

CorrectHDL: Agentic HDL Design with LLMs Leveraging High-Level Synthesis as Reference

Kangwei Xu¹, Grace Li Zhang², Ulf Schlichtmann¹, Bing Li³

¹ Technical University of Munich (TUM), Munich, Germany

² Technical University of Darmstadt, Darmstadt, Germany

³ Technical University of Ilmenau, Ilmenau, Germany

Abstract

Large Language Models (LLMs) have demonstrated remarkable potential in hardware front-end design using hardware description languages (HDLs). However, their inherent tendency toward hallucination often introduces functional errors into the generated HDL designs. To address this issue, we propose the framework *CorrectHDL* that leverages high-level synthesis (HLS) results as functional references to correct potential errors in LLM-generated HDL designs. The input to the proposed framework is a C/C++ program that specifies the target circuit's functionality. The program is provided to an LLM to directly generate an HDL design, whose syntax errors are repaired using a Retrieval-Augmented Generation (RAG) mechanism. The functional correctness of the LLM-generated circuit is iteratively improved by comparing its simulated behavior with an HLS reference design produced by conventional HLS tools, which ensures the functional correctness of the result but can lead to suboptimal area and power efficiency. Experimental results demonstrate that circuits generated by the proposed framework achieve significantly better area and power efficiency than conventional HLS designs and approach the quality of human-engineered circuits. Meanwhile, the correctness of the resulting HDL implementation is maintained, highlighting the effectiveness and potential of agentic HDL design leveraging the generative capabilities of LLMs and the rigor of traditional correctness-driven IC design flows. This work is open-sourced at the following link: <https://github.com/AgenticHDL/CorrectHDL>

1 Introduction

With the growing complexity of integrated circuits, hardware engineers are required to invest increasing effort in HDL design [1]. Traditional hardware design flows typically begin with engineers interpreting algorithm specifications and then manually coding them into corresponding HDL designs. As shown in Fig. 1(a), this process is highly labor-intensive: developers should write HDL designs by hand, perform simulation, and debug errors to ensure functional correctness, resulting in a time-consuming and costly process [2].

Recently, large language models (LLMs) have demonstrated significant potential in enhancing the efficiency of HDL design [7–21]. By learning expert design patterns, LLMs can generate HDL designs directly from specifications with significantly reduced manual effort [22–30]. Chip-Chat [7] presents the first systematic study of conversational LLMs for interactively co-designing an 8-bit accumulator microprocessor under real-world constraints. Rome [8] introduces hierarchical prompting and an automated generation pipeline that enables LLMs to reliably produce multi-level HDL designs. GPT4AIGChip [9] leverages LLMs for AI accelerator design by decoupling circuit modules with in-context learning. RTLFixer [10] introduces a RAG-based debugging flow that repairs syntax errors in HDL designs. AssertLLM [31] generates functional assertions from specifications, while *CorrectBench* [32, 33] employs LLMs to build a hybrid evaluation platform with a self-correction mechanism. The frontier of LLM-assisted hardware design is further advanced in

High-Level Synthesis (HLS) [34–40], C2HLSC [34–36], HLS-Repair [37], and HLSPilot [38] use LLMs to iteratively refactor C/C++ programs into HLS-compatible versions with tool-guided feedback.

Despite the advances of HDL design with LLMs, the inherent hallucinations of LLMs frequently introduce syntactic and, more critically, functional errors into the generated HDL designs [7]. While syntax issues can often be detected by compilers and fixed by iterative repair loops [10], functional debugging remains fundamentally challenging. Current LLM-based flows lack a bit-accurate reference at the same abstraction level as the generated HDL, forcing engineers to manually interpret algorithm specifications, analyze simulation waveforms, and infer intended behaviors to locate functional bugs. Consequently, the effort required to validate and correct LLM-generated designs can partially offset the benefits of automated code generation, thereby limiting the practical adoption of LLMs for producing functionally correct HDL in industry [25, 26].

To address the challenge above, we introduce a novel agentic design framework, *CorrectHDL*, for LLM-assisted HDL design and debugging using the results of high-level synthesis (HLS) as a bit-accurate reference to overcome functional errors. To the best of our knowledge, this is the first work leveraging the generative capabilities of LLMs and the rigor of traditional correctness-driven synthesis flows to support LLM-assisted HDL design and debugging. The key contributions of this paper are summarized as follows:

- An agentic HDL design framework, *CorrectHDL*, is proposed and released as open source. By leveraging HLS-generated HDL as a functional reference, *CorrectHDL* implements an end-to-end design flow that covers HDL generation, runtime profiling, differential verification, iterative repair, and PPA evaluation, enabling accurate and efficient hardware design.
- To mitigate attention dilution in LLMs, complex C/C++ algorithms are decomposed into LLM-friendly submodules through a structured C/C++ decomposition strategy, which also provides a basis for subsequent differential debugging.
- By leveraging the adaptive learning capability of LLMs, a Retrieval-Augmented Generation mechanism is introduced to correct syntax errors in the LLM-generated HDL designs, increasing the compilation pass rate by an average of 15.49%.
- To address functional errors caused by LLM hallucinations, a differential verification and repair loop is established using HLS-generated HDL as the functional reference, improving the functional pass rate of LLM-generated HDL by 28.05%.
- To efficiently debug the LLM-generated top-level HDL design, submodule boundary instrumentation is combined with a backward slicing technique to locate errors from the system level down to individual modules, thereby bringing the overall functional pass rate improvement to 38.54%. Experimental results also demonstrate that the proposed *CorrectHDL* achieves an average area reduction of 24.83% and power reduction of 26.98% compared with conventional HLS designs and approaches the quality of human-engineered circuits.



Figure 1: Traditional manual HDL design flow.

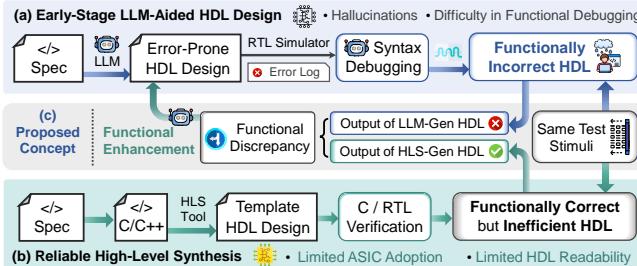


Figure 2: (a) LLM-aided HDL design flow; (b) High-Level Synthesis (HLS) flow; (c) Our proposed concept.

The paper is organized as follows. Section 2 provides the background and motivation. Section 3 details the framework. Section 4 shows the experimental results, and Section 5 concludes the paper.

2 Motivation and Concept

To address the functional errors in the HDL design generated by LLMs, an effective approach is to use a reference HDL design to guide the code generation process of LLMs. Such a functional reference should be able to establish the mapping from a high-level description to HDL that accurately reflects the intended circuit behavior. The performance, power, and area (PPA) of such a reference are not critical, because only the HDL generated by LLMs is used in real circuit design. To create such a reference, the appropriate format of the high-level description and a verified toolchain that can generate an HDL design from this description should be identified.

