TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Yuan, C., Fan, X., Liu, X., & Zhang, G. Q. (2026). Artificial Intelligence (AI) for Semiconductor Packaging. *IEEE Access*, *14*, 155421-155432. https://doi.org/10.1109/ACCESS.2026.3740841

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 September 2026, accepted 30 September 2026, date of publication 6 October 2026,
date of current version 14 October 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3740841

PERSPECTIVE

Artificial Intelligence (AI) for Semiconductor Packaging

CADMUS CHANG-ANN YUAN[1], (Member, IEEE), XUEJUN FAN[2], (Fellow, IEEE), XINYU LIU[3], AND GUOQI ZHANG[4], (Fellow, IEEE)

[1]Department of Mechanical and Computer-Aided Engineering, Feng Chia University, Taichung 407102, Taiwan
[2]Department of Mechanical Engineering, Lamar University, Beaumont, TX 77710, USA
[3]Department of Industrial and System Engineering, Lamar University, Beaumont, TX 77710, USA
[4]Department of Microelectronics, Delft University of Technology, 2628 CD Delft, The Netherlands

Corresponding authors: Cadmus Chang-Ann Yuan (cayuan@o365.fcu.edu.tw) and Xuejun Fan (xuejun.fan@lamar.edu)

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant NSTC 115-2221-E-035-052.

ABSTRACT Artificial intelligence is increasingly being applied to semiconductor packaging design, manufacturing, and reliability. However, most existing models remain isolated and problem-specific surrogates trained on small and narrowly defined datasets with limited numbers of inputs and outputs. This perspective presents a strategic roadmap toward scalable, transferable, and adaptive intelligence for semiconductor packaging. We identify the scarcity, fragmentation, and inconsistent representation of diverse data as the central barriers to AI adoption in semiconductor packaging. To address these challenges, we call for a coordinated global effort to establish shared data infrastructure with standardized representations, comprehensive metadata, traceable provenance, and quantified uncertainty. We propose simulation as a data factory, using consistent multi-fidelity simulations complemented by experimental measurements, manufacturing observations, and field data to systematically expand the scale and coverage of semiconductor packaging datasets. Building on this infrastructure, we envision a pre-trained, modular, and extensible foundation model that learns transferable representations and adapts to downstream tasks across co-design, manufacturing, and reliability. Ultimately, this roadmap seeks to advance semiconductor packaging AI toward transferable, scalable, adaptive, and continuously validated engineering intelligence.

INDEX TERMS Artificial intelligence (AI), data, foundation model, semiconductor packaging, simulation.

I. INTRODUCTION

Over the past decade, rapid advances in artificial intelligence (AI) have reshaped scientific research, industrial practice, and numerous aspects of everyday life. This progress has been driven by the coordinated advancement of three foundational pillars: data, models, and compute, as illustrated in Fig. 1. Large and diverse datasets provide the information required for learning, mathematical models extract and organize patterns from these data, and computing resources enable increasingly complex training and inference. Sustained progress in AI therefore depends on the synergistic development of all three pillars [1], [2].

The modern deep learning era was catalyzed by the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2012) [3], [4]. AlexNet provided a landmark demonstration of the synergy among data, models, and compute [4]. By combining a deep convolutional neural network with a large, manually annotated image dataset and graphics processing units (GPUs) for parallel computation, it substantially outperformed decades of conventional hand-crafted feature-based approaches in image recognition.

Subsequently, large language models (LLMs) have ushered in a new era of AI. By combining internet-scale datasets, Transformer-based architectures [5], and massive computing clusters, LLMs have demonstrated that continuous scaling of data, models, and compute leads to exponential leaps in artificial intelligence [6]. More recently, LLM-based systems have begun to evolve beyond text generation toward agentic AI [7]. These systems are increasingly designed to perform reasoning-like operations, plan multistep tasks, use external tools, and interact with digital environments with varying

The associate editor coordinating the review of this manuscript and approving it for publication was Zhiguo Qian.

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

degrees of autonomy. AI is also accelerating into robotics and embodied intelligence [1], [8]. In this frontier, physical AI platforms and robot foundation models [8] are pushing digital intelligence out of virtual environments and into the physical world. This transition, moving from digital pattern recognition to physical interaction and autonomous problem-solving, represents a critical evolution for engineering disciplines that deal with physical manufacturing and hardware.

Underpinning this entire AI revolution is the rapid advancement of semiconductor technologies. GPUs and specialized accelerators have provided the essential computational foundation for modern AI by enabling massive parallel processing capabilities. As conventional transistor scaling under Moore's law encounters increasing physical and economic constraints, advanced packaging has become a critical approach to sustaining system-level performance growth. Emerging technologies including 2.5D-based interposer, 3D stacking, hybrid bonding, micro-bumps, through-silicon vias (TSVs), co-packaged optics, and chiplet-based architectures collectively enable high-density heterogeneous integration at the system-in-package scale [9]. At the same time, this shift toward greater integration complexity creates major challenges. As highlighted in Fig. 1, co-design, smart manufacturing, and reliability represent three interconnected challenge domains in modern semiconductor packaging [10].

To address these complex bottlenecks, a rapidly growing body of research, now comprising several hundred publications, has explored the application of artificial intelligence and machine learning to semiconductor packaging [11]. These studies span a broad range of domains, including signal and power integrity, thermal management, structural analysis, manufacturing process optimization, defect detection, and reliability prediction. Digital twins [12] and multiphysics co-design frameworks [13] have also been proposed to connect package design, manufacturing data, and lifecycle performance. However, AI adoption in semiconductor packaging remains at an early and highly fragmented stage. Nearly all existing studies are based on narrowly defined datasets in which only a small number of design, material, or processing parameters are varied, often within limited ranges. Although these problem-specific models can achieve satisfactory predictive accuracy within their training domains and provide useful computational acceleration, their performance largely reflects interpolation within narrowly sampled parameter spaces. Consequently, they cannot adequately capture the variability in package architectures, material systems, manufacturing conditions, loading conditions, and failure mechanisms. More importantly, they rarely demonstrate generalization beyond the specific problems for which they were developed, within or across different package configurations, process conditions, material systems, loading conditions, and failure modes. Thus, despite the rapidly increasing number of publications, current AI applications in semiconductor packaging remain predominantly isolated, problem-specific demonstrations rather than scalable and generalizable engineering intelligence.

Motivated by these limitations, this perspective moves beyond a conventional literature review and presents a strategic framework for developing scalable, transferable, and physics-informed AI for semiconductor packaging [14]. First, we examine the technological evolution of semiconductor packaging and identify the need for standardized representations of package architecture, materials, processes, and failure mechanisms. Second, we argue that the principal barrier to broader AI adoption is not simply the availability of algorithms, but the scarcity, fragmentation, and proprietary nature of high-quality data. We therefore discuss strategies for developing shared data infrastructures and benchmark datasets, emphasizing coordinated multi-fidelity, physics-based simulation as a scalable source of training data.

