Deep Science Publishing
https://doi.org/10.70593/978-93-49910-47-8_3

DeepScience
Open Access Books

Chapter 3: Leveraging data engineering principles to streamline semiconductor research and development pipelines

3.1. Introduction

The Research and Development (R&D) of semiconductor technologies is a multibillion-dollar effort for the worldwide semiconductor industry. To keep track of the rapid advance in semiconductor technology and enhance productivity, there is a need to streamline semiconductor research R&D pipelines without compromising research fidelity and flexibility. Existing Data Engineering principles in handling datasets encompassing scientific data provenance and manuscript rewrite pipeline are employed to streamline semiconductor R&D pipelines. A semantic provenance data model, structured data integration pipelines, and distributed workflows are demonstrated and discussed with respect to three relevant semiconductor R&D tasks: site-specific ion implantation for high-performance embedded non-volatile memory, atomic layer doped gate technology for sub-5 nm node FinFETs, and pitch-scaled extreme ultraviolet lithography implementation for 0.5 nm node logic technologies. The results show the potential of Data Engineering principles for semiconductor data management and research productivity enhancement by addressing the unique needs and preferences of semiconductor R&D.

Research and Development (R&D) is the foundation of advances in semiconductor technology, which drives computer and smartphone performance improvements and low-cost ubiquitous electronics and has wide implications for other fields such as healthcare and information technology. Semiconductor technologies generally advance on the order of 2–3 years, involving multibillion-dollar investments in worldwide efforts by semiconductor and equipment manufacturers, fabrication plants, and research institutions. During this extended time, new semiconductor devices typically comprise a

handful of new processing technologies and precise control of many device structures, doping, and chemical composition parameters. R&D is the exercise of empirically exploring, implementing, optimizing, and demonstrating new technologies step by step, often with little prior knowledge of success. The general requirement for semiconductor technologies is that they be manufacturable, which implies extensive investigation of conveying, mass production, contamination, yield, and material transition issues (Begum & Chowdary, 2024; Ketelaars, n.d.; Raghunathan et al., 2024).

The semiconductor industry is faced with an impending crisis at its 5 nm technology node. There is a compelling need to keep dedicated R&D in wafer-scale fabrication without large upfront investments in factories, chemicals, and process equipment, or exposing trade secrets. Currently, R&D typically relies on legacy solutions like spreadsheets and proprietary solutions from equipment vendors, with custom database-derived solutions playing a minor role. Such solutions are hardwired to semiconductors and result in data silos. There is a need for a general-purpose solution that can be implemented within a year and is easily adaptable to many related tasks beyond semiconductors (Xu et al., 2023; Schiller & Larochelle, 2024; Schilling-Wilhelmi et al., 2024).

Fig 3.1: Data Engineer Camp

3.1.1. Background and Significance

There has been an incredible amount of speculation regarding the future of semiconductors. Moore's Law obviously cannot continue into the future indefinitely. Nevertheless, most analysts agree that future innovation will require the semiconductor

manufacturing industry to adapt to completely new physics within the next five- to ten-years. During this transition era, however, the industry still has to meet the enormous market pressures that previously drove innovation. While it is hoped that the industry will emerge from this transition with new and improved technology, tools, and infrastructure, this new "super-era" will take billions of dollars and take years of R&D. Until then, designing and manufacturing 10 nm nodes presents a daunting challenge. In addition, the semiconductor R&D industry is currently at a disadvantage in part due to the relative lack of infrastructure and fabrication capabilities.

There are many challenges facing the semiconductor R&D ingredient supply chain. First, the supply chain is longer and much more complex than assumed in the past. Disruptions from COVID-19 reveal that even with many options, continent-wide disruptions will have far-reaching effects. A complete supply chain microchip in the US from first principles utilizing expert skill sets will take decades. Without extreme intervention, there will be no alternative to Asian dominance. Second, far more needs to be understood for reliable designs, as over-engineering will blindly slow all innovation. The broader climate crisis will worsen many underlying aging multi-national issues such as poverty, inequality, and energy/disruption instability. The critical path will clearly present an opportunity for radically different chip designs and alternatives to traditional advancement curves. Thus, there is a need for a product shift to radically different chip architectures and approaches.

