Meta Silicon Infrastructure and evolution with AI

Authors : Salina Dbritto, Rupa Raghavan, Farishta Mahzoz

I : ABSTRACT

Meta is developing its own specialized silicon to support AI Infrastructure. Moving into the AI domain necessitated a shift in thinking, perspective, and a paradigm change in our approach to test and validate the AI systems and platforms. Where previously a focus on individual component validation and infrastructure was adequate, we now need to broaden our infrastructure scope to accommodate AI systems, transitioning from a component-centric to a system-level outlook. This involves integrating computing, networking, and liquid cooling strategies, and constructing infrastructure in labs that not only support component validation but also explore methods to assess integrated racks and interoperability capabilities.

II : INTRODUCTION

Shifting from component-based to AI system-based validation infrastructure poses significant challenges, and we are crafting solutions to address the needs of Meta's AI infrastructure scale. The silicon automation infrastructure is designed to be portable and scalable, allowing for seamless integration across various development and validation phases and environments. This includes pre-Silicon validation in the emulation environments, post-Silicon validation in the engineering labs, data centers during NPI phase, deployment in the fleet during MP phase, and ODM/Vendor qualifications.By standardizing our automation tools across multiple geographical regions and platforms, we provide engineers with a consistent user experience, ultimately increasing developer velocity. Additionally, we have shifted many of our validation efforts to earlier stages of the NPI cycle, minimizing issues that may arise in the fleet. The robust infrastructure, standard automated tools, and processes have played a crucial role in improving and enhancing our infra capabilities to support integrated AI systems for next-generation Silicon programs

.

III: SILICON INFRA

AI systems validation with Meta custom Silicon happens in emulation environments, engineering labs, data centers, and at ODM's/Vendor locations. In order to ensure effective validation signals at an accelerated pace across all these phases, the tests are developed to be environment agnostic. The Silicon Automation infrastructure is built to be scalable and portable across different geographical regions, platforms and validation environments.

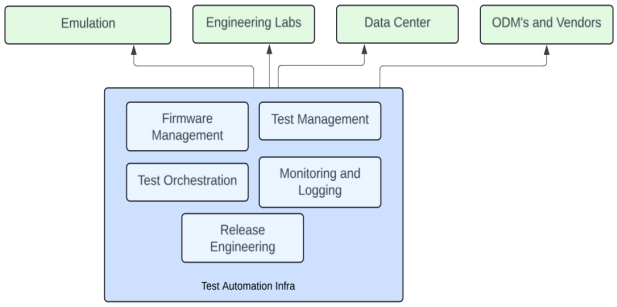

Figure (1)

The automation infrastructure has enabled us to have 5x fully remote, standardized, and seamlessly accessible labs that greatly improves developer velocity and increases the test execution volume by >10x compared to previous programs.

Our primary focus is to left-shift the validation workflows and enable E2E AI workloads as early as possible in the New Product Introduction (NPI) phase and this is achieved by the infrastructure that allows developers to validate these tests in any given environment and seamlessly port it over to another. This approach ensures that the tests are consistent, high quality, and provides the necessary signals from early on in the product life cycle.

A robust continuous integration release engineering flow gates the SW/FW releases and maintains the health of the systems and enables better utilization. Over 95% of the Silicon tests are automatically error categorized and this allows for faster debug/triage flows.

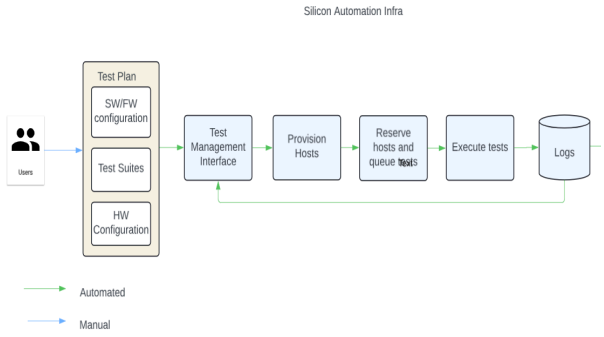

Figure (2)

IV : ENGLAB INFRASTRUCTURE

Before NPI validation, over 1 million tests were executed during the bring-up and post-silicon phases in our engineering labs, showcasing the rigorous and thorough validation process. These tests were conducted using internal tools and processes, ensuring consistency, efficiency, and control over the testing environment. The 2.5x increase in programs from 2020-2022 to 2022-2024 indicates a significant expansion of the silicon portfolio. Additionally, the 3x increase in lab users suggests that more teams and individuals are utilizing the lab's resources, indicating increased collaboration and knowledge sharing across the organization.The 2.5x increase in lab locations implies that the infrastructure is expanding to accommodate the growing needs of the organization, providing more opportunities for teams to work on new AI projects and initiatives. Furthermore, the 6x increase in devices brought up compared to 2020-2022 demonstrates the team's ability to adapt to new AI technologies and scale up operations to meet the demands of the rapidly evolving industry.