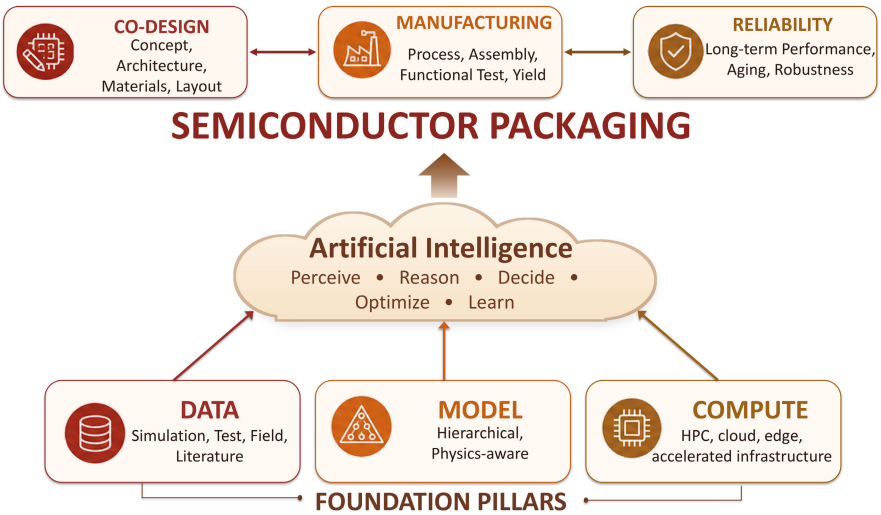


FIGURE 1. The three foundational pillars of AI (data, model, and compute) and the vertical applications in semiconductor packaging.

Finally, inspired by the development of foundation models in other fields, we envision a packaging foundation model capable of learning transferable representations from heterogeneous data and adapting to multiple domains in co-design, manufacturing, and reliability. Through this framework, the perspective provides a roadmap toward generalizable, adaptive, and trustworthy AI for semiconductor packaging.

II. SEMICONDUCTOR PACKAGING TECHNOLOGY EVOLUTION

Semiconductor packaging serves as the bridge that connects nanometer-scale on-chip transistors to the macroscopic printed circuit board (PCB). Since the inception of the integrated circuit (IC) in the late 1950s, packaging has been an integral and indispensable step in the manufacturing of every semiconductor device [10]. Far from being a mere passive protective enclosure, electronic packaging defines the ultimate electrical, thermal, and mechanical interface of the chip(s). It translates massive densities of microscopic input/output (I/O) signals into usable macroscopic connections, orchestrates the dissipation of extreme heat fluxes, and shields silicon structures from environmental and mechanical degradation.

Over the past decades, as summarized in Fig. 2 [14], packaging has undergone a continuous evolution of steadily increasing interconnect density, integration complexity, and functional heterogeneity. Wire-bonding and lead-frame technologies developed in the 1950s and 1960s remain widely utilized today as legacy technologies due to their cost-effectiveness, high reliability, and manufacturing maturity. The 1970s witnessed a revolutionary advancement with the introduction of surface mount technology (SMT), which fundamentally transitioned board-level connections to reflow soldering processes. Subsequently, the 1990s marked a golden era for electronic packaging. During this decade, the emergence of ball grid arrays (BGA) shifted interconnects from the chip's perimeter to its entire surface area. Concurrently, the commercialization of flip-chip technology and the development of wafer-level packaging (WLP) laid the groundwork for the advanced packaging of the 21st century. As the industry progressed into the 2000s and beyond, technologies such as fan-out wafer-level packaging (FOWLP), silicon interposers, TSVs and bridges, and 3D stacking propelled semiconductor packaging into an entirely new stage. Today, in the 2020s, advanced packaging has reached unprecedented levels of sophistication, driven by 2.5D/3D architectures, heterogeneous chiplet integration, and co-packaged optics.

At its core, the history of semiconductor packaging is the evolution of interconnect technologies [10], [14]. From early wire bonding to C4 (Controlled Collapse Chip Connection) bumping, redistribution layers (RDL), plated through holes (PTH), through silicon vias (TSVs), micro-bumps, embedded silicon bridges, and advanced hybrid bonding, each successive interconnect generation has expanded bandwidth, shrunk pitch, reduced parasitic resistance and capacitance delays, and enabled new integration topologies. Through a systematic combination of these interconnect methods, together with continuous innovations in substrate materials, die-attach films, underfills, dielectrics, molding compounds, and thermal interface materials, packaging engineers have continually bridged the widening gap between transistor scaling at the silicon level and the physical constraints of printed circuit boards.

However, as advanced packaging transitions into the era of high-density heterogeneous integration, it fundamentally transforms into a tightly coupled, multiphysics system-level engineering problem over an expanding design space with limited prior experience [10]. Accelerated by the escalating computational demands of modern AI and unrelenting time-to-market pressure, this technological transition has exposed critical engineering bottlenecks that demand rapid screening of candidate designs. As illustrated in Fig. 1, these challenges are concentrated within three core domains:

Co-Design (Concept, Architecture, Materials, Layout): Designing a modern package requires the concurrent optimization of conflicting multiphysics objectives. Engineers must simultaneously balance signal and power integrity, thermal management, structural integrity, electromigration, and, where applicable, high speed optical communication. For example, modifying a substrate layout to improve signal routing may inadvertently exacerbate thermal hotspots or induce mechanical stress, rendering traditional sequential design methodologies increasingly ineffective [13].

Smart Manufacturing (Process, Assembly, Functional Test, Yield): Advanced integration involves dozens of heterogeneous fabrication steps, from precise wafer thinning to sub-micron hybrid bonding. Maintaining high yield requires adaptive process window control, rigorous equipment qualification, and intelligent defect detection. Consequently, the manufacturing floor must transition from static recipe execution to dynamic, feedback-driven assembly [12].

Reliability (Long-term Performance, Aging, Robustness): Packaged devices are subjected to a complex spectrum of operational and environmental stresses throughout their assembly process and prolonged field operation. Reliability is governed by interactions among multiple physical domains, including thermal cycling, thermomechanical loading, mechanical shock and vibration, current stressing, and humidity or moisture exposure. Ensuring long-term robustness against diverse degradation mechanisms, ranging from structural fatigue and material creep to electromigration and moisture-induced delamination, demands integrated risk assessment across design margins, material property selection, and mission-critical operating environments [15], [16], [17].

Interconnect technologies provide a natural and physically grounded organizing backbone for representing semiconductor packaging systems. Each interconnect type, whether a wire bond, solder bump, through silicon via, micro-bump, embedded bridge, or hybrid bond, serves as a structural anchor that links package architectures with compatible

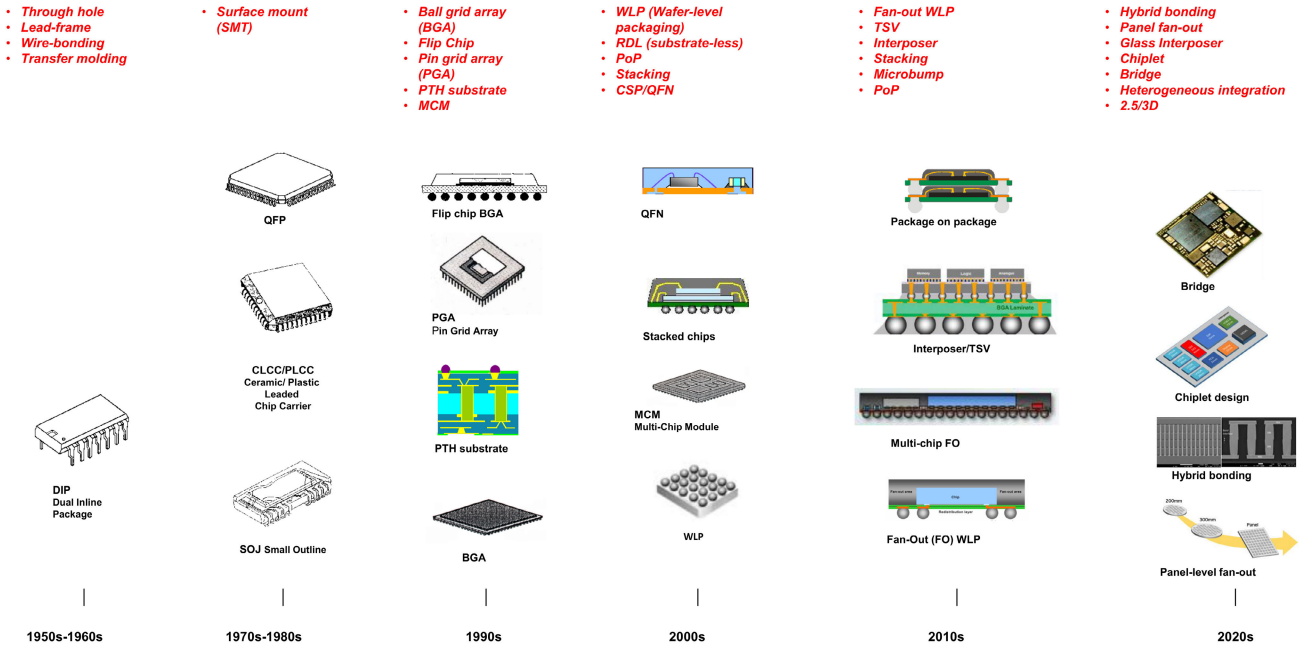


FIGURE 2. Evolution of semiconductor packaging technologies.