3.2. Overview of Semiconductor R&D

Research and Development (R&D) is a crucial component in the semiconductor industry and plays a key role in a company's competitiveness. Companies must develop new technology newer than competitors and bring it to production as fast and as freely from problems as possible. Given the high complexity and cost of developing new processes for production, knowledge gained during the R&D stage of new processes is generally recorded in a set of specifications called a Process Recipe. Although this obtained technology consists of a series of instructions representing the best understood condition to produce acceptable products, it is often subject to large variability in production execution. Technology transfer from R&D to production is finding suitable technological solutions to possible problems while adapting to the new environment of production, which is complicated in the semiconductor industry by the hijack of the R&D process by the production process.

Within a semiconductor factory, it is unavoidable for R&D and production processes to coexist in the same business enterprises. Products in their initial stages of technology development go through the R&D process, while those designated acceptable would move to production for mass fabrication. Semi-finished products in the R&D process are

firstly passed to queue waiting for processing and made subject to scheduled and arcane access to fabrication facilities. The combination of meandering product flows and rigid transport times gives rise to increased production time and loss of schedule integrity. The integration of both endeavors leads to the adoption and hijack of the R&D process by the production.

Consequently, the collection of production-related data stored in the business enterprise serves the accountabilities of the organizations who maintain them. The EDA and CAD programs being used for chip design generally do not record any of R&D-specific design information; the statistics of production yield improvement efforts are often incomprehensible even to those who are experienced R&D personnel. These data pools give rise to a wide-scattered pondering of process legends. The R&D of a semiconductor product often begins before the start of its design. Celebration of the first numeric yield of a process technology brings up excitement of a process runner, while it unveils a nightmare before the designer. The production and R&D teams need to cooperate closely to design and test schemes for the development of compact batch designs that are suitable for physical astronomy.Their R&D process efficiency enhancement initiatives are conducted via centralized data engineering teams under the corporate office, hence ensuring that interviewees from diverse positions and with various experiences participate. The interviewees comprise a mix of top executives, a director, managers, senior/lead staff, and junior staff. Such diversity allows for a comprehensive understanding of the semiconductor R&D function. At least three people with direct R&D process management experience are interviewed from each company to maximally ensure the validity of the research findings.

3.2.1. Research design

This study adopts a hybrid qualitative and quantitative approach. The qualitative approach comprises interviews to gauge a deeper understanding of the phenomena under investigation. The quantitative approach provides the means to statistically validate proposed models and frameworks that support the qualitative findings. In-depth interviews are conducted with executives, managers, and senior staff at semiconductor companies to identify data engineering practices facilitating R&D process efficiency, surge product outputs, and systematically investigate the sources of process inefficiencies and waste in order to absorb NPI workload increase. The interviews are semi-structured with questions prepared in advance to cover certain themes of interest, but open-ended enough to offer the interviewee freedom to elaborate on issues that arise. This enables either party to probe deeper on a topic of mutual interest. Each interviewee is asked questions calmly, attentively, and patiently and is given sufficient time to answer without interruptions. Interview questions are grouped into five themes: general

semiconductor R&D process efficiency and data challenges, operational definition of semiconductor R&D process efficiency, data engineering processes supporting semiconductor R&D process efficiency, R&D process inefficiencies and waste in semiconductor NPI, and semiconductor R&D process efficiency enhancement recommendations.

A total of 41 interview candidates are invited. All interview invitations are accepted; thus, all candidates become interviewees. The interviewees work at one of the seven global top 15 semiconductor companies: Intel, TSMC, Samsung Electronics, NVIDIA, Broadcom, Qualcomm, and Texas Instruments.

