 records across six lifecycle-phase queries and two source types (journals and conferences). After removing duplicate entries (4,425 records removed) and 79 version pairs (co-occurring in journals and conferences), 14,551 unique papers (5,531 from journals, 9,020 from conferences) entered ontology-based screening, of which 6,559 matched at least one ontology class (an AI method OR a chip-development task). Of these, 707 reached the high-confidence threshold, that is, an AI method co-occurring with a chip-design task. A targeted filter removed 34 generative-adversarial-network false positives caused by acronym lexical collision ("GAN"/"GaN"), leaving a 673-paper high-confidence corpus (276 journal papers, 397 conference papers). Manual curation removed survey and review papers (5 articles) and 8 false positives. The final synthesis corpus comprises 660 papers.

Retrieval covered 8 journals and 17 conferences spanning the full lifecycle: design automation (DAC, ICCAD, DATE, ASP-DAC); fabrication and device (IEDM, SISPAD, ASMC, ISSM, Symposium on VLSI Technology); packaging (ECTC, ITHERM); and test, reliability, and in-field operation (ITC, VTS, Asian Test Symposium, IRPS, IPFA). Conference venues were selected to encompass the entire lifecycle rather than the design stage alone, so including proceedings would not introduce venue-driven bias favoring any single lifecycle stage.

From that dataset, and to improve readability, we summarized the final corpus in 53 selected articles displayed in per-stage tables in **Section 3**, chosen by a criterion that combines top-cited papers, method diversity, and new (2026) papers surfaced regardless of citation count. Additionally, the summary contains a few editorial promotions for topical variety (Figure 2). All outputs—Scopus queries, retrievals per phase, the classified corpus, the 660-paper high-confidence subset with affiliations and citations, and the curated shortlist—are available in the companion data repository. The source of every curation decision (survey removals, manual false-positive exclusions, stage overrides, and editorial promotions) is documented in analysis/generate_stage_shortlist.py in the repository. Claude Code (Opus 4.6; Anthropic PBC, United States) was used to generate Python code for data retrieval, processing, and visualization, along with usage documentation. We contend that a fully automated data pipeline, including visualization, improves overall reproducibility. Generated code was executed on a Linux virtual machine running Ubuntu (22.04 LTS, Canonical Ltd., United Kingdom) using Parallels Desktop.

It is important to highlight potential methodological limitations when interpreting the findings:

(i) Single-database retrieval. We searched Scopus only. PRISMA 2020 recommends two or more databases to reduce coverage gaps; we chose Scopus for its breadth across semiconductor engineering venues and for its uniform source-type metadata, on which the journal/conference comparison reported in this review depends.

(ii) Preprints are excluded. The review includes peer-reviewed journal articles and conference proceedings, and purposely excludes preprint servers. We posit that preprint servers such as arXiv would increase the risk of duplication and provide research that is still pending peer review. Since this is a rapidly emerging area of research, preprints can induce a high probability of noise in the analyses.

(iii) Title-only classification. The ontology-based classifier matches terms against article titles. We argue this favors precision, since titles name a method when that method is the paper's contribution, whereas abstracts tend to name methods that are used as baselines or benchmarks. We validated the impact by retrieving abstracts independently for 631 corpus papers (92.7% coverage of the corpus when considering the 673-paper set) from open bibliographic sources (OpenAlex, Semantic Scholar). Of these, 588 are works whose abstract exhibits both an AI method and a chip-design task.

(iv) Ontology substring collisions. The ontology-based classifier matches terms against titles. Substring-collision patterns (that is, acronyms appearing in the middle of words) appeared during manual screening and are documented in the code.

(v) Editorial curation. Beyond automated retrieval, per-stage summaries contain editorial decisions. These decisions are individually justified in the code but remain the authors' decisions. A different editor could produce different summaries from the same corpus. We argue that the resulting analysis would not differ substantially.

(vi) Language. Queries were in English; work published

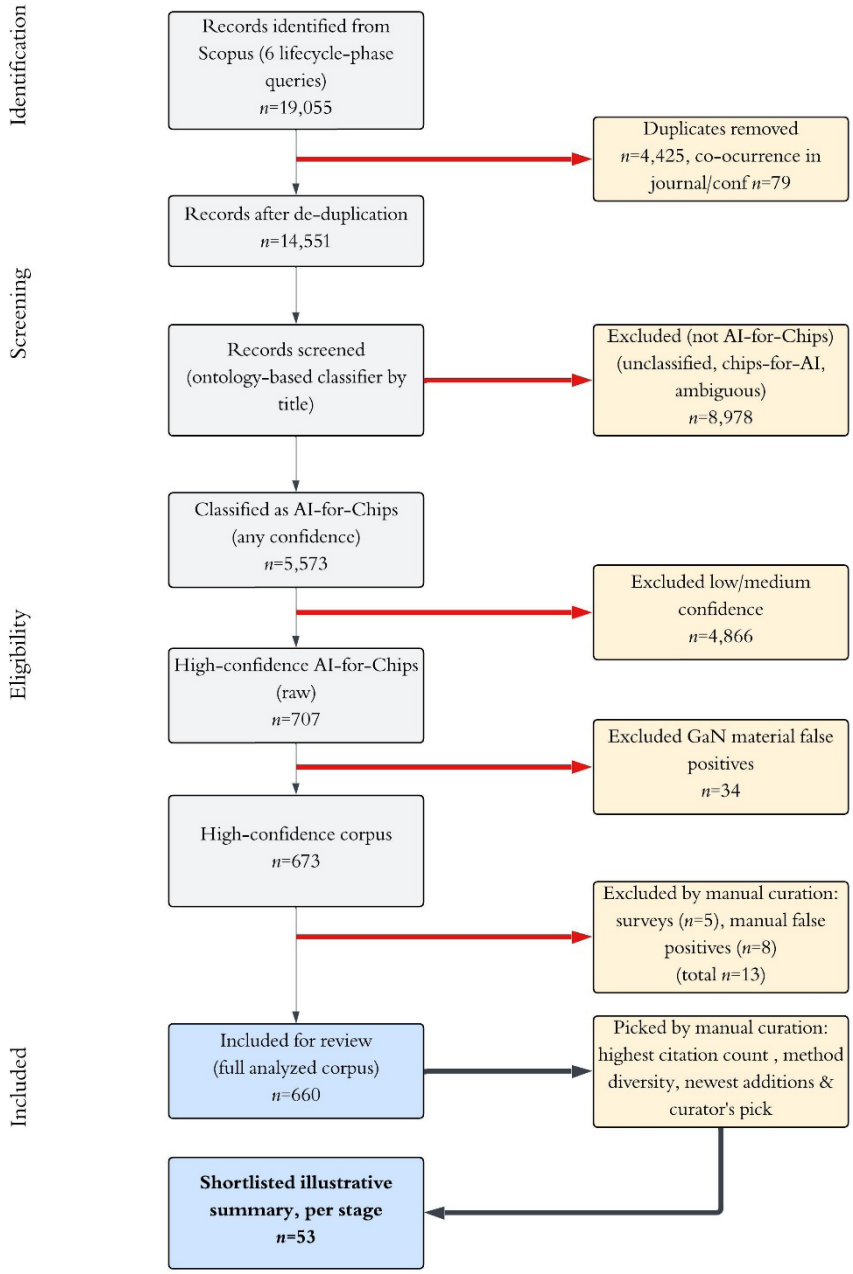


Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of this study
Abbreviations: AI: Artificial intelligence; conf: Conference.

in Chinese, Japanese, Korean, or other languages is thus under-represented.

Claude Code (Opus 4.6) was used to generate Python scripts for programmatically composing queries to the Scopus API. Queries and responses are saved in JSON and CSV format, all available in the data repository.

2.1. Ontology search

Our data pipeline produced thousands of unique articles. A preliminary inspection of titles revealed numerous instances of a subject–object inversion effect: many of the retrieved papers focused on chip architectures designed for AI workloads rather than AI methods applied to chip design problems. This precision problem arises from a fundamental ambiguity in the literature's vocabulary.

Traditional search methods rely on keyword occurrence in a document.[13] However, this approach has a flaw: identical terms may be used in different contexts, leading to the retrieval of irrelevant data. In this work, this shortcoming promptly surfaced: identical AI-related terms describe both computational methods applied to EDA tasks (we call this "AI for chips") and hardware devices that implement such methods (we call these "chips for AI"). Keyword-based filters do not distinguish between these two roles, leading to significant contamination when evaluating literature on AI usage in semiconductor engineering workflows. Consider two papers that both contain the term *neural network*:

- "Neural network classifier-based OPC with imbalanced training data"[14]; here, a neural network is a tool applied to optical proximity correction, that is, a lithography-related task.
- "DSP-Efficient Hardware Acceleration of Convolutional Neural Network Inference on FPGAs"[15]; here, the neural network is the workload that programmable logic is designed to execute.

Both papers would match any keyword query combining *neural network* with semiconductor terms. However, the former uses neural networks to enhance photolithography, whereas the latter addresses executing inference on signal-processing-efficient hardware. The distinction lies in the role the concept plays: is the AI method the instrument or the object? Keyword co-occurrence cannot accurately decode this directionality. To address this problem, we constructed a domain ontology to perform the article search.

A domain ontology is a hierarchically structured set of terms that describe a domain and can serve as a skeletal foundation for a knowledge base.[16] An ontology encodes the roles that concepts play relative to each other, not just their co-occurrence. Thus, ontology-based search is a semantic search method that uses an ontology to understand the intent of a query and the context of the information, rather than just matching keywords. The goal of ontology-based semantic search is to maximize precision, defined as the number of relevant documents retrieved with respect to the total number of documents retrieved.[17] In AI for chips, AI methods are the tool, and chip-design tasks (e.g., floorplanning, timing closure, yield prediction) are the target. In "chips for AI," the chip is the artifact used to run AI workloads (e.g., training, inference). With an ontology, we enforce this directionality, enabling a programmatic method to filter bibliographic data and reducing the need for time-consuming, manual screening for large, unfiltered datasets.

As shown, the retrieval and classification problem this review faces is not one of keyword frequency or occurrence but of semantic role. As shown above, terms such as neural network appear with equal propriety in a paper that applies a neural network to optical proximity correction and in one that designs an accelerator to execute neural network inference. The two papers share vocabulary and differ in what that vocabulary *denotes*. No amount of keyword refinement can separate them, because the distinguishing information is the relation between the terms rather than the terms themselves. Retrieval that matches keywords independently cannot represent such a relation.

2.1.1. Structure

Four concept classes carry the directional distinction, comprising 42 subclasses between them:

- AI method (12 subclasses): the computational method, considered as an instrument: deep learning, reinforcement learning, GNNs, foundation models, and so on.
- Chip-development task (20 subclasses): the semiconductor-engineering problem the method is applied to: placement, routing, lithography optimization, yield prediction, reliability analysis, and so on.
- Hardware artifact (5 subclasses): chips designed to execute AI: accelerators, systolic arrays, in-memory computing.
- AI workload (5 subclasses): the computation such silicon executes: training, inference, quantization.

A fifth class enumerates the six phases of the semiconductor lifecycle used throughout this review: design, fabrication, packaging, transit, in-field operation, and disposal. Two relations connect the concept classes; an AI method may be *applied to* a chip-development task, and a hardware artifact may be *optimized for* an AI workload. The first is the AI-for-chips relation; the second

is the chips-for-AI relation. Each class carries a set of lexical terms—423 in total—which act as the realizations by which its presence in a document is detected.

The ontology is defined once and serves three purposes. It fixes the conceptual classes and the relations that guide classification. It supplies the lexicon used for entity detection and serves as the source from which the Scopus database query vocabulary is generated. Deriving the query and classifier from one definition eliminates the vocabulary drift that arises when a search string and a coding scheme are maintained separately, and it means a reader who disagrees with a certain choice can alter the definition and re-derive both at once.

2.1.2. Detection and classification

Detection is lexical and kept deliberately conservative. A class is present in a document when any of its forms occurs in the paper title, matched on word boundaries with regular inflection. Matching keywords as bare substrings produces systematic false positives on short acronyms embedded in ordinary words, and we identified and corrected such cases during this revision. Then, for a document d, detection yields four propositions:

(i) M(d): An AI method is present
(ii) A(d): A hardware artifact is present
(iii) T(d): A chip task is present
(iv) W(d): An AI workload is present

Directionality is thus decided by a total function from the 16 possible assignments of these four propositions to a classification and a confidence level, given in full in Table 2. Three properties we argue are worth highlighting:

- It is deterministic: Identical input yields identical output. There are no tuned parameters, no thresholds, and no trained components, so the procedure is reproducible exactly.
- The corpus inclusion rule is only one assignment: A document enters the high-confidence AI-for-chips corpus if and only if M ∧ T ∧ ¬A ∧ ¬W, that is, an AI method and a chip-development task are both present, and neither a hardware artifact nor an AI workload is. This formalizes the high-confidence threshold. As shown, it is a logical predicate, not a numerical

Table 2. Complete mapping from the four ontology propositions to a classification and a confidence level

M	T	A	W	Classification	Confidence	No. of papers	Percentage (%)
0	0	0	0	Unclassified[a]	Low	7,776	53.4
0	0	0	1	Chips_for_AI	Medium	14	0.1
0	0	1	0	Chips_for_AI	Medium	197	1.4
0	0	1	1	Chips_for_AI	High	5	0.0
0	1	0	0	AI_for_chips	Low	4,860	33.4
0	1	0	1	AI_for_chips	Low	6	0.0
0	1	1	0	Unclassified[a]	Low	57	0.4
0	1	1	1	Unclassified[a]	Low	0	0.0
1	0	0	0	Ambiguous	Low	809	5.6
1	0	0	1	Ambiguous	Low	12	0.1
1	0	1	0	Chips_for_AI	Medium	78	0.5
1	0	1	1	Chips_for_AI	High	8	0.1
1	1	0	0	AI_for_chips[b]	High	707	4.9
1	1	0	1	Unclassified[a]	Low	2	0.0
1	1	1	0	Both	Medium	19	0.1
1	1	1	1	Both	Medium	1	0.0

Notes: [a]Four assignments are not determined by the four propositions and fall through to a heuristic fallback described in the text. Together, they hold 7,835 records, but 7,776 are empty assignments, where no ontology class matched and the record is simply out of scope. The three substantive fallback cells hold 59 records between them. None contributes to the corpus, which is strictly confined to M ∧ T ∧ ¬A ∧ ¬W. [b] The high-confidence artificial intelligence (AI)-for-chips corpus. M: AI method present; T: Chip-development task present; A: Hardware artifact present; W: AI workload present. All 16 assignments are covered. Paper counts are the 14,551 records that passed screening and were taken from the classifier's own output; the 12 deterministic rows reproduce its recorded classification and confidence for every record, with no disagreements.

threshold, and it does not require calibration. Thus, the confidence levels are not probabilities: it is flagged as high where the pattern is unambiguous, medium where two detections coexist, low where the evidence is one-sided. Only the high-confidence region forms the corpus; the remainder is reported only for transparency.

