

AI-Driven Hardware Verification:
From Automation to Autonomy

Hamid Shojaei

DISTINGUISHED ENGINEER

cadence®

Chips Are the Bottleneck of the AI Boom

“

The world will need **more chips** than ever before as we push toward AGI.”

Sam Altman,
OpenAI CEO



“

The synergy between advanced chips and AI research is what will ultimately make AGI possible.”

Demis Hassabis,
DeepMind CEO



It takes 2+ years and hundreds of engineers to build a chip.
And it is increasing.

Hair on Fire

Verification: The Key Bottleneck for Building New Chips

Hardware × AI × Systems

Growing Pain:

Verification effort increased by 5X in the last decade with increasing chip complexity



TIME

>60%

OF THE SCHEDULE

TALENT

3 in 5

HEADCOUNT

RISK

76%

RISK OF RESPIN

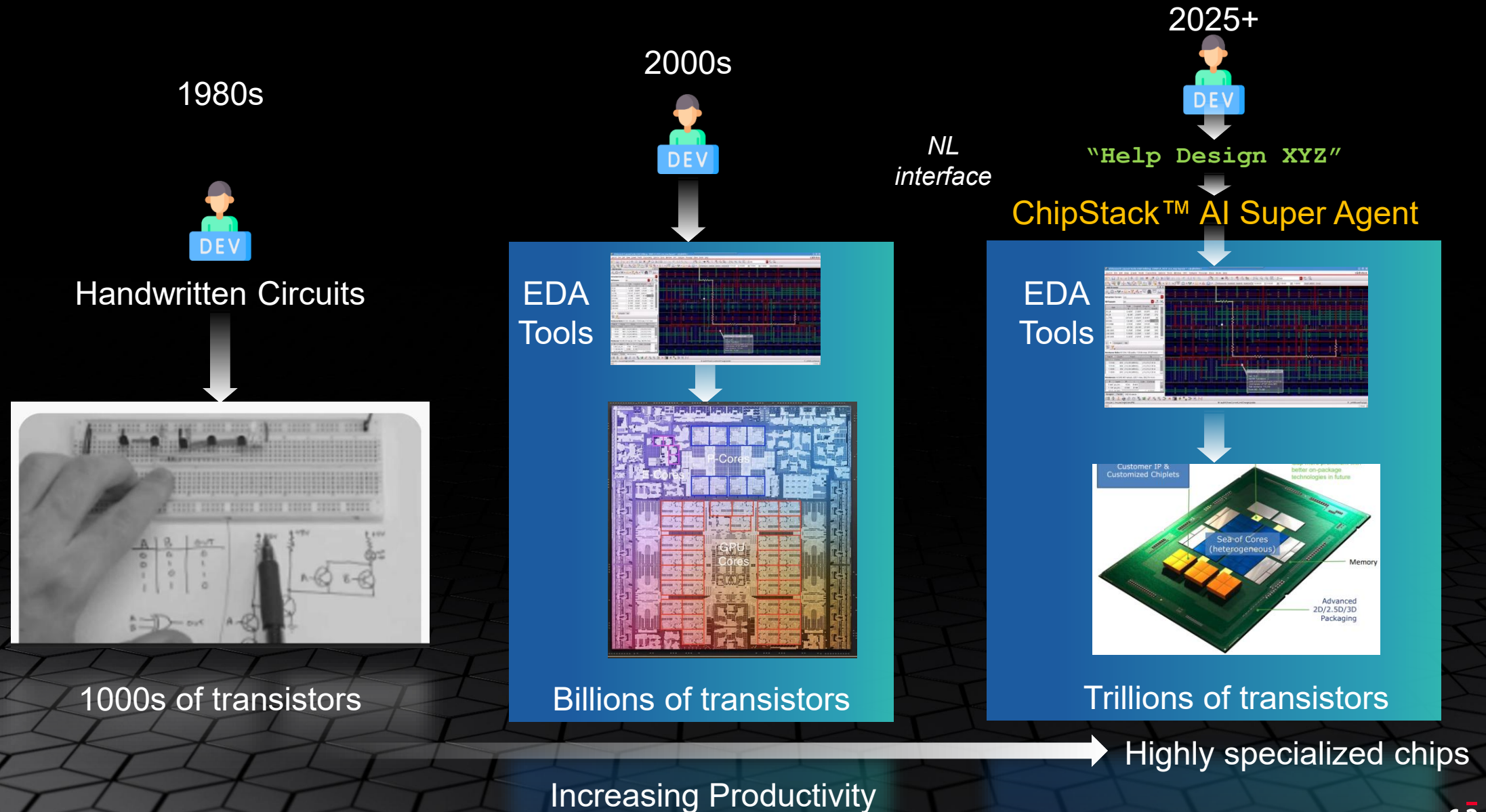
COST

>\$30M

COST OF RESPIN

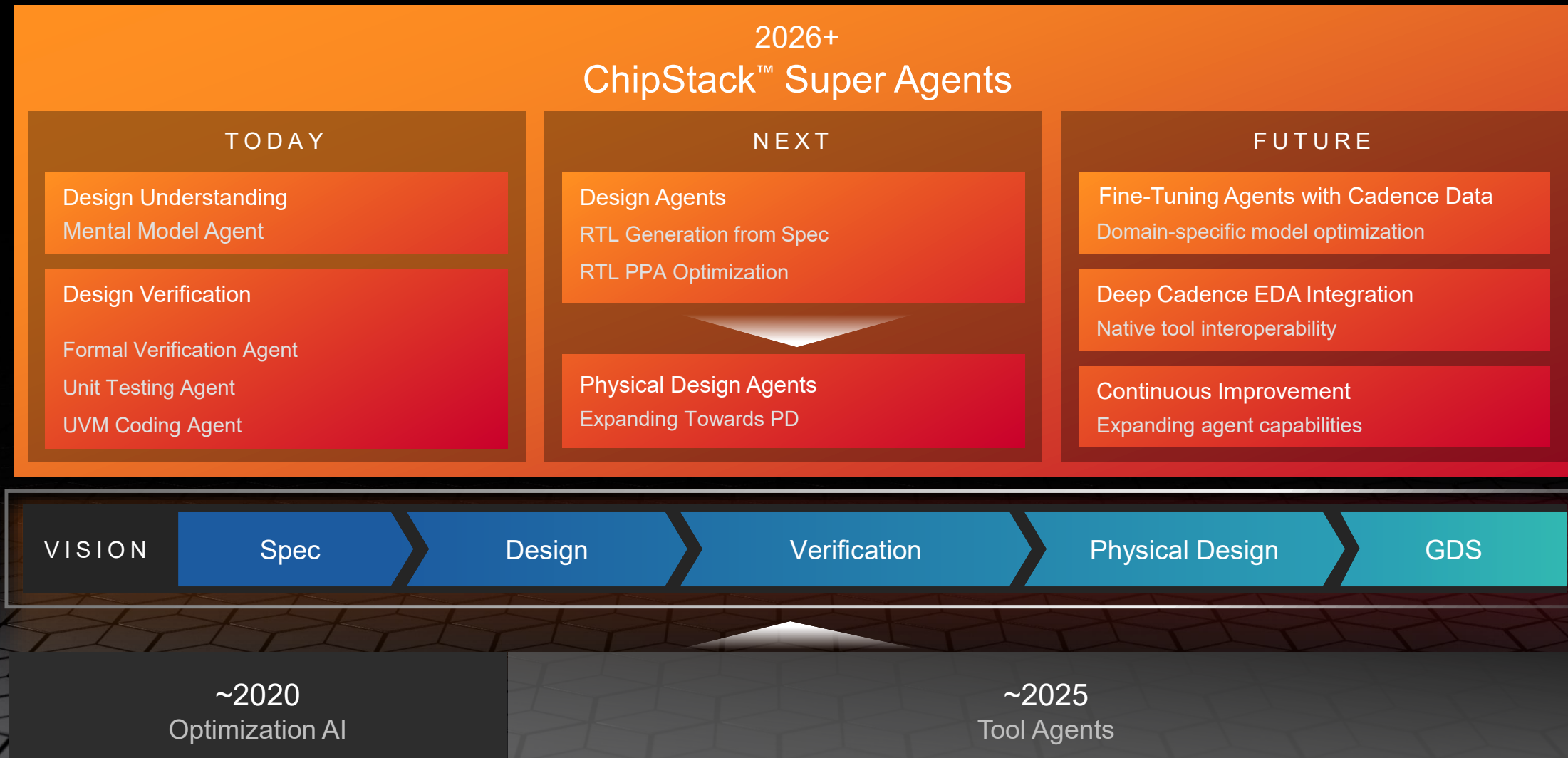
Solution: AI-Powered Chip Verification Engineers

Agentic Design Automation: Opportunity of the Decade



Cadence AI Journey

From Spec to GDS with AI Super Agents



Technology

ChipStack™ as an AI Chip Design Engineer

Give a team of virtual engineers to chip companies!