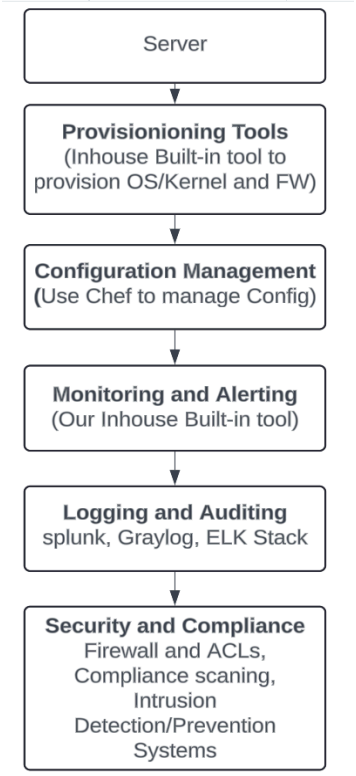

Figure (3)

Figure (3) shows the various tools and systems used in the bring up and management of servers at Meta. The provisioning tools are used to initially configure and set up the servers, while the configuration management tools are used to manage and maintain the servers' configurations over time. The monitoring and alerting tools are used to track the health and performance of the servers, while the logging and auditing tools are used to collect and analyze log data. Finally, the security and compliance tools are used to ensure the security and compliance of the servers with relevant regulations and standards.

V: AI EVOLUTION

Previously, our primary focus was on component validation, which involved evaluating the performance of individual units as devices under test. For instance, we would test a switch in a chamber to assess its power and performance. However, this approach proved insufficient as we encountered numerous issues in the fleet related to link flakiness, leading us to shift towards interop setups. Interop setups are focused on bench top setups with network and compute components and evaluating links and SerDes issues. While this method was efficient, it was not enough to service integrated AI racks which

are massive systems with compute, network, rack management, liquid cooling solutions all compacted as a single system.

AI systems consist of integrated compute, network, rack management, TOR all in a single rack. With switch bank and accelerator back, switch backend is connected to accelerators with common backplane as shown in Figure(4) In order to facilitate system-level validation, we had to establish the necessary infrastructure that combines network and compute workflows.

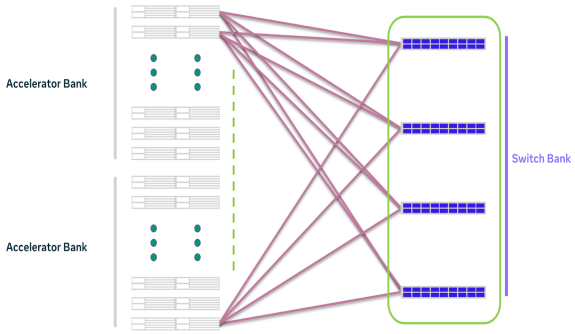

Figure (4)

Figure(5) shows examples of system integration test scenarios.

Case	Description
Link stability tests	Identifying all links within a given AI rack are healthy
System health	Identifying system health signals to determine every component is usable
Traffic tests	Running traffic tests and verifying counters
leak detection	Analyzing leak detection solution due to liquid cooling racks with AI racks
Pre silicon	Test scenarios at AI system level before Silicon arrives
Post silicon	Test scenarios at AI system level after silicon arrival

Figure (5)

Inorder to build automation for system integration tests, traditional methods were not enough which were component focused. We had to combine network and compute orchestrators to exercise rack level validation cases as shown in Figure(6)

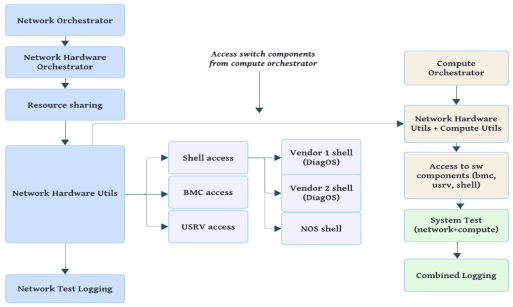

Figure (6)

Network orchestrator: A tool designed to facilitate the execution of network tests, with a focus on accessing switch components to assess BMC and USRV interfaces, as well as evaluating power and performance. Compute orchestrators, on the other hand, are focused on launching compute tests. For AI rack level validation, it is necessary to interface with both network and compute components in order to build rack level traffic tests.

A common diagnostic approach is used across different platform types, with integrated solutions built for provisioning, firmware, asset management, physical lab spaces and tooling stacks. Ensuring backplane stability in integrated racks is crucial, and building common integrated tooling for network and compute has allowed us to achieve rack level tests.

VI: CONCLUSION

In summary, we want to emphasize that while silicon development is rapidly expanding and essential for our success, it's crucial to develop tooling and test infrastructure that supports the latest programs. We need to shift our focus from solely validating components to building comprehensive solutions across the stack, including frameworks, asset management, physical lab spaces, firmware management, CI flows, and more. This will enable us to effectively support the development of integrated AI rack solutions, which are critical for the successful deployment of new silicon technologies.