OpenABC-D: A Large-Scale Dataset For Machine Learning Guided Integrated Circuit Synthesis

Animesh Basak Chowdhury
New York University

Benjamin Tan
New York University

Ramesh Karri
New York University

Siddharth Garg
New York University

Abstract

Logic synthesis is a challenging and widely-researched combinatorial optimization problem during integrated circuit (IC) design. It transforms a high-level description of hardware in a programming language like Verilog into an optimized digital circuit netlist, a network of interconnected Boolean logic gates, that implements the function. Spurred by the success of ML in solving combinatorial and graph problems in other domains, there is growing interest in the design of ML-guided logic synthesis tools. Yet, there are no standard datasets or prototypical learning tasks defined for this problem domain. Here, we describe OpenABC-D, a large-scale, labeled dataset produced by synthesizing open source designs with a leading open-source logic synthesis tool and illustrate its use in developing, evaluating and benchmarking ML-guided logic synthesis. OpenABC-D has intermediate and final outputs in the form of 870,000 And-Inverter-Graphs (AIGs) produced from 1500 synthesis runs plus labels such as the optimized node counts, and delay. We define a generic learning problem on this dataset and benchmark existing solutions for it. The codes related to dataset creation and benchmark models are available at `https://github.com/NYU-MLDA/OpenABC.git`. The dataset generated is available at `https://archive.nyu.edu/handle/2451/63311`.

1 Introduction

Complex integrated circuits (ICs) can have over a billion transistors making hand-design impossible. Hence, the IC industry relies on electronic design automation (EDA) tools that progressively transform a high-level hardware into a layout ready for IC fabrication. EDA tools let designers focus on describing function at a high-level using a hardware description language (HDL) like Verilog, without worrying about low-level implementation of the IC. Increasing design complexity and scalability challenges in the design flow has raised interest in machine learning (ML) for EDA [1].

The first step in EDA is logic synthesis. Logic synthesis transforms an HDL program into a functionally equivalent graph (netlist) of Boolean logic gates while attempting to minimize metrics such as area, power, and delay. Since logic synthesis is the first in a sequence of EDA steps that yields the final IC layout, the quality of its output impacts the size, power, and speed of the final IC. Even the simplest version of this problem, logic minimization, is Σ_p^2-Hard [2, 3].[1] Commercial logic synthesis tools use heuristics developed by academia and industry [4]. State-of-the-art in logic synthesis applies a *sequence* of logic minimization heuristics to transform a sum-of-products (SOP) or an and-inverter-graph (AIG) representation. Common heuristics remove redundant nodes, refactor Boolean formulas, and simplify node representations. The *order* in which these heuristics are applied – the

[1] Σ_p^2-Hard problems are hard even with access to an oracle solver for NP-complete problems.

Preprint.

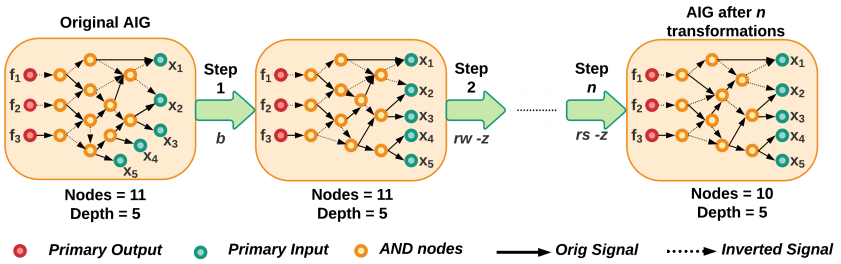


Figure 1: Logic synthesis optimizations on And-Inverter Graphs (AIG).

synthesis recipe – is critical to the quality of results. Designers use synthesis recipes that either work well for a range of inputs or have to hand-tune them by trial-and-error.

The success of deep learning methods in solving a range of combinatorial and graph problems has spurred interest in ML-guided logic synthesis [5, 6, 7, 8, 9, 10]. However, they report results on small datasets and solve different versions of the problem. As a consequence, benchmarking and comparing SoTA solutions, especially on "real-world" designs, is challenging. This is because there is no comprehensive, labeled dataset of publicly available and prototype problems that can serve as benchmarks.

This motivates us to create a realistic and feature-rich dataset for logic synthesis that is of interest to both EDA researchers (as the first large-scale dataset in this application domain) and ML researchers (who could be interested dealing with a specialized domain of structured graph data). Our work presents a dataset of 29 open source designs with 870,000 data samples. The EDA community can use this data to train ML models for a range of logic synthesis optimization tasks. This dataset can augment data for other problems in EDA and hopefully be useful to the graph datasets community.

1.1 Overview of Logic Synthesis

A digital hardware intellectual property (IP) block is designed using an HDL like Verilog or VHDL. Specifying functionality at this abstraction is typically called behavior-level or register transfer level (RTL) design. Logic synthesis takes an RTL implementation and outputs an (optimized) gate-level netlist representation of the design that can be mapped to a standard cell library (i.e., technology mapping). To meet a designer-specified area and delay overhead, optimization takes place before and after technology mapping. In this work, we focus on technology independent logic optimization. This combinatorial optimization problem uses simple Boolean logic gates.

A gate-level netlist is a Boolean function with binary-valued inputs and uses logical operations like AND, NOT, and XOR. This netlist can be represented canonically (e.g., as a truth table) or in other formats, like AIGs and majority-inverter graphs (MIG). AIG is a directed acyclic graph (DAG) representation with 2-input AND function (nodes) and NOT function (dotted edges). The AIG is popular since it scales and can compactly represent industrial-sized designs; it is used in the state-of-the-art open source logic synthesis tool, ABC [11]. AIG allows structural optimizations like cut enumeration, Boolean implication and DAG-based heuristics. The following are fundamental sub-graph optimizations supported by ABC (ABC's commands are in parentheses):

1. Balance (b) is a depth-optimization step to minimize the delay of a design. Given an AIG representation in the form of a DAG, *balance* applies tree-balancing transformations using associative and commutative properties on logic function.
2. Rewrite (rw, rw -z) is DAG-aware logic rewriting heuristic that does template pattern matching on sub-trees and replaces them with equivalent logic functions. Rewrite heuristic algorithm uses a k-way cut enumeration (k varies from 4-8) with the objective of finding an optimized representation of the sub-tree. In ABC, zero-cost variant (rw -z) does not immediately reduce the number of nodes of the DAG. The transformed structure can be optimized using other heuristics.
3. Refactor (rf, rf -z) can potentially change a large part of the netlist without caring about logic sharing. The method traverses iteratively on all nodes in the netlist, computes maximum fan out free cones and replaces them with equivalent functions if it improves cost (e.g., reduce number of nodes). Zero cost variant is available.

4. Re-substitution (rs, rs -z) optimizes by representing the logical function of a node using logic functions of existing nodes. Typically, *k* nodes are introduced to represent the function and compared against the number of redundant nodes that are no longer required. k determines the size of the sub-circuit that can be replaced. Re-substitution improves logic sharing.

Logic synthesis of a gate-level netlist is a sequential decision process applying sub-graph optimization heuristics in a non-trivial combination to obtain an optimized design. We term the fundamental sub-graph level optimizations as *synthesis transformations*. IC designers develop a sequence of synthesis transformations to get an optimized netlist. Technology mapping then results in a circuit satisfying quality of result (QoR) in terms of area, delay, and power consumption. We call sequences of synthesis transformations a *synthesis recipe*. The objective of one recipe is to reduce the number of nodes and depth of the DAG network to directly correlate to minimizing area and delay [8].

1.2 Motivation

Problem statement: Given a gate-level netlist as AIG graph representing a set of boolean functionalities, determine the sequence of sub-graph optimization steps generating the optimal AIG representing same functionalities. The computational complexity of the problem is Σ_p^2-Hard.

