Standardizing Data Formats for Open-Source Model Sharing and Benchmarking

Dr. Ioannis Savidis

Department of Electrical and Computer Engineering

Drexel University

DAC BoF 2025

Drexel UNIVERSITY

Design & Analysis Lab

What's Needed for ML in EDA to Thrive

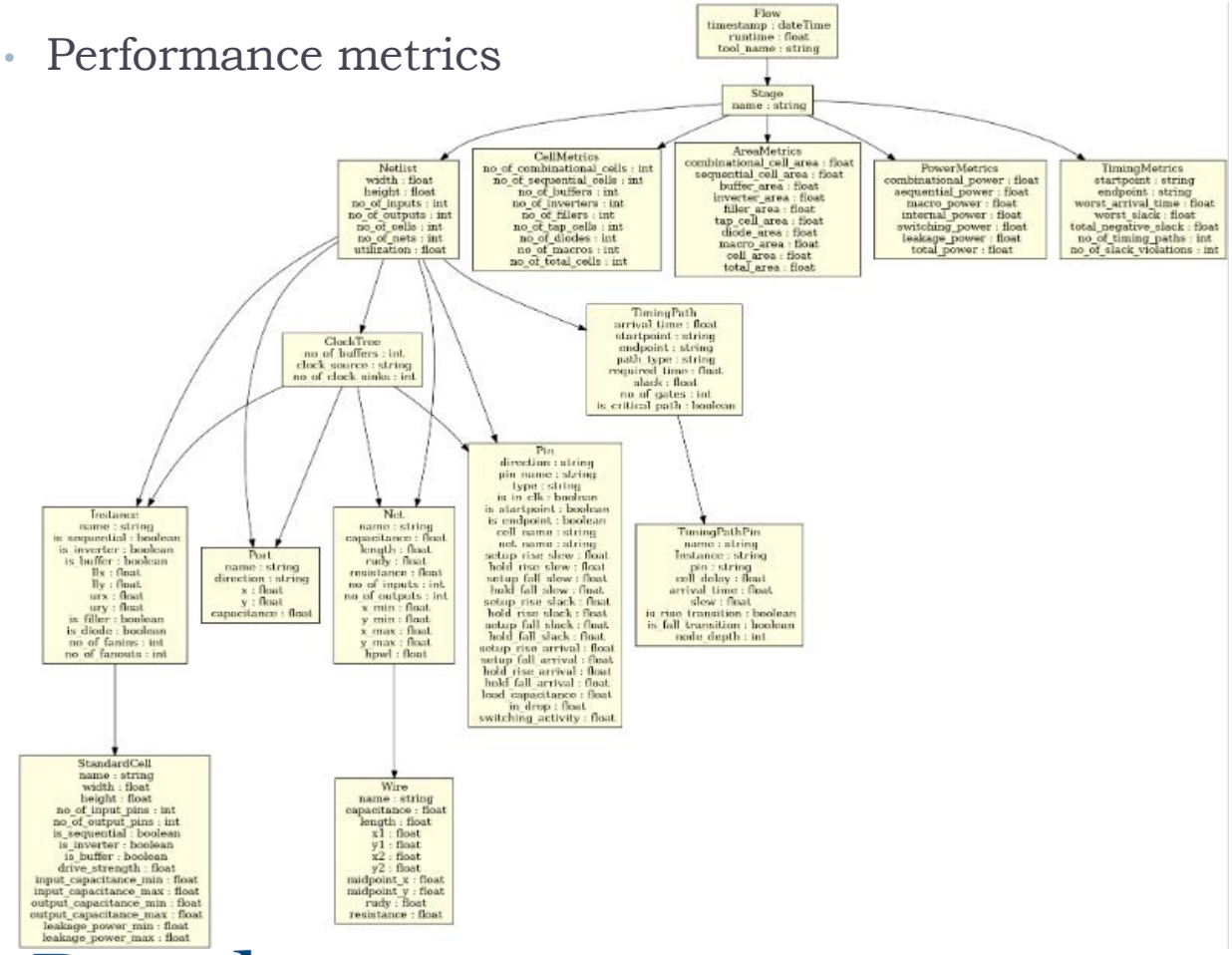
- Advancing ML in physical design needs standardized, open datasets and consistent practices.

