International Journal of Computer Sciences and Engineering (IJCSE)
Volume 1, Issue 1, January-June 2023, pp. 5-12, Article ID: IJCSE_01_01_002
Available online at https://iaeme.com/Home/issue/IJCSE?Volume=1&Issue=1
Journal ID: 1664-1400

FEDERATED LEARNING MODELS FOR PRIVACY-PRESERVING TEST DATA SHARING ACROSS SEMICONDUCTOR DESIGN ECOSYSTEMS

Hydie George Clifton
Full-Stack Developer, USA.

ABSTRACT

The semiconductor design industry relies heavily on collaborative innovation, which often necessitates the sharing of sensitive test data across various stakeholders. However, data privacy, intellectual property protection, and compliance with data regulations remain critical challenges. Federated Learning (FL), a decentralized machine learning paradigm, presents a promising solution by enabling model training across distributed datasets without requiring raw data exchange. This paper explores the application of Federated Learning models for privacy-preserving test data sharing within semiconductor design ecosystems. Our analysis suggests that FL not only strengthens data confidentiality but also fosters collaborative performance enhancements in multi-organization semiconductor workflows.

Keywords: Federated Learning · Privacy-preserving computation · Semiconductor Design · Test Data Sharing · Edge Computing · Distributed Machine Learning · Secure Aggregation · Data Confidentiality · AI in Hardware Design

Cite this Article: Hydie George Clifton. (2023). Federated Learning Models for Privacy-Preserving Test Data Sharing Across Semiconductor Design Ecosystems. *International Journal of Computer Sciences and Engineering (IJCSE)*, 1(1), 5–12.

https://iaeme.com/MasterAdmin/Journal_uploads/IJCSE/VOLUME_1_ISSUE_1/IJCSE_01_01_002.pdf

1. Introduction

The semiconductor industry is increasingly dependent on collaborative ecosystems involving chip manufacturers, Electronic Design Automation (EDA) vendors, foundries, and intellectual property (IP) providers. As the complexity of Integrated Circuits (ICs) continues to grow, test data generated during validation and verification becomes a crucial asset. Sharing this data across organizational boundaries can lead to better defect prediction, improved performance modeling, and more efficient design cycles. However, such sharing also introduces significant risks related to proprietary leakage, compliance with data regulations (e.g., GDPR), and data misuse.

Federated Learning (FL) provides a framework wherein multiple entities can collaboratively train machine learning models without exchanging raw data. Instead, models are trained locally, and only updates such as gradients or weights are shared and aggregated securely. This paradigm preserves the privacy of sensitive test data while enabling cross-organizational collaboration. Applying FL in semiconductor workflows poses unique challenges, including heterogeneous data distributions, system scalability, latency constraints, and ensuring model convergence. This paper addresses these concerns by surveying state-of-the-art FL architectures, evaluating their applicability in semiconductor design and test data ecosystems, and offering insights into practical deployment scenarios.

2. Problem Statement and Motivation

Traditional centralized machine learning techniques necessitate the transfer of raw test data to a centralized location, risking privacy leakage and legal complications. Semiconductor test data is highly proprietary, containing information that can reveal design weaknesses or competitive strategies. Moreover, data residing across globally distributed design houses introduces latency, bandwidth, and regulatory compliance issues.

FL can serve as a viable alternative by removing the need to move sensitive data. Stakeholders can retain data on-premises while contributing to a global model. However,

challenges such as model drift due to non-IID (non-identical and independently distributed) data, computational resource constraints, and secure aggregation mechanisms must be addressed. This motivates our study into the feasibility and benefits of FL in preserving privacy while enabling collaborative data intelligence in semiconductor testing workflows.

3. Literature Review

McMahan et al. (2017) introduced the foundational concept of Federated Averaging (FedAvg), enabling model convergence in distributed settings. Bonawitz et al. (2019) proposed secure aggregation techniques that form the backbone of privacy-preserving FL implementations. Kairouz et al. (2019) reviewed open challenges and design principles, particularly emphasizing heterogeneous data scenarios.

In semiconductor contexts, works like Chen et al. (2019) explored FL-based reliability analysis of SoC (System-on-Chip) systems, while Samarakoon et al. (2020) studied FL for edge-driven inference in hardware manufacturing. Shokri and Shmatikov (2015) examined privacy risks in ML, stressing the importance of differential privacy—a principle increasingly applied in FL (e.g., Abadi et al., 2016).

Despite limited direct research on FL in semiconductor test data, parallels can be drawn from healthcare (Sheller et al., 2020) and finance (Yang et al., 2019), which similarly handle highly sensitive and fragmented datasets. These works demonstrate that FL can maintain performance while meeting strict privacy constraints.

