MOSS: Multi-Modal Representation Learning on Sequential Circuits

Mingjun Wang[1,2,3,4], Bin Sun[1,3], Jianan Mu[1,3], Feng Gu[1,2,3,4], Boyu Han[5], Tianmeng Yang[6], Xinyu Zhang[1,3], Silin Liu[1,3], Yihan Wen[7], Hui Wang[4], Jun Gao[4], Zhiteng Chao[1,3,4], Husheng Han[1,3], Zizhen Liu[1,3], Shengwen Liang[1,3], Jing Ye[1,3,4], Bei Yu[2], Xiaowei Li[1,3], Huawei Li[1,3]

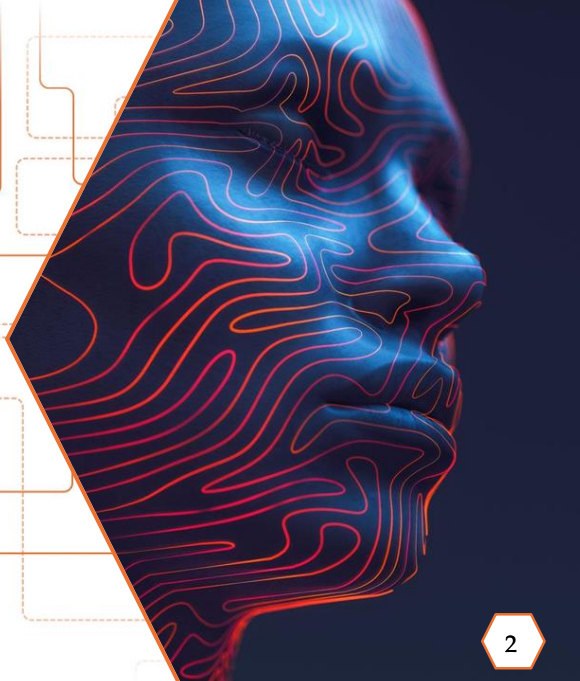
[1]SKLP, ICT, CAS [2]CUHK [3]UCAS [4]CASTEST Co., Ltd. [5]Stanford University
[6]Peking University [7]Beijing University of Technology

62 THE CHIPS TO SYSTEMS CONFERENCE

SPONSORED BY CEDA sigda

Outline

1. Introduction
2. Algorithms
3. Experimental Results



Introduction

62 THE CHIPS TO SYSTEMS CONFERENCE

SPONSORED BY CEDA IEEE Council on Electronic Design Automation sigda ACM

Background and Motivation

- **Rise of AI in EDA:** Deep learning has shown remarkable progress in Electronic Design Automation in recent years, significantly influencing tasks such as synthesis, placement, routing, and testing.
- **Importance of Circuit Representation Learning:** As neural networks become deeply integrated in EDA workflows, effective representation learning is critical for downstream predictions like timing, power, and testability.
- **Challenges in Sequential Circuits:** Modeling sequential circuits is more difficult than combinational ones due to their statefulness and feedback paths. Existing GNN-based approaches often suffer from long-range dependency issues in large circuits.
- **Multi-Modal Fusion:**
  - Large Language Models (LLMs) analyzing RTL may overlook gate-level details.
  - GNNs on netlists face difficulty scaling to large sequential designs.
  - Fusing these two modalities leverages both functional abstraction (from RTL) and structural accuracy (from gate-level netlists).

Overview of MOSS

- **Key Idea:** Introduce MOSS, a framework that combines GNNs and LLMs for multi-modal representation of sequential circuits:
  - LLM extracts higher-level functional embeddings (especially around DFFs).
  - GNN captures structural and timing dependencies across standard-cell netlists.
- **Main Contributions:**
  - **DFF Anchors**: DFFs are treated as anchors to reduce long-range dependencies in large designs.
  - **Adaptive Aggregator**: Custom aggregator for diverse standard cells.
  - **Two-Phase Propagation**: Forward and backward phases model temporal feedback in sequential circuits.
  - **Local-Global Alignment**: Node-level tasks (toggle rates, timing) plus global functionality alignment with RTL.
- **Performance Gains:** The proposed approach significantly improves accuracy in both functionality and performance predictions (e.g., arrival time, toggle rate), especially for large circuits.