Over the past decades, the EDA community has proposed a wide range of solutions for circuit design based on abstract behavioral or algorithmic descriptions, among which high-level synthesis (HLS) has emerged as the most prominent approach for translating C/C++ programs into HDL designs [2, 3]. Modern HLS toolchains are mature and highly reliable, achieving notable success in both industry and academia [4, 5]. Unlike direct software-level simulation, the HLS-generated circuit reference preserves precise hardware semantics, accurately handling customized bit widths, type conversions, bit truncation/rounding modes. This eliminates the semantic ambiguities inherent in software simulation. However, the underlying compilation flows of HLS tools typically rely on fixed templates and coarse-grained pragma tuning, leading to suboptimal area and power efficiency of the synthesized circuit [44, 45]. In addition, the HDL designs generated by HLS are often difficult to interpret, which makes late-stage Engineering Change Orders (ECOs) challenging in ASIC-oriented physical design [46]. Consequently, for high-performance designs, engineers still prefer to manually craft HDL based on the algorithm descriptions to meet stringent PPA requirements [48].

Given the ability of HLS to generate functionally correct HDL designs, our agentic design framework, *CorrectHDL*, takes advantage of the results of HLS to guide LLMs in automatic HDL generation. In this framework, the generative capabilities of LLMs are leveraged to produce flexible HDL, while the HLS tool provides a bit-accurate golden reference to guide the functional correctness of LLM-generated HDL. Since LLMs do not use predefined circuit templates but rely on abstracted design knowledge, they can flexibly create circuits that outperform HLS-generated designs in terms of PPA and approach the quality of human-engineered circuits.

```
#pragma design_goal area (a) C++ Code (161 Lines)
1: void aes256 (aes256_context *ctx, int k[32], int buf[16])...
2:   for (i = 0;i < 32;++i) {
3:     ctx->enckey[i] = k[i];ctx->deckey[i] = k[i];
4:   for (i = 8; --i;) { expandEncKey(ctx->deckey, &rcon); }
5:   addRoundKey_cpy(buf,ctx->enckey,ctx->key);.....161: }

1: module aes256 (b) LLM-Generated HDL (386 Lines)
2:   input clk, start, input [255:0] key_in, output...;
3:   reg [7:0] buf[0:15], enckey[0:31], key[0:31]; reg [7:0] rcon;...
4:   always @ (posedge clk or negedge rst_n) begin
5:     if (!rst_n) begin busy='1'b0; i='4'd0; rcon='8'd1; end
6:   else if (start && !busy) begin
7:     load_state(key_in,buf); load_key(block_in,enckey); Error
8:     load_key(key_in,key); addRound_cpy(.....386: endmodule

1: module aes256 (c) HLS-Generated HDL (1879 Lines)
2:   clk, rst, ctx_key_triosy_lz, ctx_enckey_triosy_lz, ctx_de...
3:   buf_rsc_zout, buf_rsc_lzout, buf_rsc_zin, buf_triosy_lz...);
4:   input [7:0] EnKey_read_rom_sbox_rom_map_1_cmp_data_out, ...
5:   wire aes_expandEncKey_1_i_aes_expandEncKey_1_i_mux_nl;...
6:   assign cpkey_mux1h_159_nl = MUX1HOT ((xor_1_cmx_sva_1[...]),
7:   (addkey_1_io_read_buf_rsc_sd_1_sval[...]));
8:   (EnKey_read_rom_sbox_rom_map_1_cmp_data_out[...]),
9:   mix_mix_xor_7_cmx_7_5_sva_1, (buf_rsci_din[...]),
10:  (ctx_key_rsci_data_in_d_mx0w1[...])..... 1879: endmodule
```

Figure 3: Examples of HDL Generation via LLM and HLS.

The concept of *CorrectHDL* is illustrated in Fig. 2. The circuit specification written in C/C++ is first processed by an HLS tool, which produces a functionally correct HDL design that serves as the golden reference. The functional behavior of the HDL generated by the LLM is then compared with that of the reference design from HLS. The discrepancy is used to guide the LLM to iteratively enhance the HDL designs to achieve functional correctness.

The input to the *CorrectHDL* framework consists of circuit descriptions expressed in C/C++, which provide a more detailed specification than the natural language inputs used in many state-of-the-art solutions [22–30]. This choice makes the framework directly applicable to system design engineers who typically work at the algorithm level. Meanwhile, natural language specifications can be automatically translated into C/C++ using LLMs, which have demonstrated high accuracy in code generation within the software community [41, 42]. Accordingly, the proposed *CorrectHDL* effectively bridges an important gap in LLM-based circuit design.

The output of the *CorrectHDL* framework is a circuit generated by the LLM. Its functionality is iteratively improved by comparing its behavior against a reference HDL produced by HLS. The correctness of this reference HDL is validated using a testbench that the HLS tool automatically derives from the original C/C++ testbench. However, this converted testbench is coupled to the tool-specific interfaces of the synthesized design, so it cannot be directly used to simulate the circuit generated by the LLM agent. To address this issue, we also employ LLMs to translate the original C/C++ testbench into an HDL testbench that is compatible with the LLM-generated HDL. Limited human effort is still required during this process to ensure the quality of the resulting testbench. Functional comparison is then performed by evaluating the output data of the LLM-generated circuit against that of the HLS reference under the same test stimuli.

Fig. 3(b) illustrates an example of an LLM-generated HDL design from the C/C++ program in Fig. 3(a), which exhibits good readability. However, limited hardware-oriented training data and LLM hallucinations often lead to functional errors (line 7). On the contrary, the code in Fig. 3(c) generated by HLS is functionally correct, but it has far inferior readability and PPA. The objective of *CorrectHDL* is thus to enhance the functional correctness of the LLM-generated HDL design using the HLS-generated HDL design as a reference.

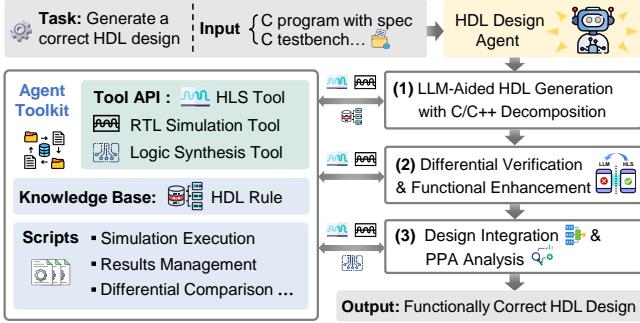


Figure 4: Proposed HDL design agent.

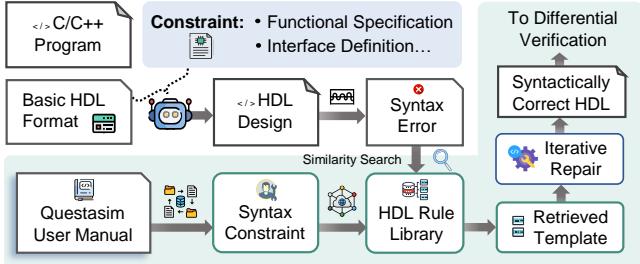


Figure 5: LLM-assisted HDL generation.