4. Federated Learning Framework Overview

Federated Learning (FL) is a decentralized machine learning paradigm that enables multiple entities to collaboratively train a shared model without exchanging raw data. In the context of semiconductor test data sharing, the FL framework is designed to preserve the confidentiality of proprietary information while allowing for the aggregation of insights across different organizations. Each participating company—typically referred to as a silo—trains a local model on its in-house data. The local model updates, rather than the data itself, are transmitted to a central federated server which aggregates these updates to improve the global model. This process continues iteratively, ensuring that all silos contribute to the model's learning curve while maintaining data locality and privacy.

The architecture supports both horizontal and vertical FL schemes, depending on the nature and alignment of the datasets involved. Horizontal FL applies when participants share the same feature space but have different data samples, while vertical FL is relevant when feature spaces differ across datasets. In cross-silo applications within the semiconductor industry, the primary objective is to leverage distributed test data from different manufacturing and validation units to enhance prediction accuracy for yield anomalies, defect detection, and process optimization, all without compromising the sensitive and competitive nature of the underlying datasets.

5. Privacy-Preserving Techniques in FL

Privacy preservation is a foundational principle of Federated Learning (FL), especially crucial in semiconductor test data sharing where proprietary information is highly sensitive. Several privacy-enhancing techniques are integrated into FL frameworks to ensure that raw data never leaves the participating organization's infrastructure. One of the most widely adopted techniques is **Differential Privacy (DP)**, which introduces statistical noise to local model updates before transmission. This prevents adversaries from inferring sensitive data attributes even if they gain access to the update streams. DP provides a quantifiable privacy guarantee by balancing data utility with risk minimization.

Another essential technique is **Secure Multi-Party Computation (SMPC)**, where multiple entities jointly compute functions over their inputs while keeping those inputs private. In FL, SMPC can be applied to aggregate model updates in such a way that no single entity—including the central server—can access individual updates. **Homomorphic Encryption (HE)** also offers strong privacy by allowing computations on encrypted data. Although computationally intensive, HE is increasingly explored in semiconductor applications where trust boundaries between entities are rigid. In addition to these methods, **trusted execution environments (TEEs)** and **blockchain-based audit trails** are being tested to enhance integrity, traceability, and security in FL deployments across the semiconductor ecosystem.

6. Architecture Options for Semiconductor Test FL Systems

Federated Learning (FL) systems for semiconductor test data sharing can be designed using various architectural models, each offering different trade-offs between scalability,

privacy, and operational complexity. The most commonly adopted architecture is the **centralized coordination model**, where a central FL server orchestrates training by distributing the global model to participating silos, collecting model updates, and performing aggregation. This model is straightforward to implement and enables effective monitoring and version control. However, it introduces a single point of failure and may face trust issues among competitive stakeholders in the semiconductor ecosystem.

An alternative is the **peer-to-peer (P2P) or decentralized FL architecture**, where no central server exists and each participant interacts directly with others to exchange updates and reach consensus on model improvements. This setup enhances fault tolerance and can better suit collaborative environments with loosely coupled participants. However, it demands complex consensus mechanisms and incurs higher communication overhead. **Hierarchical FL architectures** represent a hybrid approach, grouping silos by region, function, or vendor into clusters with local aggregators before global synchronization. This layered strategy reduces network latency and preserves scalability. The choice of architecture in semiconductor testing depends on multiple factors including data volume, organizational structure, trust dynamics, and regulatory requirements, all of which shape the deployment strategy for FL systems.