material systems, fabrication processes, geometric attributes, loading conditions, and potential failure mechanisms. Organizing packaging data according to interconnect types and integration hierarchies therefore reflects the physical construction of packaging systems and provides an interpretable basis for data annotation, comparison, and model development [14].

Interconnects alone, however, cannot fully represent the complexity of advanced packaging. A comprehensive data framework must be multi-dimensional and relational, connecting the structural hierarchy of interconnects with package architectures, geometric features, material properties, manufacturing process histories, operating and environmental conditions, measured responses, and degradation mechanisms. Such a framework should also preserve relationships across multiple scales, from local material interfaces and interconnects to package-, module-, and system-level behavior. Establishing this scalable, extensible, and physics-informed representation is an essential step toward transforming fragmented packaging data into reusable engineering knowledge and, ultimately, supporting AI models that can generalize across architectures, processes, materials, and reliability problems.

III. DATA SCARCITY: THE CENTRAL BOTTLENECK

A. DATA SCARCITY

The physically informed representation outlined in Section II defines how semiconductor packaging data can be structured. Furthermore, developing transferable and scalable AI requires data at sufficient scale and diversity. For computer vision community, when Caltech-101 was introduced in 2004, it contained approximately 9,000 images distributed across 101 object categories and represented one of the leading image-recognition benchmarks of its time [18]. At this scale, however, deep convolutional networks had not yet demonstrated a decisive advantage, and conventional approaches combining handcrafted visual features with classifiers such as support vector machines remained highly competitive [19], [20], [21]. A fundamental change occurred with the introduction of ImageNet in 2009. The full ImageNet database eventually expanded to more than 14 million annotated images organized into over 20,000 categories according to the WordNet hierarchy [3], [22]. This represented an increase of more than three orders of magnitude in the number of images and approximately two orders of magnitude in the number of categories relative to Caltech-101. When a large ImageNet subset was used in the 2012 ImageNet Large Scale Visual Recognition Challenge, AlexNet combined large-scale data with a deep convolutional architecture and GPU computing to substantially outperform conventional handcrafted feature-based approaches [4]. This breakthrough resulted from the synergy among data, models, and compute. Nevertheless, the expansion in dataset scale was indispensable: it provided the diversity and coverage needed for deep neural networks to learn hierarchical visual representations directly from data instead of fitting only a small number of predefined categories.

Against this benchmark, the datasets currently available for semiconductor packaging remain extremely limited. Although recent reviews document a rapid increase in machine learning applications involving signal and power integrity, thermal management, structural analysis,

manufacturing, and reliability prediction [11], the underlying datasets are generally small, low-dimensional, and confined to individual structures or tasks. As summarized in Table 1, even the comparatively large datasets reported in the literature cover only narrowly defined physical domains. For example, Lau and Chiang [15] reported approximately 9,000 finite element design cases for fatigue life prediction in a wafer-level package, but the entire design space was defined by only four geometric input parameters and a single scalar output. The multi-objective flip-chip co-design study by Rafiee et al.

TABLE 1. Representative machine-learning applications and dataset characteristics in semiconductor packaging.

Study	Method (data source)	Application and physics	Input / Output features	Dataset size and sampling unit	Reported metric
[13] Rafiee & Hasheminasab (2026)	RF + XGBoost surrogate → NSGA-II (FEA + full-wave EM data)	Flip-chip package co-design; thermal resistance, warpage, ELK stress, bump & BGA fatigue, insertion loss	13 / 6	500 independent design cases (LHS)	R^2 (test) 0.42–1.00 across six objectives
Lau & Chiang [15]; Panigrahy et al. [16]	ANN / RNN / SVR / KRR / KNN / RF (FEA + Coffin–Manson data)	Wafer-level packaging (WLP) solder joint; thermal-cycling reliability life	4 / 1	up to 9,000 FEM design cases (4 design parameters; datasets swept 256→9,000) [15]; earlier study 576 & 1,296 [16]	R^2 0.97–0.996 (train), 0.94–0.96 (cross-val); MAE < 20 cycles
[23] Wu & Wang (2023)	Fully-connected NN, 2 hidden layers (FEA data)	2.5D/3D RDL; anisotropic equivalent (elastic) material properties	15×15-pixel image / 15	~1,000 FEA-labeled blocks (one 21.6 mm RDL, sliced 200×200)	Reaction-force error 1.72–2.81% (best 0.076%); no R^2
[24] Mondal & Suhling (2024)	LSTM, 2 hidden layers × 50 (DAMASK simulation)	SAC305 solder joint; orientation-dependent stress–strain	sequence / curve	10,000 simulated stress–strain curves (one per single-crystal orientation); one SAC305 solder, 26 points/curve	R^2 0.974, RMSE 1.16 MPa, MAE 0.87 MPa
[25] Kim & Lee (2025)	MDP + P2R heuristic (deep RL as baseline); linear regression	Chiplet placement & routing; signal integrity (eye diagram)	n.r. / 1	Not reported	Eye opening 0.869 UI (P2R); no R^2
[26] Shanmugam & Maganti (2024)	ANN / XGBoost / LightGBM / KNN (experimental data)	Microchannel heat sink; thermal performance under non-uniform heat load	n.r. / 6	560 experimental data points (2 heat-sink designs: 279 + 288)	XGBoost mean R^2 0.98, MAE 2.1; all models R^2 0.83–0.99
[27] Rafiee & Agin (2026)	RF + XGBoost surrogate → NSGA-II (3D FEA data)	2.5D flip-chip (Si/glass interposer); thermal resistance, solder strain energy	29 / 4	500 independent design cases (400 train / 100 test)	R^2 (test) 0.61–1.00
[28] Cheng & Ma (2023)	ANN (FEA data; validated vs Shadow-Moiré)	FCCSP; process-induced warpage	6 / 1	216–810 FEM design cases (subdomain sampling) + 64 validation	Test MSE 8.0±2.9 μm; no R^2
[29] Majd et al. (2020)	Deep sequence models: MLFF/LSTM/GRU/RNN/CNN/SLFF (experimental)	Packaged chip; in-situ stress distribution (delamination prognostics)	sequence / field	4,450 / 4,444 correlated measurements (2 test vehicles / TQFP packages, 700 cycles)	MSE (cost) ~0.04; no R^2/RMSE
[30] Yao & Yang (2022)	Physics-based nested ANN (FEM data)	FOWLP; post-cure reliability (creep strain, von Mises stress, warpage)	n.r. / 3	59 EMC material sets (experimental DMA inputs)	MAPE 4.99–14.89% / R^2 0.78–0.95
[31] Yuan, Hong & Li (2026)	Ensemble-DNN AI surrogate (GA + K-PCA; FEM data)	PBGA solder joint; fatigue-life prediction	3 / 1	64 FEM design cases (4^3 full factorial)	RMSE ~12–24 cycles (val. error <1%)
Yuan, Fan & Fan (2020) [32]	Gated neural network (two-step; integrable with ANN/RNN/LSTM); experimental aging data	LED chip packaging & luminaire; spectral power distribution under thermal-aging degradation	n.r. / spectrum	30 packaging cases (5 temperatures × 6 currents; 9 train / 21 val)	CCT diff ≤ 2.6%, colour-coordinate diff ≤ 1.0%; LSTM-gated best
Yuan (2022) [33]	Hybrid offline/online adaptive surrogate (NN root models + GA; online update); experimental data	UVLED module in-service response; radiation-power reduction over time (degradation / reliability)	sequence / sequence	3 sets × 14 UVLED specimens (online window n = 3)	Prediction-capability index PCp (avg. difference < 2)