3 Agentic HDL Design with LLMs Leveraging HLS as Reference

The overall workflow of the proposed HDL Design Agent is shown in Fig. 4. Given a C/C++ program with its specification and a corresponding C/C++ testbench from the HLS benchmark suite [52, 53], the agent executes a three-stage process to produce a functionally correct HDL design. The agent interacts with a toolkit comprising tool APIs (HLS, RTL simulation, logic synthesis), a knowledge base of HDL syntax rules, and automation scripts for simulation and comparison. In stage (1), described in Section 3.1, the C/C++ program is decomposed into submodules according to the rules in Section 3.1.1. For each submodule, the LLM generates a corresponding HDL module (Design Under Test, DUT), and syntax errors are repaired through the RAG mechanism. In stage (2), described in Section 3.2, each C/C++ submodule is synthesized by the HLS tool into a functionally correct HDL as the golden reference. The DUT and the golden reference are then simulated under the same stimuli. During this process, HLS tools can automatically translate the original C/C++-based testbench into an equivalent HDL testbench to validate the HLS-generated HDL design, whereas the testbench for the LLM-generated HDL design is adapted from the original C/C++ testbench using the LLM and then verified by humans. Any detected functional mismatches are fed back to the LLM for iterative enhancement. In stage (3), described in Section 3.3, all verified LLM-generated submodules are instantiated into a top-level design, and differential verification is repeated to ensure integration correctness.

3.1 LLM-Assisted HDL Design Generation

3.1.1 C/C++ Program Decomposition. Directly feeding an entire C/C++ program with long context into the LLM tends to dilute attention, preventing it from capturing fine-grained hardware semantics, e.g., customized bit widths, type conversions, etc. This often leads to deep functional errors that are difficult to debug. To mitigate this issue, a divide-and-conquer strategy based on C/C++ decomposition is adopted before HDL generation. The original C/C++ program is decomposed into smaller, LLM-friendly C/C++ submodules, thereby improving the quality of the initial HDL design generation.

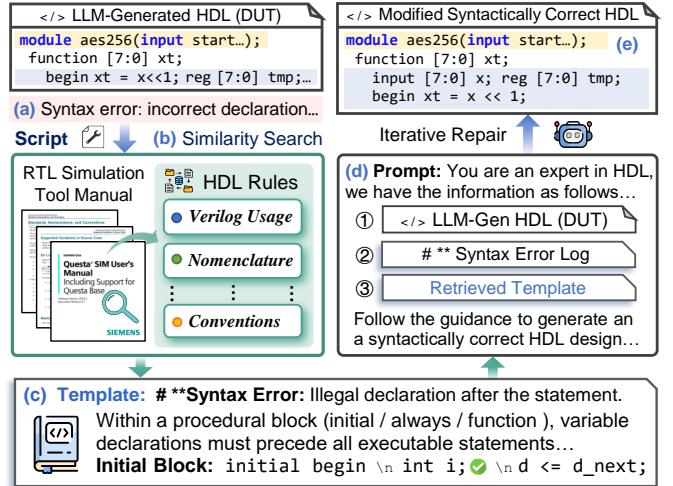


Figure 6: Syntax repair for LLM-generated HDL using RAG.

The decomposition follows three strict rules: (1) Function-Level Granularity: Decomposition is performed along function boundaries, rather than by arbitrary line-level splitting. (2) Explicit I/O definition: Each submodule uses fixed-width scalars or static arrays as interfaces (e.g., `int8_t state[16]`), with all bit widths explicitly specified. (3) Single and Clear Semantics: Each submodule should have clear semantics (e.g., `SubBytes()` performs only SBox substitution).

To ensure correctness after decomposition, a differential verification step is performed. All decomposed C/C++ submodules are re-integrated into a top-level C/C++ program and verified against the original C/C++ testbench. The decomposition is considered successful only when the output results match exactly.

3.1.2 LLM-Assisted HDL Design Generation. After decomposition, the agent proceeds to the HDL generation phase as shown in Fig. 5. For each C/C++ submodule, we generate two specification files in natural language using the LLM. (1) Functional Specification: A detailed description of the functionality and expected hardware behavior of the C/C++ submodule. (2) Interface Definition: Precisely maps C/C++ parameters to Verilog ports, including names, directions, and exact bit widths. With these two specification files, the LLM can better understand the design constraints of individual submodules and produce valid HDL designs from the C/C++ snippet effectively.

To guide the LLM toward producing effective HDL designs, we also define the design, and formatting constraints as follows: (1) Separating control logic from the datapath and using synchronous reset conventions; (2) Optimization strategies for the circuits, e.g., inserting pipeline stages; (3) Formatting constraints such as enclosing the resulting HDL between triple backticks to facilitate HDL extraction from the output of the LLM using a script. The C/C++ submodule code, functional specification, interface definition, design and formatting constraints are jointly provided to the LLM, guiding it to generate the corresponding HDL design of a submodule.

Due to LLM hallucinations, the HDL design generated for a submodule often fails syntax compilation. As shown in Fig. 5, these errors can be repaired by a retrieval-augmented generation (RAG) mechanism, which involves three stages: First, an external rule library is created containing HDL syntax rule templates. Each template includes error logs that may be triggered during syntax compilation, and a summary of the corresponding repair rule extracted from the user manual. Second, when an LLM-generated HDL design fails compilation, the error log is embedded and then matched against the templates in a rule library using a sentence transformer

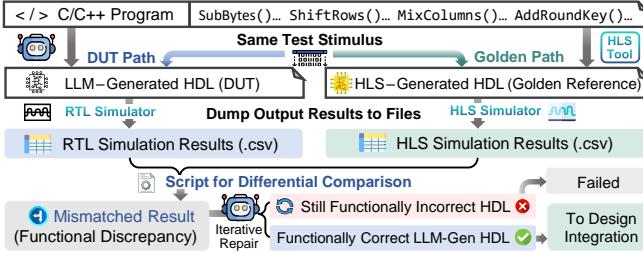


Figure 7: Differential verification for LLM-generated HDL.

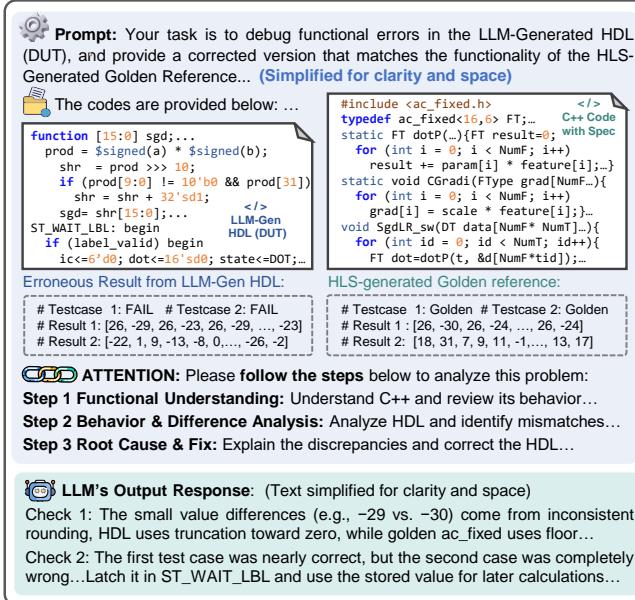


Figure 8: Functional enhancement in LLM-generated HDL.

