Enhancing Reasoning to Adapt Large Language Models for Domain-Specific Applications

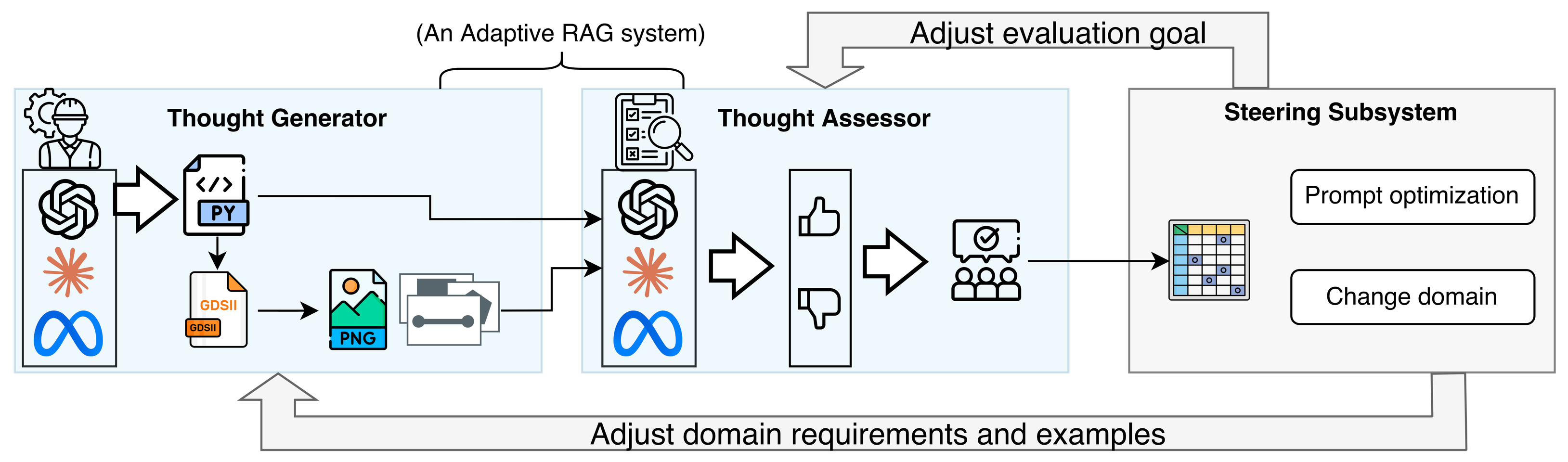
Bo Wen[1], Xin Zhang[1,2]

[1] IBM T. J. Watson Research Center, Yorktown Heights, NY, USA [2] MIT-IBM Watson AI Lab, Cambridge, MA

Introduction

- AI systems need to adapt quickly to new domains without extensive retraining
- Current LLMs struggle with spatial reasoning and applying domain knowledge
- Growing need for flexible AI architectures that can leverage existing models
- We propose a novel approach inspired by neuroscience to enhance LLM adaptability
- Aim to bridge the gap between general-purpose AI and specialized applications