Challenges and Motivation (Ref. to Fig.1 in Paper)

- **Challenge 1:** Prediction accuracy degrades on large sequential circuits due to long-range dependencies in GNN-based methods.
- **Challenge 2:** Infinite or extremely large truth tables in sequential circuits make it hard to form efficient functional supervision.
- **Motivation:**
  - DFFs can partition a sequential design naturally.
  - RTL offers higher-level semantics; combining it with gate-level netlist data can alleviate learning complexity.

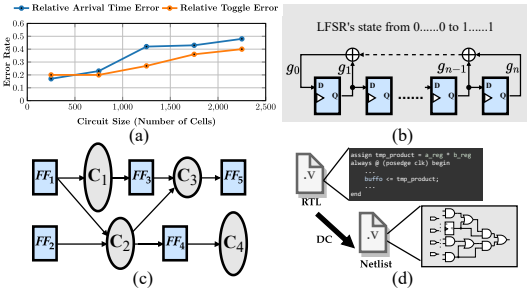


Fig.1 (Illustration of challenges and motivations in sequential circuits)

Algorithms

SPONSORED BY CEDA sigda

MOSS Framework

- **Multi-Modal Fusion of LLM and GNN:**
  - LLM provides semantically rich embeddings for critical nodes (e.g., DFFs).
  - GNN models structural dependencies in the standard-cell netlist.
- **Local-Global Alignment:**
  - Node-level tasks (e.g., toggle rate, arrival time).
  - Global netlist-RTL alignment ensures functional consistency.
- **Multi-Task Learning:** Joint optimization over multiple objectives (timing, power, functional equivalence).

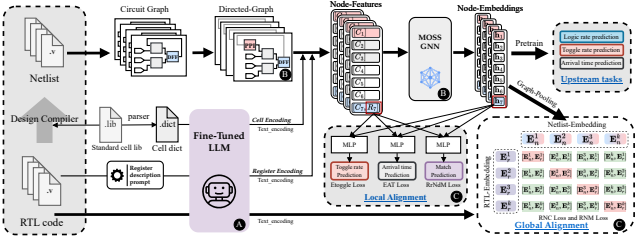


Fig.2 (Schematic of the MOSS multi-modal framework)

Semantic Enhancement via LLM

- **LLM Fine-Tuning:**
  - Trained on over 31k RTL designs for better understanding of register behavior and circuit semantics.
- **Node Feature Integration:**
  - **DFFs:** Combine corresponding register description (from RTL) with standard cell descriptions to obtain high-level functional features.
  - **Other Cells:** Get embeddings from LLM based on cell type, then concatenate with structural features.
- **Advantages:**
  - **Reduce Long-Range Dependencies:** Strengthen modeling around DFF anchor points.
  - **Improve Functional Supervision:** Leverage RTL's global semantic embedding to guide GNN training.

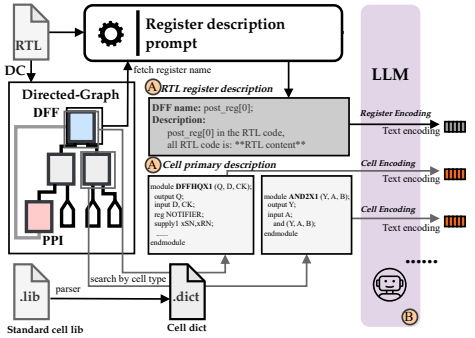


Fig.3 (Illustration of LLM-based feature extraction process)

Adaptive Aggregator and Two-Phase Propagation

Adaptive Aggregator:

- Standard cells exhibit diverse functions. A clustering-based approach (DBSCAN + hierarchical) groups similar cells.
- Each cluster uses a dedicated attention-based aggregator to capture unique interactions more effectively.

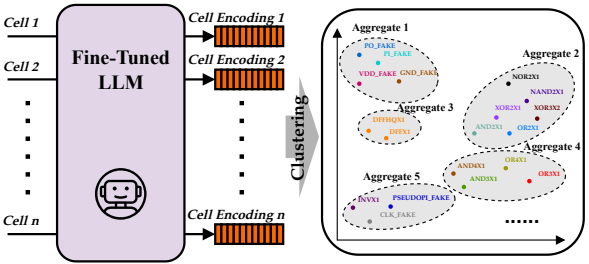


Fig.5 (Adaptive aggregator design)

Adaptive Aggregator and Two-Phase Propagation

Two-Phase Asynchronous Propagation:

- **Forward Phase:** Signal flows from PIs through combinational logic to DFF inputs.
- **Backward Phase (Turnaround):** Outputs from DFFs feed back into the circuit, capturing sequential feedback loops.
- Multiple iterative rounds refine the stateful behavior within sequential circuits.

