The Expanding Scale of Semiconductor Test Data

Chetan Arvind Patil
Contributing Writer I EPDT

Modern semiconductor manufacturing operates as a data-intensive ecosystem. Every wafer sort, functional test, burn-in and system-level qualification produces large, multi-domain datasets capturing device performance, reliability and process behaviour. On automated test equipment (ATE) platforms, thousands of parametric measurements are also collected for each device-under-test (DUT) across multiple operating conditions - including current and voltage characteristics, leakage distributions, timing margins, signal integrity metrics, embedded memory test signatures, etc.

As multi-site testing becomes commonplace, and OSATs integrate advanced test cell controllers with manufacturing execution systems and yield management platforms, data generated per product family can easily reach TBytes/day. This data trend will only increase, with more than 1 trillion semiconductor units shipped annually, each needing to undergo testing.

Alongside that, increasing complexity of silicon architectures and package arrangements, adoption of heterogeneous integration and use of stacked dies have all significantly expanded test specifications, resulting in thousands of tests per test program - driving both the volume and granularity of captured data. This test is also continuously streamed and aggregated into analytics platforms for yield learning, process correlation and reliability modelling. With AI-driven analytics now interpreting patterns across millions of devices, the test floor effectively functions as a live telemetry layer for manufacturing intelligence.

As test data scales and dimensionality grows, so does its value and exposure area. Each log, signature and dataset now carries embedded information about product architecture, design margin and process behaviour. Understanding the composition, flow and access paths of this data securely is therefore no longer optional, but a prerequisite for assured manufacturing integrity.

Why test data security matters in semiconductor manufacturing

Semiconductor testing generates huge data volumes. 300mm logic wafers can contain 60,000 die, each measured across hundreds/thousands of parameters under varying voltages and temperatures. This can result in 5GBytes of raw data per wafer at wafer sort. A high-volume test floor processing 30,000 wafers monthly may produce 150TBytes over that period, even before compression or post-processing. Furthermore, when combined with final test and system-level validation, a single product family can easily exceed 0.5PBytes of test data each quarter. Large assembly/test operations handling multiple product lines can accumulate multi-PByte datasets annually. As test data flows through ATE systems, manufacturing execution systems, yield management platforms and AI analytics engines, it is repeatedly replicated and reformatted. To ensure data is backed up, the lot file may coexist at multiple storage locations. Each replication not only increases the chance of corruption or inconsistency, but also expands the surface for unauthorised access.

The sheer scale and sensitivity of semiconductor test data make security paramount. A single breach can compromise design IP information, alter yield baselines, or contaminate AI models driving process adjustments. Thus, protecting this data ensures that every analysis, yield trend and model output reflects authentic device behaviour, preserving manufacturing integrity and upholding confidence across the semiconductor supply chain.

Data theft surfaces

Semiconductor test data moves through a complex, multi-domain ecosystem - originating at ATE rigs on the test floor, passing through fab-level MES and yield-management servers, transitioning to OSAT data systems for assembly and final test, before being consumed by analytics or AI platforms. Each transfer/transformation introduces