Success of ML algorithms has prompted researchers to re-examine logic synthesis, where most of the fundamental research happened in 1990s and 2000s [12, 13, 14]. The overarching question is: *can past experience lead to informed decision-making for future problem instances?* So far, EDA engineers use years of experience from past synthesis runs to intuit a good synthesis recipe for new IPs. To scale up designs and enable faster design sign-off, researchers formulate learning tasks in the logic synthesis domain and propose ML algorithms to solve them, e.g.,
To predict synthesis recipe quality: [5] proposes a classification model to determine "angel" and "devil" synthesis recipes, i.e., identify if a recipe will generate a good quality design for an IP by training a model on data from a few synthesis runs.
To predict the "best" optimizer: [8] proposes an ML model that identifies how a sub-circuit should be represented (AIG/MIG) to match sub-circuit and synthesis recipes. The work demonstrated state-of-the-art results compared to baseline ABC synthesis results.
To predict the best synthesis recipes (reinforcement learning (RL)-guided): Recent work [7, 6, 10, 9] formulates the problem as a Markov Decision Process (MDP), where the future AIG depends on current AIG (the state) and synthesis transformation (the action). Past synthesis transformations do not impact the transition to new AIG state. The state-action transition yields a deterministic result for an action. However for the agent to be effective across diverse hardware IPs, it needs to explore a wide range of state-action pairs before being used in "exploitation" phase on unseen data.

While promising, prior work suffers from the inappropriateness of an apples-to-apples comparison of the various proposed techniques in the absence of a standardized dataset, configurations and benchmarks. We observed that prior work is evaluated on different data sets, making it difficult for researchers to understand the effectiveness of each approach (Appendix B). Additionally, there is no clear explanation on the choice of benchmarks considered for experiments. To address this shortcoming, we thus present a three-fold contribution:
1. OpenABC-Dataset (OpenABC-D): We release OpenABC-D, a large-scale synthesis dataset comprising of 870,000 data samples by running 1500 synthesis recipes on 29 open source hardware IPs and preserving intermediate stage and final AIGs.
2. Data Generation Framework: We provide an open source framework that can generate labeled data by performing synthesis runs on hardware IPs using different synthesis recipes.
3. Benchmarking ML models: We benchmark the performance of simple graph models on our learning task using OpenABC-D.

2 The Data Generation Pipeline

For wider access, we focus on open source EDA platforms as key to advancing ML research in EDA. Thus, we propose the OpenABC-D framework using open source EDA tools, thus making it freely available for anyone to generate data. Note, however, that this framework requires substantial compute hours to generate synthesis data, considerable preprocessing of intermediate-stage data, and converting it to compatible formats for applying ML models. Thus, we applied the OpenABC-D framework on a set of open source IPs and make this data available to the community. We highlight

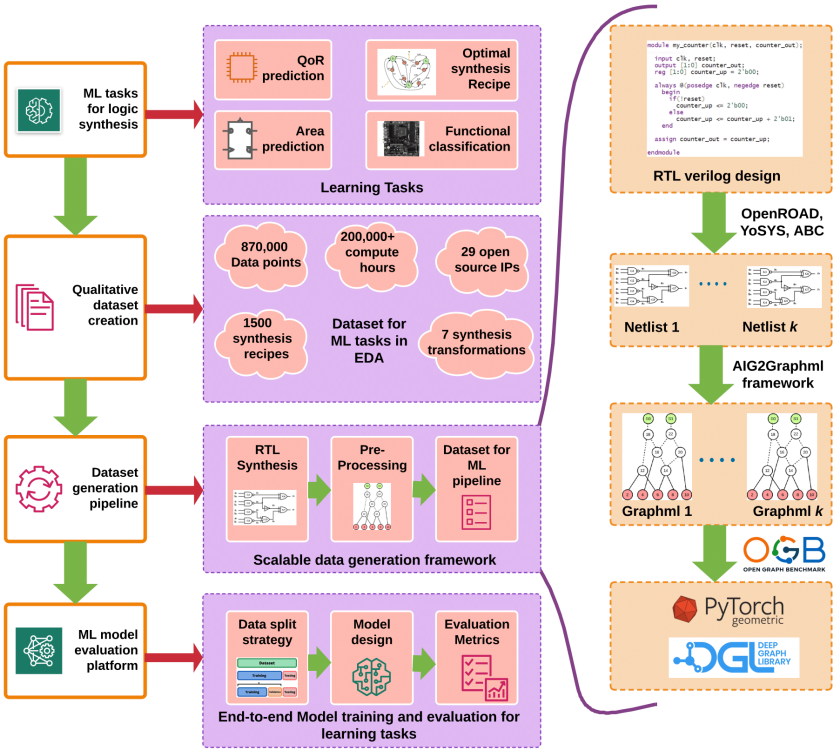


Figure 2: OpenABC-D framework

our contributions and challenges for developing OpenABC-D (Fig. 2): an end-to-end large-scale data generation framework for augmenting ML research in circuit synthesis.

2.1 Open Source Tools

We use OpenROAD v1.0 [15] EDA to perform logic synthesis; it uses Yosys [16] as the front-end engine (currently v0.9). Yosys performs logic synthesis in conjunction with ABC [11]. It can generate a logic minimized netlist for a desired QoR. We use networkx v2.6 for graph processing. For ML frameworks on graph structured data, we use pytorch v1.9 and pytorch-geometric v1.7.0. We collect area and timing of the AIG post-technology mapping using NanGate 45nm technology library and "5K_heavy" wireload model. The dataset generation pipeline has three stages: (1) RTL synthesis, (2) Graph-level processing, and (3) Preprocessing for ML.

2.2 Register Transfer Level (RTL) Synthesis

In this stage, we take the specification of IPs (in Verilog/VHDL) and perform logic synthesis. Yosys performs optimization on the sequential part of the IP and passes on the combinational part to ABC for logic optimization and technology mapping. First, ABC structurally hashes the combinational design to create AIGs. Post-hashing, a user-provided synthesis recipe performs tech-independent optimization. Our framework allows two options: 1) automated synthesis script generation and 2) user-defined synthesis scripts. A synthesis script is a synthesis recipe with additional operations to save intermediate-stage AIGs. The AIGs generated by ABC are in the `BENCH` file format [17]. The intermediate and final AIGs have different graph structures despite having the same function. Synthesis transformations affect parts of the graph differently, yielding a diverse collection of graph structured data. We prepared $K = 1500$ synthesis recipes each having $L = 20$ synthesis transformations.

2.3 Graph-Level Processing

We wrote a gate-level netlist parser that takes a `BENCH` file (containing the AIG representation of an IP) as input and generates a corresponding `GRAPHML` file. While generating the `GRAPHML` format of the design, we preserve the fundamental characteristics of the AIGs. Nodes in the AIG are 2-input

IP	Characteristics of Benchmarks						
	PI	PO	N	E	I	D	Function
spi [18]	254	238	4219	8676	5524	35	Serial peripheral interface
i2c[18]	177	128	1169	2466	1188	15	Bidirectional serial bus protocol
ss_pcm[18]	104	90	462	896	434	10	Single slot PCM
usb_phy[18]	132	90	487	1064	513	10	USB PHY 1.1
sasc[18]	135	125	613	1351	788	9	Simple asynch serial controller
wb_dma[18]	828	702	4587	9876	4768	29	Wishbone DMA/Bridge
simple_spi[18]	164	132	930	1992	1084	12	MC68HC11E based SPI interface
pci[18]	3429	3157	19547	42251	25719	29	PCI controller
wb_conmax[18]	2122	2075	47840	97755	42138	24	WISHBONE Conmax
ethernet[18]	10731	10422	67164	144750	86799	34	Ethernet IP core
ac97_ctrl[18]	2339	2137	11464	25065	14326	11	Wishbone ac97
mem_ctrl[18]	1187	962	16307	37146	18092	36	Wishbone mem controller
bp_be[19]	11592	8413	82514	173441	109608	86	Black parrot RISCV processor engine
vga_lcd[18]	17322	17063	105334	227731	141037	23	Wishbone enhanced VGA/LCD controller
des3_area[18]	303	64	4971	10006	4686	30	DES3 encrypt/decrypt
aes[18]	683	529	28925	58379	20494	27	AES (LUT-based)
sha256[20]	1943	1042	15816	32674	18459	76	SHA256 hash
aes_xcrypt[21]	1975	1805	45840	93485	36180	43	AES-128/192/256
aes_secworks[22]	3087	2604	40778	84160	45391	42	AES-128 (simple)
fir[20]	410	351	4558	9467	5696	47	FIR filter
iir[20]	494	441	6978	14397	8596	73	IIR filter
jpeg[18]	4962	4789	114771	234331	146080	40	JPEG encoder
idft[20]	37603	37419	241552	520523	317210	43	Inverse DFT
dft[20]	37597	37417	245046	527509	322206	43	DFT design
tv80[18]	636	361	11328	23017	11653	54	TV80 8-Bit Microprocessor
tiny_rocket[15]	4561	4181	52315	108811	67410	80	32-bit tiny riscv core
fpu[23]	632	409	29623	59655	37142	819	OpenSparc T1 floating point unit
picosoc[23]	11302	10797	82945	176687	107637	43	SoC with PicoRV32 riscv
dynamic_node[15]	2708	2575	18094	38763	23377	33	OpenPiton NoC architecture

Table 1: Open source IP characteristics (unoptimized). Primary Inputs (PI), Primary outputs (PO), 'Nodes (N), Edges (E), 'Inverted edges (I), Netlist Depth (D). Color code: Communication/Bus protocol, Controller, Crypto, DSP, Processor, Processor+control, Control+Communication

AND gates and the edges in the AIG are either inverters or buffers. We define Node type and number of incoming inverted edges as two node-based features and the edge type as the one edge feature. For each IP, we save $K \times L = 30,000$ graph structures.