[50]. The retrieval process can be formalized as follows:

$$E = \text{embed}(\text{error_log}), \quad R_i = \text{embed}(\text{rule_template}_i), \quad (1)$$

where E and R_i denote the embedding vectors of the error log and the i th rule template. Given an error log E and a template R , the rule template with the highest similarity is selected as follows:

$$\text{sim}(E, R) = \frac{E \cdot R}{\|E\| \|R\|}, \quad R^* = \arg \max_{R_i \in \mathcal{R}} \text{sim}(E, R_i), \quad (2)$$

where \mathcal{R} denotes all templates in the library, R^* represents the template with the highest similarity. Third, once the most relevant rule template is retrieved, it is incorporated into the LLM prompt to guide the repair of the HDL design into a syntactically correct version. An example of repairing an LLM-generated HDL design is given in Fig. 6. The agent uses compiler logs to retrieve repair rules and then prompts the LLM to repair the HDL, and repeats this process until compilation succeeds or the predefined iteration limit is reached.

3.2 Differential Verification and Functional Enhancement

After syntax errors in LLM-generated HDL designs are corrected as described in Section 3.1.2, we leverage the HLS-generated HDL as a golden functional reference to enhance its functional correctness, as shown in Fig. 7. For each C/C++ submodule, the HLS tool synthesizes a bit-accurate and functionally correct HDL design as a reference. Despite its suboptimal area/power efficiency and limited readability, its functional behavior is highly reliable [2–5] and thus serves as an effective golden reference for verifying LLM-generated design.

TABLE 1: Typical Functional Discrepancies in LLM-Generated HDL and the Corresponding Corrections Driven by *CorrectHDL*.

Task	DUT Output	HLS Reference	Functional Discrepancy	Corresponding Correction
Edge	[−3, 66, ..., −250, 108]	[−31, 664, ..., −2479, 1081]	Numeric scaling bug	Correct shift scaling
Markov	[5, 22, 17, ..., 9, 14, 14]	[5, 22, 17, ..., 9, 14, 15]	Boundary write-back bug	Commit delayed by one cycle
DMM	[262144, 0, 4096, ...]	[262144, 4096, 133120, ...]	Reset or delay bug	Decouple load and compute
AES	1: [xx, xx, xx, ..., xx, xx] 1: [8e, a2, b7, ..., 60, 89] 2: [8e, a2, b7, ..., 60, 89] 2: [6b, c1, be, ..., ef, 63]	Lag due to synchronous assignment timing bug	Align output register with next state or delay valid	
KMP	[0, 8, 0, ..., 8, 0, 8]	[7, 64, 9, ..., 59, 0, 560]	Prefix skipped by slicing	Compare full constant
N-body	[1, 0, 1, 1, 1, ..., 0, 1, 1]	[1, 1, 0, 1, 1, ..., 1, 0, 1]	Inter-beat misalignment	Use local blocking temps
BFS	[1, 0, 3, 0, 2, 1, 0, 1, ...]	[1, 1, 4, 4, 6, 7, 7, 8, ...]	State retention across calls	Add per-call initialization

To enable bit-accurate comparison, the LLM-generated HDL design as DUT and the HLS-generated HDL design as golden reference are simulated under identical test stimuli derived from the C/C++ testbench. Modern HLS tools can automatically translate the C/C++ testbench into an equivalent HDL testbench to test the HLS-generated HDL. The corresponding HDL testbench for the LLM-generated HDL is adapted from the same C/C++ testbench via the LLM with human verification to ensure its correctness in validating the LLM-generated HDL design. The new testbench is examined by human engineers to ensure its quality.

The agent then initiates a parallel simulation flow shown in Fig. 7, where in the DUT Path, the LLM-generated HDL design is simulated in an RTL simulation environment, and in the Golden Path, the HLS-generated HDL design is simulated in the HLS simulation environment. During both simulations, all output ports are recorded and dumped into result files. A comparison script conducts automated differential checking between these files. If all outputs match, the DUT is considered functionally correct and collected for subsequent integration. If any mismatch is observed, the DUT is marked as functionally incorrect, and the functional repair loop is triggered.

The repair loop uses the mismatched results as precise feedback to guide the LLM in the functional repair of the DUT. A structured prompt is constructed that contains the current (incorrect) Verilog code of the DUT, the C/C++ program that defines the intended functional behavior, and a mismatch log produced by differential comparison. This mismatch log explicitly records the test stimulus, the actual DUT output, and the expected golden reference for each failing test case. In addition, a step-wise reasoning method inspired by Chain-of-Thought is adopted to repair the LLM-generated HDL design. As shown in Fig. 8, the prompt explicitly instructs the LLM to strengthen functional understanding by analyzing the C/C++ program and summarizing the intended behavior. It then conducts behavior and difference analysis through reviewing the HDL design, correlating it with the expected functionalities, and identifying the source of the mismatches. The root cause of errors is identified and fixed by explaining the simulation discrepancy. With such concrete feedback, the LLM can then perform targeted debugging rather than blind editing. If mismatches remain, the verification–feedback–repair iteration is repeated until functional equivalence is achieved or the predefined iteration limit is reached.

A concrete example of the LLM-driven reasoning chain is shown in Fig. 8. Guided by this structured prompt, the LLM first identifies the small value differences stemming from inconsistent rounding semantics between the DUT and the golden fixed-point implementation. It then proposes a revised HDL version that corrects the rounding behavior. In the second check, the LLM observes that the first test case is almost correct but the second is completely wrong, and traces this to a missing state retention in the control logic. It proposes latching the intermediate result in the ST_WAIT_LBL state

TABLE 2: Comparison of *CorrectHDL* with Baselines in Simulation Pass Rate, and with HLS in Area and Power Efficiency.