Reported metrics are not directly comparable across studies, which use different tasks, data-partitioning strategies, and validation protocols. n.r., not reported.
Abbreviations: ANN, artificial neural network; CNN, convolutional neural network; DNN, deep neural network; FEA/FEM, finite element analysis/method; GA, genetic algorithm; GRU, gated recurrent unit; KNN, k-nearest neighbours; K-PCA, kernel principal component analysis; KRR, kernel ridge regression; LSTM, long short-term memory; MAE, mean absolute error; MAPE, mean absolute percentage error; MDP, Markov decision process; MSE, mean squared error; NSGA-II, non-dominated sorting genetic algorithm II; R^2, coefficient of determination; RF, random forest; RL, reinforcement learning; RMSE, root-mean-square error; RNN, recurrent neural network; SPD, spectral power distribution; SVR, support vector regression; UI, unit interval; XGBoost, extreme gradient boosting.

spanned only 13 input parameters across only 500 design points [13]. Wu et al. [23] generated approximately 1,000 finite element-labeled samples from local copper distributions within a single RDL structure. Mondal and Suhling [24] generated 10,000 simulated stress-strain curves for single crystal SAC305 solder, but all curves represented the same material and differed only in crystallographic orientation.

In contrast, even a relatively simple semiconductor package involves dozens of coupled variables, while advanced heterogeneous packages may require the representation of hundreds of geometric, material, process, layout, loading, environmental, and state variables. The semiconductor packaging community currently lacks even a Caltech-101-like benchmark: a public or broadly accessible, standardized, and sufficiently diverse dataset that supports reproducible model development and comparison across research groups.

This severe data scarcity is deeply rooted in the physical and commercial realities of the microelectronics ecosystem. Unlike internet texts or images, which can be collected and annotated at very large scales, experimental packaging data are expensive and time consuming to generate. High-fidelity manufacturing records and field failure data are also sparse, fragmented, and frequently protected by confidentiality and intellectual property restrictions. Data describing a single package lifecycle, spanning design, fabrication, assembly, testing, accelerated qualification, field operation, and failure analysis, are often distributed across integrated device manufacturers, fabless design houses, outsourced semiconductor assembly and test providers, material and equipment suppliers, and end users. Even when data are available, they frequently lack standardized metadata, complete process histories, quantified uncertainty, negative or failed cases, and consistent definitions of inputs and outputs. Failure datasets are particularly imbalanced because successfully operating units vastly outnumber observed failures, while destructive characterization often prevents repeated measurements on the same specimen. Thus, the challenge is not data volume alone, but also insufficient physical diversity, inconsistent representation, limited traceability, restricted accessibility, and uneven data quality.

B. SIMULATION AS A DATA FACTORY

Recent advances in embodied AI and robotics offer a useful blueprint for addressing this challenge. Like semiconductor packaging, robotics deals with complex physical interactions and faces severe natural data scarcity. To understand this challenge, it is helpful to contrast general robotics with autonomous driving. Autonomous driving benefits from the continuous generation of synchronized visual, spatial, and motion data during real world vehicle operation [34], [35]. General purpose robots, however, do not have an equivalent source of physical interaction data. Internet videos can provide useful visual representations and demonstrations, but they rarely contain synchronized information [36], [37]. Consequently, internet-scale visual data alone are insufficient for learning many physical interactions.

To address this limitation, the robotics community shifted from passively searching for data to actively manufacturing it through a ''data factory'' approach [8]. Human teleoperation provides synchronized observations, actions, and system responses [38], while cross-institutional initiatives such as Open X-Embodiment aggregate heterogeneous robotic experiences into shared training resources [39]. These efforts have improved cross-task and cross-platform transfer, but physical data collection remains slow, expensive, and difficult to scale. Robotics has therefore increasingly incorporated physics-based simulation into the data generation pipeline. Platforms such as Isaac Gym allow large numbers of physical interactions to be simulated in parallel, with systematic variations in objects, environments, actions, and operating conditions [40]. Although simulated data cannot perfectly reproduce physical reality, simulation provides scale, controllability, repeatability, and access to complete state information. When calibrated and combined with physical observations, it becomes a powerful mechanism for expanding the coverage and diversity of training data. Training on these artificially engineered datasets has dramatically increased success rates across diverse robotic tasks, convincing researchers that with enough structured physical data, embodied AI can generalize broadly across new tasks and embodiments [8].

The central lesson is directly applicable to semiconductor packaging: when large-scale physical datasets cannot be naturally accumulated, they must be systematically constructed through coordinated data generation strategies. We therefore propose ''simulation as a data factory,'' a shift from isolated, problem-specific analyses toward automated, systematic, consistent, and large-scale computational data generation. Such a data factory would integrate finite element analysis, computational fluid dynamics, computational electromagnetics, multiscale modeling, physics-based analytical formulations, and validated reduced-order models. Rather than relying on exhaustive parameter sweeps, which become computationally infeasible in high-dimensional design spaces, the data factory should combine design of experiments methods, space filling sampling, active learning, uncertainty-guided refinement [41], physics-informed neural network with operators [42], finite element-integrated neural network [43], Latin hypercube sampling and rare event sampling [27]. These methods would enable systematic exploration of package architectures, interconnect technologies, geometric configurations, material systems, manufacturing process histories, operating and environmental conditions, and failure mechanisms. In this manner, simulation would no longer serve only as a tool for analyzing individual designs, but also as a scalable infrastructure for producing the diverse, physically grounded data required to train transferable AI models.

Because generating foundation model-scale datasets through high-fidelity multiphysics simulation would be computationally prohibitive, the data factory should adopt a multi-fidelity hierarchy. Large volumes of lower-cost data

can be generated using analytical solutions, reduced-order models, simplified geometries, and coarse numerical simulations to provide broad coverage of the design space and physically consistent trends. Smaller sets of high-fidelity simulations can then resolve strongly coupled, localized, and nonlinear behavior. Experimental measurements, manufacturing observations, and field data should be used to calibrate and validate the multi-fidelity. Importantly, every data record should include its governing assumptions, constitutive models, boundary and loading conditions, numerical resolution, calibration status, fidelity level, provenance, and uncertainty estimates. This metadata is essential for preventing model and simulation biases from being silently propagated into subsequent AI models.

An important part of the data infrastructure is the dataset for package substrate, which serves as the electrical, mechanical, and thermal platform for semiconductor packaging. Across the semiconductor industry, thousands of substrate designs have been developed for processors, memory devices, networking chips, mobile electronics, automotive, and high performance computing components. Each design embodies a unique combination of routing topology, layer stack-up, geometrical dimensions, power distribution architecture, and manufacturing constraints. Yet these designs are currently dispersed across individual organizations and are rarely organized into a structured research resource. A coordinated community effort to construct a large-scale substrate design database could therefore provide a valuable starting point for packaging AI. Once sufficiently large, such a database would support a wide range of investigations, including pattern discovery, design clustering, representation learning, manufacturability assessment, performance prediction, and design space exploration. More importantly, these substrate designs could serve as the basis for automated electrical, thermal, mechanical, and reliability simulations conducted through the proposed data factory framework, generating vast amounts of physically consistent multiphysics data. Simulations of different physical domains differ in their assumptions, boundary and loading conditions, and spatial and temporal scales. Because these conditions depend on the same design, manufacturing process, and operating point, they are correlated rather than independent. This correlation is ensured in the training data by deriving the conditions of every domain from a consistent description of the design and its physical state.