3.3. Data Engineering Principles

Along with increased computing power and smartness, the amount of data produced in any organization or discipline is rising. Such advancement has given rise to the concept of data pipelines, which comprise a collection of operations performed on data. A data pipeline is a set of jobs, each of which implements a single task on data that produces data in a predefined data structure format and pushes it to the next job. Such data

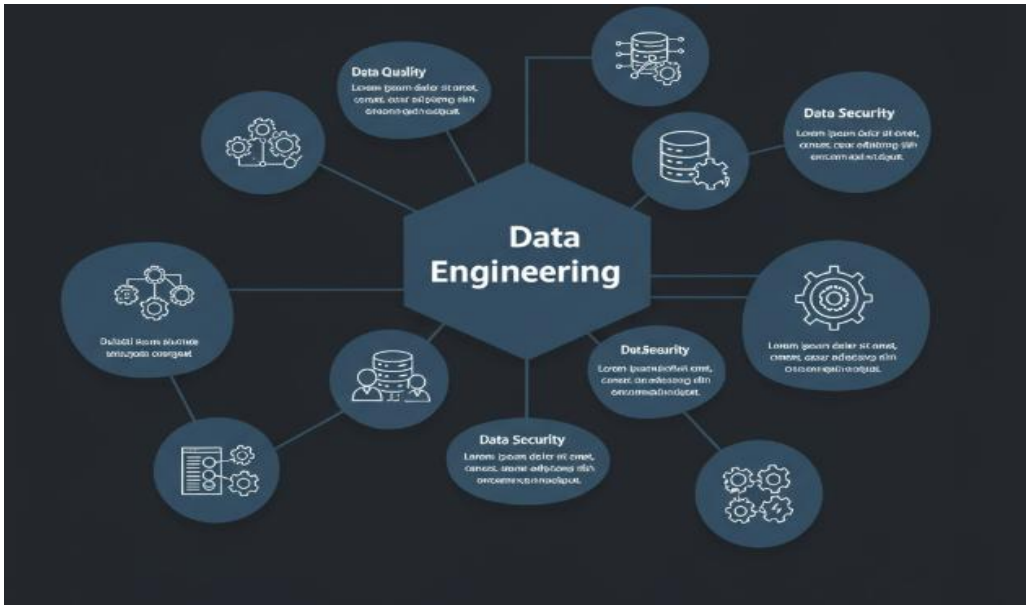

Fig 3.2: Data Engineering Principles

pipelines are usually established using a combination of services and distributed systems, and streamlining them with considerations is often an engineering challenge. This is partially because, in scientific disciplines such as semiconductor R&D, models and pipelines are rapidly evolving as new automated tests and optimization strategies are

introduced and existing ones are improved. This has led to disambiguation in both variable naming in the context of data preparation and misalignment between data preparation and data inference models. Hence, when the data preparation step is revisited, it is critical and time-consuming to realign (regenerate) the top half of the pipelines in data science workbooks. There is a desire to alleviate the engineering burden by decoupling the `data pipeline' model from the `computational method' model so that changes in either side will not impact the implementation on the other side. As a first step towards realization, a ``pipeline scripting language'' is devised. A novel mathematical principle called data homogenization is also introduced to ensure that the new meta-data pipeline models are equivalent (i.e. redundant) to the original data pipeline models.

3.3.1. Definition of Data Engineering

Data engineering has received much attention in academia and industry over the past several years. Data engineering refers to a wide range of problems related to data wrangling and describes the process of preparing data for a later phase such as data analysis or data mining . The process involves acquiring, cleaning, and integrating data and forming a feature set. Data collection addresses issues such as crawling to obtain raw data for data preparation. Data preparation is known very broadly and is much less well defined.