MODERN UI/UX

New way to collaborate with AI agents for DV

DEEP INTEGRATIONS

Use with your workflow and existing EDA tools

PRIVACY

Deploy on-premises or on-cloud

VERSATILE

Formal or simulation-based verification

Simulation Test Plan Scenarios

For all scenarios: [Generate](#)

Shared Code Testpoints

Select all [Add scenarios](#)

- ☒ BasicPushAndRead   
- ☐ UpdateAfterReadWithHaz   
- ☐ ShrinkOperationAndCredi   
- ☐ ConcurrentPushAndUpda   
- ☐ BackpressureAndFlowCor   
- ☐ ResetOutputsDefaultChec   
- ☐ ResetNoSpuriousActivity   

BasicPushAndRead

Description	Code	Results
CODE  		
Expand module		
<pre>196 196 task automatic test_BasicPu 197 197 fork 198 198 begin : timeout_block 199 199 #(PER_TASK_TIMEOUT); 200 200 \$display("Time: %0t, 201 201 \$display("Time: %0t, Stimuli is not observ test.", \$time); 202 202 failed_tests.push_back 203 203 end : timeout_block 204 204 begin : test_block 205 205 // Purpose: Verify th push and indexed read opera credit-based flow control, accurate credit reporting a 205+ // Purpose: Verify th push and indexed read credit-based flow con accurate credit repon</pre>		

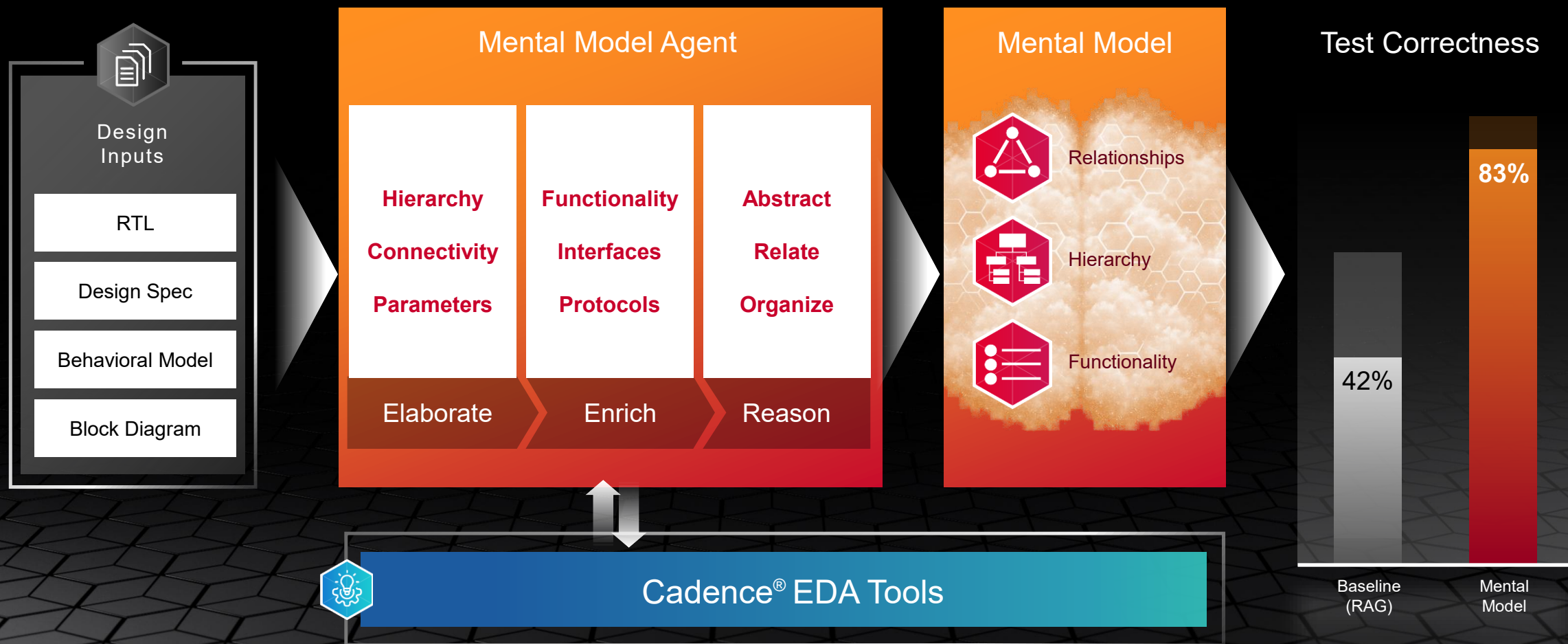
1 scenario selected [Gen](#)

 Tell ChipStack what you want to update a

The ChipStack AI Super Agent - How Does It Work?