This massive data generation requires a meticulously structured, systematically organized, and self-consistent framework. We therefore call for cross-organization and international effort to establish an open, shared pre-training database. Constructing a database of this scale and scope will require sustained international collaboration among universities, research institutes, semiconductor manufacturers, outsourced semiconductor assembly and test providers (OSATs), material and equipment suppliers, software developers, and standards organizations.

A practical strategy in developing a shared data platform is to establish a sequence of specifically defined but extensible pilot projects. An initial project could focus on a single, well-characterized failure mechanism, such as solder joint fatigue under thermal cycling or package warpage. Common definitions of inputs, outputs, metadata, validation procedures, and uncertainty measures would enable the resulting dataset to serve as a reproducible benchmark. The same data structure and collaboration framework can then be expanded to additional failure modes and physical domains. In this manner, a series of focused international pilot projects could gradually evolve into a shared, multidomain pre-training resource for semiconductor packaging.

A scalable data ecosystem also requires a practical mechanism for balancing community sharing with industrial confidentiality. Rather than relying on a single fully centralized database, the packaging community could establish a shared pre-training resource containing open or appropriately anonymized simulation data, benchmark problems, standardized metadata, and selected experimental measurements. Individual organizations could then adapt pre-trained models using proprietary design, manufacturing, qualification, and field data without disclosing sensitive information. Federated learning, secure data environments, and standardized model-evaluation protocols may provide additional pathways for extracting collective value from distributed data while preserving intellectual property. Ultimately, the objective of the proposed data ecosystem extends beyond the creation of larger datasets. The goal is to establish a sufficiently broad and physically grounded pre-training resource from which transferable representations of semiconductor packaging can emerge. A large-scale semiconductor packaging data infrastructure could enable the transition from isolated surrogate models to foundation models capable of adapting across design, manufacturing, and reliability applications.

IV. PARADIGM SHIFT TOWARD A FOUNDATION MODEL

A broad range of machine learning algorithms have been explored for semiconductor packaging applications across electrical, thermal, mechanical, manufacturing, and reliability domains, as summarized in Table 1 [13], [15], [16], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. These studies demonstrate that data-driven models can accelerate repetitive simulations and support optimization within predefined design spaces. However, most existing models remain specialized surrogates trained for a fixed package configuration, a limited number of variables, and one or a few predefined outputs. Developing more such isolated models will not, by itself, lead to transferable and scalable intelligence. What is needed is a unified representation that transforms fragmented, problem-specific data into transferable knowledge across package architectures, physical domains, and engineering tasks.

To move beyond problem-specific surrogate modeling, we envision a foundation model for semiconductor packaging trained on the large-scale data infrastructure proposed in Section III. The term foundation model is used here in a functional rather than strictly architectural sense. It does

not imply that a single monolithic neural network replaces conventional solvers or performs uniformly well across every physical domain. Instead, it refers to a pre-trained, modular, and extensible framework that learns consistent representations from diverse packaging data and can be adapted to multiple downstream tasks. Such a foundation model would serve not merely as a multiphysics prediction engine, but as a scalable representation and adaptation framework for semiconductor packaging. It would learn transferable representations from heterogeneous data and support downstream tasks across co-design, manufacturing, and reliability.

The transition from isolated surrogate modeling to a foundation model framework fundamentally changes the model development workflow. Rather than selecting several algorithms for a small dataset and identifying the one with the lowest test error, foundation model development would be viewed as a continuous process involving data orchestration, pre-training, fine-tuning, post-training, uncertainty calibration, decision-oriented adaptation, and continual learning. Fig. 3 illustrates both this development process and the modular internal structure of the foundation model, which combines graph, spatial, and parameter encoders with task-specific solvers. As shown in Fig. 3, the task-specific solvers are not conventional numerical solvers but a set of models, one specialized to each physical domain, that together form the foundation model. All these models share a common representation of the design, including its geometry, materials, and operating point, defined consistently across the different physical domains. Large-scale pre-training would rely primarily on the multi-fidelity synthetic data generated by simulation, allowing the model to learn consistent representations across package architectures, materials, processes, loading conditions, physical domains, and fidelity levels. The pre-trained model could then be fine-tuned using higher-fidelity simulations and empirical data for specific package families, manufacturing processes, material systems, or failure mechanisms [44]. A newly introduced material system must first be characterized through the selection of an appropriate constitutive model and the experimental identification and validation of its parameters. Depending on whether this constitutive model lies within the scope of the pre-trained foundation model, the new material is either represented directly by its identified parameters or accommodated through transfer learning or reinforcement learning. Post-training could further incorporate engineering design rules, physical constraints, expert knowledge, and task-specific performance requirements [45]. Depending on the application, reinforcement learning may support sequential design and simulator-in-the-loop decision-making, while active learning, Bayesian optimization, and conventional constrained optimization may provide more efficient and interpretable alternatives for data acquisition and design space exploration [46].

The validity of this framework depends on the physical grounding and traceability of its training data. Simulation workflows should therefore be calibrated and validated to a degree commensurate with their intended role in model training. Large volumes of analytical, reduced-order, or lower-fidelity simulation data may provide broad coverage of the design space and physically consistent trends, whereas smaller sets of high-fidelity simulations and experimental observations should be used for calibration, correction, and validation [47]. Each data record should retain its fidelity level, governing assumptions, constitutive models, boundary and loading conditions, numerical resolution, calibration status, provenance, and uncertainty estimates, allowing individual records to be filtered, sampled, or weighted according to their confidence and intended use. Model predictions should be continuously evaluated against higher-fidelity simulations, experimental measurements, manufacturing observations, and, where available, field data [48]. The resulting discrepancies should guide targeted data generation, model updating, and uncertainty recalibration. This closed-loop validation process is essential to prevent simulation errors and dataset biases from being systematically amplified and embedded in the foundation model.

A practical first stage in developing such an ambitious model could focus on one response or failure mode, such as solder joint fatigue under thermal cycling or package warpage. The pilot dataset should cover a sufficiently broad and systematically defined range of package geometries, material combinations, manufacturing conditions, and loading profiles. The objective would be to demonstrate generalization and transferability within and across package configurations. The key requirement is not dataset size alone, but the construction of consistent, sufficiently comprehensive, hierarchically structured and multi-fidelity data under common representation and validation protocols. Subsequent stages could extend the framework to related responses within the same physical domain and then progressively incorporate additional domains through modular multi-task learning. Tightly coupled electrical, thermal, mechanical, manufacturing, and reliability predictions would be pursued after cross-domain transfer has been shown to improve predictive accuracy, data efficiency, or generalization relative to independently trained models. Whether knowledge can be transferred from a source domain to a target domain is first judged a priori from physical knowledge. This judgment is then verified by training the target-domain model with and without the source-domain data and comparing the resulting accuracy. When the source-domain data improve the accuracy, the knowledge is transferred to the target domain. When they reduce the accuracy, negative transfer has occurred, and the two domains are modeled independently.