The data cleaning section includes schema and instance-level data cleaning issues. At the schema level, the focus is on integrity constraints. On the other hand, the instance level contains many problems such as missing data, canonicalization, data integration, and data description. Some problems are single-source (e.g., missing data, canonicalization) and others are multi-source (e.g., entity resolution, data integration). Two phases of the CRISP-DM framework specifically related to data engineering are highlighted by data understanding and data preparation. Data understanding includes data description, verification of data quality, data collection and data exploration. Data preparation includes data selection, cleaning tasks, data integration, and feature engineering.

A comprehensive classification of data engineering tasks is presented tuning up several different aspects. A taxonomy split of tasks between missing-data and non-missing-data tasks is presented, with further breakdowns into classes (e.g., different classes of missing data). The tasks have been refined to three levels: level C (conceive), level B (believe), and level A (analyze). Level C relates to the organization of data, especially data parsing and data integration. Level B covers topics under data quality and data organization, while level A includes feature engineering problems and a semantic description of features.

3.3.2. Importance in R&D

For semiconductor companies, design and manufacturing processes used to be far apart. With the development of substantial metal-oxide field-effect transistors for which both fabrication and design require nanoscale processes, the need to overlap chemical mechanical planarization design and its experimental character has emerged. To optimize the overlap, it is essential to understand how design may affect manufacturing and how manufacturing may affect design. Such a feedback loop needs to be extended to other areas of design, this is why semiconductor manufacturing processes are pulling downstream semiconductor product design closer than ever before. To explore the feasibility of a feedback loop between the understanding of design data used by design engineer and manufacturing assays implemented by manufacturing engineer, knowledge data model information of parabolic expected direction has been extracted from both sides as the first trial by a unity team consisting of both design engineers and manufacturing engineers. Because design plays various roles in producing radically different device performances in dramatically different shapes, this feedback loop spans both on-device and on-chip manners. In terms of a parabola that curves upward and knots the chosen parameter lengths of design flows together with practical usage times of manufacturing assays, this feedback loop has two knots. One knot consists of four individual axioms that link the common pipeline of manufacturing and the common pipeline of design—can the fits have the correct order based on slope and shape on curves? The other knot consists of heuristic information datums which highlight key parameters to explain design diverging effects on and thus expedite modification trials in cybersecurity and tensor design. Most semiconductor companies follow an R&D model to develop manufacturing processes for new semiconductor devices. While no existing products are manufactured, it can be called a discovery fab. In a discovery fab, various processes are explored in conjunction with two or more actively investigated and manufactured device designs. Uncertainty may arise from the recursive and concurrent operations of multiple teams and discovery fab discovering optimization. There are two types of product and process uncertainty. One is unavoidable uncertainty, which comes from the technology being explored. The other is avoidable uncertainty, which arises from familiar technology re-invented, lack of anticipation of variations in quantities and ordering windows, poor acceptance or proliferation of emergent research results, and diminished focus on problem solving. Aggressive solutions may be pursued or opportunistically short-term fixes may be adopted to ameliorate delivery challenges by the discovery fab before considering wide ranging changes to the R&D model.

3.4. Current Challenges in Semiconductor R&D

The unprecedented rapid growth of semiconductor applications, spurred by machine learning, artificial intelligence, and the internet of things (IoT), has been well-documented. However, this breakout in applications comes at a time when the semiconductor industry faces a critical shortage of engineers. The production of semiconductors is a complex process that relies heavily on several engineers specializing in distinct tasks. Merely producing an integrated chip (IC) in the fabrication facility requires detailed inputs from a wide range of engineers, performing simulation and modeling tasks spanning across multiple disciplines. The design task, arguably the most critical step in the IC value chain, is carried out exclusively by highly-skilled engineers, leading to increasingly sophisticated and complicated designs fed into the fabrication facility (fabl). There remains an enormous talent gap in the semiconductor industry regarding design, fabrication, packaging, process development and characterization, yield improvement, and test. Streamlined semiconductor research and development (R&D) pipelines can shorten the design cycle and help mitigate design complexity. Datasets generated in recent years from simulation and fabrication must be analyzed to recover key performance indicators (KPIs) that are physically interpretable. This requires both knowledge of physics-driven models and domain expertise in semiconductor devices, materials, and process technology.