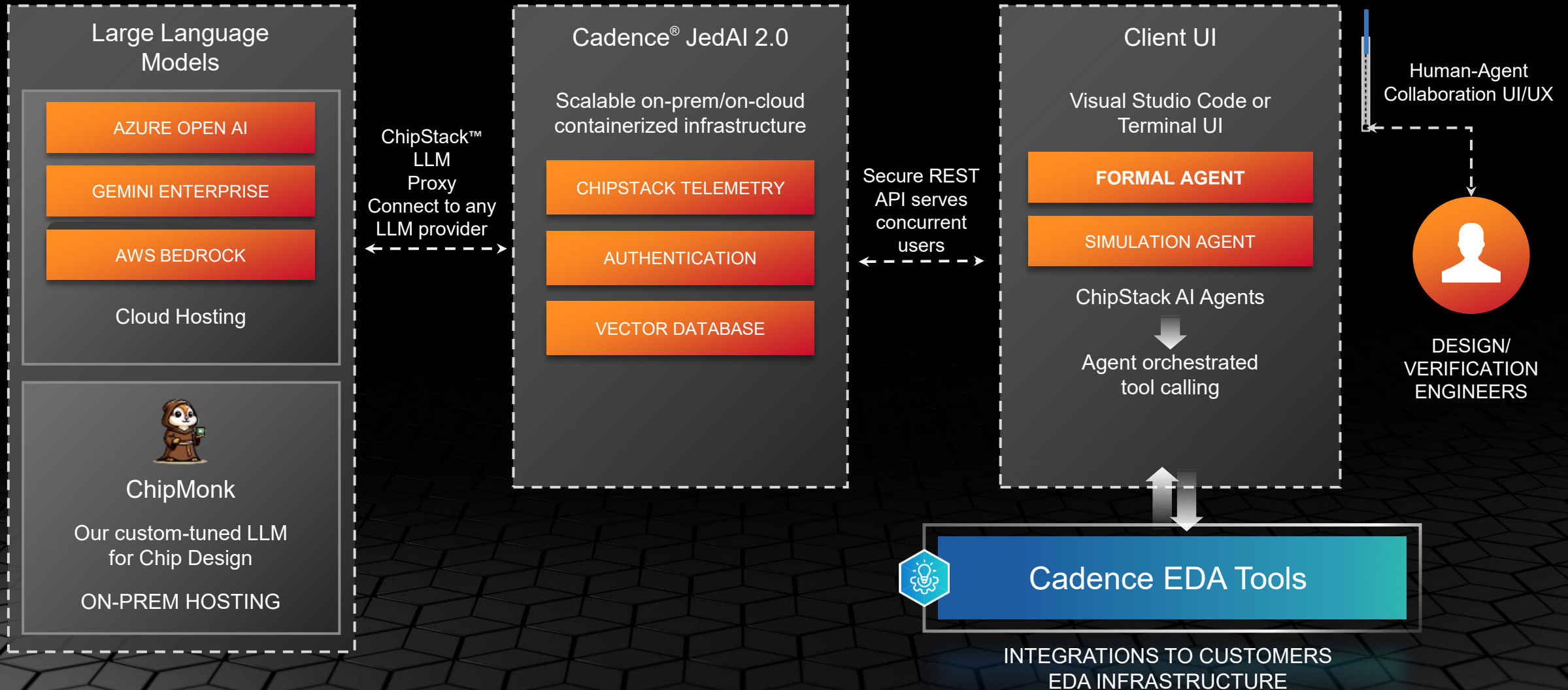
Mental Model: Single Source of Truth for Agentic Workflows