Benchmark	Functional Simulation Pass Rate (%)						HDL Readability (Code Lines)		Area Overhead (μm^2)			Power Consumption (mW)		
	LLM→HDL [29] (NL as Input)	LLM→HDL (C/C++ as Input)	LLM→HDL (w/ Task Decomp.)		Proposed CorrectHDL (w/ HLS as Refer.)									
	Submod.	Top	Submod.	Top	HLS-Gen HDL [54]	Prop. Correct-HDL	Prop. Correct-HDL (%)	HLS-Gen HDL [54]	Prop. Correct-HDL	Prop. Correct-HDL (%)				
Advanced Encryption Standard	6.25	12.5	48.44	18.75	76.56	62.5	1879	386	9591	7836	18.30	4.30	2.92	32.09
Breadth-First Search	12.5	6.25	28.13	12.5	65.63	56.25	1392	234	29181	24101	17.41	9.98	8.08	19.04
Edge Detection Algorithm	0	0	27.08	6.25	64.58	37.5	4753	388	35422	30543	13.77	16.04	13.67	14.78
Median Filter	18.75	25	53.13	31.25	90.63	75	1351	213	11207	8249	26.39	4.84	3.61	25.41
Dense Matrix Multiplication	12.5	31.25	46.88	37.5	78.13	62.5	568	179	33044	21061	36.26	16.32	13.01	20.28
Knuth-Morris-Pratt	25	18.75	40.63	31.25	81.25	56.25	966	238	32631	20320	37.73	3.84	2.39	37.76
MIPS Processor	0	0	35.42	6.25	58.33	31.25	3033	397	30698	24483	20.25	14.35	10.53	26.62
N-body Molecular Dynamics	6.75	18.75	39.58	18.75	75.08	50	1086	270	25444	16007	37.09	5.89	3.41	42.11
Dynamic Sequence Alignment	12.5	12.5	33.33	25	60.42	43.75	2998	325	23138	18104	21.76	10.05	7.69	23.48
Secure Hash Algorithm	0	6.25	41.67	18.75	79.17	50	4421	420	22729	20898	8.06	9.40	7.81	16.91
Spam Filtering	12.5	18.75	43.75	31.25	84.38	56.25	1396	183	17571	11475	34.69	6.40	4.18	34.69
Markov Model Probabilities	6.25	6.25	34.38	12.5	65.63	37.5	1388	292	38598	28487	26.20	18.86	13.1	30.54

* The simulation pass rate is calculated based on 16 independent generation rounds by the LLM for each task.

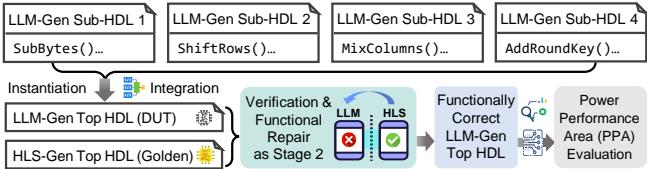


Figure 9: Design integration.

and reusing the stored value in subsequent calculations. After resimulation, the repaired DUT matches the golden outputs on all test cases, demonstrating how the agent can systematically analyze and repair functional discrepancies with reliable functional references. To better describe the types of functional errors handled by our framework, Table 1 summarizes several typical functional discrepancies observed between the DUT and the golden reference. For each task, the table lists representative DUT and HLS golden outputs and highlights how the LLM identifies the underlying root cause, such as numeric scaling errors, reset or lag issues, or state-retention faults, and generates the corresponding fixes, including correcting fixed-point scaling, decoupling load and compute operations, or adding per-call initialization. These examples indicate that the encountered bugs are not simple syntactic issues but functional errors that are difficult to debug without a golden functional reference, highlighting its essential role in the repair process.

3.3 Design Integration

After the LLM-generated HDL designs of all submodules pass differential verification, they are integrated into a functionally correct top-level HDL design. To enable correct instantiation and wiring, the LLM is prompted with the original C/C++ program from Section 3.1.1 and all verified HDL designs from Section 3.2 together with their interface definitions from Section 3.1.2. Using this information, function calls and variable transfers in the C/C++ code are systematically mapped to HDL module instances and signal connections, ensuring that data dependencies and interface contracts are preserved. The result is an LLM-generated top-level HDL design that implements the behavior of the original C/C++ program.

As shown in Fig. 9, the integrated top-level HDL design may still contain errors that are not exposed at the submodule level, such as misconnected ports or misaligned valid signals. To validate the top-level behavior, the original C/C++ program is then synthesized by the HLS tool into an HLS-generated top HDL design, which serves as the golden reference. Differential verification is then performed at the top level, as in Section 3.2, to detect integration errors.

To efficiently locate integration errors, boundary signal instrumentation is inserted at all submodule interfaces in the top-level design. During verification, the agent compares not only the final outputs but also these internal boundary signals against the golden reference. Guided by this structured feedback, the LLM performs targeted repair by using the backward slicing technique to trace the origin of the discrepancy [51]. The repair process is repeated until functional equivalence is achieved or the iteration limit is reached.

4 Experimental Results

To evaluate the quality of HDL designs generated by the proposed *CorrectHDL*, we applied it to 12 real-world tasks [52, 53] and measured the syntax compilation and functional simulation pass rate of the generated HDL designs. We also evaluated the area and power efficiency of the HDL designs produced by *CorrectHDL* and compared them with those generated by traditional HLS tools [54] and manually written designs from open-source implementations [56, 57].

In the evaluation, the Catapult HLS tool and QuestaSim RTL simulator [54, 55] were used to generate and simulate the HDL designs. The area and power of the HDL designs were evaluated by synthesizing them using Design Compiler with the Nangate 45nm library. In *CorrectHDL*, the GPT-5 model was employed as the LLM via OpenAI APIs [49]. To ensure reliability and robustness, each task was repeated n times. $n = 16$ was used in the experiments. For each time, the LLM was queried three times to iteratively correct the HDL based on tool feedback. The pass rate is computed as Pass Rate (%) = m/n , where m represents the number of successfully generated HDL designs and n denotes the total number of HDL designs.

Table 2 compares the HDL designs generated by *CorrectHDL* with those directly generated using the LLM with C/C++ codes as inputs and with HLS-generated HDL designs. Column 1 lists the benchmark names. Column 2 reports the simulation pass rate of directly prompting the LLM with natural language [30]. This setting leads to low simulation pass rates, indicating that most generated HDL designs exhibit functional errors. Column 3 shows that prompting the LLM with the C/C++ program still results in low simulation pass rates, although it performs better on average than natural language prompts because C/C++ more precisely captures the underlying algorithmic behavior. With the C/C++ decomposition step in *CorrectHDL*, the simulation pass rate can be enhanced, as shown in columns 4-5, where column 4 reports the average pass rate of decomposed submodules and column 5 shows the pass rate of the top design by integrating such submodules together by the LLM. Columns 6-7

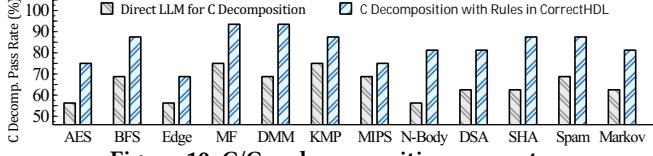


Figure 10: C/C++ decomposition pass rate.

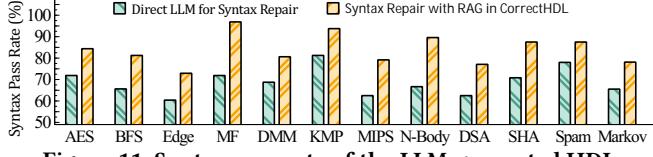


Figure 11: Syntax pass rate of the LLM-generated HDL.