Traditionally, understanding and quantifying semiconductor characteristics depend on a simulation and modeling task pipeline that may last weeks. With advanced scaling technology nodes, the enormous design and fabrication complexity leads to a cascade of those tasks, making extracting indications of physical properties from the datasets increasingly complicated. Engineering scripts are often used to scrape data from raw output files and represent them in human-readable formats. Those operations vary significantly across disciplines but can be modeled as a simple RDF pipeline. Nevertheless, the R&D pipeline bottleneck relies on the agile responses to fast-paced engineering requests of critical KPIs, which require in-depth semiconductor knowledge and are hard to break down into a single step. Although process engineers and physicists are able to quickly finger out error distributions and potential hardware problems from the probe round file (PRF) through decades of experience, such operations are hard to generalize. Existing machine learning (ML) schemes have difficulty bridging the interpretability chasm owing to the lack of domain knowledge.

3.4.1. Data Silos

Semiconductor data is inherently heterogeneous: while wafer data is structured and high-volume, observations on semiconductor devices may take the form of unstructured, low-volume images, documents, or simulations. Although a few high-throughput platforms

exist, R&D pipelines are still dominated by small-scale, artisanal efforts continuing a culture of silos. These heterogeneity and scale issues make it painfully obvious that bridges are needed to prepare and pass the correct content from one isolated case to the next. More than any discrete analytics task, this need is common across an incredible variety of industry sectors and practices.

Difficulties of integrating siloed systems and of spreading resources and data across them have been well-studied. Contemporary software development offers substantial tools for these tasks, beginning inter alia with ESB integrations, which enable event-driven communication and sharing of lineage in between systems, accelerating on-the-fly integration—particularly for data streams. In the long-term, robust solutions that imitate the natural forms of practice risk compromising usability and configurability, and thus must tread a nuanced route. They can take the burden of maintaining the cobweb out of the hands of engineers, allowing both rapid development and thorough grasp by users. Advanced and mature elements of software development practice from other domains are often easily ported over, to substantial first-order effect.

Data sharing should be distinguished from transfer and integration, both system-to-system actions. Architecturally, with regards to system composability, data can be shared directly or via a middle filtering layer. The former, with raw data streaming from one silo to another, should only be applied to low-level systems, such as a server- or file-based database, with application extraction and storage operations controlled by developers. In such cases storing raw data in standard databases allows cheap compliance without vendor lock-in. However, because it is cheaper to filter and refine singular streams and to do so in silos, sharing by architecture is a far gentler approach to implement into high-level systems, preventing leaks of trailing extensions.

3.4.2. Inefficient Workflows

Research and Development (R&D) in semiconductor technology is an intensive process that requires a combination of design creativity and graduate student effort. Developing new ideas can take years, with research reports being handed off between graduate students through their roles as teaching assistants or raters. Experiments performed on process technology may be obtained months after designs are submitted to the foundry. Working with all the stakeholders in design, test, packages, and infrastructure is a logistical challenge to get reports back to designers in the laboratory, especially for early process development experiments. When these challenges are added to the pressure to demonstrate results to funding entities and managerial stakeholders, the chances of success diminish.

Many research groups independently develop internal applications and a plethora of unified, harmonized software infrastructure to meet the needs of their individual work. Disk storage keeps mounting with a slew of scripts, patches, and specific applications. Each novice programmer has their unique vision of what is best, leaving rounded corners on eventual applications. The choice of language tools to work with is even more daunting early on because everything is open. The engineering industry standard tools are also costly and come with enormous overhead costs. Learning the proprietary languages from a fresh trainee can create burdens of months of unproductive time, which is cheerfully discarded as they move on from their project.