ChipStack Super Agent is Deeply Grounded in EDA Tools



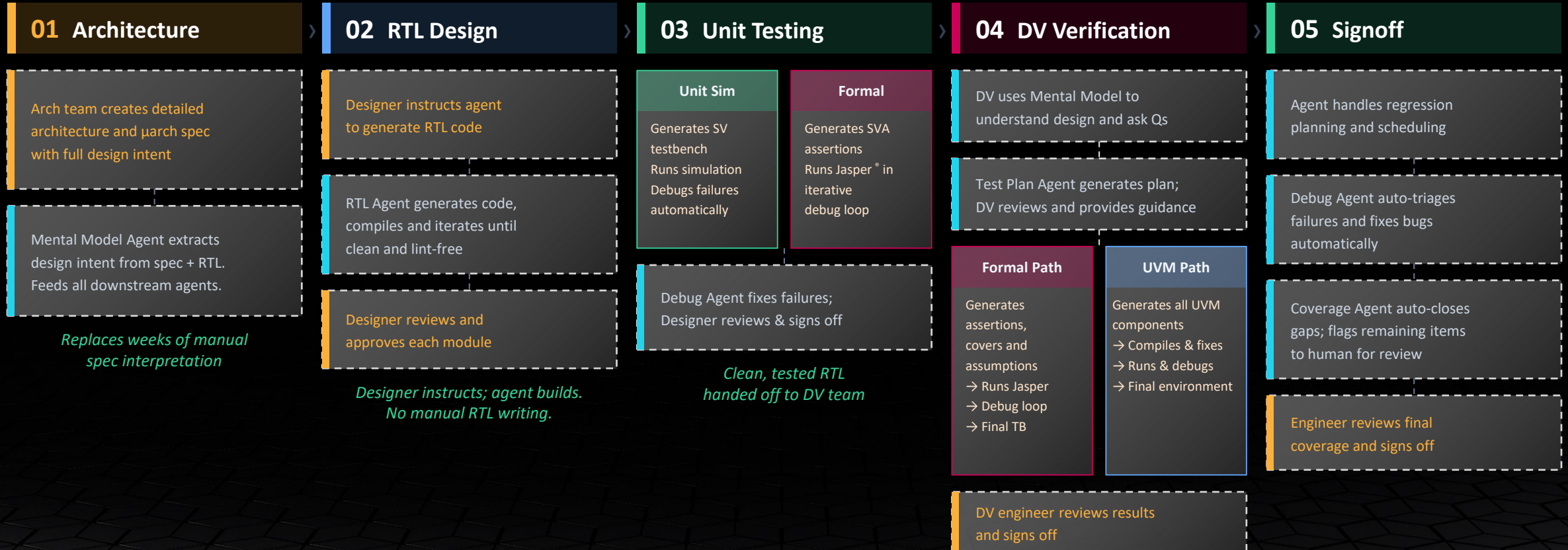
Product Architecture: Built to Work Anywhere



The New D&V Methodology with ChipStack

● Human — Architect, Review, Guide

● AI Agent — Generate, Compile, Debug, Fix



The Future: Humans architect, review and guide. Agents handle the tedious work — generating, compiling, debugging, and closing coverage. **No more writing boilerplate.**

The Future: One DV Super-Agent, End-to-End

● Human — sets intent, reviews, signs off

● DV Super-Agent — plans, decides, orchestrates

● Sub-Agents — execute the work in parallel

INPUT ▶ ANY OF THESE TRIGGERS A FULL DV RUN

Design Spec

User Prompt

RTL Change

DV SUPER-AGENT

PLAN · DECOMPOSE · ROUTE · RUN · MERGE · DEBUG · CLOSE

- Decides what runs in formal, simulation, or UVM
- Generates one unified test plan from formal + sim
- Merges coverage from formal + sim into a single DB
- Auto-debugg failures across both engines

ORCHESTRATES ▶ SPECIALIST SUB-AGENTS, EXECUTING IN PARALLEL

Mental Model

design intent

Formal

assertions

Simulation

dynamic runs

UVM

env + tests

Test Plan

unified plan

Debug

trriage + fix

Coverage

merge + close

DELIVERS ▶ END-TO-END ARTIFACTS, READY TO SHIP

Test Plans

Testbenches

Run Scripts

Coverage DB

Debug Triage

Live Dashboard

Virtual Senior DV Engineer: humans set intent → super-agent runs the full verification loop → **test plans, testbenches, scripts, coverage, dashboards** delivered end-to-end.