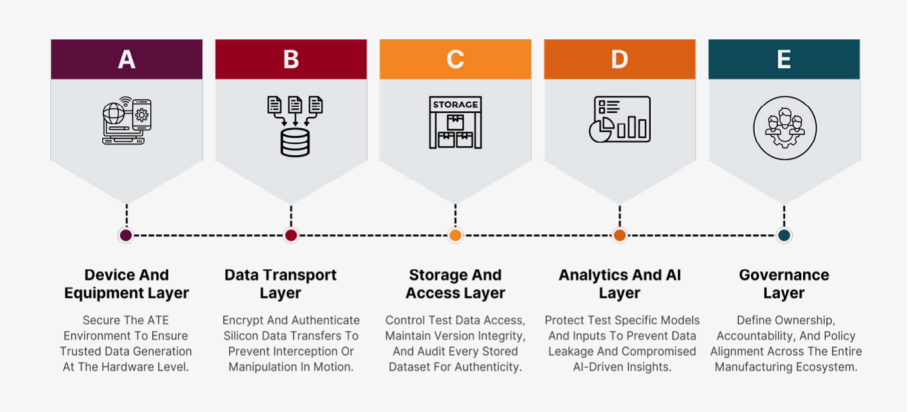

Figure 1: Protection layer coordination

Layer	Core Objective	Practical Impact
Device & Equipment	Secure test hardware, firmware and local control software at source.	Prevents unauthorised code execution, limits physical tampering and establishes hardware-rooted trust in ATE environments.
Data Transport	Protect confidentiality/integrity of data as it moves between manufacturing systems.	Maintains encryption and authentication across tester, MES and OSAT interfaces to prevent interception/manipulation during transfer.
Storage & Access	Enforce secure storage, access control and data provenance for all test results.	Provides traceability, detects unauthorised edits, and preserves audit-ready records that verify product and process integrity.
Analytics & AI	Safeguard aggregated datasets and derived models used for yield learning and process optimisation.	Protects against data leakage, model tampering, or cross-product inference that could reveal proprietary device behaviour.
Governance	Define ownership, accountability, and compliance alignment across all participating entities.	Ensures consistent policy enforcement and transparent responsibility from test floor to analytics (following NIST IR-8546 principles).

Table 1: Protection layer framework

distinct threat surfaces that expand the exposure footprint of manufacturing data. Key vulnerabilities are located at:

- **ATE -** Equipment controllers often operate on general-purpose operating systems with vendor-specific middleware. Outdated patches or unsecured service ports can enable malware intrusion or unauthorised remote sessions. Firmware tampering or misuse of diagnostic commands may alter test limits or corrupt bin data.
- **Fab -** Legacy tool controllers and shared networks can expose recipe parameters or metrology data if interfaces are left unpatched or unsecured.
- **MES -** Weak authentication between servers and external connectors can allow unauthorised queries or data extraction. Integration bridges often bypass IT monitoring, creating hidden exposure points.
- **OSAT -** Shared VPN or FTP channels, as well as multi-customer infrastructure, increase risks of cross-contamination or inadvertent exposure via misconfigured access controls.
- **Analytics/AI -** Aggregated multi-product datasets stored in cloud or hybrid analytics platforms are high-value targets. Data contamination, unauthorised model training, or exfiltration of derived feature maps can expose design-sensitive patterns and compromise the reliability of analytics.

The interconnected nature of these elements means a single weak link can compromise the entire test data chain. An ATE connectivity vulnerability can propagate upstream into MES databases or downstream into analytics models, undermining yield accuracy and data confidentiality. Effective risk management therefore requires visibility across every hand-off point, understanding not only how test data is generated, but also how it moves, transforms and is ultimately consumed. Recognising threat surfaces is the 1st step towards building a resilient protection framework that secures manufacturing intelligence.

Establishing practical protection layers

Securing semiconductor test data requires a defence architecture built across every manufacturing stage. Everything from ATE hardware/firmware to analytics platforms and governance policies serves a specific function in maintaining confidentiality, integrity and availability. Because test data moves across multiple systems and organisations, any single weakness can expose sensitive electrical signatures or corrupt yield analytics. Coordinated, multi-tier protection models ensure every point in the data pipeline is both visible and controlled. Table 1 illustrates how coordinated protection layers can collectively enhance data trust and operational stability.

Effective implementation of these layers requires active coordination among IT teams, product engineers and manufacturing operations. Security must be designed into workflows, not added after deployment. When executed effectively, this approach transforms test data from a potential vulnerability into a verified manufacturing intelligence source. It builds trust across fabs, OSATs and analytics partners, ensuring every dataset and AI-derived insight originates from authentic, tamper-free information.

AI-era challenges to test data integrity

AI-driven analytics is transforming semiconductor test data. Models trained on wafer sort, burn-in and final test datasets now reveal correlations that once required extensive manual analysis. These models enhance yield prediction and adaptive test control. However, when shared across fabs and OSATs, they can unintentionally expose design behaviour or process signatures if not adequately secured. Model leakage is thereby at critical risk.

Feature embeddings and learned parameters can retain statistical fingerprints of device architecture and process variation. Studies have demonstrated that model inversion attacks can recover sensitive data with >70% accuracy, effectively turning the model itself into a channel for IP exposure. Other threats include training data manipulation (which can bias prediction models), plus inference theft (where external AI services deduce process behaviour via repeated queries). Privacy-preserving techniques (like differential privacy and federated learning) help reduce exposure, but may limit diagnostic precision.

As semiconductor manufacturing moves into the AI-native era, test data becomes both the engine of optimisation and a critical vulnerability. Every parametric signature, fail map and yield correlation will continue to feed AI models guiding process and product decisions. Securing this data, from tester output to model training, is thus essential to protect design integrity and prevent misuse of device behaviour patterns.