3.5. Data Integration Strategies

Effective and efficient integration of services has significant effects on the success of all data platforms. Integrating a service not only means programming interfaces relevant to the service but also understanding the logic of how the service operates, which is especially demanding for interdisciplinary fields like semiconductor research, where data engineers and all other stakeholders are typically from weakly connected domains. In addition, the data domains of services are also highly heterogeneous, resulting in different syntactical definitions over standard frameworks released for integration. This section focuses on two main aspects: Integration interfaces developed to integrate data ingestion, harmonization, and push to data platform services, and a syntactic approach that helps to automatically map data streams of various data formats to the one expected by the target service.

Particularly, the design of how to connect third-party adapters well with sensors and pre-process the adapter memories before being sent off will not be covered. To ensure extensibility, flexibility, and minimal maintenance during integration and service updates, all adapters closely follow the formal service structure defined in their corresponding configuration files. This configuration structure is RDF-based and contains rich semantic information, describing the core logic span and the data input-output schema of each service. Using these configurations, the streams from their previous steps can be converted. Firstly, the transformation steps (if any) defined in the configuration will be checked against each new stream. The original sample values collected in raw processing and storage technologies will be modified to the final model classes or formats defined in each service and prepared for submission. It is complex but has a seamlessly smooth pipeline when implemented. The built transformation models can be periodically saved and loaded with incoming streams in a minute. A messaging system enforces collaboration in releasing pre-processing services to adapt all kinds of data domains still without professional programming.

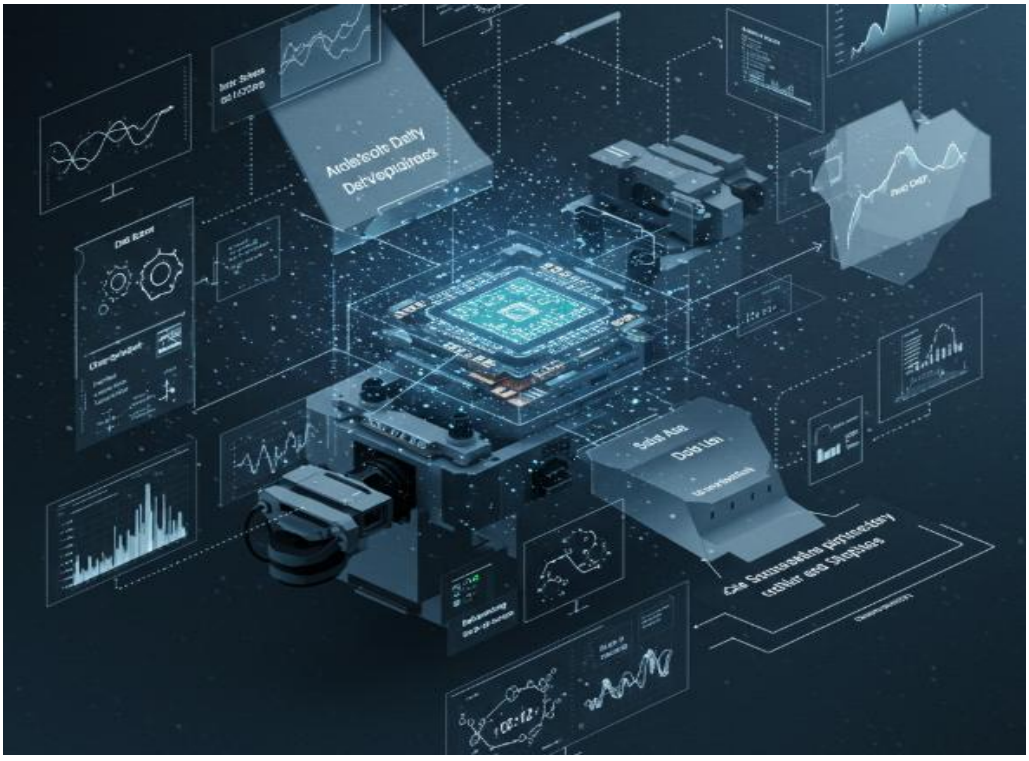

Fig : Leveraging Data Engineering Principles to Streamline Semiconductor Research and Development Pipelines