Graph neural networks provide one particularly attractive, although not exclusive, approach for architecture-aware representation [49], [50], [51]. Semiconductor packages possess structured and hierarchical connectivity defined by dies, layers, interfaces, interconnects, and substrates, etc. In a graph representation, nodes may correspond to components, material regions, interfaces, or interconnects, while edges describe their electrical, thermal, mechanical, geometric,

or manufacturing relationships. Message passing can propagate local information through physically connected components, allowing the model to learn how local components and interactions influence system-level behavior. A practical implementation may combine a graph encoder for connectivity, a convolutional or spatial encoder for layouts and physical fields, and parameter encoders for geometry, materials, processes, and loading conditions. These structural graph representations can also be extended to manufacturing histories by using spatial-temporal or dynamic graph models, in which process steps, sensor traces, loading histories, and time-varying material states are encoded as temporal node or edge features rather than only as static design parameters [52].

The input to such a foundation model would be multi-modal and hierarchical. It may include package architecture and interconnect topology; two- or three-dimensional layouts; geometric dimensions; layer stack-up; material properties; manufacturing processes; electrical, thermal, mechanical, moisture, and environmental boundary conditions; assembly and operational loading histories; and available experimental, inspection, and field observations. These data may appear as design tables, CAD layouts, images, physical fields, time series, process records, or textual specifications. Appropriate encoders can be used for different modalities, followed by fusion into a shared representation conditioned on package architecture, physical domain, data fidelity, and provenance. The outputs of the model would be conditioned on the requested engineering task. They may include scalar quantities such as thermal resistance, peak temperature, insertion loss, warpage, interfacial stress, fatigue life, yield probability, and process margin; etc. The model should also provide uncertainty estimates, sensitivity information, and an indication of whether the requested case lies within its validated applicability domain. For multi-objective package design, the foundation model would rapidly predict a vector of electrical, thermal, mechanical, manufacturing, and reliability metrics for candidate designs. A separate constrained optimization layer could use these predictions to construct a Pareto front, representing the set of nondominated designs for which improving one objective would necessarily compromise at least one other objective. Product specifications and manufacturing requirements could be imposed as hard constraints, while application-dependent preferences could be represented through objective weights, acceptable ranges, reference points, or human ranking of candidate solutions. The model would therefore identify feasible trade-offs and candidate designs, while engineers retain responsibility for defining design intent and selecting among alternatives.

The proposed foundation model for semiconductor packaging could operate either as a standalone engineering model or as part of a broader LLM-enabled engineering AI system. In a standalone mode, it would encode package architectures, interconnect hierarchies, materials, manufacturing processes, loading conditions, and physical fields into transferable representations. It could then support quantitative multiphysics prediction, uncertainty estimation, applicability-domain assessment, and design space exploration. When integrated with an LLM, the packaging foundation model would provide domain-specific physical intelligence, while the LLM would enable natural language interaction and workflow orchestration. The LLM could translate engineering requirements into structured variables, constraints, and objectives, and retrieve relevant information from packaging databases and technical documents. It could then invoke appropriate models, solvers, and optimization

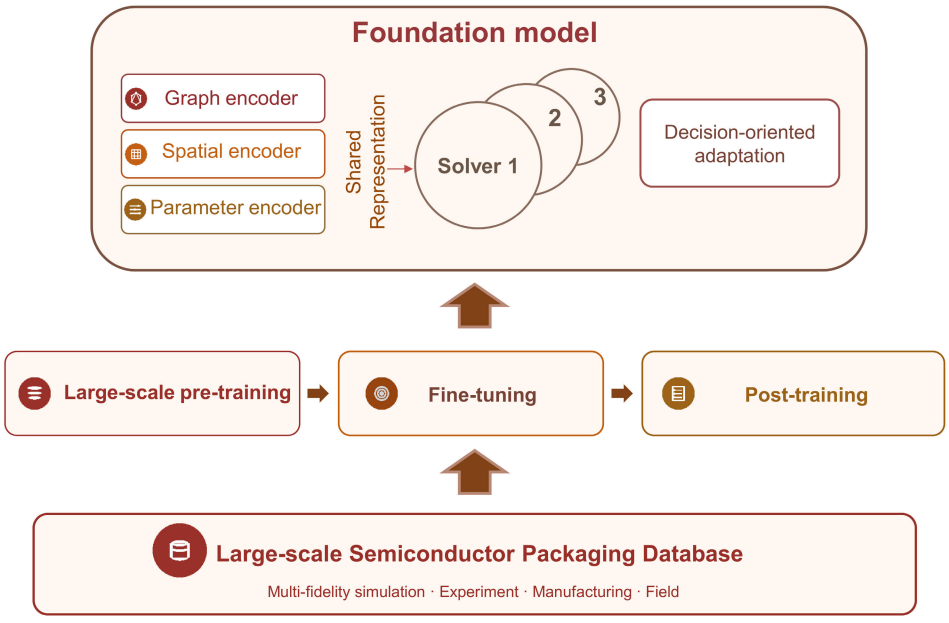


FIGURE 3. Development of a foundation model for semiconductor packaging through pre-training, fine-tuning, and post-training.

tools and synthesize the results into engineering reports and explanations. This integration would make the predictive and analytical capabilities of the packaging foundation model more accessible and adaptable to different engineering tasks. In the longer term, increasingly capable general purpose open-weights LLMs could serve as agentic orchestration platforms that coordinate the packaging foundation model with data resources, validated solvers, surrogate models, and optimization tools within closed-loop engineering workflows [7]. Such a framework could integrate database retrieval, simulation, model inference, uncertainty assessment, active learning, constrained optimization, and targeted experimental validation, while retaining the packaging foundation model as the primary engine for domain-specific representation and quantitative physical prediction.

V. CONCLUDING REMARKS

Artificial intelligence has demonstrated considerable potential to accelerate semiconductor packaging design, manufacturing, and reliability assessment. However, most existing studies remain isolated demonstrations based on narrowly defined datasets, fixed package configurations, and limited numbers of input variables and outputs. Developing more such problem-specific models will not, by itself, lead to transferable and scalable intelligence.

The central bottleneck is data scarcity, which extends far beyond insufficient data volume. It also encompasses limited physical diversity, inconsistent data representation, incomplete metadata, restricted accessibility, uneven data quality, and inadequate characterization of fidelity and uncertainty. Recent advances in embodied AI and robotics provide a useful blueprint for addressing this challenge. When large-scale physical data cannot be naturally accumulated, they must be systematically generated. We therefore propose simulation as a data factory, representing a shift from isolated, problem-specific analyses toward coordinated, automated, and large-scale computational data generation.

Because generating foundation model-scale datasets through high-fidelity multiphysics simulation would be computationally prohibitive, the proposed data factory should adopt a multi-fidelity hierarchy. Analytical formulations, reduced-order models, simplified numerical simulations, physics-informed learning, neural operators, and finite element integrated learning could provide broad and computationally efficient coverage of the design space. High-fidelity simulations, experimental measurements, manufacturing observations, and field data would provide calibration, correction, and validation. Simulation would thus evolve from a tool for analyzing individual designs into a scalable infrastructure for generating diverse, physically grounded data for transferable AI models.

Realizing this vision requires more than large-scale data generation. We therefore call for a coordinated international and cross-organizational effort to establish shared pre-training resources with standardized data representations, comprehensive metadata, traceable provenance, explicit fidelity labels, quantified uncertainty, and common validation protocols. An actionable starting point would be a sequence of focused pilot projects addressing well-characterized responses or failure mechanisms and evaluating generalization across package configurations, materials, processes, and loading conditions.

Building on this evolving data infrastructure, a foundation model for semiconductor packaging could be progressively developed as a pre-trained, modular, and extensible representation and adaptation framework, rather than merely a multiphysics prediction engine. Its development would involve data orchestration, pre-training, fine-tuning, post-training, uncertainty calibration, decision-oriented adaptation, and continual learning. Graph neural networks offer one attractive, although not exclusive, architecture for representing the hierarchical structures and physical connectivity of packaging systems.