2.4 Preprocessing for ML

This stage involves preparing the circuit data for use with any ML framework. We use pytorch-geometric APIs to create data samples using AIG graphs, synthesis recipes, and area/delay data from synthesis runs. The dataset is available using a customized dataloader for easy handling for preprocessing, labeling, and transformation. We also provide a script that helps partition the dataset (e.g., into train/test) based on user specified learning tasks. Our aim is to help the EDA community to focus on their domain problems without investing time on data generation and labeling.

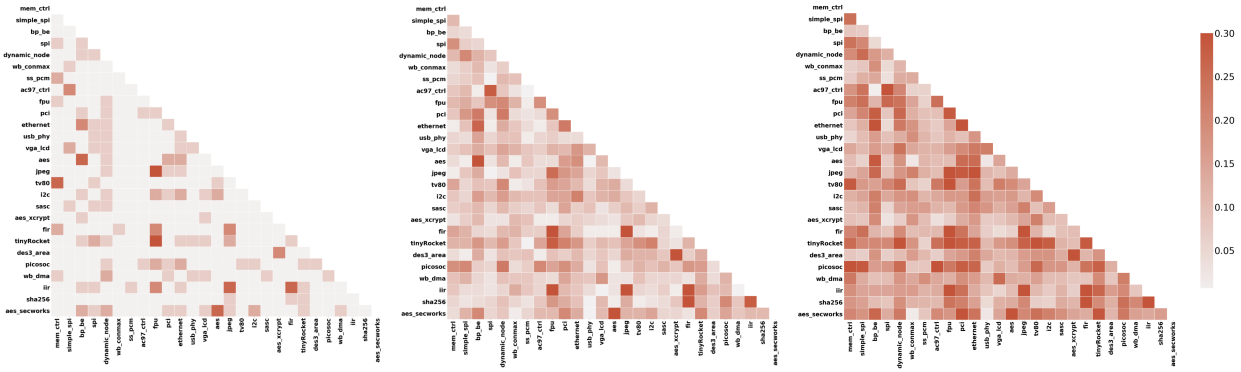
3 OpenABC-D Characteristics

To produce the OpenABC-dataset, we use 29 open source IPs with a wide range of functions. In the absence of a preexisting dataset like ImageNet that represents many classes of IPs, we hand-curated IPs of different functionalities from MIT LL labs CEP [20], OpenCores [18], and IWLS [24]. These benchmarks are more complex and functionally diverse compared to ISCAS [25, 26] and EPFL [27] benchmarks that are used in prior work and represent large, industrial-sized IP. In the context of EDA, functionally diverse IPs should have diverse AIG structures (e.g., tree-like, balanced, and

designIP_synthesisID_stepID.pt						
Graph	**Structural features**		**Synthesis recipe**		**Labels**	
Connectivity	Node	Edge	Recipe ID	Step ID	Individual	Final
Adjacency matrix of graph	Type (PI/PO/AND), # Incoming inverters (0/1/2)	Type (Buffer: 0, Inverter: 1)	ID (0-1499), Recipe array (length 20)	Index of intermediate step of recipe	#PIs, #POs,#nodes, #inverters, #edges, depth,IP name	#nodes, area, delay of final AIG

Table 2: Data sample description of OpenABC-D generated dataset

skewed) that mimic the distribution of real hardware designs. Table 1 summarizes the structural and functional characteristics of the data after synthesis (without optimizations). These designs are functionally diverse – bus communication protocols, computing processors, digital signal processing cores, cryptographic accelerators and system controllers.



(a) Top 1% (b) Top 5% (c) Top 10%

Figure 3: Correlation plots amongst top $k\%$ synthesis recipes various IPs. Darker colors indicate higher similarity between the top synthesis recipes for the pair of IPs.

We run this data through 1500 different synthesis recipes, each having 20 transformation steps, saving the AIG after each transformation step (i.e., producing 20 AIGs per recipe, per IP). The synthesis recipes were prepared by randomly sampling from the set of synthesis transformations, assuming a uniform distribution. We ran synthesis using server-grade Intel processors for $200,000+$ computation hours to generate the labeled data. We analyzed the top k synthesis recipes by varying $k = 15$, 75 and 150 (1% to 10%). We found that the similarity of the top recipes is less than 30% (Fig. 3). The synthesis recipes are diverse and relate to the graph structures and sequence of transformations in the recipes. We describe characteristics of each data sample and naming conventions.

Encoding the AIGs: As described in §2.3, we convert the `BENCH` to `GRAPHML`. We store the graph as an adjacency matrix preserving the node and edge features. We use two node-level features and numerically encode them: (1) node type and (2) number of predecessor inverter edges. For edges, we used binary encoding: 0 for original and 1 for inverted signal.

Encoding the Synthesis Recipe: We encode each synthesis transformation that ABC supports and create a vector for all 1500 synthesis recipes. We tag and identify each synthesis recipe with a *synthesis ID* as a 20-dimensional vector.

Preparing Data Samples: We create each data sample by combining the AIG encoding and synthesis flow encoding and add label information about the sample. We mark each data sample as `designIP_synthesisID_stepID.pt`. Table 2 presents the information of each sample. The sample's name indicates the IP, the applied synthesis recipe, and the state of AIG (initial/intermediate/final), e.g., `aes_syn149_step15.pt` represents the intermediate AIG obtained after applying the first 15 synthesis transformations of recipe ID 149. The labels include: # of primary inputs and outputs; information about the AIG including # of nodes, inverted edges and depth; IP function; # of nodes in the AIG after applying the recipe; area and delay post-technology mapping. The quality of the result for every IP is different for each recipes (as illustrated in Fig. 4).