3.5.1. ETL Processes

Streamlined semiconductor Device researcher, foundation experimental data are collected in various databases and analysis results are generated in various formats such as spreadsheets, figures, images, and textual articles. Aggregating and processing them to extract the data to which numerically processed descriptions can be attributed to facilitate reusability, and enhancing their compatibility for machine learning is an essential function to be implemented in a standardized architecture. This functionality can be termed as the extract-transform-load (ETL) process.

In semiconductor Device research, the ETL process typically extracts analysis results from unstandardized files, migrates to existing standardized databases, and transforms the data into compatible data shapes for machine learning. In the extraction phase, unstandardized files produced by the foundation's Device exploration analysis results are analyzed, and a set of existing analysis templates that contains basic operations is prepared. In the transformation phase, the downloaded data and Analysis ID are analyzed, and corresponding local files are selectively cleaned. After cleaning, the data is uploaded to the designated database in a range parameterized by the user. This ETL

utility is easy to implement and customize. All variables and constant parameters can be controlled upon implementation. Customization of input and output plugin is also readily achievable by developers in deep learning, especially related to text similarity search. As it is based on markdown format and relational database implementation, it is highly flexible for database size and layout.

3.5.2. Data Lakes vs. Data Warehouses

Data lakes (DLs) have emerged as a new class of storage architectures to manage massive volumes of heterogeneous, potentially fast-changing data, the so-called big data. They are different from both traditional databases and data warehouses (DWs). In contrast to a DW, which requires a strong pre-processing phase consisting in cleaning, transforming, and aggregating data before storage, DLs keep data in raw format. Stored data can be structured, semi-structured, or unstructured. Similarly, professionally formulated datasets may be mixed with impossible quality raw data and personal notes. In comparison to databases and warehouses, DLs have relaxed constraints for file structure, format, and also for query languages and use. Integration of a new format into a DL can be performed without affecting formerly stored data. Conversely, traditional DWs can render useless, misplaced, or exaggeratedly organized a wealth of intriguing information, despite its enormous potential. This is a crucial constraint in experimental sciences; it is not unusual that research questions evolve significantly over time. It is not infrequent that one archive previously stored measurements, analyses, and formulas, only to find them barely useful at a later date, when a DL repository might have supported on-the-fly production of new datasets. DLs flatten the data organization schema; scientists, engineers, and analysts may work at the same abstraction, scale, or granularity levels. Nevertheless, DLs have to embed adequate content description and retrieval (CR) functionalities and representations. Users require a CBR-like mechanism to help them in data querying. Currently, there are two fundamentally different mechanisms.

Content-based retrieval might seem a good approach. This strategy relies on several data signatures or fingerprints derived from lower-level information to discriminate between different sources. Instead of delving deep into a DL, only looking at the fingerprints might help create an idea for relevant datasets. Such signatures are often too expensive to compute. The search space may become enormously large, making it intractable to explore all combinations of datasets. Therefore, looking for fingerprints may help if they are well-implemented in lower-dimensional spaces generated by fast and efficient tools.

3.6. Conclusion

Characterization of emerging semiconductor devices is essential to their successful integration into advanced technology nodes. The growing complexity of these devices has rendered traditional one-off device characterization impractical, even in high-throughput semiconductor foundries. Meanwhile, chip or module-level failure analysis, often involving huge volumes of data, requires a more sophisticated approach. A framework that captures the end-to-end characterization process in a formal way is needed. This framework must enable intuitive and faster characterizations by non-expert users, seamless sharing and optimization of effective characterizers across applications, and the realization of intelligent data-fusion engines that autonomously operate entire characterization pipelines.