The domain ontology is implemented as a Python class in slr_ontology.py and enforced at runtime in classify_scopus.py. Although previous work on ontologies for the semiconductor and product lifecycle domains exists[18-21], to the best of our knowledge, this work represents the first ontology defined in research on the use of AI methods for chip design and development across the entire lifecycle.

2.2. Patent analysis

We gauged industry-side activity on AI-for-chips by analyzing intellectual property (IP) creation. By running a companion analysis on Google Patents[22], we applied a patent-side criterion similar to the one applied for academic articles. In this analysis—executed by analysis/patent_analysis.py—a patent qualifies as AI-for-chips only if (i) its Cooperative Patent Classification (CPC) codes (https://www.epo.org/en/searching-for-patents/helpful-resources/first-time-here/classification/cpc) include both a chip-design/EDA class (G06F30, G06F17/50, G06F115, G06F119, G06F11/22, or G01R31/28) and an ML class (any subclass of G06N), and (ii) its title explicitly names an AI method—one of AI, ML, neural network, deep learning, reinforcement learning, GNN, generative adversarial network, convolutional neural network, recurrent neural network, transformer-based neural network, LLM, variational autoencoder, Bayesian optimization, Gaussian process, active/self-supervised/semi-supervised learning, or diffusion model. The CPC-conjunction filter follows our ontology's dual requirement that a publication co-mentions an AI method and a chip-design task; the title-keyword filter removes the residual false positives we observed on CPC-only filtering (AI-accelerator chip patents, semiconductor-equipment patents, and simulation patents unrelated to chip design). Nevertheless, this criterion may undercount companies whose patent-drafting style omits method names from titles (e.g., "Method for …" or "System and apparatus …") to avoid false positives.

3. Results and discussion

3.1. Evolution of the use of artificial intelligence in semiconductor engineering, 2015-2026

Across the corpus, deep learning (23.8% of papers) and classical ML (21.8%) were the two most frequently named method families in AI-for-chips research, followed by generic ML learned models (12.9%); Bayesian and probabilistic methods (11.6%); foundation models and LLMs (11.1%); GNNs (10.8%); and reinforcement learning (10.1%).

When projecting methods' use over time (Figure 3), the dominant method family changed twice within the window observed. Classical ML led in four of the first five years, peaking at 60% of papers in 2017. Deep learning overtook it in 2020 and led for five years, peaking at 38% in 2023. Foundation models and LLMs then displaced both: absent from the corpus entirely between 2015 and 2022, they reached 2.8% in 2023, 9.8% in 2024, and 29.3% in 2025, becoming the single most frequently named method in both 2025 and 2026. We are aware of no comparable transition elsewhere in the corpus, where a method moved from absence to field leadership in three publication years.

Additionally, we observed that method families were unevenly distributed across the lifecycle (Figure 4). Foundation models were concentrated in the design stage (15.2% of design papers) and were comparatively less popular elsewhere: 9.4% in packaging, 6.9% in fabrication, 5.6% in in-field, and 3.8% in transit. The later lifecycle stages continued to rely on classical ML, which led in fabrication (30%), packaging (25%), and transit (38%).

Figure 4 does not fully capture certain factors, such as the comparatively recent emergence of LLMs shown in Figure 3. Therefore, we repeated the same graph as in Figure 4 but this time for the period 2023–2026 (Figure 5), which better illustrates two effects: the impact of LLM/foundation models in the design stage, and the substantial increase in the adoption of GNNs in the transit stage, largely driven by the successful use of GNNs in hardware-Trojan (HT) detection (see **Section 3.5**).

Last but not least, on method trends, we analyzed the preference for AI methods across different chip development-related tasks, as per the ontology keywords. Figure 6 presents these results, including the popularity (in number of papers published) of deep learning in analog design. Register transfer level (RTL) and synthesis grouped the digital end of the workflow, and one effect is worth noting: LLMs were not evenly distributed across this class. Of the 40 LLM-tagged papers, most named RTL, Verilog, or code generation in their titles; one named logic synthesis. The result is therefore better understood as language models being applied to hardware description language code—a text-generation task, where their aptitude is largely unsurprising—than as evidence that they have made their way into the numerical optimization that logic synthesis involves.

3.2. Design stage

Design appears to be the most diverse stage in methods

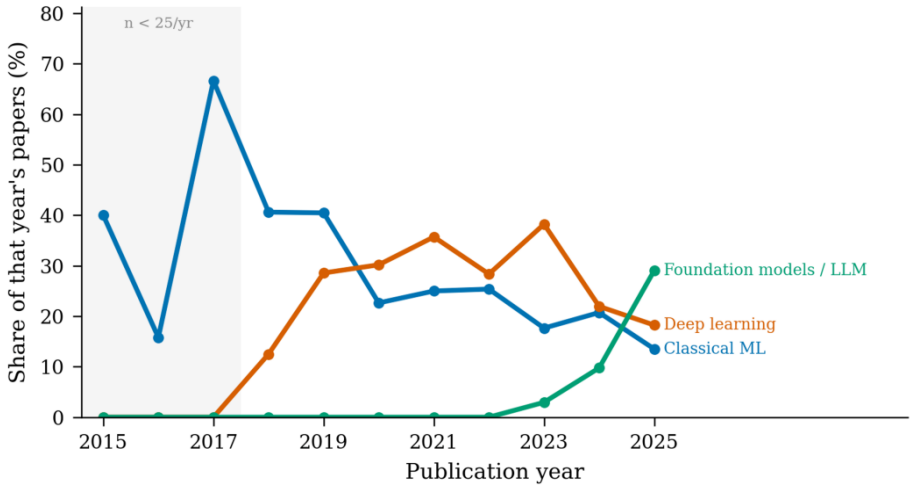


Figure 3. AI method families, 2015–2026. Share of each year's papers naming the three successively dominant families; the leading family changes twice.
Note: Scopus indexing for 2026 was partial at retrieval, so it was removed from the graph for clarity.
Abbreviations: AI: Artificial intelligence; LLM: Large language model; ML: Machine learning.

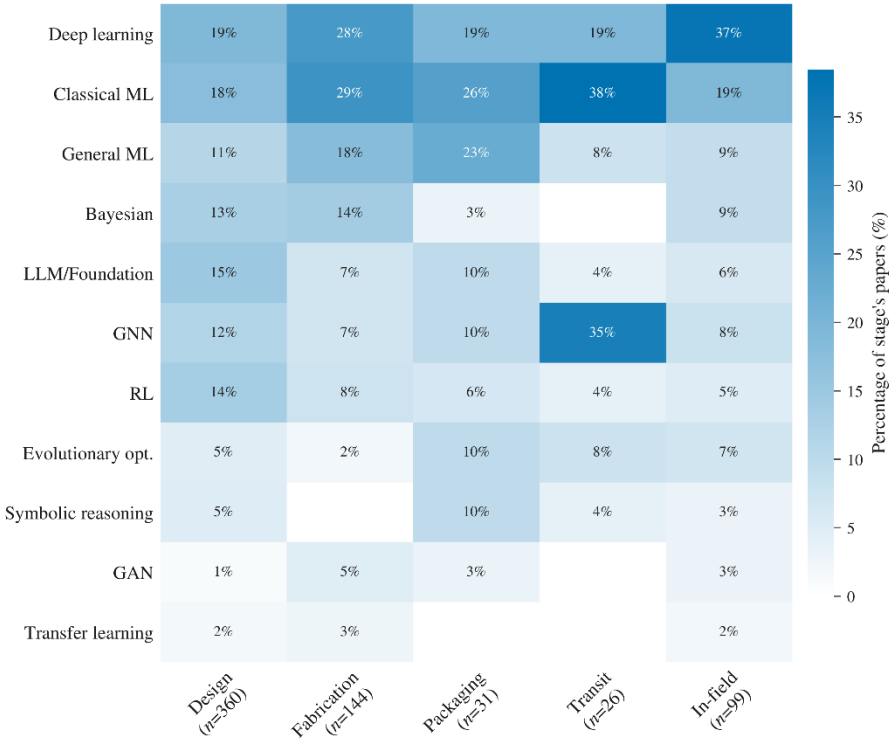


Figure 4. AI method versus lifecycle stage
Note that columns may not sum to 100%, given that certain articles may name more than one method.
Abbreviations: AI: Artificial intelligence; LLM: Large language model; ML: Machine learning.

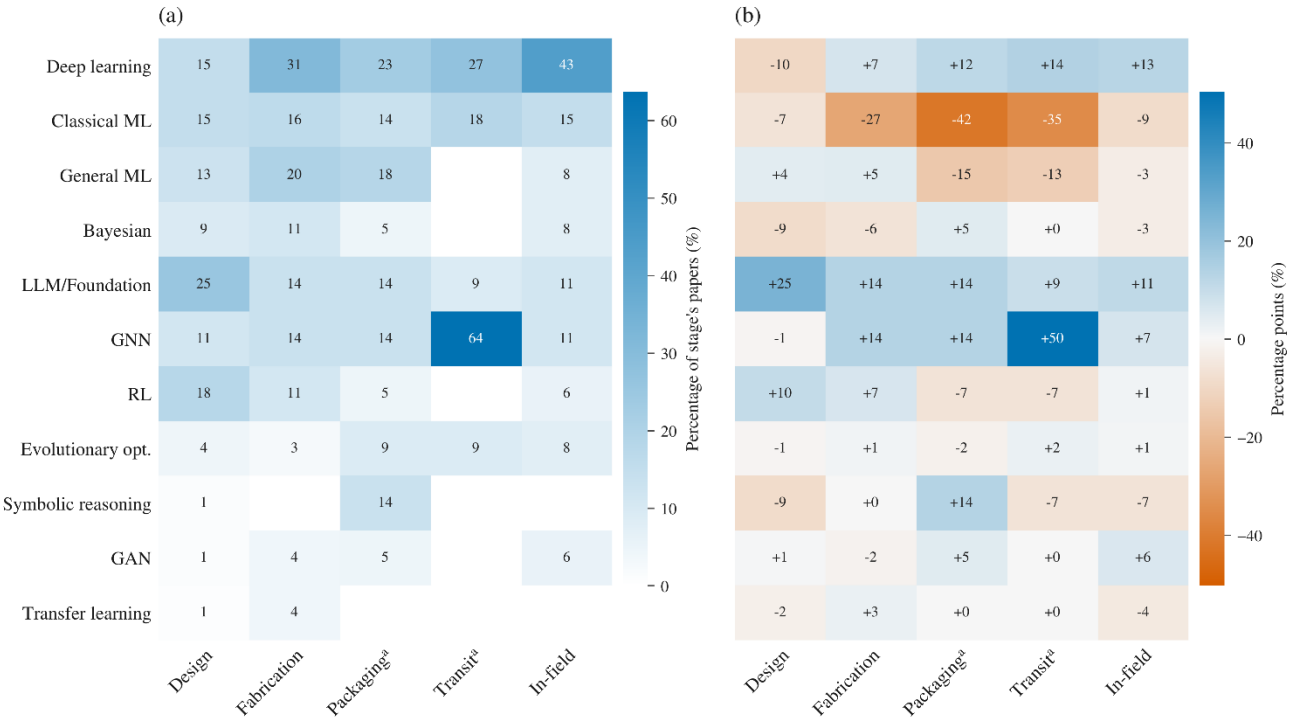


Figure 5. AI method family versus lifecycle stage: (a) share within stage, 2023–2026, (b) changes versus 2015–2022
Note: [a] Packaging and transit stages have fewer than 25 papers in recent years.
Abbreviations: AI: Artificial intelligence; LLM: Large language model; ML: Machine learning.