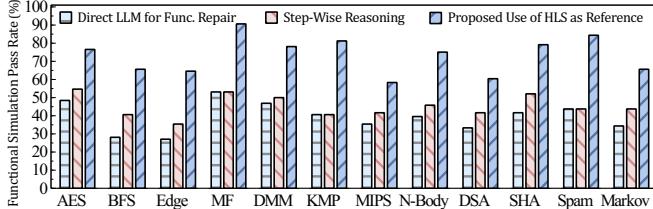


Figure 12: Functional pass rate of the LLM-generated sub-HDL.

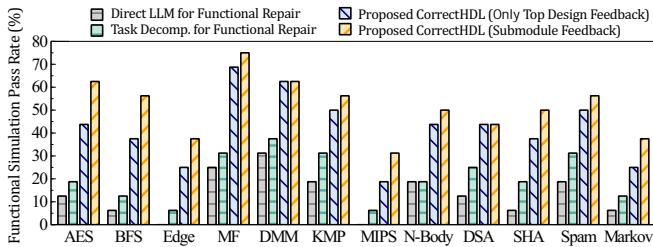


Figure 13: Functional pass rate of the LLM-generated top HDL.

show the functional simulation rate of submodules and top design of *CorrectHDL*, which demonstrates significant improvements in functional correctness of the LLM-generated HDL.

To compare the quality of HDL code generated by *CorrectHDL* and the traditional HLS tool, columns 8–9 in Table 2 report readability and maintainability of HDL code, measured by the number of lines of HDL code. This comparison shows *CorrectHDL* consistently generates more readable HDL than HLS. Columns 10–15 further compare area and power under identical frequency, showing that *CorrectHDL* produces HDL designs with significantly lower area and power consumption than HLS-generated HDL designs.

To demonstrate the effectiveness of the rule-based C/C++ decomposition in *CorrectHDL*, we compare the proposed decomposition method against direct LLM-based decomposition. As shown in Fig. 10, the proposed approach achieves an average improvement of 18.19% in decomposition pass rate. To assess the effectiveness of syntax repair for LLM-generated HDL designs, we compare the compilation pass rate achieved by *CorrectHDL* (with RAG) against directly using the LLM. As shown in Fig. 11, *CorrectHDL* improves the compilation pass rate by an average of 15.49%.

To evaluate the quality of functional repair of submodules with *CorrectHDL*, Fig. 12 compares the functional simulation pass rates of submodules using *CorrectHDL*, using the LLM directly, and using the LLM with step-wise reasoning for functional repair. Step-wise reasoning improves the pass rate by an average of 5.90% over direct LLM usage, while *CorrectHDL*, using HLS as a functional reference, further improves the pass rate by an average of 28.05%.

To demonstrate the top-level functional simulation pass rate of the HDL design with *CorrectHDL*, which incorporates submodule feedback from the HLS-generated reference during the final HDL

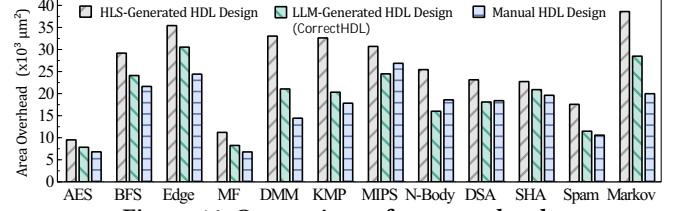


Figure 14: Comparison of area overhead.

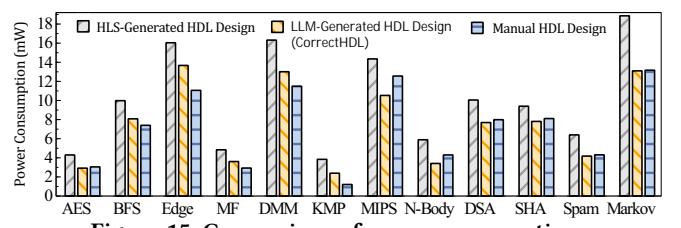


Figure 15: Comparison of power consumption.

design generation, it is compared with three other settings, as shown in Fig. 13. *Direct LLM Baseline*: The LLM is used directly to generate and debug the top-level HDL design. *LLM with task decomposition*: The C/C++ program is decomposed into LLM-friendly submodules to generate and debug the top-level HDL design. *CorrectHDL with Only Top-Design Feedback*: Submodule-level feedback is disabled. Fig. 13 shows that, by incorporating an HLS reference, task decomposition, and submodule boundary instrumentation, *CorrectHDL* outperforms the three settings with average gains of 38.54%, 30.73%, and 9.35% in functional simulation pass rate, respectively.

To demonstrate the area and power efficiency of HDL designs generated by *CorrectHDL*, we compare them with those generated by the traditional HLS tool and by manual design from open-source implementations [56, 57]. For fairness, all HLS designs are compiled using `#pragma design_goal area` to explicitly target minimum area, while the open source designs are adapted so that their bit widths, array sizes, and other parameters are consistent with the other two approaches. As shown in Fig. 14 and Fig. 15, area overhead and power consumption are evaluated under identical frequency. The synthesized circuits generated by *CorrectHDL* achieve an average area reduction of 24.83% and power reduction of 26.98% compared with HLS-generated designs. Moreover, the HDL designs generated by *CorrectHDL* approach the quality of human-engineered circuits in many cases, while for a few benchmarks, *CorrectHDL* produces even better designs than those written by human engineers.

5 Conclusion

In this paper, we have proposed *CorrectHDL*, an agentic LLM-assisted HDL design framework that leverages HLS-generated HDL as a functional reference. Starting from a C/C++ program, complex algorithms are decomposed into LLM-friendly submodules whose HDL implementations are generated by the LLM. Syntax errors are corrected via RAG, and functional discrepancies are resolved through differential verification against the HLS-generated golden reference. Experimental results on 12 real-world benchmarks show that *CorrectHDL* significantly improves both syntax and functional pass rates of LLM-generated HDL, while achieving lower area and power than HLS-generated designs and approaching the efficiency of manually crafted designs. Future work will integrate C/C++ program generation from natural language to establish a complete LLM-assisted workflow from natural language to HDL design. The LLM-generated circuits will also be optimized further to explore their potential in matching and even surpassing human-engineered designs.