SOLOMON Architecture

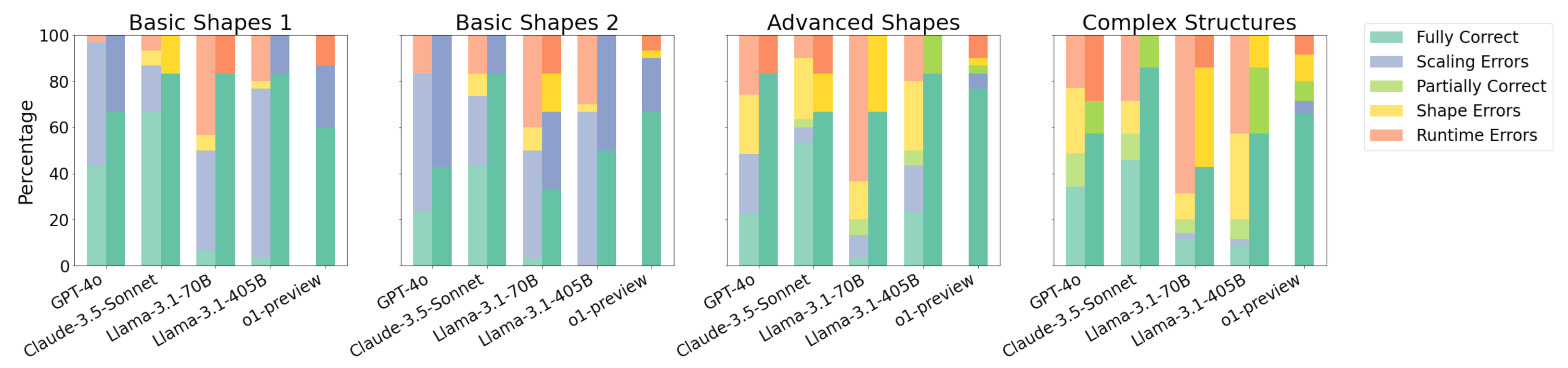


- SOLOMON = System for Optimizing Language Outputs through Multi-agent Oversight Networks
- Inspired by Brain-like AGI and Free Energy Principle theories
- Key components:
  - Thought Generators: Diverse pool of LLMs generating ideas
  - Thought Assessor: Analyzes proposed thoughts to refine output
  - Steering Subsystem: Human-operated component controlling attention
- Enables flexible AI systems for diverse specialized contexts without fine-tuning

Experiments

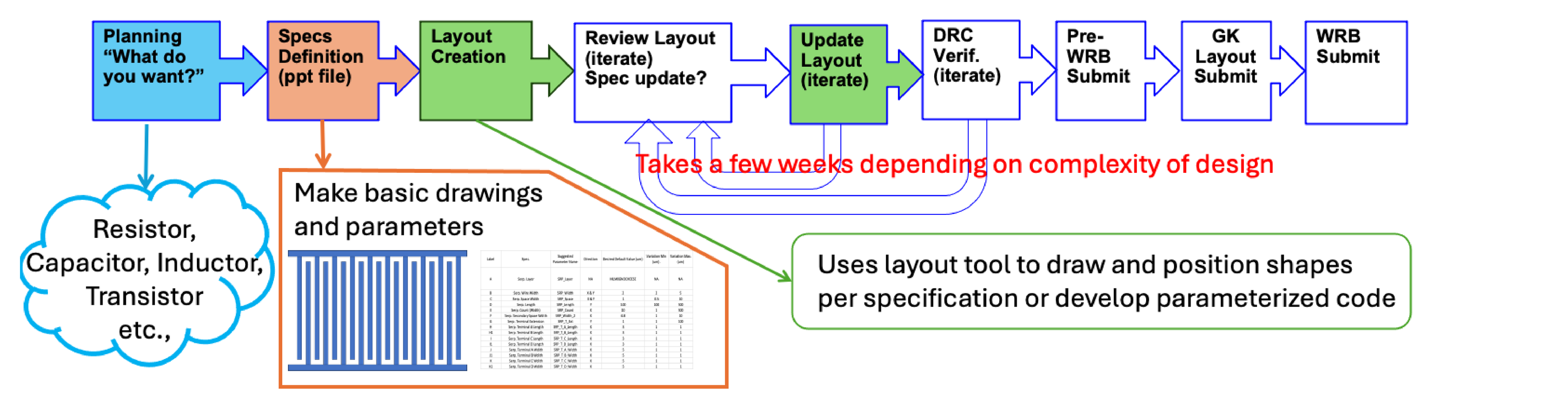
- Dataset of 25 layout design tasks across four categories:
  - Basic 1: Circle, Donut, Oval, Square, Triangle, Grid
  - Basic 2: Heptagon, Octagon, Trapezoid, Hexagon, Pentagon, Text
  - Advanced: Arrow, SquareArray, Serpentine, RoundedSquare, Spiral, BasicLayout
  - Complex Structures: RectangleWithText, MicrofluidicChip, ViaConnection, FiducialCircle, ComplexLayout, DLDChip, FinFET
- Evaluation process:
  - LLMs generate Python code to create desired GDSII output
  - Then converted to PNG images and organized into a table for human evaluation
  - Categorized into: correct, scaling error, partially correct, shape error, runtime error
- Baseline experiment with 5 LLMs: GPT-4o, Claude-3.5-Sonnet, Llama-3.1-70B, Llama-3.1-405B, o1-preview
- SOLOMON instances created using thoughts from 4 baseline LLMs (except o1-preview)