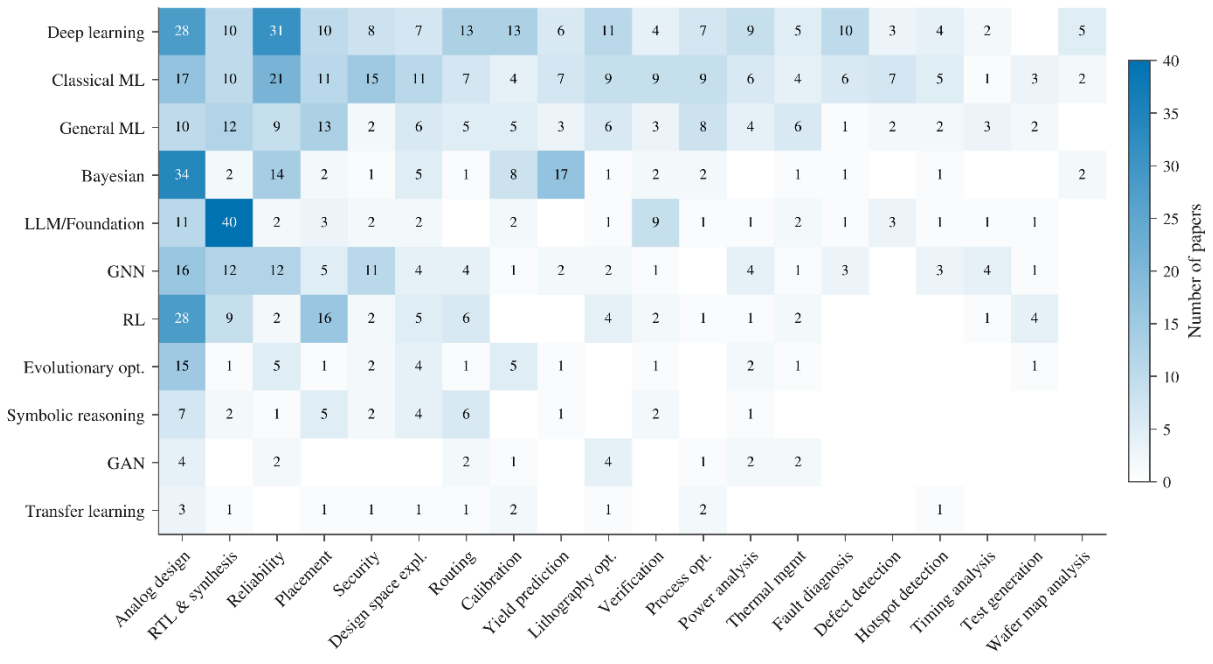


Figure 6. AI method versus chip development task
Abbreviations: AI: Artificial intelligence; Expl.: Exploration; GAN: Generative adversarial network; GNN: Graph neural network; LLM: Large language model; Mgmt: Management; ML: Machine learning; Opt.: Optimization; RL: Reinforcement learning; RTL: Register transfer level.

and tasks. We acknowledge that this diversity reflects our lifecycle definition (**Figure 1**), where design, in our taxonomy, includes a wide breadth of activities: from design exploration to simulation and sign-off. However, a closer look reveals a counterintuitive pattern in this stage's data: analog design tasks appeared most active by citation and frequency. At first glance, digital design appeared more amenable to AI, to algorithmic manipulation, scripting, and automation, taking into account that digital designs build upon a clean hierarchy from standard cell to circuit, module, and system-on-chip.[23] We posit that digital designs' "scriptability" is precisely what makes them a poor target for ML/AI. Where classical, iterative algorithms and deterministic tools already work well, the value of adding AI appears comparatively small. We observed that decades of EDA research have progressively matured the digital design flow into mature automated toolchains. This results in little room for AI methods to outperform the baseline. Analog, by contrast, has resisted such maturation and automation. Analog sizing is a craft that historically depends on designers' intuition, iterative simulation, topology-specific optimizations, and figure-of-merit trade-offs that do not easily map to cost functions.[11] This challenge appears to explain why AI-assisted analog design has produced more research, as its baseline remains time-consuming and human-centered; therefore, even a modest model that yields 10–100× improvements is considered a win. Thus, it is not surprising that the most cited subfield is analog, with papers employing circuit sizing using artificial neural network surrogates[24,25], Bayesian optimization[26], and reinforcement learning.[27,28] The 2024–2026 period appears to match the rising popularity of LLMs, working at different levels: RTL generation (VeriGen[29]), tool orchestration (ChatEDA[30]), analog layout (LayoutCopilot[31]), and analog topology (Atelier[32]). That said, active research continues on the digital side, with a recurring theme of physical-design automation: DREAMPlace[33] for placement, GoodFloorplan[34] for floorplanning, and IronMan-Pro[35] for high-level synthesis. Design-technology co-optimization (DTCO) appeared only in complementary-field-effect-transistor (FET) cell synthesis[36], suggesting the academic ML community has not yet shown strong interest in this direction. Table 3 shows the shortlist of articles for the design stage.

Table 3. Design stage (illustrative summary)

Year	Reference	Method	Cites	Title	Summary + Key results
2020	24	DL	127	"An ANN Assisted Optimization System for Analog Design"	Genetic-algorithm sizing for analog circuits where a neural network stands in for slow SPICE simulations; about 4× faster.
2021	36	Symbolic reasoning	44	"Complementary-FET Standard Cell Synthesis Framework for DTCO"	Constraint-solver-based place-and-route for Complementary-FET standard cells; 21% area reduction over conventional 4.5-track designs. (Symbolic; no learned ML model).
2021	33	DL	107	"DREAMPlace: Deep Learning Toolkit-Enabled GPU Acceleration for Modern Placement"	Recasts classical chip-placement algorithms as neural-network training in PyTorch, exploiting GPU acceleration; 40× faster with no quality loss.
2022	34	GCN; RL	48	"GoodFloorplan: Graph Convolutional Network (GCN) and RL-Based Floorplanning"	Treats chip-block floorplanning as a sequential decision problem solved by GNNs and RL; beats heuristic floorplanners on area and wirelength.
2022	27	RL	60	"Automated Design of Analog Circuits Using Reinforcement Learning"	RL framework that sizes analog circuits in a modern 16nm process; 38× more efficient than prior methods, with layouts that pass downstream design checks.
2022	37	Evolutionary optimization	60	"An Analog Circuit Design and Optimization System With Rule-Guided Genetic Algorithm"	Genetic algorithm with mutations guided by analog-design rules vs. random tweaks; 1.5–3.3× faster on op-amps and oscillators.
2022	25	Classical ML	106	"An Efficient Analog Circuit Sizing Method Based on ML-Assisted Optimization"	Critiques common pitfalls in ML-assisted analog sizing—training cost and accumulating prediction errors—that prior work obscures by testing on narrower problems.

(*Cont'd...*)

Table 3. (*Continued*)

Year	Reference	Method	Cites	Title	Summary + Key results
2022	26	Bayesian optimization/ Probabilities	114	"An Efficient Batch-Constrained Bayesian Optimization Approach for Analog Circuit Design"	Combines different acquisition strategies inside one Bayesian optimizer for analog circuit sizing; 15–74× faster than prior batch optimizers.
2023	38	DL	9	"Worst-case Power Integrity Prediction Using Convolutional Neural Network (CNN)"	CNN-based worst-case power integrity prediction on PDNs; 25–69× speedup at 0.6%–1.0% error for dynamic noise, 24–64× at 0.2%–1.1% for bump current.
2023	39	DL	43	"A New Compact MOSFET Model Based on ANN With Unified Topology"	Neural network replaces standard BSIM4 transistor model for analog simulation at 0.18μm; converges better on amplifier and voltage-regulator simulations.
2023	35	DL; GNN; RL	51	"IronMan-Pro: Multiobjective DSE in HLS via RL"	RL + GNNs predict resource use and timing in high-level synthesis; ~10× more accurate and 400× faster than prior search. Open-source.
2024	30	LLM/Foundation	116	"ChatEDA: A Large Language Model Powered Autonomous Agent for EDA"	Language-model agent that turns plain-English design requests into chip-design tool scripts; outperforms GPT-4 at orchestrating an end-to-end EDA flow.
2024	29	LLM/Foundation	141	"VeriGen: A Large Language Model for Verilog Code Generation"	Open-source language model fine-tuned on Verilog code from GitHub and textbooks; generates more syntactically correct hardware description code than GPT-3.5.
2025	31	LLM/Foundation	22	"LayoutCopilot: LLM-Powered Multi-agent Framework for IC Layout"	Multi-agent language-model assistant for analog layout that turns spoken design input into commands for Cadence Virtuoso, lowering the tool's steep learning curve.
2026	40	LLM/Foundation	0	"Understanding and Mitigating Errors of LLM-Generated RTL Code"	Delves into why language models produce broken Verilog, then fixes the errors with three tools: a domain knowledge base, a rule-checker on the spec, and an iterative simulate-and-debug loop. 98.1% pass rate on a standard Verilog benchmark.
2026	32	LLM/Foundation	4	"Atelier: Automated Analog Circuit Design via Multiple LLM Agents"	Splits analog circuit design across several language-model agents using a small knowledge base instead of expensive fine-tuning.
2026	28	RL	5	"AnaCraft: Duel-Play Probabilistic-Model-Based RL for Analog Sizing"	First RL method that pits two agents against each other to size analog circuits robust to manufacturing and temperature variation; 3× fewer simulations needed.

Abbreviations: CNN: Convolutional neural network; DL: Deep learning; EDA: Electronic design automation; FET: Field-effect transistor; GCN: Graph convolutional network; GNN: Graph neural network; GPU: Graphics processing unit; LLM: Large language model; ML: Machine learning; PDN: Power distribution network; RL: Reinforcement learning; SPICE: Simulation program with integrated circuit emphasis.

3.3. Fabrication stage

Articles on the fabrication stage group around two task families: mask/lithography optimization and wafer-level defect analytics. Computational lithography shows a strong presence. GAN-OPC 2020[41] established AI for optical proximity correction (OPC), followed by Neural-ILT 2.0, which predicts photomasks and refines them via inverse lithography (ILT)[42]; DevelSet's transformer using graphics processing unit (GPU) kernels to optimize photomasks[43], and INN-ILT 2026's physics-embedded invertible neural network (INN).[44] Wafer-map and yield work is dominated by classical ML and early deep learning (defect detection using WM-811K dataset as benchmark[45], a deep learning yield predictor[46], and a support vector machine surrogate[47]). We added a "curator's pick" article on AI for ALD. Gaussian process surrogate (e.g., Gaussian process regression) models for process-parameter optimization, as demonstrated by Seo *et al.*[48], can replace expensive computational fluid dynamics simulation with a learned model and let design-of-experiments search the parameter space at orders-of-magnitude lower cost. Pattern-loading effects—the way local topography distorts deposition rate—appear as an apt problem for graph or CNN surrogates over the wafer layout. Virtual metrology, already mature

in chemical mechanical planarization and etch, remains somewhat unexplored for ALD and could remove the need for expensive *ex situ* thickness measurement. Precursor-chemistry design portrays another opportunity: property prediction of candidate chemicals using molecular GNNs could screen for self-limiting behavior, thermal stability, and byproduct volatility before any wafer is exposed. Finally, real-time closed-loop control driven by *in situ* optical data with an ML controller in the feedback path would convert today's recipe-driven ALD into an adaptive process. One plausible reason why none of these is yet visible in the journal literature is that the data resides with equipment makers and is guarded as trade secrets. Another contributing factor is that the problems are multi-disciplinary in ways that favor industrial teams over academic groups, and there is no public ALD benchmark analogous to WM-811K for wafer maps. We consider the work by Seo *et al.*[48] as the first foothold rather than an outlier. As fabrication data infrastructure matures through silicon lifecycle management (SLM) platforms and process telemetry standardization, we posit that AI-for-ALD is one of the subfields most likely to go active in the next few years. Table 4 shows the shortlisted articles for the fabrication stage.

Table 4. Fabrication stage (illustrative summary)

Year	Reference	Method	Cites	Title	Summary + Key results
2016	47	Bayesian; classical ML	35	"SVM Surrogate Model-Based Method for Parametric Yield Optimization"	ML surrogate (support vector machine) stands in for slow SPICE simulations during Monte Carlo yield optimization; validated against STMicroelectronics industrial reference flow.
2019	46	Classical ML	36	"Wafer Map Yield Prediction via ML for Productivity Improvement"	Deep learning predicts yield from a wafer-map layout before production starts, helping designers pick layouts that improve productivity by ~9%.
2020	41	GAN	90	"GAN-OPC: Mask Optimization with Lithography-Guided GAN"	GAN, pre-trained on inverse-lithography solutions, proposes near-optimal photomask shapes; cuts the iterations needed by conventional optical-proximity correction.
2021	49	Classical ML	55	"ML-Based Detection Method for Wafer Test Induced Defects"	Distinguishes dies that failed because test probe damaged them from those that failed because of real fabrication defects, spotting patterns in probe order; 97% accuracy across six real products.
2021	45	Deep learning	59	"Data Augmentation and Mask R-CNN Instance Segmentation for Wafer Defects"	Image-segmentation network classifies wafers showing more than one defect type at once; 97.7% accuracy on the standard wafer-map benchmark (WM-811K).
2022	42	Deep learning	39	"Neural-ILT 2.0: Domain-Specific Multitask NN for ILT"	One neural network predicts a photomask and refines it via inverse lithography; 15–30× faster turnaround than prior methods, with simpler mask shapes.
2023	43	Deep learning	23	"DevelSet: Deep Neural Level Set for Instant Mask Optimization"	Combines level-set mask representation, a transformer network, and GPU kernels to optimize photomasks in around one second, matching state-of-the-art printability.
2023	50	Classical ML	29	"Linear Regression + KNN for Latent Wafer Defect Identification"	Two-stage model (linear regression for feature extraction, k-nearest-neighbors for classification) screens out latent defects that would cause early-life chip failures, separating them from harmless process variation; 32% more accurate than prior ML by handling the rarity of true defects.
2025	48	Classical ML	4	"Deriving optimal ALD process conditions using ML"	Neural network plus Gaussian-process surrogate predicts thin-film uniformity in atomic layer deposition; ~18× faster than fluid-dynamics simulation, and closes the loop from prediction to optimal process conditions. Curator's pick.
2026	44	Deep learning	0	"INN-ILT: Inverse Lithography via Invertible Neural Network"	Single invertible neural network—with lithography physics built in—handles both the forward (mask → wafer image) and inverse (wafer target → mask) directions of photomask design.

Abbreviations: GAN: Generative adversarial network; GPU: Graphics processing unit; ML: Machine learning; SPICE: Simulation program with integrated circuit emphasis.