- Key Requirements

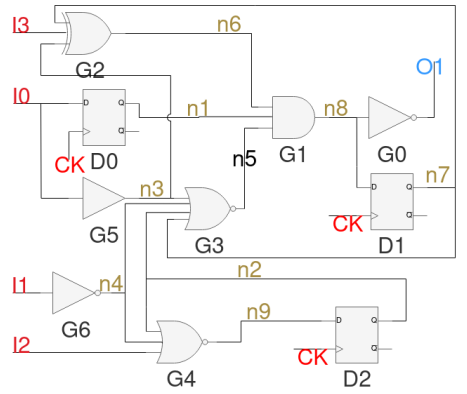
Reproducible Dataset	Consistent Data Handling	Seamless Data Sharing
Ensure datasets are publicly available, well-documented, and easy to reuse across studies.	Adopt uniform preprocessing and labeling conventions to allow fair benchmarking and result comparison.	Use common formats and metadata to simplify sharing, integration, and scaling across tools and teams.

EDA Schema

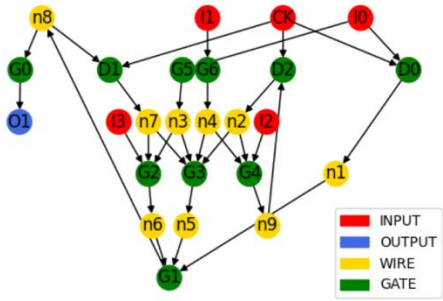
- Property graph data-model schema incorporating
  - Structural data of the circuit
  - Performance metrics



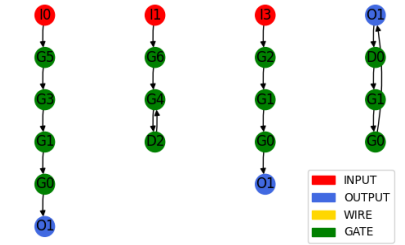
Entity Relationship Diagram



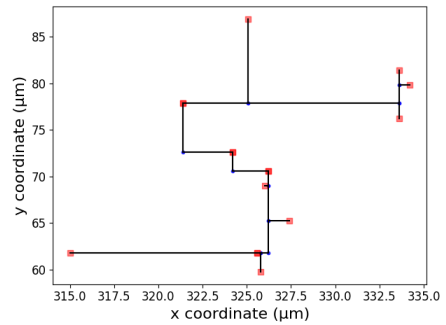
Circuit

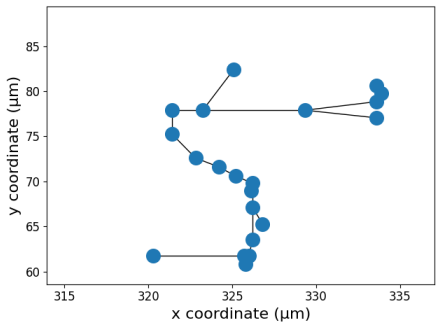


Netlist Graph



Timing Path Graph





Net Graph

EDA-schema Open Dataset

- Open Dataset v1.1:
  - Toolset: OpenROAD
  - PDKs: sky130hd
  - Circuits: IWLS05 benchmark circuits
  - Parameter setup/Constraints
    - Clock latency: 0.01
    - Clock IO delay: 0.01
    - Utilization: 30%
    - Aspect ratio: 1.0
  - Dataset size: 23 GB
  - Time to load all netlist entities: 28 mins 23 secs
  - link

	Barely Fail (id=000001)		Barely Pass (id=000002)	
Circuit	Target Clock Period (ns)	Worst Slack (ns)	Target Clock Period (ns)	Worst Slack (ns)
ac97_ctrl	3	-0.1823	3.25	0.1333
aes_core	3.5	-0.1999	3.75	0.0458
des3_area	3.75	-0.0119	4	0.1823
i2c	2.25	-0.1594	2.5	0.0658
mem_ctrl	5	-0.0863	5.25	0.1420
pci	3.75	-0.1326	4	0.1548
sasc	1.75	-0.0992	2	0.1752
simple_spi	2	-0.1452	2.25	0.1232
spi	4.25	-0.1739	4.5	0.2488
ss_pcm	1.625	-0.0144	1.75	0.1107
systemcaes	4.75	-0.1644	5	0.0050
systemcdes	4	-0.0069	4.25	0.3052
tv80	7	-0.0294	7.25	0.1227
usb_funct	2.5	-0.0917	2.75	0.1018
usb_phy	1.75	-0.0449	2	0.1821
wb_dma	3	-0.0746	3.25	0.0708

EDA-schema Expansion Update (Dataset v2.0)

- Extending dataset to multiple PDKs: Skywater 130nm, ASAP 7nm, NanGate45nm

- Parameter Selection
  - **Set 1:**
    Uses fixed, default values from OpenROAD
    (e.g., clock latency = 0.01 ns, transition = 0.25 ns)

TABLE IV: Constraints and parameters utilized to generate the dataset. (OpenROAD default)