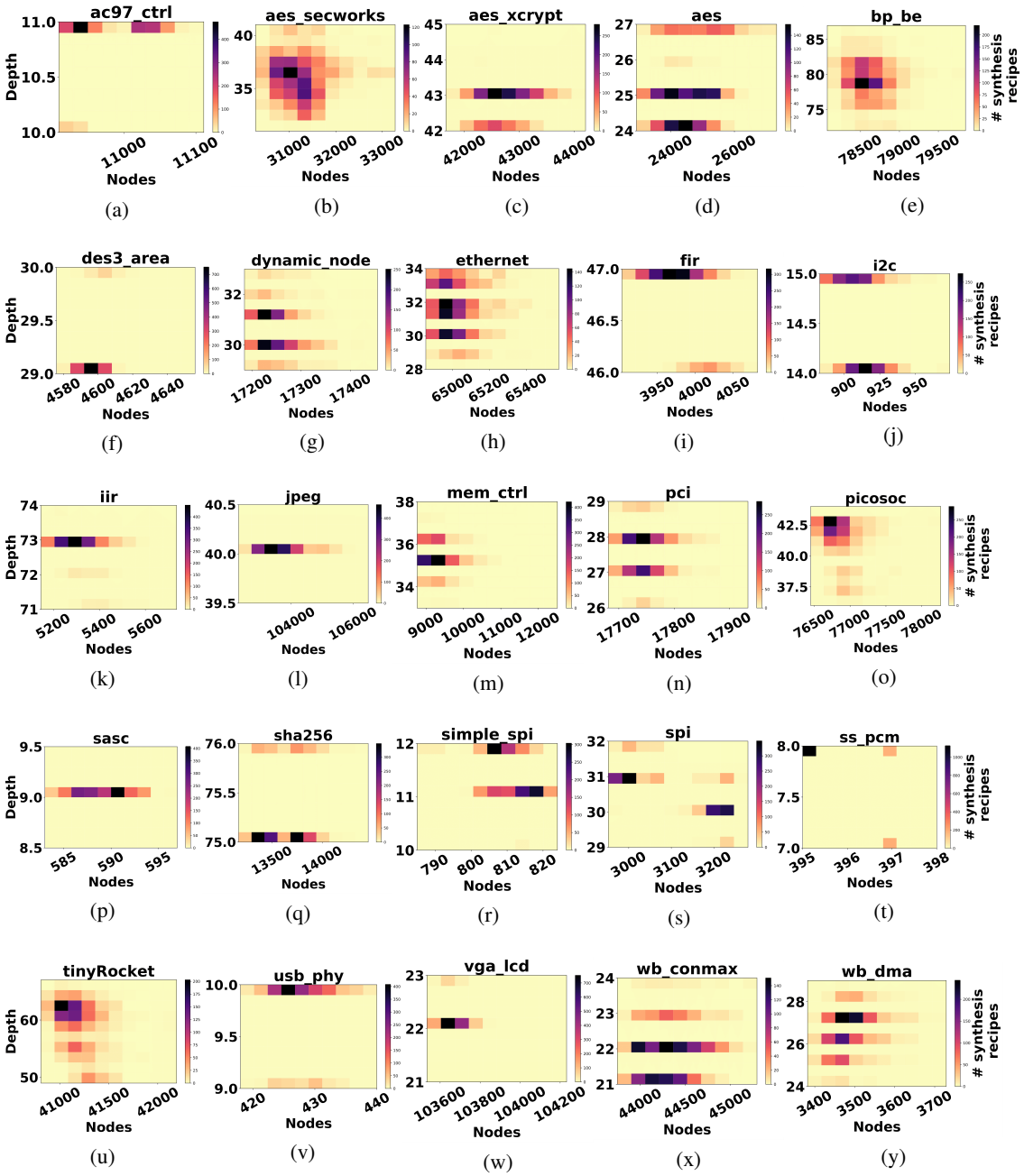


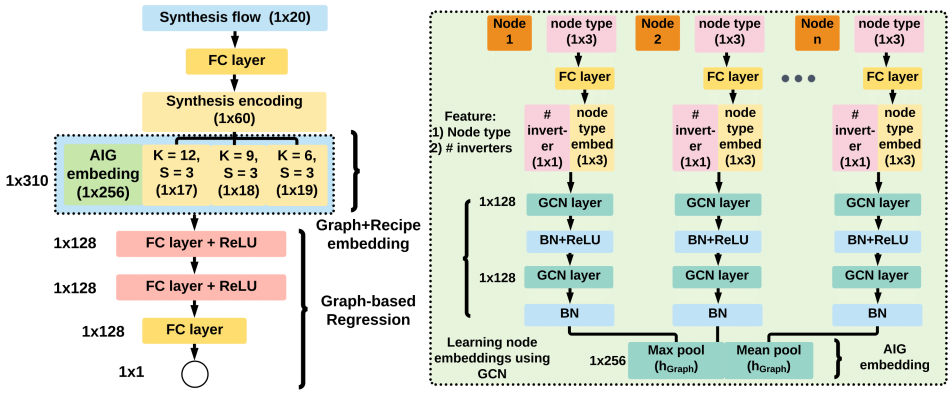
Figure 4: Heatmaps of the synthesis quality (logic depth vs. # of nodes) after after using 1500 synthesis recipes. Darker colors represent more recipes achieving that synthesis result.

4 Benchmarking Learning on OpenABC-D

The OpenABC-D netlist dataset can be used as a benchmarking dataset for ML-based EDA tasks. An example of an ML task for logic synthesis is supervised learning for predicting the quality of the synthesis result (e.g., % of nodes optimized, longest path in the design). We now demonstrate the use of OpenABC-D on this task.

Net	AIG Embedding				Recipe Encoding				FC Layers		
	I	L1	L2	Pool	I	# filters	kernels	stride	# layers	architecture	dr
Net1	4	128	128	Max+Mean	60	3	6,9,12	3	3	310-128-128-1	0
Net2	4	64	64	Max+Mean	60	4	12,15,18,21	3	4	190-512-512-512-1	0
Net3	4	64	64	Max+Mean	60	4	21,24,27,30	3	4	178-512-512-512-1	0,2

Table 3: Hyperparameters for the QoR Prediction Models. I: Input dimension, dr: dropout ratio. L1, L2: dimension of GCN layers



(a) Baseline model

(b) AIG embedding network

Figure 5: Graph convolution network for synthesis recipe QoR prediction

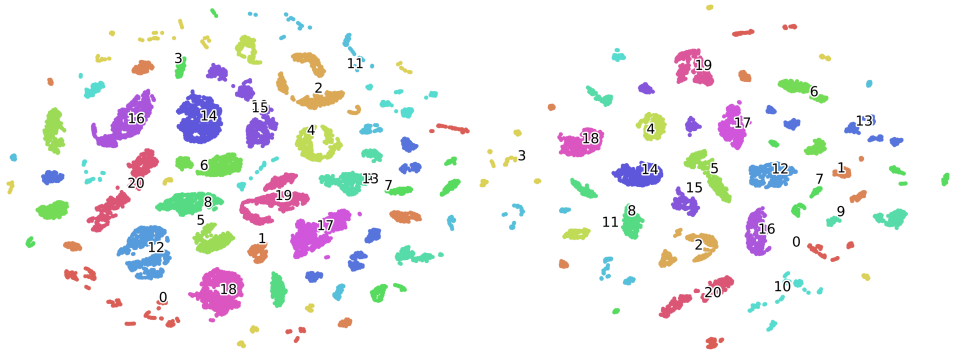
4.1 Example Task: Predicting the Quality of a Synthesis Recipe

Determining the "best" synthesis recipe is a challenge in logic synthesis. Logic minimization involves finding a transformation sequence that leads to an optimized AIG. Our empirical observations (Fig. 4) on the impact of synthesis recipes on different IPs show that there is no single synthesis recipe that works well on all IPs. As synthesis runs are computationally costly, predicting QoR of a given synthesis recipe on a given IP can help designers find better recipes in a short time. For example, consider the normalized number of AIG nodes remaining after applying a synthesis recipe as the QoR. One task variant is described next (more variants discussed in Appendix).

Variant 1: Predict QoR of unseen synthesis recipes Given an IP and synthesis recipe, can we predict the quality of the synthesis result? We train a model using all IP netlists and the AIG outputs of 1000 synthesis recipes used on those netlists (training dataset of $29 \times 1000 = 29000$ samples). To evaluate the model, we test if it predicts # of nodes in the AIG after synthesis with each of the 500 remaining unseen recipes. This mimics the scenario when existing expert guided synthesis recipes have been tried out on IP blocks and QoR prediction is required for new synthesis recipes (to pick the best since synthesis of all options is time consuming).

Model architecture and hyperparameters: For all task variants, we train the model to do graph-level predictions using a synthesis recipe encoding. To benchmark the performance of graph convolution networks (GCN), we considered a simple architecture (Fig. 5a). Here, the AIG is input to a two-layer GCN. The GCN learns node-level embeddings. Graph-level embedding is generated by a readout across all nodes in the graph. In our case, the readout is a global max pooling and average pooling. For synthesis recipe encoding, we pass the numerically encoded synthesis recipe through a linear layer and follow it with a set of filters of 1D convolution layer. The kernel size and stride length are tunable hyper-parameters. We concatenate graph-level and synthesis recipe embeddings and pass it through a set of fully connected layers to perform regression. In Table 3, we show three configurations of the model (hyperparameter settings). For all settings, we used batchsize=64 and initial learning rate=0.001. We trained all the networks for 80 epochs. We used Adam optimizer for our experiments.