SOLOMON Performance



- SOLOMON significantly improved performance across all LLMs
- Notable reductions in runtime errors and scaling errors
- Enhanced ability to handle shape errors and partial correctness issues
- SOLOMON instances achieved comparable or superior results to o1-preview
- Claude-based SOLOMON surpassed o1-preview in 3 categories

Semiconductor Layout Design Process



- Layout design is a critical step in semiconductor manufacturing
- Requires precise spatial reasoning and domain knowledge application

Via Connection Case Study

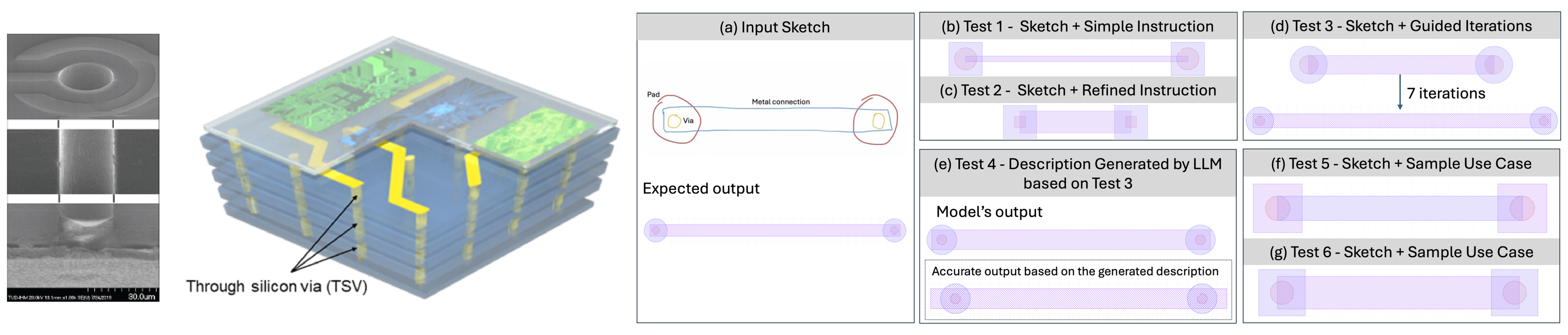


Figure 1: Left: SEM cross-section of a through-silicon via (TSV) structure. Middle: 3D schematic of TSVs in a stacked semiconductor device. Right: Sketch input and ChatGPT-generated outputs for the via connection experiment. The sketch depicts a desired layout with two vias connected by a metal layer and circular contact pads on top. The outputs show the progression of ChatGPT's understanding and refinement of the layout based on iterative feedback and context provided by the user.

- Via connections create 3D electrical pathways between chip layers
- Challenges: positioning and sizing to avoid short/open circuits
- LLMs struggle to translate domain knowledge from its memory into practical design requirements
- Highlights need for enhanced reasoning capabilities in LLMs