References

- [1] Tinghuan Chen, Grace Li Zhang, Bei Yu, Bing Li, Ulf Schlichtmann, "Machine Learning in Advanced IC Design: A Methodological Survey," *Design & Test*, 2023.
- [2] Jason Cong, Jason Lau, Gai Liu, Stephen Neuendorffer, Peichen Pan, Kees Vissers, Zhiru Zhang, "FPGA HLS Today: Successes, Challenges, and Opportunities," *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 2022.
- [3] Jason Cong, Bin Liu, Stephen Neuendorffer, Juanjo Noguera, Kees Vissers, Zhiru Zhang, "High-Level Synthesis for FPGAs: From Prototyping to Deployment," *IEEE Transactions on CAD of Integrated Circuits and Systems (TCAD)*, 2011.
- [4] Jason Cong, Zhenman Fang, Muhan Huang, Libo Wang, and Di Wu, "CPU-FPGA Coscheduling for Big Data Applications," *IEEE Design & Test*, 2018.
- [5] Atefeh Sohrabizadeh, Cody Hao Yu, Min Gao, and Jason Cong, "AutoDSE: Enabling Software Programmers to Design Efficient FPGA Accelerators," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 2022.
- [6] Shailja Thakur, Jason Blocklove, Hammond Pearce, Benjamin Tan, Siddharth Garg, Ramesh Karri, "AutoChip: Automating HDL Generation Using LLM Feedback," *Arxiv Preprint: 2311.04887*, 2023.
- [7] Jason Blocklove, Siddharth Garg, Ramesh Karri, Hammond Pearce, "Chip-Chat: Challenges and Opportunities in Conversational Hardware Design," *ACM/IEEE International Symposium on Machine Learning for CAD (MLCAD)*, 2023.
- [8] Andre Nakkab, Sai Qian Zhang, Ramesh Karri, Siddharth Garg, "Rome was Not Built in a Single Step: Hierarchical Prompting for LLM-based Chip Design," *IEEE International Symposium on Machine Learning for CAD (MLCAD)*, 2024.
- [9] Yonggan Fu, Yongan Zhang, Zhongzhi Yu, Sixu Li, Zhifan Ye, Chaojian Li, Cheng Wan, Yingyan Celine Lin, "GPT4AIGChip: Towards Next-Generation AI Accelerator Design Automation via Large Language Models," *IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, 2023.
- [10] Yun-Da Tsai, Mingjie Liu, Haoxing Ren, "Automatically Fixing RTL Syntax Errors with Large Language Model," *ACM Design Automation Conference (DAC)*, 2024.
- [11] Kangwei Xu, Denis Schwachhofer, Jason Blocklove, Ilia Polian, Peter Domanski, Dirk Pflüger, Siddharth Garg, Ramesh Karri, Ozgur Sinanoglu, Johann Knechtel, Zhiyuan Yan, Bing Li, Grace Li Zhang, Xunzhao Yin, Cheng Zhuo, Ulf Schlichtmann, "Paradigm-Based Automatic HDL Code Generation Using LLMs," *ACM/IEEE International Symposium on Quality Electronic Design (ISQED)*, 2025.
- [12] Chandan K. Jha, M. Hassan, K. Qayyum, S. Ahmadi-Pour, K. Xu, R. Qiu, J. Blocklove, L. Collini, A. Nakkab, U. Schlichtmann, Grace L. Zhang, R. Karri, B. Li, S. Garg, R. Drechsler, "Large Language Models (LLMs) for Verification, Testing, and Design," *IEEE European Test Symposium (ETS)*, 2025.
- [13] Bing-Yue Wu, Utsav Sharma, Austin Rovinski, Vidya A Chhabria, "OpenROAD Agent: An Intelligent Self-Correcting Script Generator for OpenROAD," *IEEE International Conference on LLM-Aided Design (ICLAD)*, 2025.
- [14] Yuan Pu, Zhoulun He, Shutong Lin, Jiajun Qin, Xinyun Zhang, Hairuo Han, Haisheng Zheng, Yuqi Jiang, Cheng Zhuo, Qi Sun, David Z. Pan, Bei Yu, "A Multi-Modal EDA Tool Documentation QA Framework Leveraging Retrieval Augmented Generation," *IEEE/ACM International Conference on CAD (ICCAD)*, 2025.
- [15] Yingbing Huang, Lily Jiaxin Wan, Hanchen Ye, Manvi Jha, Jinghua Wang, Yuhong Li, Xiaofan Zhang, Deming Chen, "New solutions on LLM acceleration, optimization, and application," *IEEE/ACM Design Automation Conference (DAC)*, 2024.
- [16] Zesong Jiang, Yuqi Sun, Qing Zhong, Mahathi Krishna, Deepak Patil, Cheng Tan, Sriram Krishnamoorthy, Jeff Zhang, "MACO: A Multi-Agent LLM-Based Hardware/Software Co-Design Framework for CGRAs," *Arxiv: 2509.13557*, 2025.
- [17] Bing-Yue Wu, Utsav Sharma, Sai Rahul Dhanvi Kankipati, Ajay Yadav, Bintu Kapil George, Sai Ritish Guntupalli, Austin Rovinski, Vidya A Chhabria, "EDA Corpus: A Large Language Model Dataset for Enhanced Interaction with OpenROAD," *IEEE International Conference on LLM-Aided Design (ICLAD)*, 2024.
- [18] Zeng Wang, Minghao Shao, Rupesh Karn, Likhitha Mankali, Jitendra Bhandari, Ramesh Karri, Ozgur Sinanoglu, Muhammad Shafique, Johann Knechtel, "SALAD: Systematic Assessment of Machine Unlearning on LLM-Aided Hardware Design," *IEEE International Symposium on Machine Learning for CAD (MLCAD)*, 2023.
- [19] Lakshmi Likhitha Mankali, Jitendra Bhandari, Manaar Alam, Ramesh Karri, Michail Maniatakos, Ozgur Sinanoglu, Johann Knechtel, "RTL-breaker: Assessing the Security of LLMs Against Backdoor Attacks on HDL Code Generation," *ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2025.
- [20] Muhammad Hassan, Sallar Ahmadi-Pour, Khushboo Qayyum, Chandan Kumar Jha, Rolf Drechsler, "LLM-guided Formal Verification Coupled with Mutation Testing," *IEEE Design, Automation & Test in Europe Conference (DATE)*, 2024.
- [21] Kangwei Xu, Ruidi Qiu, Zhiyuan Yan, Grace Li Zhang, Ulf Schlichtmann, Bing Li, "LLM-Aided Efficient Hardware Design Automation," *Arxiv: 2410.18582*, 2024.
- [22] Yongan Zhang, Yonggan Fu, Zhongzhi Yu, Kevin Zhao, Cheng Wan, Chaojian Li, Yingyan Celine Lin, "An Automated Data Generation and Validation Flow for LLM-Assisted Hardware Design," *Design Automation Conference (DAC)*, 2024.
- [23] Shailja Thakur, Baleegh Ahmad, Hammond Pearce, Benjamin Tan, B. D. Gavitt, Ramesh Karri, Siddharth Garg, "A Large Language Model for Verilog Code Generation," *ACM Transactions on Design Automation of Electronic Systems*, 2024.
- [24] Mingjie Liu, Nathaniel Pinckney, Brucek Khailany, Haoxing Ren, "VerilogEval: Evaluating large language models for verilog code generation," *IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, 2023.
- [25] Zhiyuan Yan, Ruidi Qiu, Ing-Chao Lin, Grace Li Zhang, Bing Li, Ulf Schlichtmann, "Enhancing Verilog Code Generation from Large Language Models via Self-Consistency," *International Symposium on Quality Electronic Design*, 2025.