Results: We consider mean squared error (MSE) metric to evaluate model effectiveness: Net1:



(a) Train samples (1000/class) (b) Test samples (500/class)

Figure 6: t-SNE plots. Labels - ac97_ctrl: 0, sasc: 1, wb_conmax: 2, ss_pcm: 3, tinyRocket: 4, i2c: 5, mem_ctrl: 6, des3_area: 7, aes_secworks: 8, simple_spi: 9, pci: 10, dynamic_node: 11, usb_phy: 12, wb_dma: 13, iir: 14, sha256: 15, aes: 16, fpu: 17, fir: 18, tv80: 19, spi: 20.

0.648 ± 0.05, Net2: 0.815 ± 0.02, Net3: 0.579 ± 0.02. We show scatter plots of inferred vs. actual values of normalized number of nodes in the optimized netlist in Appendix D. We observe that GCNs plus synthesis encodings consistently showed good results on most benchmarks as seen in Fig. 7, 8 and 9). The scatter plots follow the trend of $y = x$ showing that QoR prediction of an unknown synthesis recipe is good. The kernel filters learn and capture properties amongst the synthesis recipe subsequences trained on which are responsible for effective performance. However, there is a slightly noticeable difference in scatter plots for IPs: `fir`, `iir` and `mem_ctrl`. Inference of net3 is better on these IPs than net1 and net2. Both net1 and net2 performs bad on `fir` and `iir`; however net1 performs better than net2 on `mem_ctrl`.

5 Discussion and further insights

Limitations Using OpenABC-D dataset, we demonstrated the use of a simple GCN model for a typical task. In any ML-guided EDA pipeline, it is important that the model is trained on dataset generated from samples of hardware IPs representing the true distribution. Our effort of creating OpenABC-D with data generated from hardware IP of different functionalities is the first step towards this. We note, however, that there is room for OpenABC-D to grow. While OpenABC-D is more comprehensive and tailored towards use with ML approaches compared to existing datasets, there remains a scarcity of industrial-scale open source hardware IPs. Therefore, it is important for an EDA engineer to be cautious about using the inference results and perform an out-of-distribution check for any unseen hardware IP.

What can a GCN learn from this data? For further insights, we examined the embeddings learned by a GCN trained on our proposed data. The netlist embedding is the lower-dimensional representation of complex IPs for tasks like area and delay prediction after technology mapping onto a cell library. OpenABC-D has labels like number of nodes, function of IP, and depth of DAG. They can be used to learn robust embeddings from the netlist graph. We trained a two-layer GCN to identify the functionality of an IP based on its netlist structure (batch size$= 64$ with Adam optimizer, learning rate$= 10^{-3}$, decay$= 10^{-2}$, and categorical cross-entropy as loss function) and generated t-SNE plots to visualize the learned embeddings in Fig. 6. We obtained 98.05% accuracy for IP classification after 35 epochs of training. Given an unknown IP, it might be possible to see if it has similarities with the known IPs, and if not, suggest when a model should be retrained with new data. In addition, the insights from graph convolutional network (GCN) embeddings suggest that the models can learn rich structural and functional information from AIGs. This shows that GCN models can be pre-trained using simple self-supervised tasks (like predicting depth of AIG) and can later be used for fine-tuning/transfer learning approaches for tasks like predicting delay of circuit.

6 Conclusion

ML-guided IC design needs standard datasets and baseline models to nurture open, reproducible research. OpenABC-D dataset and benchmarking models will help the ML for EDA community towards an open, standardized evaluation pipeline (like ImageNET [28] for images, GLUE [29] for language, LibriSpeech [30] for speech). Creating such a large-scale domain-specific dataset involves substantial effort in generating labelled data, pre-processing it, and making it available in standard formats that are ingestible by standard ML evaluation pipelines. Creating OpenABC-D dataset required 200,000+ hours of computational resources (see Table 5). We discussed an important prototypical and fundamental task in logic synthesis: predicting the QoR quality of a synthesis recipe for a given IP and benchmarked simple GCN models across a diverse set of hardware IPs.

References

[1] Guyue Huang, Jingbo Hu, Yifan He, Jialong Liu, Mingyuan Ma, Zhaoyang Shen, Juejian Wu, Yuanfan Xu, Hengrui Zhang, Kai Zhong, Xuefei Ning, Yuzhe Ma, Haoyu Yang, Bei Yu, Huazhong Yang, and Yu Wang. Machine Learning for Electronic Design Automation: A Survey. *ACM Transactions on Design Automation of Electronic Systems*, 26(5):40:1–40:46, June 2021.

[2] Christopher Umans. Hardness of Approximating Σ_p^2 Minimization Problems. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 465–474, 1999.

[3] David Buchfuhrer and Christopher Umans. The complexity of Boolean formula minimization. *Journal of Computer and System Sciences*, 77(1):142–153, 2011.

[4] Luca Amarú, Patrick Vuillod, Jiong Luo, and Janet Olson. Logic optimization and synthesis: Trends and directions in industry. In *IEEE Design, Automation & Test in Europe Conference (DATE)*, pages 1303–1305, 2017.

[5] Cunxi Yu, Houping Xiao, and Giovanni De Micheli. Developing synthesis flows without human knowledge. In *ACM/IEEE Design Automation Conference (DAC)*, pages 1–6, 2018.

[6] Abdelrahman Hosny, Soheil Hashemi, Mohamed Shalan, and Sherief Reda. DRiLLS: Deep reinforcement learning for logic synthesis. In *IEEE Asia South Pacific Design Automation Conf.*, pages 581–586, 2020.

[7] Winston Haaswijk, Edo Collins, Benoit Seguin, Mathias Soeken, Frédéric Kaplan, Sabine Süsstrunk, and Giovanni De Micheli. Deep learning for logic optimization algorithms. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4, 2018.

[8] Walter Lau Neto, Max Austin, Scott Temple, Luca Amaru, Xifan Tang, and Pierre-Emmanuel Gaillardon. LSOracle: a logic synthesis framework driven by artificial intelligence: Invited paper. In *ACM/IEEE International Conference on Computer-Aided Design (ICCAD)*, pages 1–6, 2019.

[9] Cunxi Yu. Flowtune: Practical multi-armed bandits in boolean optimization. In *ACM/IEEE International Conference On Computer Aided Design (ICCAD)*, pages 1–9, 2020.

[10] Keren Zhu, Mingjie Liu, Hao Chen, Zheng Zhao, and David Z. Pan. Exploring logic optimizations with reinforcement learning and graph convolutional network. In *ACM/IEEE Workshop on Machine Learning for CAD (MLCAD)*, pages 145–150, 2020.

[11] Robert Brayton and Alan Mishchenko. ABC: An Academic Industrial-Strength Verification Tool. In Tayssir Touili, Byron Cook, and Paul Jackson, editors, *Computer Aided Verification*, pages 24–40, 2010.

[12] Robert K Brayton, Richard Rudell, Alberto Sangiovanni-Vincentelli, and Albert R Wang. Mis: A multiple-level logic optimization system. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 6(6):1062–1081, 1987.

[13] Robert K Brayton, Gary D Hachtel, and Alberto L Sangiovanni-Vincentelli. Multilevel logic synthesis. *Proceedings of the IEEE*, 78(2):264–300, 1990.

[14] Alan Mishchenko, Satrajit Chatterjee, and Robert Brayton. DAG-aware aig rewriting: A fresh look at combinational logic synthesis. In *ACM/IEEE Design Automation Conference*, pages 532–535, 2006.

[15] T Ajayi, D Blaauw, TB Chan, CK Cheng, VA Chhabria, DK Choo, M Coltella, S Dobre, R Dreslinski, M Fogaça, et al. Openroad: Toward a self-driving, open-source digital layout implementation tool chain. *Proc. GOMACTECH*, pages 1105–1110, 2019.

[16] Clifford Wolf. Yosys open SYnthesis suite. (`http://www.clifford.at/yosys/`).

[17] F. Brglez, D. Bryan, and K. Kozminski. Combinational profiles of sequential benchmark circuits. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1929–1934 vol.3, May 1989.

[18] Opencores hardware RTL designs. (`https://opencores.org/`).

[19] Black Parrot SoC. (https://github.com/black-parrot/black-parrot).

[20] MIT Common Evaluation Platform(CEP). (https://github.com/mit-ll/CEP).

