How Agentic AI is Reinventing Chip Design and Verification

William Wang

Team

William Wang
Founder, CEO, and Chairman

Mellichamp Chair Professor of AI, UCSB +
Amazon AWS Bedrock / Amazon Q (2022-2024)

IEEE Laplace Award, BCS Karen Spärck Jones Award,
NSF CAREER Award, IEEE AI's 10 to Watch,
DARPA Young Faculty Award. PhD @ CMU

$25M previous gifts and contracts from Meta, Google,
Amazon, Intel, JP Morgan, Adobe, NVIDIA, IBM,
CISCO, Apple, etc.
Helped Apple shipped MGIE, and Google's Gemini.

Press: Venture Beat, Business Insider, Wired, GeekWire, Guardian, Fast
Company, Fortune, Scientific American etc.

ChipAgents: Board of Advisors

Wally Rhines
ex-CEO, Mentor Graphics
(Siemens EDA)

Raul Camposano
ex-CTO, Synopsys

Jack Harding
ex-CEO, Cadence

"ChipAgents demonstrates the major impact that AI can have on the full range of integrated circuit design tasks". "I've met with three major semiconductor companies that have done competitive assessments of AI-based design solutions and ChipAgents is the number one choice at all three". - Wally Rhines
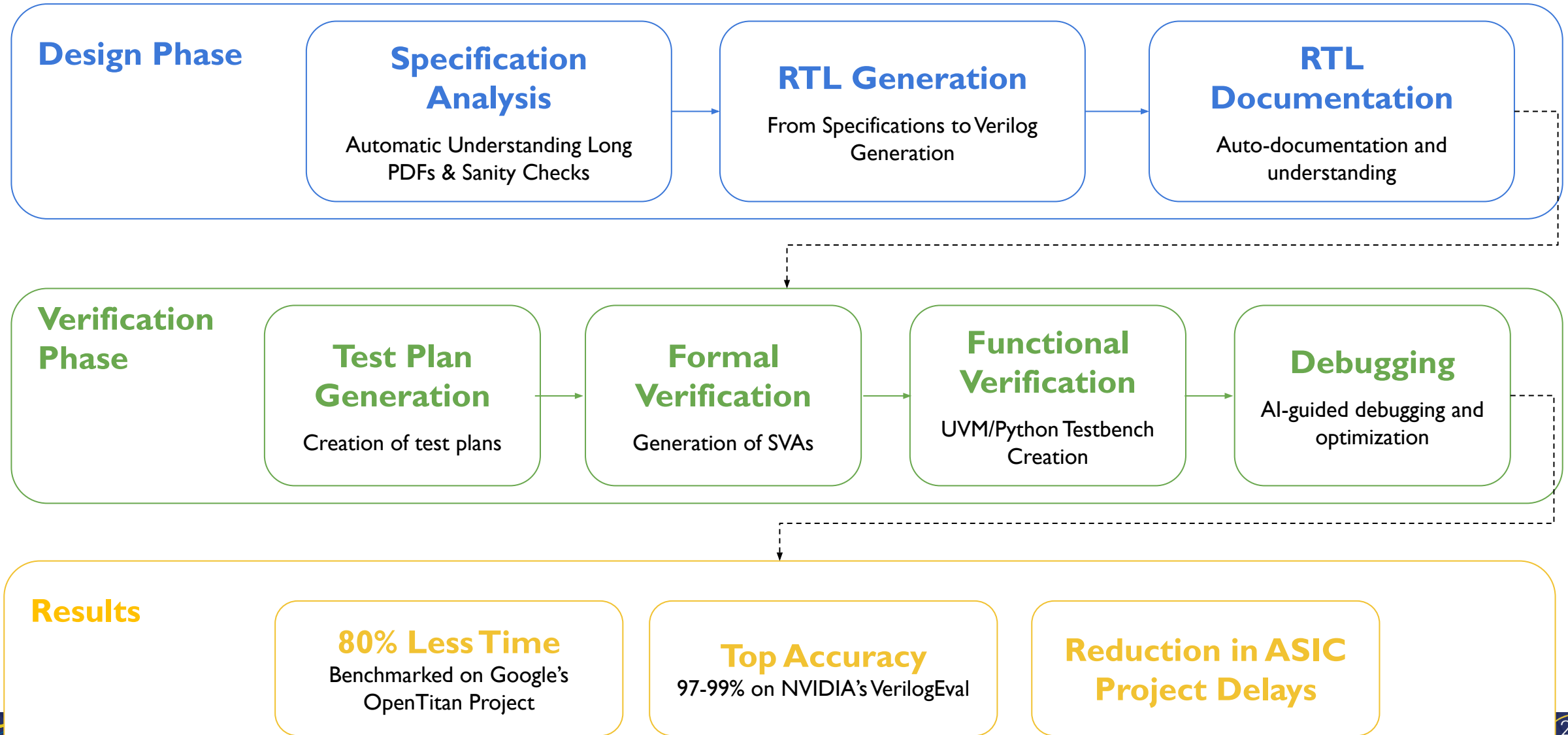
- Series A Company, Backed by Bessemer Venture Partners, MediaTek, Micron, and Ericsson etc.