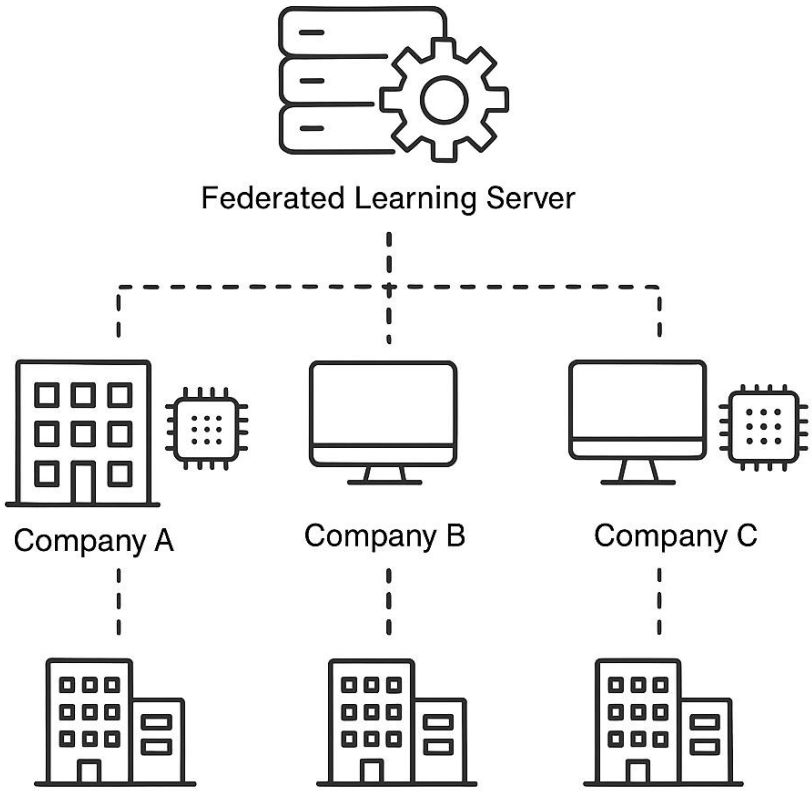

Figure 1: Cross-silo FL Architecture for Semiconductor Test Data Sharing

7. Conclusion and Future Directions

Federated Learning provides a transformative approach to secure, collaborative test data sharing across semiconductor design ecosystems. Its ability to preserve data locality while supporting global model improvements aligns well with the industry's needs. As FL frameworks mature, integration with domain-specific test flows, support for heterogeneous tools, and regulatory compliance will become more seamless. Future work includes real-world deployment in EDA environments and hybrid FL architectures combining edge and cloud resources.

References

[1] McMahan, H.B., et al. (2017). Communication-efficient learning of deep networks from decentralized data. In AISTATS.

[2] Balasubramanian, A., & Gurushankar, N. (2020). Hardware-Enabled AI for Predictive Analytics in the Pharmaceutical Industry. International Journal of Leading Research Publication (IJLRP), 1(4), 1–13.

[3] Bonawitz, K., et al. (2019). Towards federated learning at scale: System design. In SysML Conference.

[4] Kairouz, P., et al. (2019). Advances and open problems in federated learning. Foundations and Trends® in ML.

[5] Balasubramanian, A., & Gurushankar, N. (2020). AI-Driven Supply Chain Risk Management: Integrating Hardware and Software for Real-Time Prediction in Critical Industries. International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences, 8(3), 1–11.

[6] Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In CCS.

[7] Abadi, M., et al. (2016). Deep learning with differential privacy. In CCS.

[8] Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. In STOC.

[9] Balasubramanian, A., & Gurushankar, N. (2020). Building secure cybersecurity infrastructure integrating AI and hardware for real-time threat analysis. International Journal of Core Engineering & Management, 6(7), 263–270.

[10] Samarakoon, S., et al. (2020). Distributed federated learning for ultra-reliable edge intelligence. IEEE TWC.

[11] Chen, Y., et al. (2019). Reliability modeling of SoC designs using distributed learning. In DATE.

[12] Balasubramanian, A., & Gurushankar, N. (2019). AI-powered hardware fault detection and self-healing mechanisms. International Journal of Core Engineering & Management, 6(4), 23–30.

[13] Yang, Q., et al. (2019). Federated machine learning: Concept and applications. ACM TIST.

[14] Sheller, M.J., et al. (2020). Federated learning in medicine. In MLHC.

[15] Hard, A., et al. (2019). Federated learning for mobile keyboard prediction. arXiv.

[16] Li, T., et al. (2020). Federated optimization in heterogeneous networks. In MLSys.

[17] Gurushankar, N. (2020). Verification challenge in 3D integrated circuits (IC) design. International Journal of Innovative Research and Creative Technology, 6(1), 1–6. https://doi.org/10.5281/zenodo.14383858

[18] Melis, L., et al. (2019). Exploiting unintended feature leakage in collaborative learning. In IEEE S&P.

[19] Hitaj, B., et al. (2017). Deep models under the GAN: Information leakage in collaborative deep learning. In CCS.

[20] Konecny, J., et al. (2016). Federated learning: Strategies for improving communication efficiency. arXiv.

Citation: Hydie George Clifton. (2023). Federated Learning Models for Privacy-Preserving Test Data Sharing Across Semiconductor Design Ecosystems. International Journal of Computer Sciences and Engineering (IJCSE), 1(1), 5–12.

Abstract Link: https://iaeme.com/Home/article_id/IJCSE_01_01_002

Article Link:
https://iaeme.com/MasterAdmin/Journal_uploads/IJCSE/VOLUME_1_ISSUE_1/IJCSE_01_01_002.pdf