[21] xie jian jiang. AES 128/256-bit symmetric block cipher. (https://github.com/crypt-xie/XCryptCore/tree/master/ciphers/aes).

[22] Joachim Strömbergson and Olof Kindgren. AES 128/256-bit symmetric block cipher. (https://github.com/secworks/aes).

[23] Jonathan Balkind, Michael McKeown, Yaosheng Fu, Tri Nguyen, Yanqi Zhou, Alexey Lavrov, Mohammad Shahrad, Adi Fuchs, Samuel Payne, Xiaohua Liang, et al. OpenPiton: An open source manycore research framework. *ACM SIGPLAN Notices*, 51(4):217–232, 2016.

[24] Christoph Albrecht. IWLS 2005 benchmarks. In *IEEE Intl Workshop for Logic Synthesis (IWLS)*, 2005.

[25] Franc Brglez and Hideo Fujiwara. A neutral netlist of 10 combinational benchmark circuits and a target translator. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, 1985.

[26] Franc Brglez, David Bryan, and Krzysztof Kozminski. Combinational profiles of sequential benchmark circuits. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1929–1934, 1989.

[27] Luca Amarú, Pierre-Emmanuel Gaillardon, and Giovanni De Micheli. The EPFL combinational benchmark suite. In *IEEE Intl Workshop on Logic & Synthesis (IWLS)*, 2015.

[28] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[29] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*, 2018.

[30] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. Librispeech: an asr corpus based on public domain audio books. In *IEEE Intl Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5206–5210, 2015.

[31] Kevin E. Murray, Oleg Petelin, Sheng Zhong, Jai Min Wang, Mohamed ElDafrawy, Jean-Philippe Legault, Eugene Sha, Aaron G. Graham, Jean Wu, Matthew J. P. Walker, Hanqing Zeng, Panagiotis Patros, Jason Luu, Kenneth B. Kent, and Vaughn Betz. VTR 8: High Performance CAD and Customizable FPGA Architecture Modelling. *ACM Trans. Reconfigurable Technol. Syst.*, 2020.

A Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] (See Section 5)
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [N/A]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 4.1, 5
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] (see Table 5)
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

B Benchmarks used in prior state-of-art

Prior work	Task Application	Netlists used Source, (# benchmarks)	#nodes in largest benchmark
Yu *et al.* [5]	Classification of synthesis flows	64bit ALU & AES-128 (2)	44045
Nato *et al.* [8]	Optimizer selection for minimization	ISCAS89 [26] & OpenPiton[23] (5)	124565
Haaswijik *et al.* [7]	Optimal synthesis recipe	DSD funcs. & MCNC[25] (5)	$\leq$2000
Hosny *et al.* [6]	Optimal synthesis recipe	arithmetic EPFL[27] (10)	176938
Yu *et al.* [9]	Optimal synthesis recipe	8 DSP funcs. from VTR[31] (8)	30003
Zhu *et al.* [10]	Optimal synthesis recipe	ISCAS85[25] (10)	2675

Table 4: Prior work on machine learning for Logic synthesis

C Computational resource usage

Purpose	Machine configuration	# threads used	# hours used	Computing hours
Data collection	4x AMD EPYC 7551 32-Core Processor, RAM: 504GB, Freq.: 2.0GHz	100	35 days = 840hrs	84,000
	Dual Intel(R) Xeon(R) CPU E5-2650 v3 RAM: 502GB, Freq. 1.8GHz	80	40 days = 960hrs	76,800
	Intel(R) Xeon(R) CPU E5-2640 RAM: 252GB, Freq. 2.5GHz	40	35 days = 840hrs	33,600
	AMD Ryzen Threadripper 2920X 12-Core Processor RAM: 32GB, Freq: 2.32GHz	16	7 days = 168hrs	2688
	8x Intel(R) Core(TM) i7-6700 RAM: 96GB, Freq: 2.10GHz (avg.)	40	55 days = 1320hrs	52,800
Model: Training and inference	Lambda-quad Intel(R) Core(TM) i9-7920X CPU, RAM: 126GB, Freq: 2.5GHz Batchsize: 4 GPU: GTX 1080i, 11GB VRAM	-	QoR prediction: 48 hrs Classification: 18 hrs	-
	Greene (High performance computing) 2x Intel Xeon Platinum 8268 RAM: 369GB, Freq. 2.9GHz Batchsize: 64 GPU: RTX 8000, 48GB VRAM	-	QoR prediction: $\tilde{3}$6 hrs Classification: $\tilde{1}$2 hrs	-

Table 5: Computational resource usage breakdown for data generation and benchmarking models

D Additional Results on GCN Models using OpenABC-D (QoR prediction)

In addition to predicting the QoR given synthesis using an unseen recipe (Variant 1, §4.1), we also considered two further task variants.

Variant 2: Predicting QoR of synthesizing unseen IPs We train a model on the QoR from synthesizing a set of smaller IPs and evaluate the model on its ability to predict the QoR of synthesizing the unseen larger IPs, given the IP and a synthesis recipe. This mimics real-world problems where synthesis runs on large IPs take weeks to complete while synthesis runs on small IPs are possible within a short time. The motivation is to understand whether models trained with data generated from small-size IPs can be used to meaningfully predict on large-size IPs. We consider 16 smaller IPs during training and 8 larger IPs during inference.

QoR Task	Test MSE on baseline networks		
	Net1	Net2	Net3
Variant1	0.648 ± 0.05	0.815 ± 0.02	$\mathbf{0.579 \pm 0.02}$
Variant2	10.59 ± 2.78	1.236 ± 0.15	$\mathbf{1.47 \pm 0.14}$
Variant3	0.588 ± 0.04	0.538 ± 0.01	$\mathbf{0.536 \pm 0.03}$

Table 6: Benchmarking GCN models for QoR prediction tasks

Results: Using MSE as metric for comparison on test dataset, we presented the results in Table 6. In train-test split strategy where IPs are unknown, the performance varied across test IPs for different networks (see Fig. 10,11 and 12). For IPs like `aes_xcrypt` and `wb_conmax`, the inference results are consistently poor indicating AIG embeddings are different from training data AIG embedding. Inference results on like `bp_be`, `tinyRocket`, and `picosoc` are close to QoR values across networks indicating the network has learnt from graph structures of training data IPs.

Variant 3: Predicting QoR on unseen IP-Synthesis Recipe Combination We train a model on the QoR achieved from a random pick of 70% of the synthesis recipes across all IPs and test the model's ability to predict the QoR given an unseen IP-recipe pair. In this use case, experts develop synthesis recipes for specific IPs and the user is interested in assessing the performance of these recipes on other IPs without running the synthesis recipes on them.

Results: Table 6 shows the results on three baseline networks. The trained model has observed IPs and synthesis recipes as standalone entities as part of training data. Inference of networks are expected to be better than previous two data-split strategies and results in Table 6 confirm this. Performance of network 3 is better indicating a large kernel size helps network learn more information about effectiveness of synthesis recipes. This is similar to results obtained in [5] and therefore sets a baseline for GCN-based networks on such tasks.