- Location: Santa Barbara, California

- ChipAgents: AI Agents for Everyone, AI Agents for Every Chip.

- Proven success: in production w. public & private customers in US, Europe, and Asia, who have shipped 15B+ chips in total.

- Benchmark Results: 99.4% pass rate for spec-to-RTL generation on NVIDIA's VerilogEval dataset.

ChipAgents™: Agentic AI Powered Design & Verification Flow

Design Phase

Specification Analysis
Automatic Understanding Long PDFs & Sanity Checks

→

RTL Generation
From Specifications to Verilog Generation

→

RTL Documentation
Auto-documentation and understanding

Verification Phase

Test Plan Generation
Creation of test plans

→

Formal Verification
Generation of SVAs

→

Functional Verification
UVM/Python Testbench Creation

→

Debugging
AI-guided debugging and optimization

Results

80% Less Time
Benchmarked on Google's OpenTitan Project

Top Accuracy
97-99% on NVIDIA's VerilogEval

Reduction in ASIC Project Delays

Agenda for Today's TechTalk

1. Overview

2. Introduction to AI Agents

Challenges in Design and Verification

We observe the following problems after working with leading companies in chip design and verification:

- Increasing complexity of RTL design and testbench.
- Lengthy cycles and verification bottlenecks.
- Limited scalability with traditional methods.
- Cost and time pressures for first-time silicon success.

ChipAgents™: AI Agents and large language models for IC

We propose novel solutions of **specialized AI Agents and large language models** to transform IC into language-based AI design: using **natural language and AI** for design, debugging, and verification.

- Reduce 80% Development Time
- Increase Accuracy and Precision
- Optimization, Predictive Analysis and Simulation
- Personalization and Customized Designs

State-of-the-Art Agentic AI Performances

Approach	VerilogEval-Human Pass Rate	**VerilogEval-v2 Pass Rate**
Generic LLM (GPT-4o)	51.3	N/A
Claude-3.5-Sonnet	75.0	72.4 (113/156)
VerilogCoder (NVIDIA)	N/A	94.2 (147/156)
MAGE	94.8	95.5 (149/156)
ChipAgents (Ours)	**99.4 (155**/156)	**97.4 (152**/156)

Seeking Alpha α

ChipAgents Achieves State-of-the-Art Results on NVIDIA's VerilogEval Benchmark

seekingalpha.com

Leaderboard

	Lite	**Verified**	Full				
Model		**% Resolved**	**Org**	**Date**	**Logs**	**Trajs**	**Site**
CodeStory Midwit Agent + swe-search		62.20		2024-12-21	✓	✓	🔗
devlo		58.20	⊙	2024-12-13	✓	✓	🔗
Emergent E1 (v2024-12-23)		57.20	e	2024-12-23	✓	✓	🔗
Gru(2024-12-08)		57.00		2024-12-08	✓	✓	🔗
EPAM AI/Run Developer Agent v20241212 + Anthopic Claude 3.5 Sonnet		55.40	\<epam\>	2024-12-12	✓	✓	🔗
Amazon Q Developer Agent (v20241202-dev)		55.00	aws	2024-12-02	✓	✓	🔗

SWE-Search: Enhancing Software Agents with Monte Carlo Tree Search and Iterative Refinement