The opposing nature of difficult-to-measure device physics and fast-paced commercial development poses challenges for semiconductor R&D. Rather than research-first proof of concept studies of potentially impactful methods, the presented work aims to share industrial-strength insights and tools to co-specify and realize semiconductor R&D pipelines capable of adapting to changing needs. These are polymer photoresist, etch selectivity characterizer, XFS reader, single-parameter interpolation, machine-learning classification, and image augmentation, as well as basic programming language principles. It should be emphasized that the proposed tools are neither unique nor groundbreaking on their own. Instead, they are intended as pipelines and interpreters at the software level for the application of modeling/construction methodologies such as finite element methods and discrete mechanics.

3.6.1. Emerging Trends

Advances in AI, machine learning, and design space exploration are providing unprecedented opportunities to accelerate semiconductor technology innovation. However, the pace of innovation is hampered by semiconductor research and development (R&D) bottlenecks in architecting, designing, fabricating, characterizing, and validating new devices and materials. Cutting-edge AI and machine learning implementations applied to semiconductor R&D are generating a surge of interest and effort across academia, industry, and the open-source community. Machine learning-enabled design exploration, cloud-enabled co-simulation, and high-throughput experimental characterization pipelines are some examples of machine learning tools developed to create, evaluate, and characterize new materials and devices. These machine learning tools and pipelines span computing and experimental domains. This section describes emerging trends in these pipelines and how the principles of data engineering can be applied to leverage, expose, and expand the impact of these tools. Semiconductor R&D pipelines for building and integrating machine learning tools are

labor-intensive, ad hoc, and often brittle. By adopting data engineering principles in the design, development, and maintenance of new machine learning pipelines, semiconductor industry players can better leverage and improve machine learning capabilities, enabling broader adoption and impact. Semiconductor production and design pipelines are not typically seen as software pipelines; rather, they are viewed as monolithic, bespoke, and non-reusable systems built by a few engineers in isolation. Current semiconductor production and design pipelines often lag in speed, flexibility, and feature richness compared to the compute domain. Existing pipelines primarily consist of one-off scripts and applications that are ad hoc; not designed to be reused or modified, requiring insider knowledge; and are difficult to maintain as historical knowledge is lost with personnel transitions. As a result, many workflows across the industry are executed on a one-off basis, with frequent regular interrupts in systems due to errors and slow speeds. Optimization-based design space exploration is used to efficiently find optimal designs. Challenges in accurate metrics of merit. Accurate performance metrics provided by analytic solvers, numerical solvers, or nonintrusive models as proxy of expensive-to-evaluate domains. Privacy or competition-sensitive intellectual property in design parameters. Inherently costly evaluations.

References

Schiller, R. J., & Larochelle, D. (2024). Data Engineering Best Practices: Architect robust and cost-effective data solutions in the cloud era. Packt Publishing Ltd.

Ketelaars, O. J. J. Moving to a decentralized organization by adopting data mesh principles: A review and proposal.

Raghunathan, S., Manukonda, K. R. R., Das, R. S., & Emmanni, P. S. (2024). Innovations in Tech Collaboration and Integration. Cari Journals USA LLC.

Begum, S., & Chowdary, F. (2024). Competitive Semiconductor Product Roadmaps. In Competitive Semiconductor Product Management: Navigating the Future of Semiconductor Technology in the AI Era (pp. 259-297). Berkeley, CA: Apress.

Xu, D., Zhang, Q., Huo, X., Wang, Y., & Yang, M. (2023). Advances in data-assisted high-throughput computations for material design. Materials Genome Engineering Advances, 1(1), e11.

Schilling-Wilhelmi, M., Ríos-García, M., Shabih, S., Gil, M. V., Miret, S., Koch, C. T., ... & Jablonka, K. M. (2024). From text to insight: large language models for materials science data extraction. arXiv preprint arXiv:2407.16867.