Dr. John A. Curreri

Madison, AL · 1-352-262-7861 · johncurreri@gmail.com · curreri.cc · github.com/j-curreri · US citizen

Summary

Computer Engineering Ph.D. and hardware validation expert. For 15 years I found and root-caused bugs in cache coherency and on-die interconnect across 12 generations of Intel server CPUs, in pre-silicon emulation and post-silicon debug, and led the debug teams that did it. Root-causing from partial evidence (machine checks, silent data corruption, transaction traces, logic-analyzer captures) is exactly the kind of expert judgment today's models lack. I now build ground-truth benchmarks that measure where LLMs succeed and where they are confidently wrong, using Claude and Claude Code daily.

Professional Experience

Intel — Senior Validation Engineer, Mar 2011 – Aug 2026 · Columbia Design Center, Austin and remote

- Validated the cache coherency and on-die interconnect of 12 generations of server CPUs, from pre-silicon emulation through post-silicon debug.
- Root-caused failures surfaced by machine checks and by software self-checking for silent data corruption, using transaction trackers and signal traces (Verdi) in emulation and oscilloscopes / logic analyzers in the lab.
- Group lead and debug lead for the interconnect/coherency team and the debug, process & closure (DPC) team.
- Led a common test plan shared across multiple server products.
- Built an automatic, agent-based failure-binning report system that reduced engineers' manual triage load.
- Maintained the team's C++ cache-coherency tool in Git; product owner and scrum master in Agile teams.

University of Florida — Research Assistant, May 2005 – Mar 2011 · NSF Center for High-Performance Reconfigurable Computing (CHREC) and HCS Lab, advisor Dr. Alan George

- Dissertation: execution productivity for high-level synthesis, with a novel runtime performance-analysis and verification tool for FPGA applications (in-circuit assertions synthesized from high-level code).
- Demonstrated a new fault-tolerant satellite architecture for NASA's New Millennium Program; built the fault-tolerance middleware and timed error propagation.

AI Evaluation & Research

RTL Bug-Reproduction Agent Benchmark — Oct 2026 · github.com/j-curreri/rtl-bug-repro-bench

- Agents must turn a bug report into a failing test for real CPU RTL: 14 upstream bugs re-inserted into the CVA6 RISC-V core, graded on fixed vs buggy core. 1,000+ episodes, 11 models.
- Top agents reproduced 88–100% from a full report; Opus 5.5 fell to 38% given only file names. OpenAI's cyber classifier ended most GPT-6.1 sessions. Agent tests also exposed real CVA6 deviations, one with security impact, reported via a GitHub security advisory.

CVE Recall Benchmark — Sep 2026 · github.com/j-curreri/cve-recall-bench

- Ground-truth benchmark for expert-level code auditing: 30 security advisories published after the models' training cutoffs, each given to the model twice (before and after the real fix). Findings are auto-scored against the lines the fix changed, and on the patched file to measure confident false alarms.
- Across 10 model configurations the best reached 56.7% recall@3, yet still flagged already-fixed code in 27–37% of patched files; security-specialised fine-tunes trailed their general-purpose bases.

Silicon Failure Triage Agent — curreri.cc/demo

- LLM agent that bins post-silicon debug logs: retrieval over a corpus of prior failures, calibrated confidence, and every tool call exposed for inspection; public demo replays recorded runs.

Validating Attribution-Graph Claims: a Two-Hop Circuit Test Suite — Oct 2026 · github.com/j-curreri/two-hop-circuit-validation

- Treated one interpretability claim about Qwen3-8B as a hypothesis and tested it causally across 49 prompts: ablating the labelled intermediate features cut median P(correct answer) from 0.97 to 0.18 (below 0.5 in 31/49) vs 0/49 for size-matched random controls; swapping features flipped the answer in 28/49.

Fine-Tuned Structured-Decision Model (Qwen3-4B) — Sep 2026

- Fine-tuned Qwen3-4B to return a calibrated probability for every option of every field in one pass; on a sealed held-out set it came within 4.3 points of Qwen3.8-27B (77.3% vs 81.6%) at 26 vs 976 ms.

Precision Metrology Automation

- AI-assisted automation for nanovolt meters, frequency counters, potentiostats, LCR meters and calibration standards over SCPI/GPIB in Python; an agent designs, runs and analyzes sandboxed experiments against real instrumentation.

LLM Systems Lab — 2026

- Run a two-GPU lab (RTX PRO 6000 Blackwell + RTX 3090) serving models from 8B to the 2.8T-parameter Kimi K3; raised K3 decode throughput 4.2× by profiling NVMe expert streaming, with output matching the exact-match oracle.

Technical Skills

- **Hardware validation:** cache coherency and on-die interconnect, machine checks, silent data corruption, pre-silicon emulation, post-silicon lab debug (scopes, logic analyzers), Verdi, test planning.
- **Hardware design:** VHDL, Verilog, high-level synthesis, FPGA platforms (Cray XD1, XtremeData XD1000, SRC-7, Nallatech H100), MSP430/ATmega/ARM boards, SCPI/GPIB instrument control, RISC-V, Verilator.
- **AI & ML:** Claude, Claude Code, Claude Agent SDK, MCP, tool-using agents, retrieval, benchmark and evaluation design, LLM fine-tuning, PyTorch, Hugging Face Transformers, TransformerLens, circuit-tracer, SGLang, vLLM, llama.cpp.
- **Languages:** Python, C, C++.

Education

Ph.D., Computer Engineering, University of Florida, 2011 (GPA 3.9)

M.S., Computer Engineering, University of Florida, 2006 (GPA 3.9)

B.S., Computer Engineering and Mathematics, University of Alabama in Huntsville, 2005 (GPA 3.977)

Selected Publications

- **J. Curreri**, S. Koehler, A. George, B. Holland, R. Garcia, “Performance Analysis Framework for High-Level Language Applications in Reconfigurable Computing,” *ACM TRETS*, 3(1), 2010.
- **J. Curreri**, G. Stitt, A. D. George, “High-level Synthesis of In-Circuit Assertions for Verification, Debugging, and Timing Analysis,” *Int. J. Reconfigurable Computing*, 2011.
- Internal paper, Intel Design and Test Technology Conference (DTTC), 2013.
- “One attribution graph is an anecdote: validating a two-hop circuit in Qwen3-8B,” curreri.cc/circuits, Oct 2026.
- “Can Coding Agents Reproduce Real RTL Bugs? A CVA6 Benchmark,” curreri.cc/rtl-bugs.pdf, Oct 2026.