3.4. Packaging stage

Packaging population spans wire-bonding prognostics and health management[51], on-chip power-delivery clustering[52], and thermal-aware 3D-IC test scheduling.[53] A clear 3D integrated circuit topic thread runs through the summary.[54-56] An article using a physics-informed neural network (PINN) surrogate for a 2.5D chiplet power distribution network is the newest entry in the summary.[57] It is worth noting that two papers[54,56] in the summary use symbolic/combinatorial optimization rather than learning-based methods; our ontology tagged them as AI via its symbolic-reasoning category, although a stricter ML-only framing would not include them. However, in this work, we used AI in the broad sense of standard reference texts, in which methods that search a space of candidate solutions belong to the field alongside ML. Russell and Norvig[58], for example, treat search, local search—simulated annealing and genetic algorithms among them—and constraint satisfaction as core problem-solving material, reaching ML only several parts later; venues such as Association for the Advancement of Artificial Intelligence (AAAI) and International Joint Conference on Artificial Intelligence (IJCAI) likewise carry standing tracks on search, constraint satisfaction and optimization. Under that definition, a paper qualifies when it applies any such method to a chip-development task. We adopted it as the primary criterion because the aim of the present study is to map the whole space of computational methods entering the semiconductor lifecycle, and because excluding these methods would leave out genuine hybrid work, such as surrogate-assisted evolutionary sizing, where learned models and a metaheuristic work side by side. Table 5 shows the shortlist of papers for the packaging stage.

Table 5. Packaging stage (illustrative summary)

Year	Reference	Method	Cites	Title	Summary + Key results
2015	52	Classical ML	11	"Energy Efficient Adaptive Clustering of On-Chip Power Delivery Systems"	Polynomial-time clustering algorithm that jointly places on-chip power converters and configures voltage regulators; up to 21% power-efficiency gain on IBM power-grid benchmarks, orders of magnitude faster than prior methods (symbolic/algorithmic; no learned ML model).
2022	53	Classical ML	7	"Frequency-Scaled Thermal-Aware Test Scheduling for 3D ICs via ML"	ML thermal model lets test scheduler scale the test clock up or down to stay within temperature limits while testing 3D-stacked chips; up to 60% test-time reduction on benchmark circuits.
2023	56	Symbolic reasoning	3	"ILP-based Substrate Routing with Mismatched Via Dimensions for FBGA"	Integer-linear-programming router for substrate inside a fine-pitch ball-grid-array package, handling via-dimension mismatches between layers; 278× solver speedup from constraint reduction (symbolic/algorithmic; not learning-based).
2023	55	Classical ML; GNN	5	"GNN-based Multi-bit Flip-Flop Clustering and Post-clustering Design Optimization"	GNN clusters single-bit flip-flops into shared multi-bit cells to cut clock-network power in 3D-stacked chips; 18% lower total power and 8.2% higher performance when plugged into a state-of-the-art 3D flow.
2023	54	Symbolic reasoning	6	"Construction of All Multilayer Monolithic RSMTs and Application to M3D ICs"	Pre-computes every minimum-wirelength routing topology on a 3D grid, building a lookup table that accelerates inter-tier-via insertion in monolithic 3D-stacked-chip routing (symbolic/algorithmic, not learning-based).
2023	51	Deep learning	27	"Deep Learning-Based Positioning Error Fault Diagnosis of Wire Bonding Equipment"	Combined deep- and convolutional-neural-network detector spots a misinstalled bond head on a wire-bonder, from position-sensor signals; predictive maintenance flags the fault before it causes scrap.
2025	59	GNN	3	"Lightweight Heterogeneous Graph Embedding for Hotspot Detection"	Lightweight GNN detects lithography hotspots more accurately than image-based detectors, with smaller models.
2026	57	Deep learning	0	"PINN Surrogate for Runtime PDN and Dynamic IR Drop Estimation"	Physics-informed neural network (PINN) analyzes the power-delivery network of a 2.5D chiplet integration at runtime; 300× faster than commercial solvers, with sub-millivolt accuracy.

Abbreviations: GNN: Graph neural network; ML: Machine learning.

3.5. Transit stage

In **Section 1** (Introduction), we defined the transit stage as the journey from a chip's departure from the foundry/assembly facility to its integration into the final system (for new designs, chips typically go back from fabrication to the design office for post-silicon validation before volume production). This journey can be complex and long, including multiple stops at assembly houses worldwide, where chips may receive rough treatment. In our data, we observed that transit-stage research is almost entirely security-focused. Every shortlisted article targets HT detection or counterfeit-integrated circuit identification. Two GNN papers[60,61] defined golden-reference-free Trojan detection; NetVGE[62] and another paper[63] addressed scalability with novel graph convolutions. GNN4HT[64] adds function classification, not just whether a Trojan is present, but also what it does. EMFORCED[65] uses electromagnetic side-channel fingerprints to detect counterfeits without decapsulation, and Narwariya *et al.*[66] proposed a zero-area-overhead age prediction from input-output-pad protection diodes for recycled-integrated circuit detection. Notably absent are ML works on processing environmental or mechanical transit telemetry (shock, temperature, humidity, electrostatic discharge), a clear gap in the literature. We contend that such environmental information during the transit stage could be of value to correlate with in-field failures and infer improvements in handling, storing, and packaging of integrated circuits. Table 6 summarizes this lifecycle stage.

Table 6. Transit stage (illustrative summary)

Year	Reference	Method	Cites	Title	Summary + Key results
2022	60	GNN	41	"Golden Reference-Free HT Localization Using GCN"	Graph convolutional network locates specific signals where a hardware Trojan has been inserted, working directly on the circuit graph without needing a clean reference design; 99.6% accuracy.
2025	61	DL; GNN	35	"Hardware Trojan Detection Using Graph Neural Networks"	GNN operating on a circuit's data-flow graph detects previously unseen hardware Trojans; 97% recall at register-transfer level (21 ms per design) and 84% at gate level.
2022	67	Classical ML	30	"ML for Board-Level Drop Response of BGA Packaging Structures"	Deep neural network replaces slow finite-element simulation for predicting how ball-grid-array solder joints deform when a populated board is dropped; over 1,000× faster, accurate across drop angles outside the training distribution.
2020	65	Classical ML	28	"EMFORCED: EM Fingerprinting for Counterfeit IC Detection"	Uses electromagnetic emissions from a chip's clock network as a "fingerprint" to flag remarked or cloned counterfeit integrated circuits at near-perfect accuracy across vendors, lot codes, and gray-market parts.
2025	64	GNN	12	"GNN4HT: Two-Stage GNN for HT Multifunctional Detection"	GNN that not only flags hardware Trojans but also classifies the malicious behavior they implement. First method to do both stages.
2024	63	GNN	6	"Netwise HT Detection Using Scalable Convolution of Graphs"	Custom graph-convolution algorithm that scales to 17,000-gate netlists and pinpoints which individual wires carry a hardware Trojan in under 0.1 s; 96% recall on the standard Trojan benchmark suite.
2025	62	GNN	3	"NetVGE: Netwise HT Detection at RTL via Variable Dependency Graphs"	Treats register-transfer-level Verilog as a graph of variable dependencies and uses knowledge-graph-style embeddings to score unfamiliar nets; 93% recall on the standard Trojan benchmark, unsupervised and without a clean reference.
2025	66	Classical ML	1	"Leveraging IO Pad Protection Diodes for Recycled IC Detection and Age Estimation"	Detects recycled integrated circuits and estimates chip age from input-output-pad diode voltage drop via polynomial regression; sensorless, zero-area-overhead.
2026	68	Classical ML	0	"AI-Enabled Image Processing for Counterfeit IC Clustering"	Renders timing data from on-chip ring oscillators as images, then clusters them with unsupervised learning to spot hardware-Trojan insertion; 95% accuracy with no clean-reference chip needed.

Abbreviations: DL: Deep learning; GNN: Graph neural network; ML: Machine learning.

3.6. In-field stage

In-field refers to the stage where the chip sits in its final system and performs the task it was designed to perform. In this stage, the chip withstands real workloads. We observed that the in-field stage is the most chip-task-diverse, covering aging and reliability prediction, thermal management, and runtime security. In this stage, physics-informed ML is the emerging theme: physics-informed long short-term memory network (PI-LSTM) for solder fatigue[69] and PINN-LSTM for coupled negative bias temperature instability and hot carrier damage aging.[70] Rahmani *et al.*[71] (post-silicon trace-signal selection) and Bathla and Vasudevan[72] (Bayesian reliability of combinational circuits) tackled validation workflows. In the work by Elnaggar *et al.*[73], runtime HT detection via changepoint anomaly analysis on performance counters complements the pre-silicon HT work we described in the transit stage. Table 7 summarizes selected works for this stage.

Table 7. In-field stage (illustrative summary)

Year	Reference	Method	Cites	Title	Summary + Key results
2017	71	Classical ML	26	"Postsilicon Trace Signal Selection Using ML Techniques"	ML model trained on simulation runs picks which signals to record on a debug bus during post-silicon validation; up to 143% better at letting designers reconstruct the rest of the chip's state.
2019	73	Classical ML	25	"HT Detection Using Changepoint-Based Anomaly Detection"	Changepoint-style anomaly detection on hardware performance counters catches hardware Trojans at runtime; over 99% detection rate on an open-source CPU running on an FPGA. Note: This paper could also reside in the transit stage.
2020	74	Classical ML/System identification	5	"Robust Identification of Thermal Models for In-Production HPC Systems"	Picks which workload traces from a live high-performance-computing cluster are usable to fit a thermal model of the chips; deep learning makes the right pick up to 96% of the time, with a model prediction error of less than 1 °C.
2023	72	Bayesian optimization/ Probabilistic	10	"Bayesian Framework for Reliability Analysis of Combinational Circuits"	Bayesian inference on a four-valued logic representation estimates how often a combinational circuit's output is wrong, replacing the standard duplicated-circuit error-rate method; predictions scale consistently with gate error probability.
2023	75	Deep learning	10	"A DL Framework for Solving Stress-based PDEs in Electromigration"	Physics-informed neural network solves the partial differential equations that govern electromigration (slow drift of metal atoms under high current) in chip interconnects; mesh-free and faster than traditional solvers at comparable accuracy.
2025	69	Classical ML	8	"Solder Joint Reliability via Physics-Informed ML"	Physics-informed long short-term memory network predicts solder-joint deformation and cycles-to-failure for power-electronics reliability, matching the accuracy of finite-element simulation.
2026	70	Deep learning	0	"NN Approach to NBTI/ HCD Coupled Failure Analysis of MOS ICs"	Physics-informed neural network models how transistors age from two coupled mechanisms—bias-temperature instability and hot-carrier damage—across a range of operating voltages; under 12% error against measured ring-oscillator frequency drift.
2026	76	Deep learning	0	"Hybrid Junction Temperature Prediction of IGBTs"	Compares random forests and several neural-network types for predicting the junction temperature of an insulated-gate bipolar transistor; the best recurrent network achieves an error of less than 0.5 K compared with the simulated data (IGBT = power semiconductor, broader AI-for-semiconductor scope).

Abbreviations: AI: Artificial intelligence; CPU: Central processing unit; FPGA: Field-programmable gate array; IGBT: Insulated-gate bipolar transistor; ML: Machine learning.

3.7. Disposal stage

The disposal stage retrieved no articles. This means no published journal work in the window we surveyed applies AI methods to dismantling, material recovery, environmentally compliant decommissioning, or secure cryptographic erasure of on-chip data.

We observed that cost may be a factor. Chips are disposed of in aggregate e-waste, where optical sorting on conveyor belts and bulk hydrometallurgical recovery outperform any plausible per-chip AI pipeline. We also posit that disposal challenges are regulatory (environmental compliance, hazardous-material handling, secure erasure) rather than computational. Together, these factors may explain why disposal is the only lifecycle stage all chips undeniably go through, with no qualifying papers in our corpus. We argue that the so-called "AI race" will sooner require addressing the environmental impact of billions of chips that will be disposed of globally in the years to come.[77]

Another interpretation of the lack of data could be that AI methods have little to offer to this lifecycle stage. AI methods are inherently data-intensive, and chips do not generate data during disposal. Nevertheless, we contend that AI could contribute to specific sub-problems. For example, ML-assisted verification of secure cryptographic erasure on non-volatile memory, chip-architecture-aware material recovery optimization, and end-of-life reuse-versus-recycle decision support driven by on-chip fingerprints. As the so-called "digital thread" of chips materializes through SLM tools, chiplet metadata, and initiatives such as the European Union Digital Product Passport regulation[78], we expect articles on AI-assisted chip disposal work to emerge.

3.8. Cross-stage coupling

The preceding sections describe the adoption of AI methods within each lifecycle stage. Nevertheless, a question we sought to answer is: how does AI-for-chips research connect *across* stages? In our pipeline, each paper was retrieved by one or more of six lifecycle phase queries, each carrying its own stage-specific vocabulary. A paper matched by two phase queries therefore engaged the search vocabulary of two lifecycle stages. Such a "multi-stage membership" was purposely discarded during the per-stage analysis; we recover it here. Our approach is a lexical proxy: we acknowledge that a paper may share the vocabulary of two stages without transferring data or models between them.

Of 660 papers in the corpus, we observed that 506 (76.7%) engage one lifecycle stage, 129 (19.5%) engage two, and 25 (3.8%) engage three. No paper in the corpus engages four or more of the six stages.

However, raw co-occurrence counts can be influenced by each stage's size. Design, fabrication, and in-field are the three largest stages (360, 199, and 166 papers), so their pairwise counts are the largest: design and in-field share 41 papers despite lying four stages apart in the lifecycle, simply because both stages have large counts. The stage-to-stage counts are given in Figure 7.