Detailed Results Example

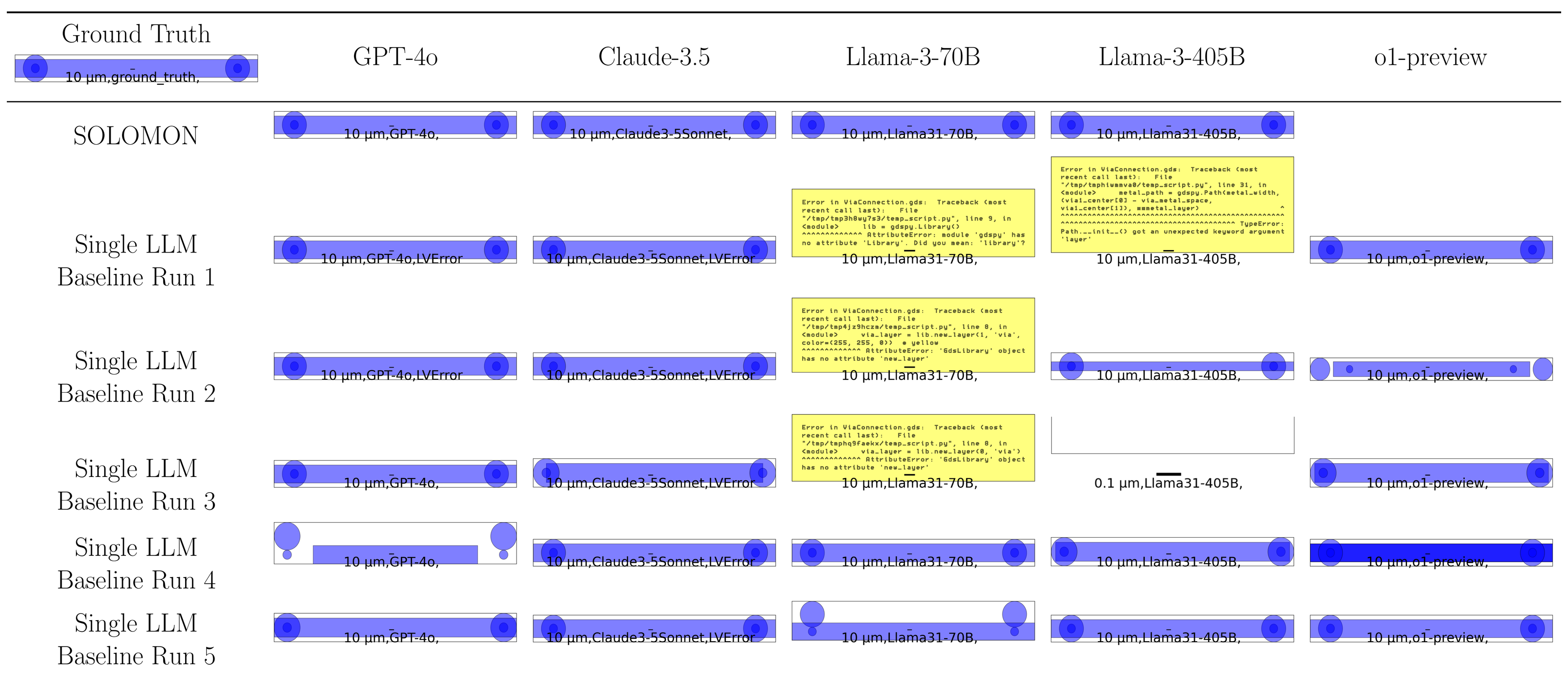


Table 1: ViaConnection Task Question: Create a design with three layers: via layer (yellow), metal layer (blue), and pad layer (red). The via radius is 10 units, pad radius is 30 units, and metal connection width is 40 units with a total length of 600 units. Position the first via at (50, 150) and the second via at (550, 150). Ensure the metal connection fully covers the vias and leaves a margin of 10 units between the edge of the metal and the pads. Leave a space of 50 units between the vias and the edges of the metal connection.

Conclusions and Future Work

- SOLOMON architecture demonstrated enhanced spatial reasoning capabilities and reduced errors
- Challenges remain in translating domain knowledge into practical design requirements
- Future research directions:
  - Developing comprehensive benchmark datasets
  - Improving multimodal input linking
  - Exploring performance with lower quality initial thoughts
- SOLOMON represents a significant step towards more adaptive AI systems for complex, domain-specific applications

Visit our GitHub repository for the complete benchmark dataset and code:
`https://github.com/wenboown/generative-ai-for-semiconductor-physical-design`

We thank Kuan Yu Hsieh for her valuable contribution and discussions.

Credit: image: `Flaticon.com`