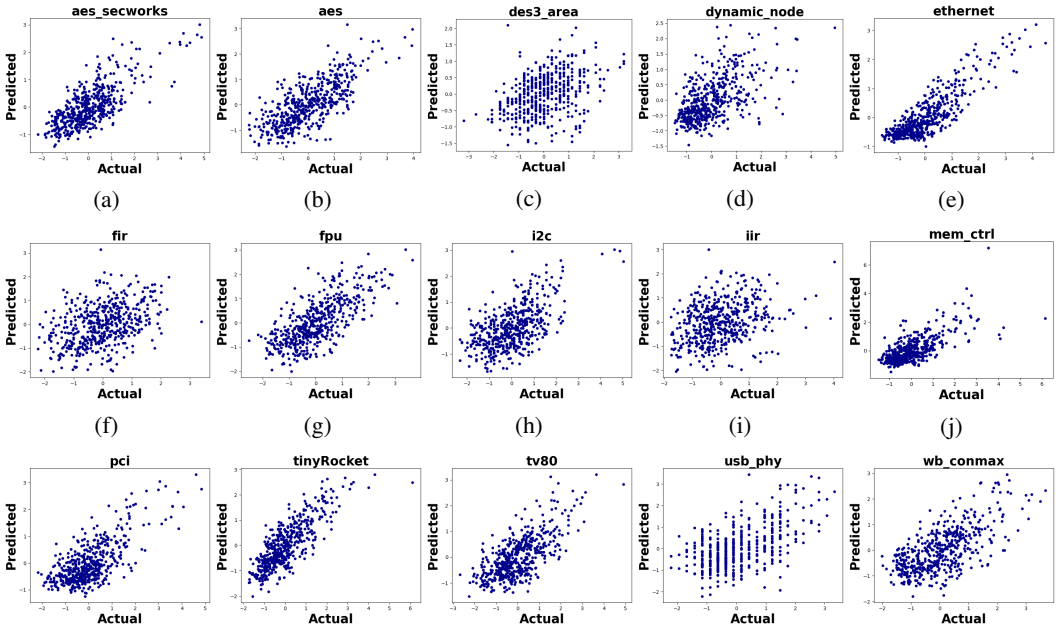


Figure 7: Net 1 for QoR Task Variant 1 (Unseen Recipe)

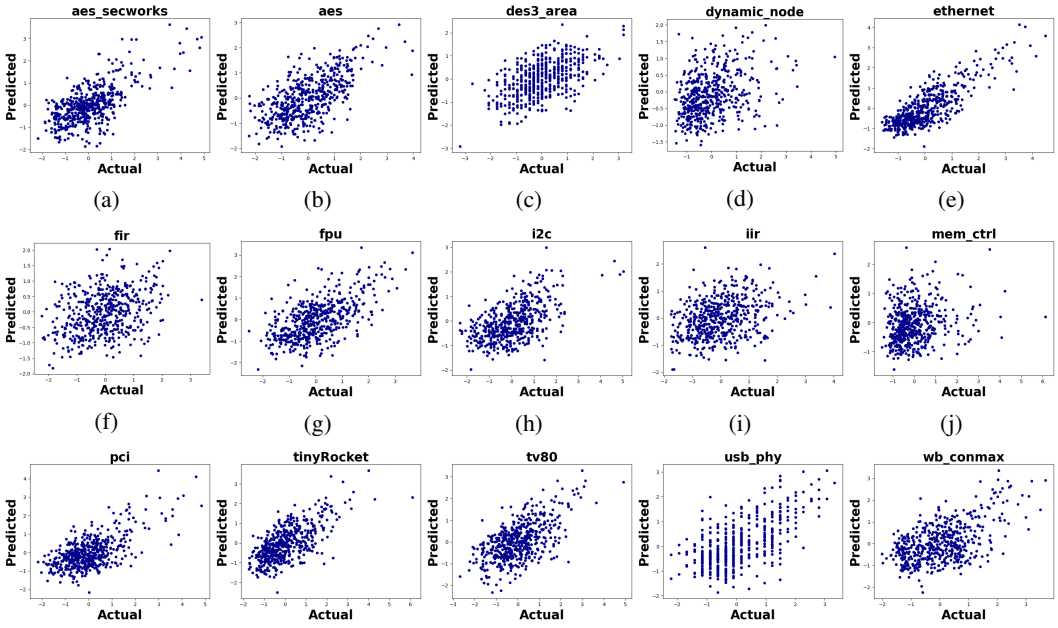


Figure 8: Net 2 for QoR Task Variant 1 (Unseen Recipe)

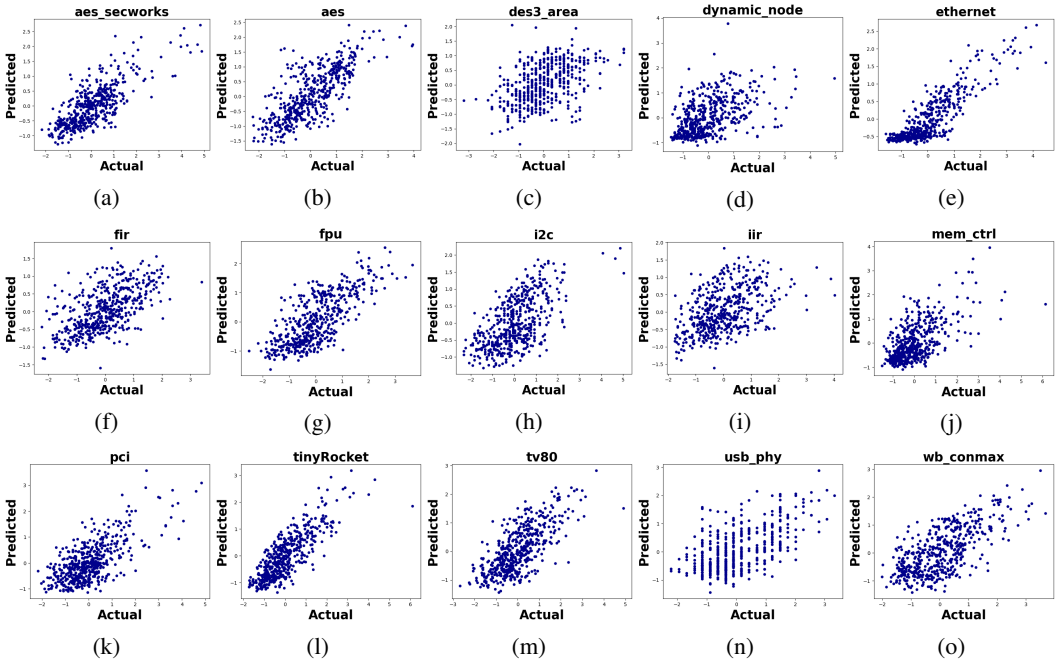
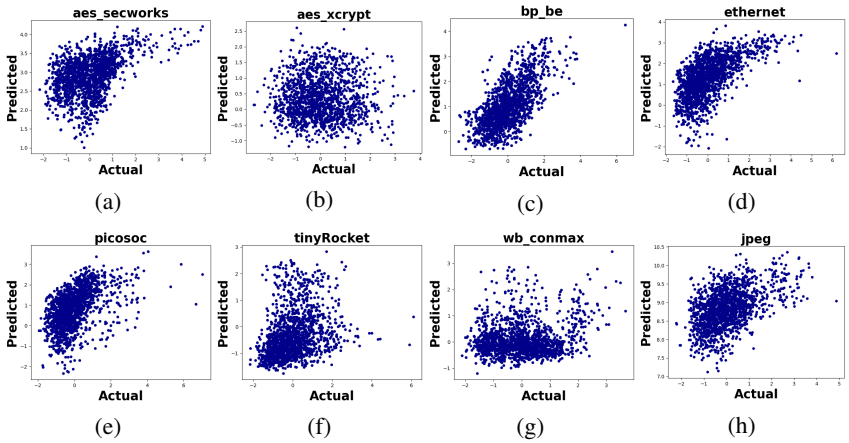


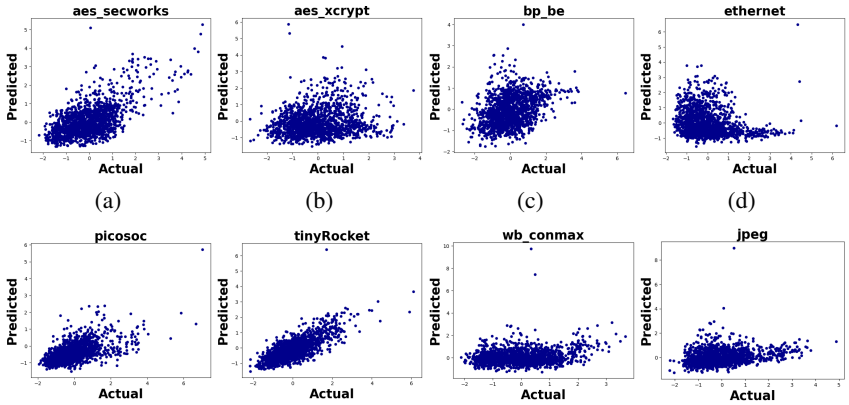
Figure 9: Net 3 for QoR Task Variant 1 (Unseen Recipe)



(a) (b) (c) (d)

(e) (f) (g) (h)