We posit that explicit cross-stage mechanisms are rarer than shared vocabulary. Only 24 papers (3.6%) use any term semantically denoting coupling: co-optimization, cross-layer, digital twin, SLM, DTCO/STCO, feedback loops, and end-to-end. The comparable figure for the full 14,551-record screened corpus (most of which is not AI-for-chips work) is 3.0%. AI-for-chips research is therefore only slightly more likely to invoke cross-stage language than the surrounding chip literature, not less. Thus, we contend that stage-locality is not distinctive to AI-for-chips research, but it characterizes the surrounding chip literature. We found the absence of inter-stage coupling quite meaningful when compared against domains where the closed loop from in-field data back to design is established practice, for example, rotating machinery[79].

We observed that the underlying dynamics are not changing. The share of papers engaging more than one stage rose from 15% in 2020 to roughly 23% by 2022 and has been flat since (21% in 2023, 23% in 2024, 24% in 2025) while the corpus more than doubled. Growth in AI-for-chips research has not been accompanied by growth in lifecycle integration. We found it relevant that a decade of methodological evolution, from classical ML through deep learning to foundation models (as discussed in **Section 3.1**), has produced no substantial change in how lifecycle stages connect.

3.9. Geographic affiliation of AI-for-chips research

We analyzed where the conversation of AI-for-chips research is happening worldwide and who is leading it. Our data show that the literature is geographically concentrated in a pattern that shifted substantially across the 2015–2025 decade: China overtook the United States as the leading national contributor to AI-for-chips research in 2024 (34 papers to 19 papers) and widened the gap in 2025 (62 to 32), and is leading in the partially indexed 2026 (30 to 18). However, the takeover is recent enough that it has not yet reversed the cumulative record: across 2015–2026 the United States (207 papers) retained a narrow overall lead over China (197 papers). We posit that such a margin is a stock accumulated almost entirely before 2021, and it has now fallen to roughly 10 papers. China's 2025 output alone (62 papers) exceeds its own combined output for the seven years from 2015 to 2021 (35 papers).

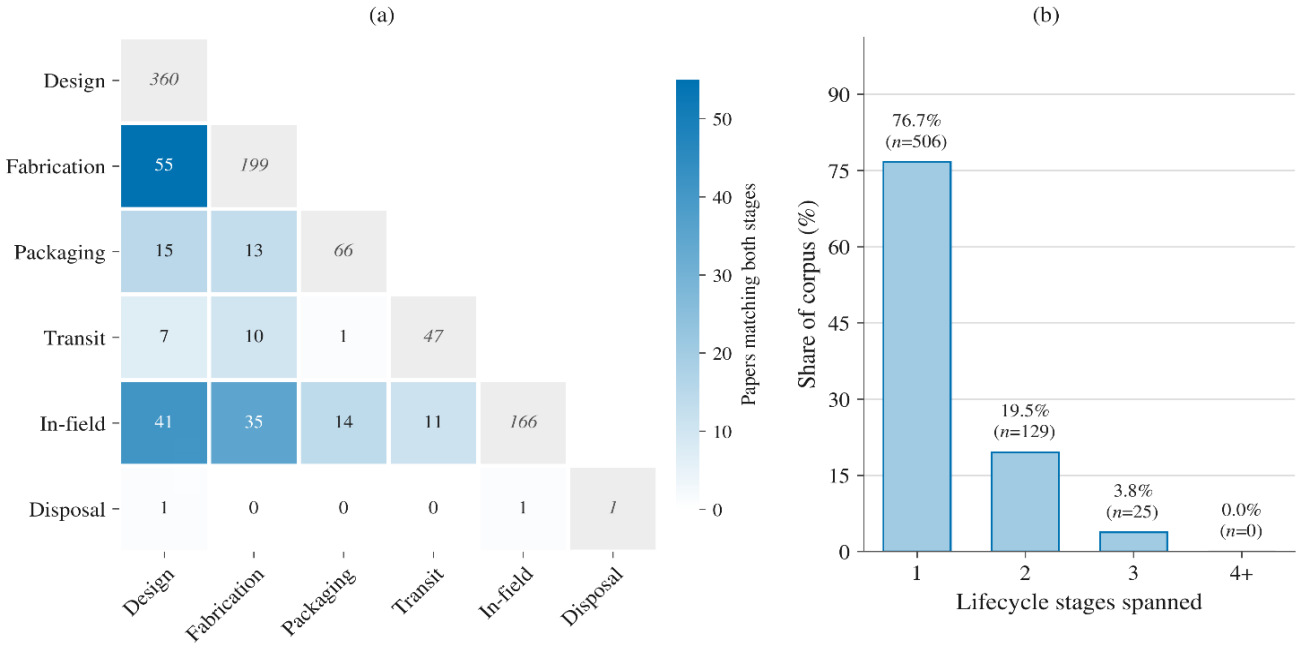


Figure 7. Cross-stage coupling across the semiconductor lifecycle. (a) Stage co-occurrence matrix: off-diagonal cells count papers whose title matches the search vocabulary of both stages; the diagonal (italic) gives the number of papers engaging each stage in total and is held outside the color scale. Co-occurrence is symmetric, so only the lower triangle is shown. (b) Number of lifecycle stages engaged by a single paper.

To analyze national contribution, we processed the high-confidence papers in our corpus with resolvable author affiliations, assigning each paper to every unique country in its affiliation list, following the *any-affiliation* rule that is standard for bibliometric leadership analysis. We excluded five papers with no resolvable affiliation. To test whether European participation should be read as national or supranational, we computed an aggregate "Europe" bloc covering all EU-27 member states plus the United Kingdom, Switzerland, and Norway, matching the framing used by the EU Chips Act[80]. Shares are reported as a country's papers over all papers in the window analyzed; we report this pooled share rather than the mean of annual shares because, in the earliest years, the corpus contains fewer than 20 papers, and a mean of annual percentages weights 2015 as heavily as 2025. Counts and shares are reported over the whole corpus, 2015–2026. The period rates in Table 8 exclude 2026, and the trend figures end at 2025, because Scopus indexing for 2026 was still partial at retrieval: a partial year is legitimate in a total but understates any per-year rate and puts a spurious dip at the end of a series.

Across the corpus, the distribution is headed by the United States (207 papers, 31.4%), China (197, 29.8%) and—treated as a bloc—Europe (71, 10.8%). A second tier follows at 4%–7% each: South Korea (43, 6.5%), Hong Kong (41, 6.2%; Scopus resolves it as an affiliation country distinct from mainland China) and Taiwan (26, 3.9%). India (24, 3.6%), Canada (13) and Japan (12) trail, followed by a long tail of emerging contributors: Iran, Turkey, Brazil, Singapore, the United Arab Emirates, and Saudi Arabia, each with one to seven papers across the decade. Table 8 provides the period comparison (Period 1: 2015–2020; Period 2: 2021–2025); Figure 8 plots the share of the main contributors.

Three findings are worth mentioning. First, five contributors grew faster than the field. AI-for-chips output as a whole grew ×3.1 between the two periods, from 27.5 to 84.2 papers per year. India (×7.2), China (×6.5), South Korea (×4.5), Hong Kong (×4.1) and Taiwan (×3.3) all exceeded that baseline, and all five gained relative weight. India's ratio is the largest but rests on three papers across Period 1 and should be read as a small-base effect; China's rests on 26 and does not (Figure 8).

Second, the United States and Europe both grew in absolute terms but shrank relatively. The United States's output slightly more than doubled between the periods (×2.2) and Europe's grew ×2.3, but both accelerations fell below the field's ×3.1, so their shares compressed: the United States by 11.6 percentage points, Europe by 3.0 percentage points. We argue this is a displacement effect by faster growth elsewhere rather than either country group withdrawing from the topic.

Table 8. National contribution to AI-for-chips research, Period 1 (P1; 2015–2020) versus Period 2 (P2; 2021–2025)

Country	Papers (2015–2026)	Share (%)	P1 rate/yr	P2 rate/yr	Growth	Δ Share (%)
United States	207	31.4	11.2	24.4	×2.2	−11.6
China	197	29.8	4.3	28.2	×6.5	+17.7
Europe (bloc)	71	10.8	3.5	8.2	×2.3	−3.0
South Korea	43	6.5	1.3	6.0	×4.5	+2.3
Hong Kong	41	6.2	1.3	5.4	×4.1	+1.6
Taiwan	26	3.9	1.2	3.8	×3.3	+0.3
India	24	3.6	0.5	3.6	×7.2	+2.5
Canada	13	2.0	0.8	1.6	×1.9	−1.1
Japan	12	1.8	1.0	1.0	×1.0	−2.4
Field total	660	100.0	27.5	84.2	**×3.1**	–

Notes: Rates are papers per year within each period; growth is the ratio between periods. Δ Share is the change in the country's share of field output between periods, in percentage points. Paper counts and shares cover 2015–2026; the period columns exclude the partially indexed 2026, which would understate the Period 2 rate.
Abbreviation: AI: Artificial intelligence.

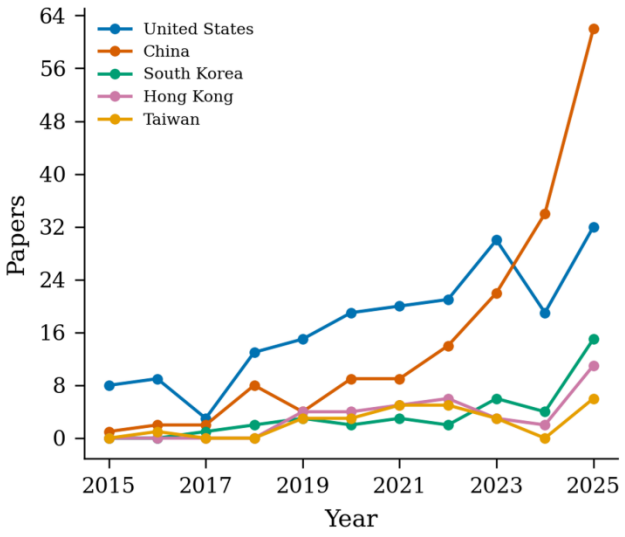


Figure 8. Top 5 contributing countries trends (in papers/year), illustrating China's rapid rise in the artificial intelligence (AI)-for-chips "race," overtaking the United States in 2024

Third, the transition appears as abrupt rather than gradual. The United States led every year from 2015 to 2023, at one point holding 40.6% of Period 1 output. Chinese output was flat at one to nine papers per year through 2021, then ramped sharply—14, 22, 34, and 62 papers from 2022 to 2025—while United States output oscillated between 19 and 32 over the same four years. We observed that the inflection aligns with, and is probably not independent of, the LLM-for-EDA wave, the GNN-based work in placement and HT detection, and the broader post-2022 surge in ML research worldwide.

We also cross-referenced the AI-for-chips subset against the full screened chip-research corpus (n = 14,551; of which 14,456 [99.3%] have resolvable affiliations), revealing two dynamics not visible in the AI-for-chips numbers alone (Figures 9–Figure 14). First, AI-for-chips grew ×3.1 between the periods while general chip research grew only ×1.2: it is a fast-growing subfield inside a comparatively static parent domain (Figure 10). Second, national participation differs between the two. China (29.8% of AI-for-chips versus 20.0% of the full corpus), Hong Kong (6.2% versus 1.7%), and the United States (31.4% versus 28.1%) were over-concentrated in AI-for-chips relative to their overall chip-research presence; Europe (10.8% versus 19.1%), Taiwan (3.9% versus 7.5%), and India (3.6% versus 5.4%) were under-concentrated (Figures 13 and 14). The asymmetry is sharpest for the United States and Europe, whose general chip-research output was flat (×1.0) to mildly declining (×0.9) between the periods, respectively, while their AI-for-chips output more than doubled. Therefore, assessing the geographic landscape through the AI-for-chips lens overstates China's overall dominance in chip development; China leads growth in the subfield but accounts for only 20.0% of the full corpus. Equivalently, it understates European and Indian engagement, which is concentrated in non-AI topics such as materials, device physics, and automotive that fall outside our inclusion criterion while remaining substantial contributions to the broader field. Figure 12 depicts the comparison of the percentage share between Periods 1 and 2 for the main

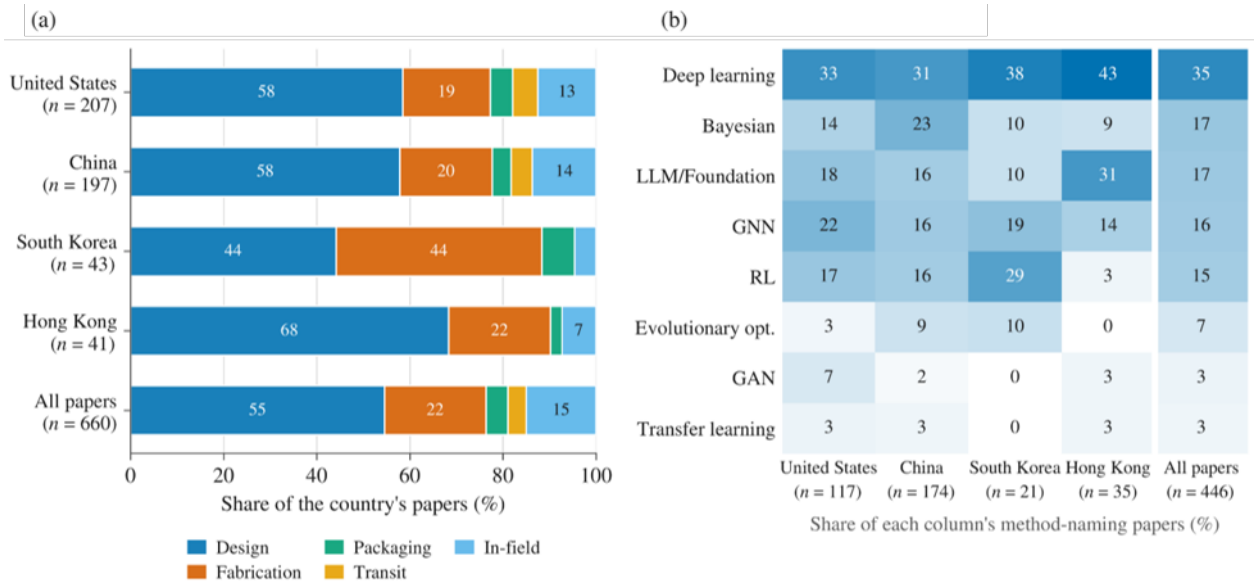


Figure 9. Research profile of the four largest national contributors. (a) Proportion of papers per lifecycle stage for the top contributing countries. (b) AI method family for top contributing countries, restricted to papers whose title names a specific method family (n given per column).
Abbreviations: AI: Artificial intelligence; GAN: Generative adversarial network; GNN: Graph neural network; LLM: Large language model; Opt.: Optimization; RL: Reinforcement learning.