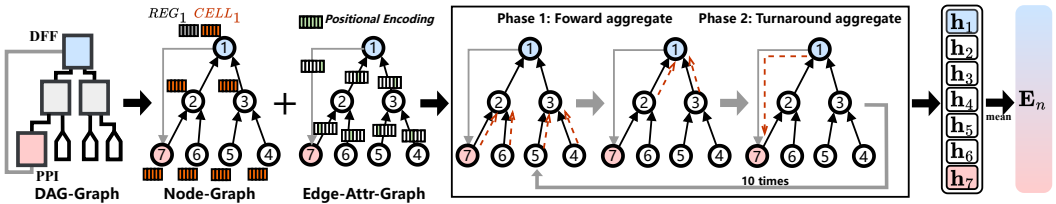


Fig.4 (Two-phase propagation scheme)

Local-Global Alignment Strategy

- **Local Alignment:**
  - Tasks such as toggle rate (Etoggle) and arrival time (EAT) are supervised to learn node-level behaviors.
  - An RTL-to-DFF matching loss aligns register descriptions with actual netlist DFF nodes.
- **Global Alignment:**
  - RTL-Netlist Contrastive Loss (RNC) and Matching Loss (RNM) ensure overall functional consistency and structural coherence.
  - These global losses are computed via pooled embeddings of the entire netlist vs. global RTL embeddings.
- **Multi-Task Objective:**

$$\mathcal{L}_{\text{total}} = \sum_i \lambda_i \mathcal{L}_{\text{task}_i}$$

Balancing multiple losses guarantees both fine-grained accuracy and holistic alignment.

Experimental Results

Experimental Setup and Metrics

- **Datasets:**
  - 31,701 RTL designs, synthesized with Synopsys Design Compiler.
  - Circuit sizes range from 100 to 5000 standard cells, ensuring diversity in scale and functionality.
- **Evaluation Metrics:**
  - **ATP (Arrival Time Prediction)**, **TRP (Toggle Rate Prediction)**, **PP (Power Prediction)**
  - **FEP (Functional Equivalence Prediction)**: Checking RTL vs. netlist functional consistency.
- **Compared Methods:**
  - **DeepSeq2 (state-of-the-art GNN)**,
  - Ablation versions of MOSS without certain components.
- **Implementation Details:**
  - PrimePower, VCS for ground truth of timing, toggle rates, power.

Quantitative Results (Ref. to Table I, Table II)

TABLE I Performance Comparison of MOSS Framework Variants

Circuit	#Cells	DeepSeq2 [14]			MOSS w/o FAA			MOSS w/o AA			MOSS w/o A			MOSS		
		ATP	TRP	PP	ATP	TRP	PP	ATP	TRP	PP	ATP	TRP	PP	ATP	TRP	PP
max_selector	278	81.4	78.7	94.6	47.0	75.8	88.6	82.3	85.2	94.5	95.4	89.4	99.9	**95.6**	**90.5**	**99.9**
pipeline_reg	610	77.6	83.6	91.4	52.2	63.6	63.4	80.5	88.3	90.2	94.2	92.1	94.1	**94.5**	**92.4**	**94.6**
prbs_generator	643	87.8	76.5	71.7	57.2	72.7	81.7	78.6	82.5	90.8	92.0	**87.4**	94.5	**93.0**	85.4	**95.1**
shift_reg_24	731	86.9	80.9	90.4	58.2	63.9	75.4	82.4	85.6	92.5	**96.2**	**90.2**	**97.6**	95.8	89.0	97.5
error_logger	812	79.5	83.2	94.3	58.5	59.0	81.3	81.2	80.2	95.3	94.5	85.4	99.5	**95.0**	**86.3**	**99.7**
signed_mac	1306	66.4	77.3	**95.6**	26.9	56.1	73.6	76.5	78.4	88.6	93.8	83.8	92.3	**94.5**	**85.3**	94.1
wb_data_mux	1364	95.7	64.3	88.6	45.3	25.5	82.6	85.4	75.6	88.5	98.8	82.9	91.2	**99.1**	**83.3**	**96.2**
mult_16x32_to_48	4144	57.6	66.6	80.1	19.3	40.1	54.1	75.2	72.3	85.4	93.9	84.8	91.5	**94.3**	**87.9**	**93.5**
Average	-	79.1	76.4	88.4	45.6	57.1	75.1	80.3	81.0	90.7	94.9	87.0	95.1	**95.2**	**87.5**	**96.3**

ATP: Arrival Time Prediction accuracy (%). **TRP**: Toggle Rate Prediction accuracy (%). **PP**: Power Prediction accuracy (%).
MOSS w/o FAA: MOSS without Feature Enhancement, Adaptive-Aggregator and Alignment.
MOSS w/o AA: MOSS without Aligment and Adaptive-Aggregator. **MOSS w/o A**: MOSS without Alignment.