Antonis Antoniades, Albert Örwall, Kexun Zhang, Yuxi Xie, Anirudh Goyal, William Wang

Software engineers operating in complex and dynamic environments must continuously adapt to evolving requirements, learn iteratively from experience, and reconsider their approaches based on new insights. However, current large language model (LLM)-based software agents often rely on rigid processes and tend to repeat ineffective actions without the capacity to evaluate their performance or adapt their strategies over time. To address these challenges, we propose SWE-Search, a multi-agent framework that integrates Monte Carlo Tree Search (MCTS) with a self-improvement mechanism to enhance software agents' performance on repository-level software tasks. SWE-Search extends traditional MCTS by incorporating a hybrid value function that leverages LLMs for both numerical value estimation and qualitative evaluation. This enables self-feedback loops where agents iteratively refine their strategies based on both quantitative numerical evaluations and qualitative natural language assessments of pursued trajectories. The framework includes a SWE-Agent for adaptive exploration, a Value Agent for iterative feedback, and a Discriminator Agent that facilitates multi-agent debate for collaborative decision-making. Applied to the SWE-bench benchmark, our approach demonstrates a 23% relative improvement in performance across five models compared to standard open-source agents without MCTS. Our analysis reveals how performance scales with increased search depth and identifies key factors that facilitate effective self-evaluation in software agents. This work highlights the potential of self-evaluation driven search techniques to enhance agent reasoning and planning in complex, dynamic software engineering environments.

Proprietary and Confidential

2025
DESIGN AND VERIFICATION
DVCON
CONFERENCE AND EXHIBITION
TAIWAN

SYSTEMS INITIATIVE

Testimonials from Customers and Partners

ChipAgents is in production with many top-20 semiconductor companies.

"... ChipAgents suggested invariant assertions have found two bugs in one of our projects." - a top Taiwanese public company.

"ChipAgents' auto-fixing testbenches is solving the right problem." - Joe Costello, ex-CEO, Cadence.

"We went from drawing schematics to RTL, and now, finally, we are onto something. This is very exciting." - Raul Camposano, ex-CTO, Synopsys.

Product Roadmap

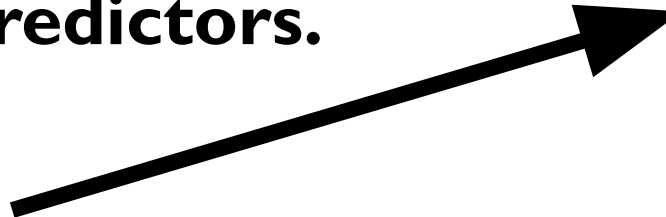| AI-Assisted Chip Design (2024)	LLM-Guided Debugging & Functional Verification (2025)	Fully Agentic AI Design & Verification (2026)
AI-guided coding for RTL (e.g., Verilog, SystemVerilog) with line-by-line support, workspace understanding, spec understanding, spec-to-code generation, automated template generation, and real-time error detection.	Automatic testbench creation, assessments, simulation, optimization, and log analysis to verify functionality, highlight errors, and provide actionable fixes. We utilize specifications, designs, and existing tests for line and functional coverage agents.	Fully automated verification agents for consolidating specs, code, logs, waveforms, and simulations into automated agents for optimizing and accelerating design and verification workflows.

What are AI Agents?
Why do we want them?

Large language models, and why they aren't enough

LLMs are next word predictors.

DAC 2025 is in [MASK]

```
module toplevel(clock,reset);

    input clock;

    input [MASK]
```

Question: What's 5 + 7? Answer: [MASK]

Probable Next Words:
San Francisco

California

America

…

#include
supercalifragilisticexpialidocious

Proprietary and Confidential

2025
DESIGN AND VERIFICATION™
DVCON
CONFERENCE AND EXHIBITION
TAIWAN

Large language models, and why they aren't enough

What they can do:

- Understand a single, short SV file

- Fix bugs in a single file

- Generate test benches for a single module

- Read and check a 5-page PDF

What they can't:

- Understand your entire design

- Find & fix bugs in a 5000-file codebase.

- Execute test plans for an entire design, improve coverage, run simulations.