- [26] Zhiyuan Yan, Bing Li, Grace Li Zhang, Ulf Schlichtmann, "VFocus: Better Verilog Generation from Large Language Model via Focused Reasoning," *IEEE SOCC*, 2025.
- [27] Shailja Thakur, Baleegh Ahmad, Zhenxing Fan, Hammond Pearce, Benjamin Tan, Ramesh Karri, Brendon Dolan-Gavitt, Siddharth Garg, "Benchmarking Large Language Models for Automated Verilog RTL Code Generation," *IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2023.
- [28] Matthew DeLorenzo, Animesh Basak Chowdhury, Vasudev Gohil, Shailja Thakur, Ramesh Karri, Siddharth Garg, Jeyavijayan Rajendran, "LLM-based High-Quality RTL Code Generation Using MCTS," *arXiv preprint: 2402.03289*, 2024.
- [29] Prithwish Basu Roy, Akashdeep Saha, Manaar Alam, Johann Knechtel, Michail Maniatakos, Ozgur Sinanoglu, Ramesh Karri, "Veritas: Deterministic Verilog Code Synthesis from LLM-Generated Conjunctive Normal Form," *arXiv: 2506.00005*.
- [30] Wenhao Sun, Bing Li, Grace Li Zhang, Xunzhao Yin, Cheng Zhuo, Ulf Schlichtmann, "Paradigm-Based Automatic HDL Code Generation Using LLMs," *ACM/IEEE International Symposium on Quality Electronic Design (ISQED)*, 2025.
- [31] Zhiyuan Yan, Wenji Fang, Mengming Li, Min Li, Shang Liu, Zhiyao Xie, Hongce Zhang, "AssertLLM: Generating and Evaluating Hardware Verification Assertions from Design Specifications via Multi-LLMs," *IEEE/ACM ASP-DAC*, 2025.
- [32] Ruidi Qiu, Grace Li Zhang, Rolf Drechsler, Ulf Schlichtmann, Bing Li, "AutoBench: Automatic Testbench Generation and Evaluation Using LLMs for HDL Design," *IEEE International Symposium on Machine Learning for CAD (MLCAD)*, 2024.
- [33] Ruidi Qiu, Grace Li Zhang, Rolf Drechsler, Ulf Schlichtmann, Bing Li, "Correct-Bench: Automatic Testbench Generation with Functional Self-Correction using LLMs for HDL Design," *Design, Automation & Test in Europe Conference*, 2025.
- [34] Luca Collini, Siddharth Garg, Ramesh Karri, "C2HLS: Can LLMs Bridge the Software-to-Hardware Design Gap?," *IEEE ICLAD*, 2024.
- [35] Luca Collini, Siddharth Garg, Ramesh Karri, "Leveraging Large Language Models to Bridge the Software-to-Hardware Design Gap," *ACM TODAES*, 2025.
- [36] Luca Collini, Andrew Hennessey, Ramesh Karri, Siddharth Garg, "Can reasoning models reason about hardware? an agentic HLS perspective," *arXiv preprint: 2503.12721*, 2025.
- [37] Kangwei Xu, Grace Li Zhang, Xunzhao Yin, Cheng Zhuo, Ulf Schlichtmann, Bing Li, "Automated C/C++ Program Repair for High-Level Synthesis via Large Language Models," *International Symposium on Machine Learning for CAD*, 2024.
- [38] Chenwei Xiong, Cheng Liu, Huawei Li, Xiaowei Li, "HLSPILOT: LLM-based High-Level Synthesis," *ACM/IEEE International Conference on CAD (ICCAD)*, 2024.
- [39] Kangwei Xu, Bing Li, Grace Li Zhang, Ulf Schlichtmann, "HLSTester: Efficient Testing of Behavioral Discrepancies with LLMs for High-Level Synthesis," *IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, 2025.
- [40] Stefan Abi-Karam, Cong Hao, "HLS-Eval: A Benchmark and Framework for Evaluating LLMs on High-Level Synthesis Design Tasks," *IEEE ICLAD*, 2025.
- [41] Zhaojiao Yu, Yilun Zhao, Arman Cohan, Xiao-Ping Zhang, "HumanEval Pro and MBPP Pro: Evaluating Large Language Models on Self-invoking Code Generation," *Association for Computational Linguistics (ACL)*, 2025.
- [42] Wenpin Hou, Zhicheng Ji, "Comparing large language models and human programmers for generating programming code," *Advanced Science*, 2025.
- [43] Kangwei Xu, Grace Li Zhang, Xunzhao Yin, Cheng Zhuo, Ulf Schlichtmann, Bing Li, "HLSRewriter: Efficient Refactoring and Optimization of C/C++ Code with LLMs for HLS," *ACM Transactions on Design Automation of Electronic Systems*, 2025.
- [44] Sakari Lahti, Panu Sjövall, Jarno Vanne, T. Hämäläinen, "Are We There Yet? A Study on the State of High-Level Synthesis," *IEEE TCAD*, 2019.
- [45] Yifan Zhang, Q. Cao, Jie Yao, Hong Jiang, "R-LDPC: Refining Behavior Descriptions in HLS to Implement High-throughput LDPC Decoder," *IEEE/ACM DATE*, 2023.
- [46] Lan Huang, Da-Lin Li, Kang-Ping Wang, Teng Gao, Adriano Tavares, "A Survey on Performance Optimization of High-Level Synthesis Tools," *JCST*, 2020.
- [47] K. Xu, et al., "Logic Design of Neural Networks for High-Throughput and Low-Power Applications," *IEEE/ACM Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2024.
- [48] R. Nane, V. Sima, C. Pilato, J. Choi, B. Fort, A. Canis, Y. Chen, H. Hsiao, S. Brown, F. Ferrandi, J. Anderson, K. Bertels, "A Survey and Evaluation of FPGA HLS Tools," *IEEE Transactions on CAD of Integrated Circuits and Systems*, 2016.
- [49] "OpenAI API." Accessed: 2025. [Online]. Available: <https://platform.openai.com/>.
- [50] N. Reimers, I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *Empirical Methods in Natural Language Processing*, 2019.
- [51] W. Eric Wong, Ruizhi Gao, Yihao Li, Rui Abreu, Franz Wotawa, "A Survey on Software Fault Localization," *IEEE Transactions on Software Engineering*, 2016.
- [52] Brandon Reagen, Robert Adolf, Yakun Sophia Shao, Gu-Yeon Wei, David Brooks, "MachSuite: Benchmarks for accelerator design and customized architectures," *IEEE International Symposium on Workload Characterization (IISWC)*, 2014.
- [53] Yuko Hara, Hiroyuki Tomiyama, Shinya Honda, Hiroaki Takada and Katsuya Ishii, "CHStone: A benchmark program suite for practical C-based high-level synthesis," *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2008.
- [54] "Siemens EDA Catapult High-Level Synthesis Tools," Accessed: 2023. [Online]. Available: <https://eda.sw.siemens.com/en-US/ic/catapult-high-level-synthesis/hls/>.
- [55] "Siemens EDA QuestaSIM RTL Simulator," Accessed: 2023. [Online]. Available: <https://eda.sw.siemens.com/en-US/ic/questa-one/simulation/questa-one-sim/>.
- [56] "OpenCores Manual HDL Designs," Accessed: 2025. [Online]. Available: <https://opencores.org/>.
- [57] "GitHub Manual HDL Designs," Accessed: 2025. [Online]. Available: <https://github.com/>.