The packaging foundation model could operate independently as the primary engine for domain-specific representation and quantitative physical prediction. It could also be integrated into an LLM-enabled engineering AI system, in which the LLM supports natural language interaction and workflow orchestration. Increasingly capable general purpose LLMs could coordinate the foundation model with data resources, validated physics-based solvers, specialized surrogate models, optimization tools, and targeted experiments within closed-loop agentic workflows. Ultimately, this roadmap seeks to advance semiconductor packaging AI from fragmented, task-specific modeling toward transferable, scalable, adaptive, and continuously validated engineering intelligence.

ACKNOWLEDGMENT

The authors gratefully acknowledge the organizers of EuroSimE 2026 (27th International Conference on Thermal, Mechanical and Multiphysics Simulation and Experiments in Microelectronics and Microsystems, April 19-22, Warsaw, Poland) for the opportunity to present the keynote, ''Artificial Intelligence (AI)-Enabled Reliability Prediction and Prognostics for Electronic Packages: A Perspective on Data, Models, and Applications'' (Cadmus Yuan, Xuejun Fan, and Xinyu Liu). The core concept and motivation of the perspective originated from the preparation and delivery of the keynote presentation and further expands and develops the perspectives introduced therein. They also thank the Associate Editor and the anonymous reviewers for their constructive comments, which helped improve this article.

REFERENCES

[1] R. Bommasani et al., ''On the opportunities and risks of foundation models,'' 2021, *arXiv:2108.07258*.
[2] J. Kaplan et al., ''Scaling laws for neural language models,'' 2020, *arXiv:2001.08361*.
[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ''ImageNet: A large-scale hierarchical image database,'' presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, doi: 10.1109/CVPR.2009.5206848.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[5] A. Vaswani, "Attention is all you need," presented at the Adv. Neural Inf. Process. Syst. (NeurIPS), 2017.

[6] T. B. Brown, "Language models are few-shot learners," presented at the Adv. Neural Inf. Process. Syst. (NeurIPS), 2020.

[7] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, "A survey on large language model based autonomous agents," *Frontiers Comput. Sci.*, vol. 18, no. 6, Dec. 2024, Art. no. 186345, doi: 10.1007/s11704-024-40231-1.

[8] J. Bjorck, "GR00T N1: An open foundation model for generalist humanoid robots," 2025, *arXiv:2503.14734*.

[9] R. V. Mahajan, W. Chen, P. Thompson, W. Bottoms, A. S. Helmy, B. Chan, M. Swaminathan, J. Shalf, C. Bailey, J. Schutt-Aine, G. Subbarayan, T. Lee, S. S. Iyer, S. Narumanchi, and R. Mongia, "The heterogeneous integration of electronic components," *Nature Rev. Electr. Eng.*, vol. 3, no. 4, pp. 254–263, Mar. 2026, doi: 10.1038/s44287-026-00270-1.

[10] J. Lau and X. Fan, *Hybrid Bonding, Advanced Substrates, Failure Mechanisms, and Thermal Management for Chiplets and Heterogeneous Integration*. Cham, Switzerland: Springer, 2025.

[11] M. Rafiee and P. Saini, "Advanced semiconductor packaging design via artificial intelligence and machine learning: A review," *Microelectron. Rel.*, vol. 181, Jun. 2026, Art. no. 116145, doi: 10.1016/j.microrel.2026.116145.

[12] G. Datta, S. S. Sharif, and Y. M. Banad, "Toward digital twins in 3D IC packaging: A critical review of physics, data, and hybrid architectures," *Electronics*, vol. 15, no. 8, p. 1740, Apr. 2026, doi: 10.3390/electronics15081740.

[13] M. Rafiee, M. Hasheminasab, F. Agin, and F. El Siss, "Machine learning-based co-design of advanced flip-chip packages: Co-optimization of package warpage, die stress, solder-joint reliability, thermal management, and signal integrity," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 16, no. 8, pp. 1823–1831, Aug. 2026, doi: 10.1109/TCPMT.2026.3674462.

[14] X. Fan, C. Yuan, and X. Liu, "Artificial intelligence (AI)-enabled reliability prediction and prognostics for electronic packages: A perspective on data, models, and applications," presented at the 27th Int. Conf. Thermal, Mech. Multiphys. Simulation Exp. Microelectronics Microsyst (EuroSimE), Warsaw, Poland, Sep. 2026.

[15] J. Lau and K.-N. Chiang, *Cu-Interconnects, Glass, and AI-Assisted Simulation for Chiplets and Heterogeneous Integration*. Singapore: Springer, 2026.

[16] S. K. Panigrahy, Y.-C. Tseng, B.-R. Lai, and K.-N. Chiang, "An overview of AI-assisted design-on-simulation technology for reliability life prediction of advanced packaging," *Materials*, vol. 14, no. 18, p. 5342, Sep. 2021, doi: 10.3390/ma14185342.

[17] C. Yuan, X. Fan, and G. Zhang, "Solder joint reliability risk estimation by AI-assisted simulation framework with genetic algorithm to optimize the initial parameters for AI models," *Materials*, vol. 14, no. 17, p. 4835, Aug. 2021, doi: 10.3390/ma14174835.

[18] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," presented at the Conf. Comput. Vis. Pattern Recognit. Workshop, Jun. 2004.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/b:visi.0000029664.99615.94.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," presented at the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jun. 2005, doi: 10.1109/CVPR.2005.177.

[22] G. A. Miller, "WordNet," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995, doi: 10.1145/219717.219748.

[23] X. Wu, Z. Wang, S. Ma, X. Chu, C. Li, W. Wang, Y. Jin, and D. Wu, "An RDL modeling and thermo-mechanical simulation method of 2.5D/3D advanced package considering the layout impact based on machine learning," *Micromachines*, vol. 14, no. 8, p. 1531, Jul. 2023, doi: 10.3390/mi14081531.

[24] D. Mondal, J. C. Suhling, E. Mirkoohi, and P. Lall, "A data-driven machine learning model for the stress-strain behavior of single grain SAC305 solder joints," presented at the IEEE 74th Electron. Compon. Technol. Conf. (ECTC), May 2024, doi: 10.1109/ECTC51529.2024.00110.

[25] H. Kim, J. Lee, S. Choi, F. Berto, T. Shin, J. Park, J. Kim, J. Yoon, B. Kim, Y. Kim, and J. Kim, "Advanced chiplet placement and routing optimization considering signal integrity," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 15, no. 11, pp. 2331–2343, Nov. 2025, doi: 10.1109/TCPMT.2025.3561039.

[26] M. Shanmugam and L. Sirisha Maganti, "Machine learning-based thermal performance study of microchannel heat sink under non-uniform heat load conditions," *Appl. Thermal Eng.*, vol. 253, Sep. 2024, Art. no. 123769, doi: 10.1016/j.applthermaleng.2024.123769.

[27] M. Rafiee, F. Agin, K. Kumar, and E. Murali, "Thermo-mechanical co-design of 2.5D flip-chip packages with silicon and glass interposers via finite element analysis and machine learning," *Microelectron. Rel.*, vol. 176, Jan. 2026, Art. no. 115983, doi: 10.1016/j.microrel.2025.115983.

[28] H.-C. Cheng, C.-L. Ma, and Y.-L. Liu, "Development of ANN-based warpage prediction model for FCCSP via subdomain sampling and Taguchi hyperparameter optimization," *Micromachines*, vol. 14, no. 7, p. 1325, Jun. 2023, doi: 10.3390/mi14071325.

[29] M. Majd, P. Meszmer, A. Prisacaru, P. J. Gromala, and B. Wunderle, "Stress prognostics for encapsulated standard packages by neural networks using data from in-situ condition monitoring during thermal shock tests," presented at the 21st Int. Conf. Thermal, 2020, doi: 10.1109/eurosime48426.2020.9152697.