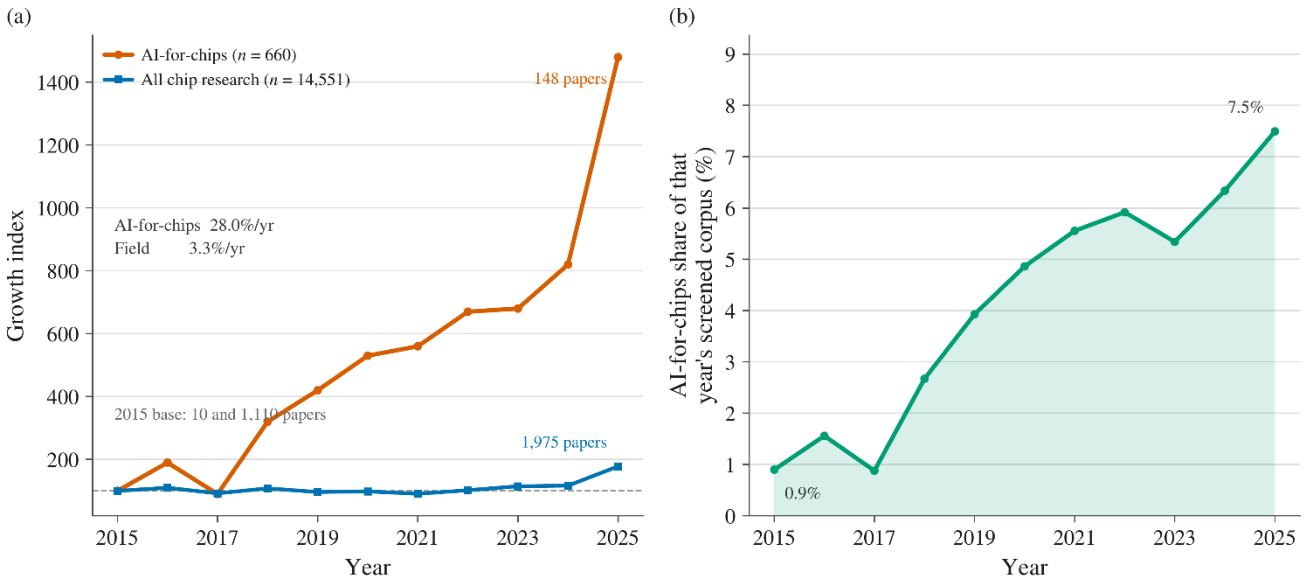


Figure 10. Research corpus growth. (a) Growth relative to 2015 baselines (2015 baseline for artificial intelligence [AI]-for-chips: 10 papers; for general chip: 1,110 papers). (b) Presence of AI-for-chips papers in the general field.

contributing countries. Furthermore, we observed that research effort in the two leading contributing countries is more similar than intuition might suggest. The United States (207 papers, 31.4%) and China (197 papers, 29.8%) are almost indistinguishable in where they work across the lifecycle—design 58% versus 58%, fabrication 19% versus 20%, in-field 13% versus 14%; see Figure 9a. AI method shares in Figure 9b appear to diverge, but we contend that such differences can be a titling convention rather than a research choice: 88.3% of China-affiliated papers name a specific method family in their title, compared with 56.5% of the United States papers, which favor the generic phrase

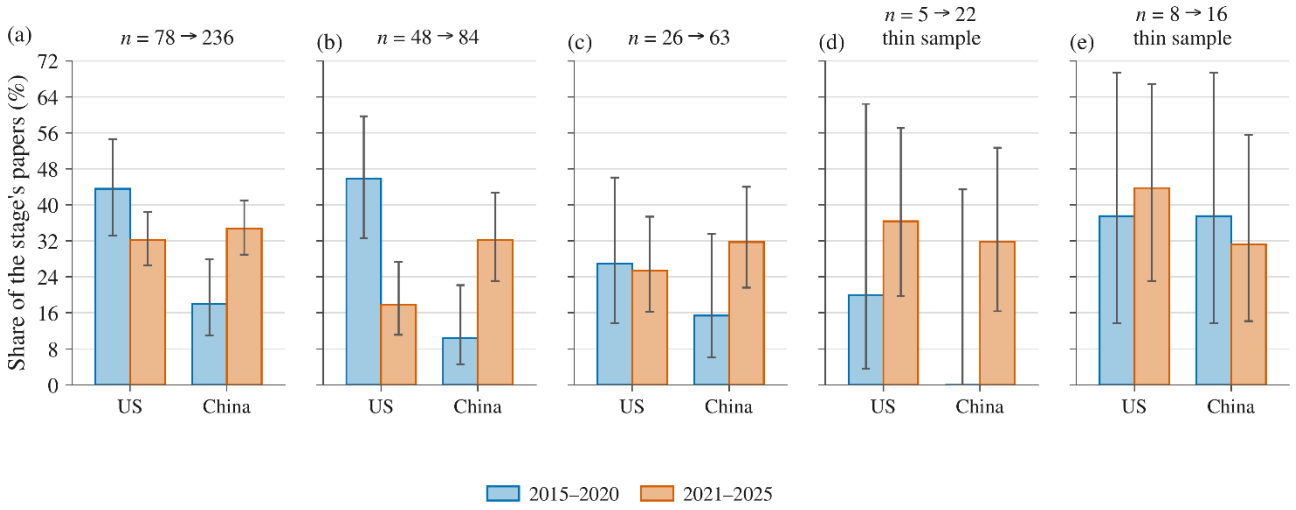


Figure 11. Period 1 (2015–2020) vs. Period 2 (2021–2025) comparison per lifecycle stage, United States (US) vs. China: (a) design, (b) fabrication, (c) in-field, (d) packaging, and (e) transit.

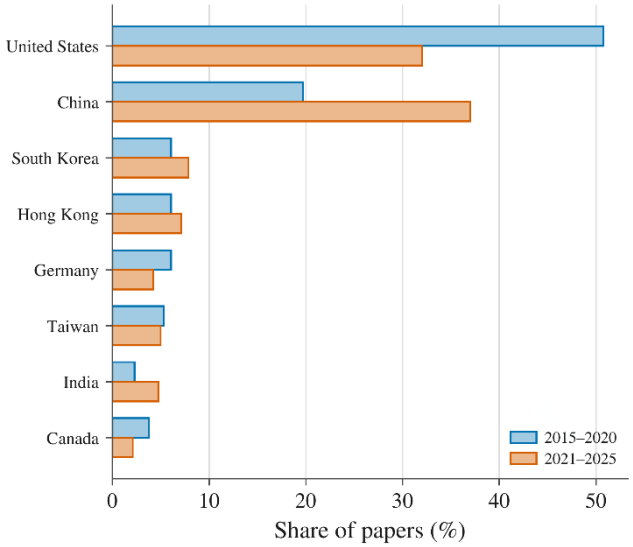


Figure 12. Comparison of contribution (in percentage with respect to the total of artificial intelligence [AI]-for-chips works published) between Period 1 (2015–2020) and Period 2 (2021–2025)

"machine learning." Restricted to papers that name a family, no major method family separates the two countries beyond noise. Notably, the two largest contributors barely work together: 2 of 660 papers carry both United States and Chinese affiliations. Figure 11 depicts a comparison, per lifecycle stage and for each period, between the United States and China.

3.10. Industrial involvement in AI-for-chips research

We gauged industry-side activity on AI-for-chips by analyzing IP creation. Running a parallel analysis on Google Patents—data retrieved August 26, 2026—we applied a patent-side criterion similar to the one applied to academic articles. In this analysis, executed by analysis/patent_analysis.py, a patent qualifies as AI-for-chips only if (i) its CPC codes include both a chip-design/EDA class (G06F30, G06F17/50, G06F115, G06F119, G06F11/22 or G01R31/28) and a ML class (any subclass of G06N), and (ii) its title explicitly names an AI method—one of AI, ML, neural network, deep learning, reinforcement learning, graph neural network, generative adversarial network, convolutional neural network, recurrent neural network,

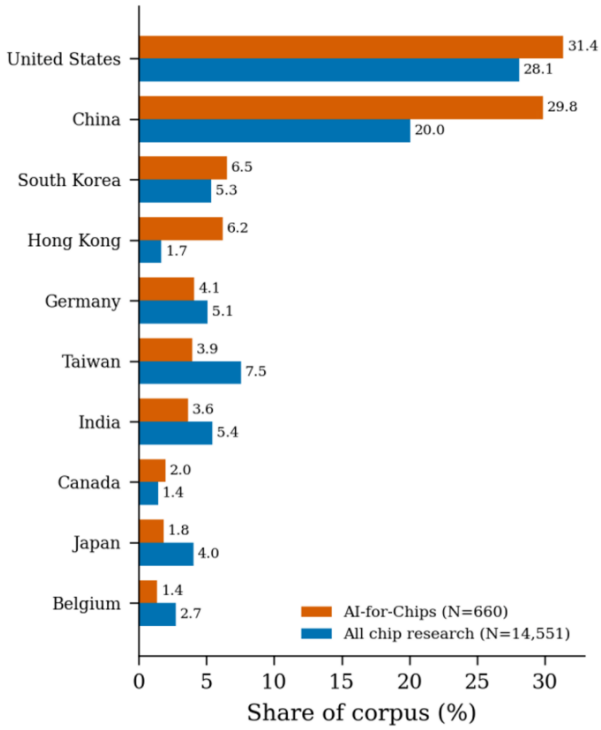


Figure 13. National contribution in artificial intelligence (AI)-for-chips vs. all chip research

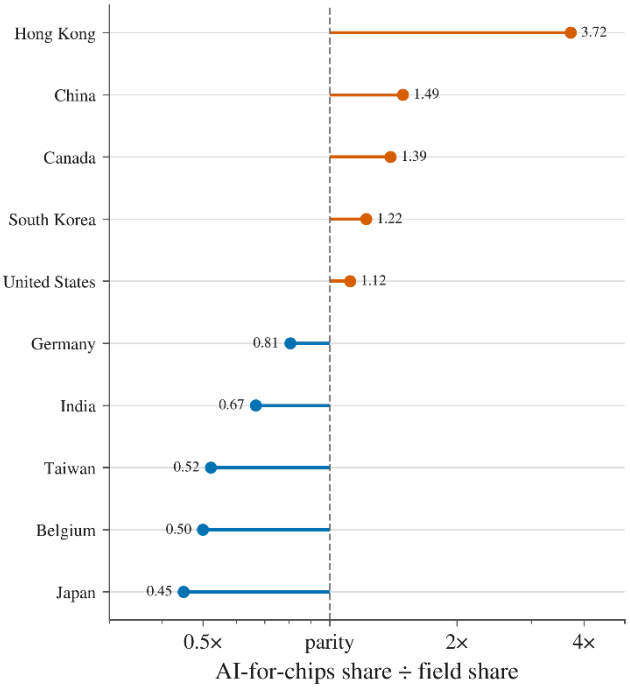


Figure 14. National specialization computed as a ratio of artificial intelligence (AI)-for-chips research over the general chip research field
Note: The ratio is computed in the following way

transformer-based learning, LLM, Bayesian optimization, Gaussian process, active/self-supervised/semi-supervised learning, or diffusion model. The CPC-conjunction filter follows our ontology's dual requirement that a publication co-mentions an AI method and a chip-design task; the title-keyword filter removes the residual false positives we observed under CPC-only filtering (AI-accelerator chip patents, semiconductor-equipment patents, and simulation patents unrelated to chip design). This criterion undercounts firms whose drafting style omits method names from titles ("Method for …", "System and apparatus …"), which appears to be a reasonable trade-off for increasing retrieval precision.

Our analysis yielded 233 distinct patent families across 20 companies, with earliest publication dates from 2015 to 2026. IBM leads with 41 families, followed by Samsung (35), Siemens (including Mentor Graphics; 34), Synopsys (25), and NVIDIA (23); the five largest holders account for two-thirds of the total. The three EDA-tool vendors—Synopsys, Cadence, and Siemens EDA—together hold 69 of the 233 families (29.6%), the largest industrial bloc, showing that a concentration of AI-for-chips in the design stage is as active on the patenting side as in the literature.

From Table 9, we observed that seven companies hold patents without appearing in the corpus at all: ASML (11), Google (6), Applied Materials (5), KLA (4), MediaTek (4), Meta (2), and Micron (2).