Models

Supervised Fine-Tuning (SFT)

Experience: Fine-tuning in-house model ChipMonk

Task

- Domain: Formal verification
- Goal: Train a model to generate SVA assertion code from natural language test-scenarios

Data

- 24,000 examples of (scenario, SVA) pairs
- Synthetically generated using GPT-4 by running our ChipStack™ flow
- Applied heuristic filters to remove low-quality data

Infrastructure

- We used Mosaic ML's Multi-Cloud Training tool (now Databricks)
- 8 GPUs, each with 40GB
- We also deployed the model on customer site

ChipMonk



Our custom-tuned LLM for formal assertion generation

Supervised Fine-Tuning (SFT)

Experience: Fine-tuning in-house model ChipMonk

Results

- Evaluated the tuned model for syntax validity

Model	Baseline	Flow-Engineered	Fine-Tuned
GPT-4 (Data Generator)	9.72%	73.07%	—
Llama 3.1-70B	16.49%	64.79%	91.71%

- Fine-tuning increased performance to ~92% (+26 percentage points)
- Smaller tuned model outperformed the larger data-generating model
- Task-specific SFT unlocked performance beyond flow-engineering

ChipMonk



Our custom-tuned LLM for formal assertion generation

Reinforcement Fine-Tuning (RFT)

Experience: VIP matching with single-turn RFT

Task: VIP Matching

- Train a model to bind DUT to verification IP (VIP)
- VIP are pre-defined components that can improve verification coverage and quality
- A VIP must be properly wired to a DUT's ports based on design features
- Time-consuming task for humans:
 - Find the right VIP for the DUT
 - Understand the VIP's many signals
 - Map VIP signals to DUT ports

Reinforcement Fine-Tuning (RFT)

Experience: VIP matching with single-turn RFT

Task: VIP Matching

- **Data**
 - Annotated by our chip-design team
 - Training: < **50** samples
 - Evaluation: **10** samples

- **Reward Model**
 - Simple string-matching with annotated labels

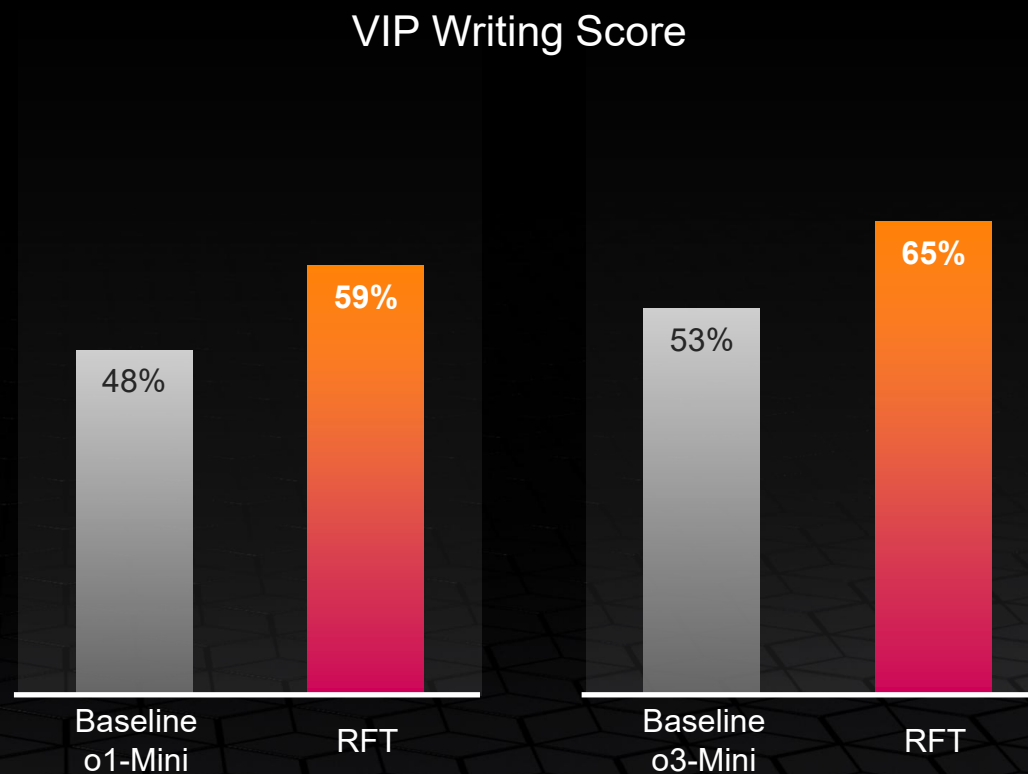
- **Platform**
 - OpenAI RFT Platform
 - Our collaboration allowed tuning o1-mini and o3-mini reasoning models
 - Tuning runs were ~4-5 hrs

