Specification Mining facing Generative AI

Goerschwin Fey
Hamburg Univ. of Tech., Germany

Harry Foster
Siemens EDA, Plano, TX, USA

Tara Ghasempuri
TalTech, Tallinn, Estonia

Badri Gopalan
Synopsys Inc, Sunnyvale, CA, USA

Jörg Müller
Cadence Design Systems Inc., Munich, Germany

Manish Pandey
Synopsys, Inc., Sunnyvale, CA, USA

***Abstract*—Specifications for complex designs and their consistency are always a headache. Automated specification mining – including but not limited to generative AI – offers attractive solutions, but there are also various unmet needs.**

***Index Terms*—Design, Specification, Documentation, Verification**

I. Opportunities

Figure 1 illustrates how Ⓐ starting from some reference documents, Ⓑ a mining approach, automatically derives Ⓒ an executable or formal specification, that then Ⓓ has to be validated with respect to the actual design and design intent. A promise is that by this, a design, its implementation, specification, and documentation can be kept consistently without diverging from each other. Moreover, manual tasks like writing assertions in a specific language or coding a testbench are more efficiently handled. The most prominent use case so far is regression testing.

Various approaches specification mining exist including simulation trace based assertion inference, source code guided extraction, or generative Artificial Intelligence (AI). However, there are unmet needs that we group into general concerns and – more recently – those regarding generative AI.

A. General Concerns

Whenever considering formal or informal specification their quality matters. For mined specifications, correctness is difficult to guarantee and depends on the reference data Ⓐ used. A major pitfall may be circular reasoning that generates false specifications from erroneous data satisfying buggy implementations. Thus, a proper independent reflection process Ⓓ – whether automated or manual, structured or AI supported – must validate the results.

Completeness is another concern. While a set of assertions may completely specify observed behavior or cover a textual reference document, underrepresented behavior cannot be mined at all. This may concern specific features missing in the input data or non-functional properties. Security belongs to this group being underdocumented and subject to unknown attacks. Thus, there is a need for methodologies to efficiently adopt security specifications across different designs. RISC-V is gaining popularity in processor design. Thus, a methodology tailored to RISC-V that translates security assertions probed on real-world attacks from one RISC-V architecture to another would be an appealing solution.

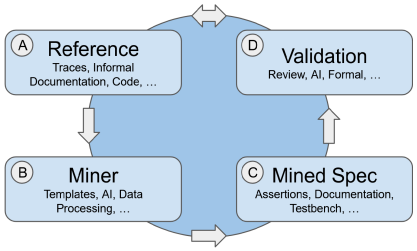

Fig. 1. Specification Mining

Users' expectations for specification mining tools must also be managed. Effective integration of mining methodologies into workflows requires user education to understand the limitations and strengths. Comprehensibility of complex specifications is intrinsically difficult and may benefit from presentation approaches beyond code, assertion languages, or text.

B. Generative AI

Generative AI can extract draft specifications from vague or informally defined inputs ranging from design documents and code to emails. Several critical gaps must be addressed.

First, generative AI must handle ambiguity and incorporate external context, such as prior project knowledge or hardware simulation traces, into its mining processes. Advances at the block or subsystem level are encouraging, but scaling to System-on-Chip (SoC) levels remains a challenge. Second, mined specification must be mathematically precise and verifiable to align with actual designs. Finally, the lack of sufficient high-quality training data or reusable methodologies is an issue. Due to the confidentiality of design data, access to real-world examples is limited. Open-source data and standardized processes are essential for robust tools.

The overall success depends on integrating AI capabilities with traditional EDA tools, which excel at analyzing hardware design details such as cones of influence and clock/reset domains. These tools can verify the structure and function of AI-generated assertions, ensuring alignment with design implementations.

II. Conclusions

Specification mining is a powerful tool. Generative AI provides opportunities even for training engineers, but is still far from being the one-catch solution. General concerns like comprehensibility, security, completeness and circular reasoning must be solved agnostic of the mining approach.

978-3-9826741-0-0/DATE25/© 2025 EDAA