[30] P. Yao, J. Yang, Y. Zhang, X. Fan, H. Chen, J. Yang, and J. Wu, "Physics-based nested-ANN approach for fan-out wafer-level package reliability prediction," presented at the IEEE 72nd Electron. Compon. Technol. Conf. (ECTC), May 2022, doi: 10.1109/ECTC51906.2022.00287.

[31] C. Yuan, J.-C. Hong, and P.-S. Li, "AI surrogate modeling for PBGA solder joint fatigue risk assessment," *Microelectron. Rel.*, vol. 176, Jan. 2026, Art. no. 115956, doi: 10.1016/j.microrel.2025.115956.

[32] C. C. A. Yuan, J. Fan, and X. Fan, "Deep machine learning of the spectral power distribution of the LED system with multiple degradation mechanisms," *J. Mech.*, vol. 37, pp. 172–183, Dec. 2020, doi: 10.1093/jom/ufaa025.

[33] C. Yuan, "An AI-based adaptive surrogate modeling method for the in-service response of UVLED modules," *Electronics*, vol. 11, no. 18, p. 2861, Sep. 2022, doi: 10.3390/electronics11182861.

[34] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "NuScenes: A multimodal dataset for autonomous driving," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020.

[35] P. Sun et al., "Scalability in perception for autonomous driving: Waymo open dataset," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020.

[36] B. Baker, I. Akkaya, P. Zhokov, J. Huizinga, J. Tang, A. Ecoffet, B. Houghton, R. Sampedro, and J. Clune, "Video PreTraining (VPT): Learning to act by watching unlabeled online videos," presented at the Adv. Neural Inf. Process. Syst. (NeurIPS), 2022.

[37] S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta, "R3M: A universal visual representation for robot manipulation," presented at the Conf. Robot Learn. (CoRL), 2022.

[38] T. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," presented at the Robot., Sci. Syst. (RSS), Jul. 2023.

[39] A. O'Neill et al., "Open X-embodiment: Robotic learning datasets and RT-X models: Open X-embodiment collaboration 0," presented at the IEEE Int. Conf. Robot. Automat. (ICRA), May 2024, doi: 10.1109/icra57147.2024.10611477.

[40] D. Dall'Alba, M. Naskręt, S. Kamińska, and P. Korzeniowski, "FF-SRL: High performance GPU-based surgical simulation for robot learning," presented at the Neural Inf. Process. Syst. (NeurIPS) Datasets Benchmarks Track, Oct. 2024. [Online]. Available: https://openreview.net/forum?id=fgFBtYgJQX_.

[41] Q. Su, C. Yuan, and K.-N. Chiang, "A small database with an adaptive data selection method for solder joint fatigue life prediction in advanced packaging," *Materials*, vol. 17, no. 16, p. 4091, Aug. 2024, doi: 10.3390/ma17164091.

[42] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," *Nature Mach. Intell.*, vol. 3, no. 3, pp. 218–229, Mar. 2021, doi: 10.1038/s42256-021-00302-5.

[43] T. Gandía-Sastre and M. Prados-Privado, "Integrating finite element data with neural networks for fatigue prediction in titanium dental implants: A proof-of-concept study," *Appl. Sci.*, vol. 15, no. 19, p. 10362, Sep. 2025, doi: 10.3390/app151910362.

[44] C. Yuan, J.-C. Hong, and P.-S. Li, "AI surrogate model for PBGA solder fatigue assessment using progressive transfer learning with multi-fidelity data," presented at the 27th Int. Conf. Thermal, Warsaw, Poland, Apr. 2026.
[45] G. Tie, "A survey on post-training of large language models," 2025, *arXiv:2503.06072*.
[46] M. S. Mahmud, X. Yang, P. Kashyap, and M. Swaminathan, "Cross-topology transfer learning using Bayesian optimization for scalable surrogate modeling of 3D packaging structures," in *Proc. IEEE 76th Electron. Compon. Technol. Conf. (ECTC)*, May 2026, pp. 486–493.
[47] Y. Zhang, Z. Gong, W. Zhou, X. Zhao, X. Zheng, and W. Yao, "Multi-fidelity surrogate modeling for temperature field prediction using deep convolution neural network," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106354, doi: 10.1016/j.engappai.2023.106354.
[48] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2346–2363, Dec. 2019, doi: 10.1109/TKDE.2018.2876857.
[49] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021, doi: 10.1109/TNNLS.2020.2978386.
[50] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2018, *arXiv:1710.10903*.
[51] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017, *arXiv:1609.02907*.
[52] S. M. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupart, "Representation learning for dynamic graphs: A survey," *J. Mach. Learn. Res.*, vol. 21, no. 70, pp. 1–73, 2019.



CADMUS CHANG-ANN YUAN (Member, IEEE) received the Ph.D. degree from the Department of Power Mechanical Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2005. He was a Postdoctoral Researcher with Delft University of Technology, Delft, The Netherlands, and with NXP Semiconductors and TNO. He is currently a Professor with the Department of Mechanical and Computer-Aided Engineering, Feng Chia University, Taichung, Taiwan. His research interests include the application of AI methods and finite element modeling to semiconductor packaging design, reliability, and manufacturing. He currently focuses on process modeling and AI-assisted design for reliability (AIDfR) for electronic packaging, with applications to 3-D heterogeneous integration packaging, panel level packaging, fan-out wafer-level packaging, and LED packaging.



XUEJUN FAN (Fellow, IEEE) is currently a Regents' Professor with Texas State University System and the Mary Ann and Lawrence E. Faust Endowed Professor with the Department of Mechanical Engineering, Lamar University, Beaumont, TX, USA. He is an IEEE Distinguished Lecturer. His research interests include the characterization, modeling, reliability of materials, components, systems for micro- and opto-electronics manufacturing, packaging, and heterogeneous integration, leading to a prolific output of five books and more than 350 technical papers. He was a recipient of several major IEEE Electronic Packaging Society Awards. He serves as a Co-chair of Chapter 14 (Modeling and Simulation) for the Heterogeneous Integration Roadmap and holds several editorial roles, serving as the Co-Editor-in-Chief for Micro-/Nano-Electronics Section in *Frontiers in Electronics* and an Associate Editor for several leading journals. He has been listed in top 2% scientist worldwide by Stanford University composite index score for career and single year, since 2020.



XINYU LIU is currently a Professor with the Department of Industrial and Systems Engineering, Lamar University, Beaumont, TX, USA, and directs the Micro-Manufacturing Research Laboratory. His research interests include micro-, meso-, and nano-scale manufacturing, precision engineering and metrology, with applications spanning advanced manufacturing and energy and infrastructure systems. He has contributed to micro-milling process modeling, tool-wear characterization, and in-situ metrology. He is a member of American Society of Mechanical Engineers (ASME), the Society of Manufacturing Engineers (SME), and International Society of Automation (ISA).



GUOQI ZHANG (Fellow, IEEE) received the Ph.D. degree in aerospace engineering from Delft University of Technology (TU Delft), Delft, The Netherlands, in 1993. He was a Professor with Eindhoven University of Technology, from 2002 to 2005. He is currently a Chair Professor of micro/nanoelectronics system integration and reliability with the Department of Microelectronics, TU Delft. Before returning to academia, he spent about 20 years with Philips, Eindhoven, The Netherlands, where he held a senior scientific and management positions, including Philips Fellow. He has authored more than 350 papers and holds more than 100 patents. His research interests include heterogeneous system integration and packaging, multi-physics and multiscale modeling, digital twin and design for reliability, and wide-bandgap semiconductor devices. He is a fellow of Netherlands Academy of Engineering, recognized for contributions to heterogeneous micro/nanoelectronics packaging, system integration, and reliability.