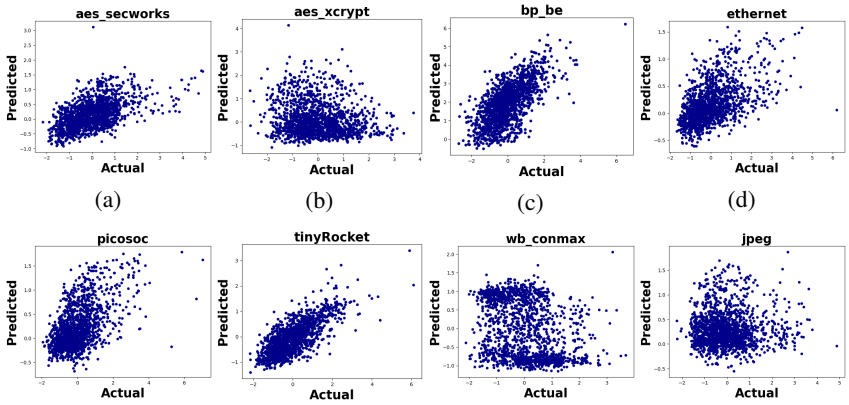
Figure 10: Net 1 for QoR Task Variant 2 (Unseen IP)



(a) (b) (c) (d)

(e) (f) (g) (h)

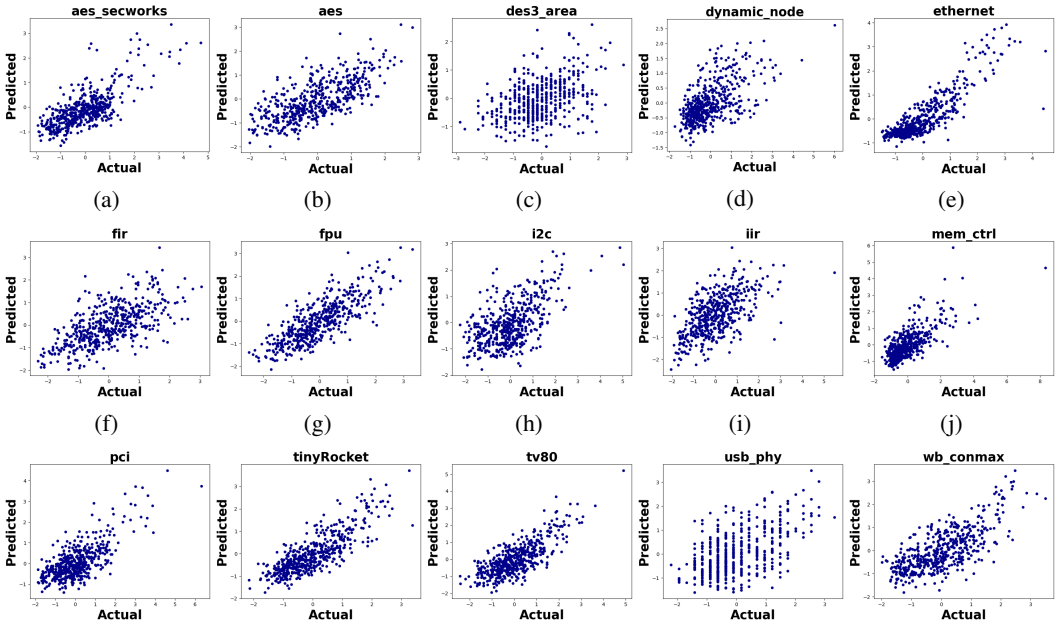
Figure 11: Net 2 for QoR Task Variant 2 (Unseen IP)



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 12: Net 2 for QoR Task Variant 2 (Unseen IP)

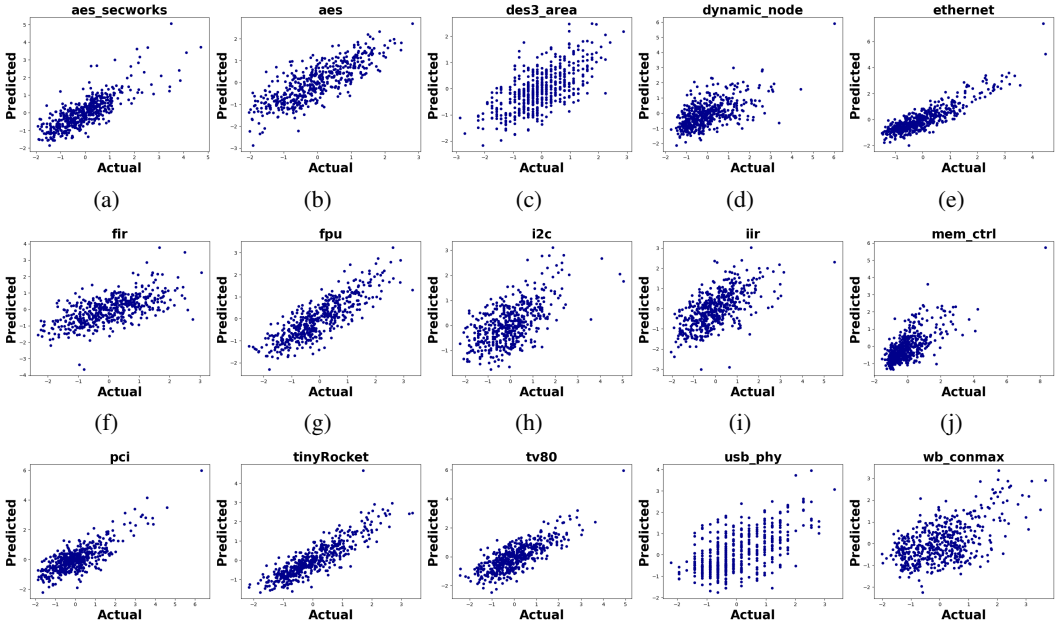


(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

Figure 13: Net 1 for QoR Task Variant 3 (Unseen IP-Recipe combination)

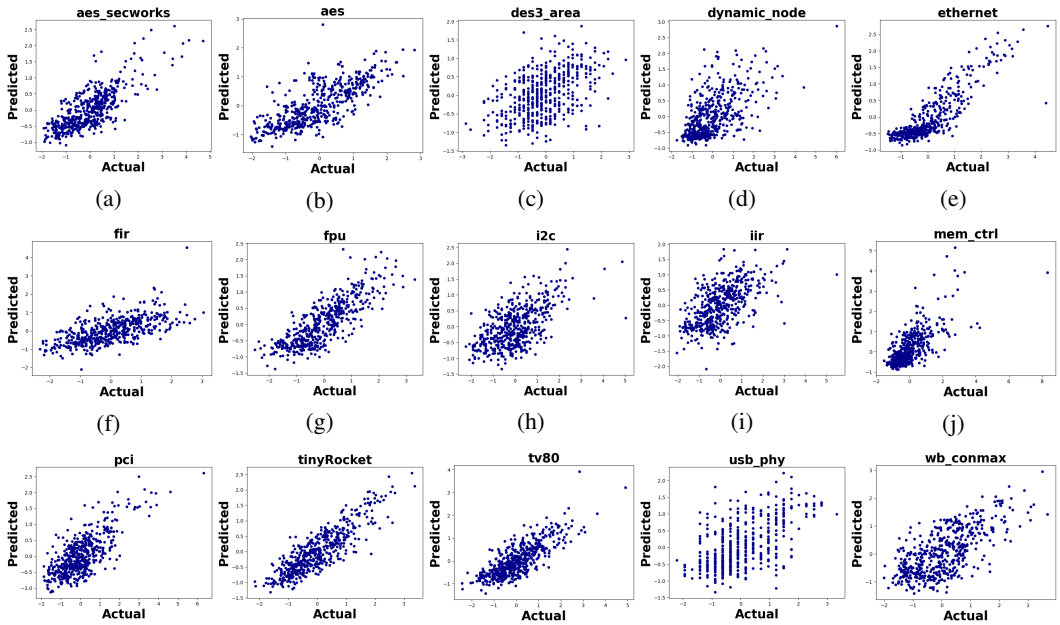


(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

Figure 14: Net 2 for QoR Task Variant 3 (Unseen IP-Recipe combination)



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

Figure 15: Net 3 for QoR Task Variant 3 (Unseen IP-Recipe combination)