Quantitative Results (Ref. to Table I, Table II)

TABLE II RTL-netlist functional equivalence prediction accuracy (FEP) on different circuit sources

Circuit	MOSS w/o FAA	MOSS w/o AA	MOSS w/o A	MOSS
github_0	4.8	19.4	24.1	**91.4**
github_1	5.3	20.3	33.6	**95.0**
github_2	10.0	23.7	32.0	**94.3**
huggingface_0	7.9	16.4	19.5	**94.1**
huggingface_1	8.7	18.3	22.9	**93.6**
huggingface_2	14.1	21.1	27.5	**93.5**
Average	8.5	19.9	26.6	**93.7**

Bold numbers indicate the best performance for each circuit
All circuits are from public repositories (GitHub and HuggingFace)

Analysis of Results

- **Arrival Time (ATP):** MOSS achieves an average of 95.2% accuracy, over 16% higher than DeepSeq2. Maintains >94% even on large (>2000-cell) designs.
- **Toggle Rate (TRP):** MOSS attains 87.5%, outperforming DeepSeq2 by about 11%. Particularly strong in complex sequential circuits like signed MAC and pipeline designs.
- **Power Prediction (PP):** Up to 96.3% accuracy, leveraging global functional features from LLM to better capture dynamic power factors.
- **Functional Equivalence (FEP):** MOSS shows around 93.7% on public GitHub/HuggingFace test sets, significantly outperforming ablation models. Highlights the critical role of global alignment.
- **Ablation Studies:** Removing feature enhancements, adaptive aggregators, or local-global alignment all degrades performance, confirming each module's importance.

Convergence Curves

- **Pre-Training Phase:**
  - Overall and sub-task losses steadily decrease, indicating stable convergence.
- **Multi-Modal Alignment Phase:**
  - RNC/RNM losses drop to very low values within several epochs, suggesting successful RTL-netlist alignment.
- GNN and LLM effectively learn matching representations, ensuring robust multi-modal embeddings.

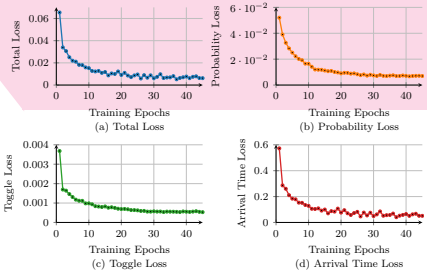


Fig.7 Loss for tasks over training epochs

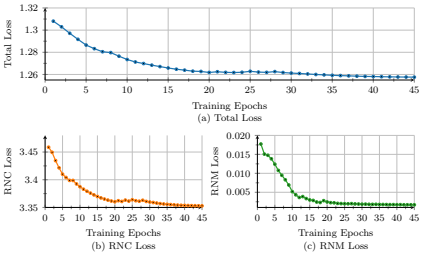


Fig.8 multi-modal alignment loss

Conclusion and Future Work

- **Key Contributions:**
  - Joint LLM-GNN multi-modal framework for large-scale sequential circuits.
  - Adaptive aggregator + two-phase asynchronous propagation to capture long-range temporal dependencies.
  - Local-global alignment for enhanced functionality and performance prediction.
- **Experimental Highlights:**
  - Superior results in arrival time, toggle rate, power, and functional consistency predictions.
  - Remains highly accurate (>94%) even for circuits over 2000 cells.
- **Future Directions:**
  - More efficient fine-tuning strategies for LLM on larger RTL repositories.
  - Extending MOSS to layout, physical design, or model checking tasks to further drive AI4EDA research.

AI

Security

Systems

EDA

Design

62 THE CHIPS TO SYSTEMS CONFERENCE

SPONSORED BY CEDA IEEE Council on Electronic Design Automation SIGDA ACM