Table 9. AI-for-chips patent families (2015–2026) versus papers in the curated corpus (n = 660) per company

Company	Patent families (n)	Papers in corpus (n)
IBM	41	4
Samsung	35	9
Siemens (including Mentor Graphics)	34	6
Synopsys	25	1
NVIDIA	23	7
ASML	11	0
Intel	10	8
Cadence	10	1
TSMC	7	3
Huawei	6	1
Google	6	0
Applied Materials	5	0
Qualcomm	4	2
KLA	4	0
MediaTek	4	0
AMD	2	1
Meta	2	0
Micron	2	0
NXP	1	2
STMicroelectronics	1	1
SK Hynix	0	3

(*Cont'd...*)

Table 9. (*Continued*)

Finally, the geographic asymmetry between the publication and patent views is worth mentioning. Filings resolve to the United States in 173 of 237 company-family records, against 14 for China, 14 for South Korea, and 11 for Taiwan; only Huawei and MediaTek appear among the Chinese and Taiwanese firms in our benchmark set. Contrasted against China's leadership of the publication record from 2024 onward (**Section 3.9**), this is difficult to read as a genuine absence. We acknowledge the retrieval is English-language, and the CPC-plus-title criterion favors filings drafted for Western jurisdictions, both limitations described earlier in this review. Our patent analysis therefore almost certainly under-represents the mainland Chinese AI-for-chips patent landscape, in which companies and research institutions are known to be active (e.g., QiMeng [or Qimeng] is an AI-powered, fully automated processor chip design system unveiled in June 2025 by the Chinese Academy of Sciences [https://qimeng-ict.github.io/]), and should be interpreted as a patent landscape seen from a Western vantage point rather than as a global count.

4. Conclusion

This systematic literature review examined 14,551 papers published across leading semiconductor journals and conference proceedings between 2015 and 2026, identifying 660 high-confidence instances of AI methods applied to semiconductor chip development. The findings describe a pattern in which pre-tapeout design stages capture most of the research attention.

The design stage, the largest stage in terms of articles retrieved, shows a preference for analog sizing tasks and describes a rapid rise of LLMs in recent works. The United States dominated this stage in 2015–2020; China dominated it in 2021–2025.

Fabrication stage articles cluster around mask/lithography optimization and wafer-level defect analytics.

The stage shows a preference for deep learning and classical ML methods. ALD shows a small but likely-to-grow presence in the stage's corpus. The United States dominated this stage in 2015–2020; China dominated it in 2021–2025.

The packaging stage depicts a growing interest in AI methods for 3D chips, with a focus on thermal management and routing. The United States led this stage in 2015–2020, although the corpus of AI-for-chips papers in this period is rather small. In 2021–2025, the United States led but was closely followed by China.

The transit and in-field stages show strong interest in anti-counterfeit measures using GNNs and fingerprinting techniques. In-field operation shows a growing focus on thermal spot detection under real workloads. The United States led the transit stage in 2015–2020, although, just like with the packaging stage, the number of AI-for-chips works in this period is rather small. In 2021–2025, the United States led more clearly, with many works addressing cybersecurity and HT detection.

The disposal stage retrieved no articles. Considering that the venues selected encompass chip development activities from end to end, this lack of data suggests that the disposal stage is still not actively included in the semiconductor development research landscape. This suggests that the semiconductor industry is not actively pursuing AI methods for sustainability. However, we argue that the rapidly rising demand for semiconductor chips has put pressure on their supply, prompting sustainable disposal from end-of-life, i.e., reuse and recycling, to address the surge in demand. Although purely sustainability-dedicated journals were outside the scope of the present study, we contend that sustainability for semiconductors warrants a dedicated track, such as special issues, most plausibly within an existing electronics, materials, and environment venue rather than as a new scientific publication channel. Such a track would provide a stable reviewer pool spanning device engineering, sustainable development, and materials treatment and recovery.

Beyond stage-by-stage reading, we observed that most of the AI-for-chips research reviewed here addresses specific problems local to a single lifecycle phase. The journey from architecture to tapeout, from fabrication to target system, appears thinly connected between lifecycle stages beyond their adjacent neighbor. No paper in our high-confidence corpus explicitly addresses the use of AI from architecture through sign-off to disposed silicon. We contend that the risk of this fragmentation is that research focuses on increasingly smaller per-stage (or stage-pairs) gains without looking into the full lifecycle benefits. This lack of interpolation between lifecycle stages opens a major research area for unified cross-stage modeling, e.g., involving data collection, data structuring and storage, and data analysis.

The economic model of integrated circuits was laid out more than half a century ago, and it showed that a complex electronic function could be realized either on one large chip with low yield and high cost, or partitioned across several smaller chips at higher per-die yield[81], with testing and assembly costs—which increase considerably for partitioned chips—acting as counterweights that determine the optimum partition. For decades, Moore's Law scaling suppressed this dilemma by making the single-die option progressively cheaper.[82] That equilibrium has been dismantled[83]; the industry has turned to partitioning, enabled by complex 3D packaging technologies. We argue that this shift has consequences for AI-for-chips research. In a chiplet-based heterogeneous system, a design decision for one die directly affects other dies in the same package across many domains, including thermal, mechanical, and electrical. Moreover, we posit that, in multi-die packages, fabricators and integrators will face bigger challenges verifying against failure scenarios that can only reveal themselves under real in-field workloads. If heterogeneous integration is to become the dominant design paradigm over the next decade, we contend that the field will need AI methods that also bridge life cycle actors more tightly together, including academia, tooling, design houses, chiplet vendors, fabs, and system integrators. Such bridging is, in our view, the most important open research direction this review reveals because it will require unprecedented levels of standardization and interoperability. More precisely, we mean data interoperability; the challenge at hand has a data problem embedded in it. The stage boundaries visible in the literature coincide with boundaries of commercial ownership: design house, foundry, assembly facilities, and operator in the field. No party seems to hold both sides of any boundary, and our data shows that ownership drops sharply across stages. However, we contend that it is also a modeling problem. Shared knowledge across boundaries and actors producing and consuming data will require structuring and encoding time-variant, expert-level knowledge as product development evolves; this will require formalizing ontologies and knowledge graphs and creating high-fidelity digital twins that incorporate more realism as computational power becomes available. That, in our view, is the realistic path toward a realistic digital thread for the semiconductor. We contend that, if there is any time in the history of complex product development to tackle this challenge, that time is now.

One interesting finding in this work is that a decade of methodological change, from classical ML through

deep learning to foundation models, has passed through this field without altering how its lifecycle stages connect. However, we acknowledge we might still be in the beginning, as our data suggests, of the "LLM takeover," so we can expect with reasonable certainty that these new, transformative technologies will create a greater impact in product development lifecycles. Nevertheless, we recognize the principal challenge for these technologies lies in jumping from the pure computational substrate where they currently live into the physical world.

As we draw closer to the end of this work, we acknowledge that peer-reviewed literature on AI-for-chips is a partial view of the broader field. Readers should therefore interpret our findings as a characterization of what the academic community has chosen to publish, not of what the AI-for-chips field is doing overall.

Last but not least, geographically, our analysis shows that research on AI-for-chip development is not only picking up pace but also driving most semiconductor research growth in several developed nations. Moreover, our data show China as the accelerator, with growth metrics that beat everyone else, positioning the country as a strong leader in a sector that will be critical for the computing needs of the years to come.

Acknowledgments

None.

Funding

None.

Conflict of interest

Jan Akmal serves as an Editorial Board Member of the journal but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Other authors declare they have no competing interests.

Author contributions

Conceptualization: Ignacio Chechile
Writing–original draft: Ignacio Chechile
Writing–review & editing: Ignacio Chechile, Jan Akmal

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data presented in this work are openly accessible at the repository https://github.com/ignaciosim/AI4Chips_SLR_data/ (release v.1.0.7), and are archived on Zenodo at https://doi.org/10.5281/zenodo.22237162.

The code and analysis/post-processing scripts supporting this article are openly accessible at the repository https://github.com/ignaciosim/AI4Chips_SLR and are archived (release v1.0.7) on Zenodo at https://doi.org/10.5281/zenodo.22237151.

References

1. Verner A, Kim H, Jeong I, *et al.* Optical overlay measurement accuracy improvement with machine learning. *Proc Metrology, Inspection, and Process Control for Microlithography XXXIV*. 2020;11325:113251Z.

   doi: 10.1117/12.2551850

2. May GS, Spanos CJ. *Fundamentals of Semiconductor Manufacturing and Process Control*. Wiley-IEEE Press; 2006.

   doi: 10.1002/0471790281

3. Neshati R. A Statistical Model to Analyze the Complexity of Semiconductor Manufacturing. *Proc 2013 Portland Int Conf Manag Eng Technol (PICMET)*. 2013:2757-2764. https://ieeexplore.ieee.org/document/6641614

4. Wu Q, Li Y, Zhao Y. The evolution of photolithography technology, process standards, and future outlook. *Proc IEEE 15th Int Conf Solid-State Integr Circuit Technol (ICSICT)*. 2020:1-5.

   doi: 10.1109/ICSICT49897.2020.9278164

5. Huang G, Hu J, He Y, *et al.* Machine Learning for Electronic Design Automation: A Survey. *ACM Trans Des Autom Electron Syst*. 2021;26(5):40.

   doi: 10.1145/3451179

6. Sánchez D, Servadei L, Kiprit GN, Wille R, Ecker W. A Comprehensive Survey on Electronic Design Automation and Graph Neural Networks: Theory and Applications. *ACM Trans Des Autom Electron Syst*. 2023;28(2):15.

   doi: 10.1145/3543853

7. Pan J, Zhou G, Chang CC, Jacobson I, Hu J, Chen Y. A Survey of Research in Large Language Models for Electronic Design Automation. *ACM Trans Des Autom Electron Syst*. 2025;30(3):34.

   doi: 10.1145/3715324

8. Ngo AT, Dey B, Halder S, De Gendt S, Wang C. Machine Learning-Based Edge Placement Error Analysis and Optimization: A Systematic Review. *IEEE Trans Semicond Manuf*. 2023;36(1):1-13.

doi: 10.1109/TSM.2022.3217326

9. Shalan M, Edwards T. Building OpenLANE: A 130nm OpenROAD-based Tapeout-Proven Flow. *Proc 39th Int Conf Comput Aided Des (ICCAD '20)*. 2020:110.

   doi: 10.1145/3400302.3415735

10. Ajayi T, Chhabria VA, Fogaça M, *et al.* Toward an Open-Source Digital Flow: First Learnings from the OpenROAD Project. *Proc 56th ACM/IEEE Design Automation Conf (DAC '19)*. 2019:76.

    doi: 10.1145/3316781.3326334

11. Scheible J, Lienig J. Automation of Analog IC Layout: Challenges and Solutions. *Proc 2015 ACM Int Symp Phys Des (ISPD '15)*. 2015:33-40.

    doi: 10.1145/2717764.2717781

12. Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.

    doi: 10.1136/bmj.n71

13. MacFarlane A, Russell-Rose T, Shokraneh F. Search strategy formulation for systematic reviews: Issues, challenges and opportunities. *Intell Syst Appl*. 2022;15:200091.

    doi: 10.1016/j.iswa.2022.200091

14. Choi S, Shim S, Shin Y. Neural Network Classifier-Based OPC with Imbalanced Training Data. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2019;38(5):938-948.

    doi: 10.1109/TCAD.2018.2824255

15. Wang D, Xu K, Guo J, Ghiasi S. DSP-Efficient Hardware Acceleration of Convolutional Neural Network Inference on FPGAs. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2020;39(12):4867-4880.

    doi: 10.1109/TCAD.2020.2968023

16. Swartout B, Patil R, Knight K, Russ T. Toward distributed use of large-scale ontologies. In: *Proc 10th Workshop on Knowledge Acquisition for Knowledge-Based Systems*. 1996. https://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html

17. Ramkumar AS, Poorna B. Ontology based semantic search: An introduction and a survey of current approaches. *Proc Int Conf Intell Comput Appl (ICICA)*. 2014:372-376.

    doi: 10.1109/ICICA.2014.82

18. Mönch L, Stehli M. An ontology for production control of semiconductor manufacturing processes. In: Schillo M, Klusch M, Müller J, Tianfield H, eds. *Multiagent System Technologies*. MATES 2003. Lecture Notes in Computer Science, vol 2831. 2003:156-167.

    doi: 10.1007/978-3-540-39869-1_14

19. Ehm H, Ramzy N, Moder P, Summerer C, Fetz S, Neau C. Digital Reference – A Semantic Web for Semiconductor Manufacturing and Supply Chains Containing Semiconductors. *Proc 2019 Winter Simulation Conf (WSC)*. 2019:2409-2418.

    doi: 10.1109/WSC40007.2019.9004831

20. Otte JN, Kiritsis D, Ali MM, *et al.* An ontological approach to representing the product life cycle. *Appl Ontol*. 2019;14(2):179-197.

    doi: 10.3233/AO-190210

21. Wawrzik FW, Grimm CG, Neumann PN. Ontology Module Suite for Electronics and Systems Engineering. IEEE DataPort; 2023.

    doi: 10.21227/88nn-eb43

22. Google LLC. Google Patents. https://patents.google.com/

23. Gajski DD, Kuhn RH. Guest editors' introduction: New VLSI tools. *Computer*. 1983;16(12):11-14.

    doi: 10.1109/MC.1983.1654264

24. Li Y, Wang Y, Li Y, Zhou R, Lin Z. An Artificial Neural Network Assisted Optimization System for Analog Design Space Exploration. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2020;39(10):2640-2653.

    doi: 10.1109/TCAD.2019.2961322

25. Budak AF, Gandara M, Shi W, Pan DZ, Sun N, Liu B. An Efficient Analog Circuit Sizing Method Based on Machine Learning Assisted Global Optimization. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2022;41(5):1209-1221.

    doi: 10.1109/TCAD.2021.3081405

26. Zhang S, Yang F, Yan C, Zhou D, Zeng X. An Efficient Batch-Constrained Bayesian Optimization Approach for Analog Circuit Synthesis via Multiobjective Acquisition Ensemble. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2022;41(1):1-14.

    doi: 10.1109/TCAD.2021.3054811

27. Settaluri K, Liu Z, Khurana R, Mirhaj A, Jain R, Nikolic B. Automated Design of Analog Circuits Using Reinforcement Learning. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2022;41(9):2794-2807.

    doi: 10.1109/TCAD.2021.3120547

28. Ahmadzadeh M, Lappas J, Wehn N, Gielen G. AnaCraft: Duel-Play Probabilistic-Model-Based Reinforcement Learning for Sample-Efficient PVT-Robust Analog Circuit Sizing Optimization. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2026;45(2):901-914.