- Cross-check a 5000-page PDF spec with your design.

What are AI Agents?

**AI Agents are Large Language Models (LLMs) which can take <u>actions</u>**

LLMs

✖ Mistakes propagate with each prediction

✖ Reactive (user driven)

✖ No intelligent context selection

Agents

✔ Self corrects by producing its own verified feedback

✔ Self directed

✔ Intelligently decides on context
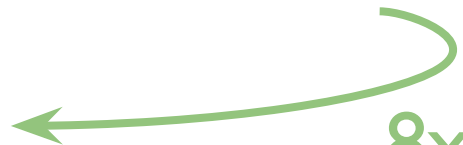
LLM Agents, and their success

SWE-Bench: Given an issue, automatically locate and fix bugs in a software repository.
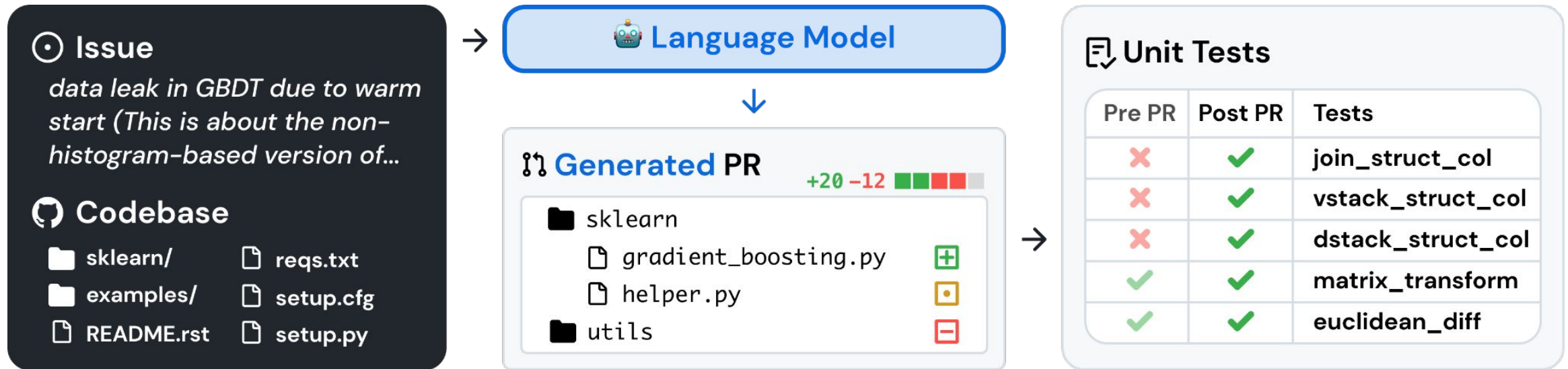
LLM Agents, and their success

SWE-Bench: Given an issue, automatically locate and fix bugs in a software repository.

LLM-Only Approach with GPT-4:   **2.8% resolve rate**

Naive LLM-Agent baseline: **23.2%**   8x better w/ agents

Amazon Q Developer Agent: **55.0%**

Agent equipped with <u>SWE-Search</u>:   **62.2% (Best in 2024)**

Our team member developed this!

LLM Agents, and their success

LLM Agents have been a huge success in software engineering.

Over 25% of Google's code is now written by AI.

3 in 4 programmers have tried AI.

17% of them use it "at all times". [WIRED]

How about chip design?

ChipAgentsBench

The first benchmark for **realistic, agent workflows** for debugging in chip design.

Much larger, harder, and more complex than existing benchmarks:

> **Existing:** Input Spec <100 lines, Single Output <100 lines, Toy Problems
> **ChipAgentsBench:**
> > Input: 2.8k SV files, >600k Lines of Code, Waveforms and Logs
> > Output: Multiple edits across the project
> > Real designs that have been taped out

We plan to open-source it to the community.

ChipAgentsBench Example

Context: The agent has access to the design, the test benches, some simulation results. There's a bug in the design that fails tests.

Issue: "Why is this test case failing? Please look at the waveform and find out."

Evaluation Criterion: Check if the agent's changes to the design can pass the tests.

Get in touch with ChipAgents!

Visit us at our Booth at DVCon Taiwan!