Reinforcement Fine-Tuning (RFT)

Experience: VIP matching with single-turn RFT

Results

- Performance improved by ~12% for both base models
- We were able to improve upon a frontier reasoning model with 60 data samples on a highly specialized task



Conclusion

Constraints and Future Vision

SPEC TO SILICON: COLLABORATIVE MULTI-AGENT SYSTEMS

WHAT MADE THIS HARD

Confidential data → Limited public benchmarks

- EDA tools are slow, expensive, stateful
- Determinism vs flexibility tradeoffs

Multi-tenant deployment (on-prem + cloud)

WHAT I'D DO DIFFERENTLY

More realistic data for fine-tuning

- Better training data from production workflows

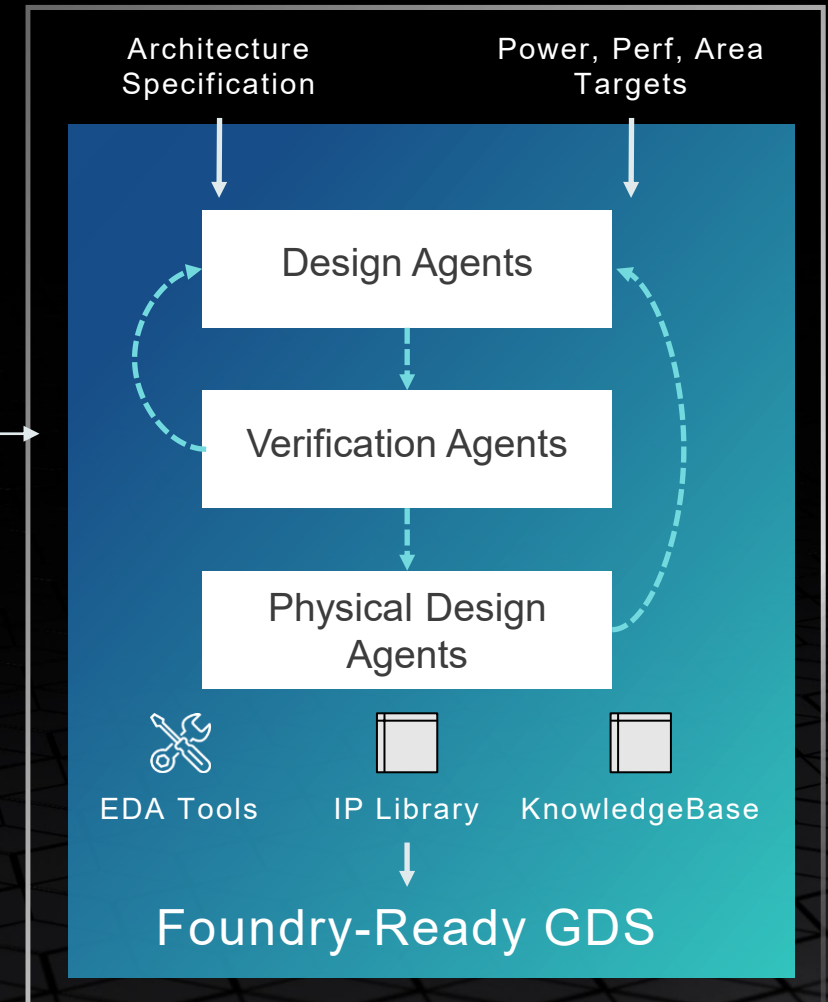
Robust agent memory system

- Long-term context retention across sessions

Deeper EDA tool integration



CHIP
DESIGNER





cā dence®

© 2026 Cadence Design Systems, Inc. All rights reserved worldwide. Cadence, the Cadence logo, and the other Cadence marks found at <https://www.cadence.com/go/trademarks> are trademarks or registered trademarks of Cadence Design Systems, Inc. Accellera and SystemC are trademarks of Accellera Systems Initiative Inc. All Arm products are registered trademarks or trademarks of Arm Limited (or its subsidiaries) in the US and/or elsewhere. All MIPI specifications are registered trademarks or service marks owned by MIPI Alliance. All PCI-SIG specifications are registered trademarks or trademarks of PCI-SIG. All other trademarks are the property of their respective owners.