    doi: 10.1109/TCAD.2025.3582175

29. Thakur S, Ahmad B, Pearce H, *et al.* VeriGen: A Large Language Model for Verilog Code Generation. *ACM Trans Des Autom Electron Syst*. 2024;29(3):46.

    doi: 10.1145/3643681

30. Wu H, He Z, Zhang X, *et al.* ChatEDA: A Large Language

Model Powered Autonomous Agent for EDA. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2024;43(10):3184-3197.

doi: 10.1109/TCAD.2024.3383347

31. Liu B, Zhang H, Gao X, *et al.* LayoutCopilot: An LLM-Powered Multiagent Collaborative Framework for Interactive Analog Layout Design. IE*EE Trans Comput Aided Des Integr Circuits Syst*. 2025;44(8):3126-3139.

doi: 10.1109/TCAD.2025.3529805

32. Shen J, Chen Z, Zhuang J, *et al.* Atelier: An Automated Analog Circuit Design Framework via Multiple Large Language Model-Based Agents. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2026;45(1):31-44.

doi: 10.1109/TCAD.2025.3573228

33. Lin Y, Jiang Z, Gu J, *et al.* DREAMPlace: Deep Learning Toolkit-Enabled GPU Acceleration for Modern VLSI Placement. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2021;40(4):748-761.

doi: 10.1109/TCAD.2020.3003843

34. Xu Q, Geng H, Chen S, *et al.* GoodFloorplan: Graph Convolutional Network and Reinforcement Learning-Based Floorplanning. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2022;41(10):3492-3502.

doi: 10.1109/TCAD.2021.3131550

35. Wu N, Xie Y, Hao C. IronMan-Pro: Multiobjective Design Space Exploration in HLS via Reinforcement Learning and Graph Neural Network-Based Modeling. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2023;42(3):900-913.

doi: 10.1109/TCAD.2022.3185540

36. Cheng C, Ho CT, Lee D, Lin B, Park D. Complementary-FET (CFET) Standard Cell Synthesis Framework for Design and System Technology Co-Optimization Using SMT. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2021;29(6):1178-1191.

doi: 10.1109/TVLSI.2021.3065639

37. Zhou R, Poechmueller P, Wang Y. An Analog Circuit Design and Optimization System With Rule-Guided Genetic Algorithm. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2022;41(12):5182-5192.

doi: 10.1109/TCAD.2022.3166637

38. Dong X, Chen Y, Chen J, *et al.* Worst-case Power Integrity Prediction Using Convolutional Neural Network. *ACM Trans Des Autom Electron Syst*. 2023;28(4):54.

doi: 10.1145/3564932

39. Wei J, Wang H, Zhao T, Jiang YL, Wan J. A New Compact MOSFET Model Based on Artificial Neural Network With Unique Data Preprocessing and Sampling Techniques. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2023;42(4):1250-1254.

doi: 10.1109/TCAD.2022.3193330

40. Zhang J, Liu C, Cheng L, Li X, Li H. Understanding and Mitigating Errors of LLM-Generated RTL Code. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2026;45(10):4876-4889.

doi: 10.1109/TCAD.2026.3661446

41. Yang H, Li S, Deng Z, Ma Y, Yu B, Young EFY. GAN-OPC: Mask Optimization with Lithography-Guided Generative Adversarial Nets. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2020;39(10):2822-2834.

doi: 10.1109/TCAD.2019.2939329

42. Jiang B, Liu L, Ma Y, Yu B, Young EFY. Neural-ILT 2.0: Migrating ILT to Domain-Specific and Multitask-Enabled Neural Network. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2022;41(8):2671-2684.

doi: 10.1109/TCAD.2021.3109556

43. Chen G, Yu Z, Liu H, Ma Y, Yu B. DevelSet: Deep Neural Level Set for Instant Mask Optimization. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2023;42(12):5020-5033.

doi: 10.1109/TCAD.2023.3286262

44. Zhu B, Fei S, Yu BX. INN-ILT: Inverse Lithography Technique via Invertible Neural Network. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2026;45(9):4163-4176.

doi: 10.1109/TCAD.2025.3650094

45. Chiu M, Chen TM. Applying Data Augmentation and Mask R-CNN-Based Instance Segmentation Method for Mixed-Type Wafer Maps Defect Patterns Classification. *IEEE Trans Semicond Manuf*. 2021;34(4):455-463.

doi: 10.1109/TSM.2021.3118922

46. Jang SJ, Kim JS, Kim T, Lee HJ, Ko S. A wafer map yield prediction based on machine learning for productivity enhancement. *IEEE Trans Semicond Manuf*. 2019;32(4):400-407.

doi: 10.1109/TSM.2019.2945482

47. Ciccazzo A, Pillo GD, Latorre V. A SVM surrogate model-based method for parametric yield optimization. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2016;35(7):1224-1228.

doi: 10.1109/TCAD.2015.2501307

48. Seo J, Hwang HS, Park SY, *et al.* Deriving optimal atomic layer deposition process conditions using machine learning. *J Ind Inf Integr*. 2025;47:100879.

doi: 10.1016/j.jii.2025.100879

49. Cheng KCC, Chen LLY, Li J, *et al.* Machine Learning-Based Detection Method for Wafer Test Induced Defects. *IEEE Trans Semicond Manuf*. 2021;34(2):161-167.

doi: 10.1109/TSM.2021.3065405

50. Huang L, Song T, Jiang T. Linear regression combined KNN

algorithm to identify latent defects for imbalance data of ICs. *Microelectron J*. 2023;131:105641.

doi: 10.1016/j.mejo.2022.105641

51. Kao SX, Chien C. Deep Learning-Based Positioning Error Fault Diagnosis of Wire Bonding Equipment and an Empirical Study for IC Packaging. *IEEE Trans Semicond Manuf*. 2023;36(4):619-628.

doi: 10.1109/TSM.2023.3243775

52. Vaisband I, Friedman EG. Energy efficient adaptive clustering of on-chip power delivery systems. *Integration*. 2015;48(1):1-9.

doi: 10.1016/j.vlsi.2014.06.003

53. Chatterjee S, Roy SK, Giri C, Rahaman H. Frequency-scaled thermal-aware test scheduling for 3D ICs using machine learning based temperature estimation. *Microelectron J*. 2022;128:105535.

doi: 10.1016/j.mejo.2022.105535

54. Dewan MI, Lin SED, Kim DH. Construction of All Multilayer Monolithic RSMTs and Its Application to Monolithic 3D IC Routing. *ACM Trans Des Autom Electron Syst*. 2024;29(1):17.

doi: 10.1145/3626958

55. Vanna-Iampikul P, Lu Y, Shim DE, Lim SK. GNN-based Multi-bit Flip-flop Clustering and Post-clustering Design Optimization for Energy-efficient 3D ICs. *ACM Trans Des Autom Electron Syst*. 2023;28(5):76.

doi: 10.1145/3588570

56. Wu JS, Pan CA, Liu YY. ILP-based Substrate Routing with Mismatched Via Dimension Consideration for Wire-bonding FBGA Package Design. *ACM Trans Des Autom Electron Syst*. 2023;28(5):77.

doi: 10.1145/3579843

57. Chen X, Ju Y, Ogrenci S, Gu J. A Physics-Informed Neural Network Surrogate for Runtime PDN and Dynamic Droop Prediction in 2.5-D Chiplet Integration. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2026;34(4):1075–1086.

doi: 10.1109/TVLSI.2025.3650633

58. Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach*. 4th ed. Pearson; 2020. http://lib.ysu.am/disciplines_bk/efdd4d1d4c2087fe1cbe03d9ced67f34.pdf

59. Yan H, Wang Y, Gao P, *et al.* A Lightweight Heterogeneous Graph Embedding Framework for Hotspot Detection. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2025;44(9):3479-3489.

doi: 10.1109/TCAD.2025.3543436

60. Yasaei R, Faezi S, Faruque MAA. Golden Reference-Free Hardware Trojan Localization Using Graph Convolutional Network. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2022;30(10):1401-1411.

doi: 10.1109/TVLSI.2022.3191683

61. Yasaei R, Chen L, Yu SY, Faruque MAA. Hardware Trojan Detection Using Graph Neural Networks. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2025;44(1):25-38.

doi: 10.1109/TCAD.2022.3178355

62. Popryho Y, Pal D, Partin-Vaisband I. NetVGE: Netwise Hardware Trojan Detection at RTL Using Variable Dependency and Knowledge Graph Embedding. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2025;44(12):4667-4678.

doi: 10.1109/TCAD.2025.3569492

63. Utyamishev D, Partin-Vaisband I. Netwise Detection of Hardware Trojans Using Scalable Convolution of Graph Embedding Clouds. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2024;43(10):3116-3128.

doi: 10.1109/TCAD.2024.3383348

64. Chen L, Dong C, Wu Q, *et al.* GNN4HT: A Two-Stage GNN-Based Approach for Hardware Trojan Multifunctional Classification. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2025;44(1):172-185.

doi: 10.1109/TCAD.2024.3428469

65. Stern A, Botero UJ, Rahman F, Forte D, Tehranipoor M. EMFORCED: EM-Based Fingerprinting Framework for Remarked and Cloned Counterfeit IC Detection Using Machine Learning Classification. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2020;28(2):363-375.

doi: 10.1109/TVLSI.2019.2949733

66. Narwariya AS, Kalanadhabhatta S, Acharyya A. Leveraging IO Pad Protection Diodes for Recycled IC Detection and Age Estimation Using Polynomial Regression. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2025;33(11):3166-3175.

doi: 10.1109/TVLSI.2025.3590317

67. Mao M, Wang W, Lu C, Jia F, Long X. Machine learning for board-level drop response of BGA packaging structure. *Microelectron Reliab*. 2022;134:114553.

doi: 10.1016/j.microrel.2022.114553

68. Ghimire A, Alkurdi M, Mohanty SP, Amsaad F. AI-enabled image processing approach for efficient clustering and identification of hardware Trojans. *Integration*. 2026;107:102628.

doi: 10.1016/j.vlsi.2025.102628

69. de Jong SDM, Ghezeljehmeidan AG, van Driel WD. Solder joint reliability predictions using physics-informed machine learning. *Microelectron Reliab*. 2025;172:115797.

doi: 10.1016/j.microrel.2025.115797

70. Wu Z, Chai Z, Liu B, Liu M. Neural network approach to NBTI/HCD coupled failure analysis of MOS ICs.

Microelectron Reliab. 2026;177:115996.

doi: 10.1016/j.microrel.2025.115996

71. Rahmani K, Ray S, Mishra P. Postsilicon Trace Signal Selection Using Machine Learning Techniques. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2017;25(2):570-580.

doi: 10.1109/TVLSI.2016.2593902

72. Bathla S, Vasudevan V. A Framework for Reliability Analysis of Combinational Circuits Using Approximate Bayesian Inference. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2023;31(4):543-554.

doi: 10.1109/TVLSI.2023.3237885

73. Elnaggar R, Chakrabarty K, Tahoori MB. Hardware Trojan Detection Using Changepoint-Based Anomaly Detection Techniques. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2019;27(12):2706-2719.

doi: 10.1109/TVLSI.2019.2925807

74. Pittino F, Diversi R, Benini L, Bartolini A. Robust Identification of Thermal Models for In-Production High-Performance-Computing Clusters with Machine Learning-Based Data Selection. *IEEE Trans Comput Aided Des Integr Circuits Syst*. 2020;39(10):2042-2054.

doi: 10.1109/TCAD.2019.2950378

75. Hou T, Zhen P, Ji Z, Chen H. A Deep Learning Framework for Solving Stress-based Partial Differential Equations in Electromigration Analysis. *ACM Trans Des Autom Electron Syst*. 2023;28(4):57.

doi: 10.1145/3567424

76. Benmahdjoub MA, Mohammed M, Trabelsi M, *et al.* Hybrid junction temperature prediction of IGBTs combining detailed electro-thermal modelling and deep learning regression. *Microelectron J*. 2026;172:107133.

doi: 10.1016/j.mejo.2026.107133

77. Hamzat AK, Murad MdS, Subeshan B, Asmatulu R, Asmatulu E. Rare earth element recycling: a review on sustainable solutions and impacts on semiconductor and chip industries. *J Mater Cycles Waste Manag*. 2025;27(5):3009-3032.

doi: 10.1007/s10163-025-02276-7

78. European Parliament and Council of the European Union. Regulation (EU) 2024/1781 of 13 June 2024 establishing a framework for the setting of ecodesign requirements for sustainable products and repealing Directive 2009/125/EC (ESPR). Published online June 2024. Accessed April 23, 2026. http://data.europa.eu/eli/reg/2024/1781/oj

79. Lu Q, Zhu D, Li M. Research on Remaining Useful Life Prediction of Bearing with Digital Twin and TSMixer. *Appl Sci*. 2026;16(13):6762.

doi: 10.3390/app16136762

80. European Parliament and Council of the European Union. Regulation (EU) 2023/1781 of 13 September 2023 establishing a framework of measures for strengthening Europe's semiconductor ecosystem and amending Regulation (EU) 2021/694 (Chips Act). Published online September 2023. Accessed April 22, 2026. http://data.europa.eu/eli/reg/2023/1781/oj

81. Noyce RN. Microelectronics. *Sci Am*. 1977;237(3):62-69.

doi: 10.1038/scientificamerican0977-62

82. Moore GE. Cramming more components onto integrated circuits. *Proc IEEE*. 1998;86(1):82-85.

doi: 10.1109/JPROC.1998.658762

83. Khan HN, Hounshell DA, Fuchs ERH. Science and research policy at the end of Moore's law. *Nat Electron*. 2018;1(1):14-21.

doi: 10.1038/s41928-017-0005-9