Parameters	**Values or Ranges**		
Target clock periods (ns)	{BF - 1ns, BF, BP, BP + 1ns}		
Aspect ratio	{0.5, 1, 1.5}		
Max utilization	{0.3, 0.4, 0.5}		
Input/Output Delay	20% of Target clock period		
Core Margin (μm)	1		
Place Density	skywater 130nm	asap 7nm	nangate 45nm
	0.6	0.6	0.3
Clock Latency (ns)	0.01		
Clock Transition (ns)	0.25		
Clock Uncertainty (ns)	skywater 130nm	asap 7nm	nangate 45nm
	0.2	0.2	0.1

  - **Set 2:**
    Scales certain parameters like clock latency and uncertainty relative to the target clock period (TCP)

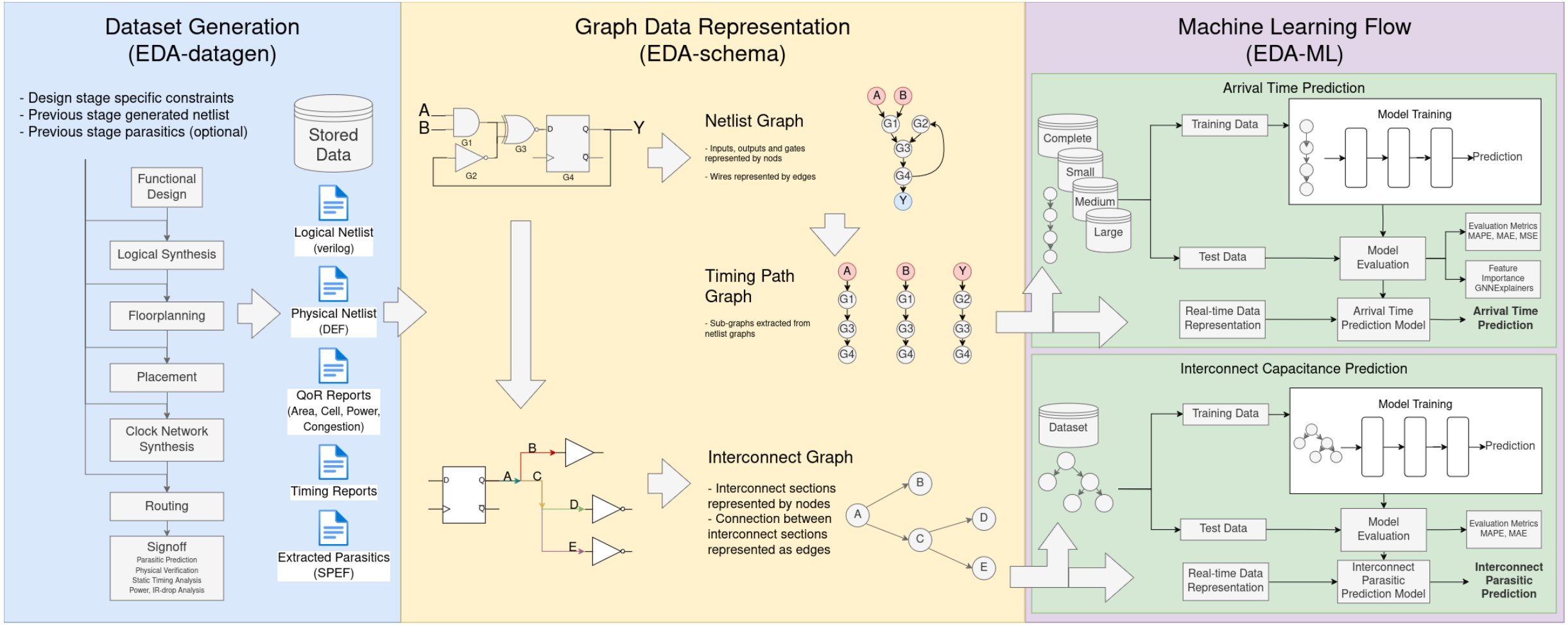
TABLE V: Constraints and parameters utilized to generate the dataset.

Parameters	**Values or Ranges**		
Target clock periods (ns)	$\{0.8 \times BF, BF, BP, 1.2 \times BP\}$		
Aspect ratio	{0.5, 1, 1.5}		
Max utilization	{0.3, 0.4, 0.5}		
Input/Output Delay	20% of Target clock period (TCP)		
Core Margin (μm)	1		
Place Density	skywater 130nm	asap 7nm	nangate 45nm
	0.6	0.6	0.3
Clock Latency (ns)	1% of TCP (capped at 25 ps)		
Clock Transition (ns)	12.5% of TCP, bounded [30 ps, 200 ps]		
Clock Uncertainty (ns)	5% of TCP		

EDA-learn

- Current research based on this flow are

  a) arrival time prediction b) interconnect capacitance prediction



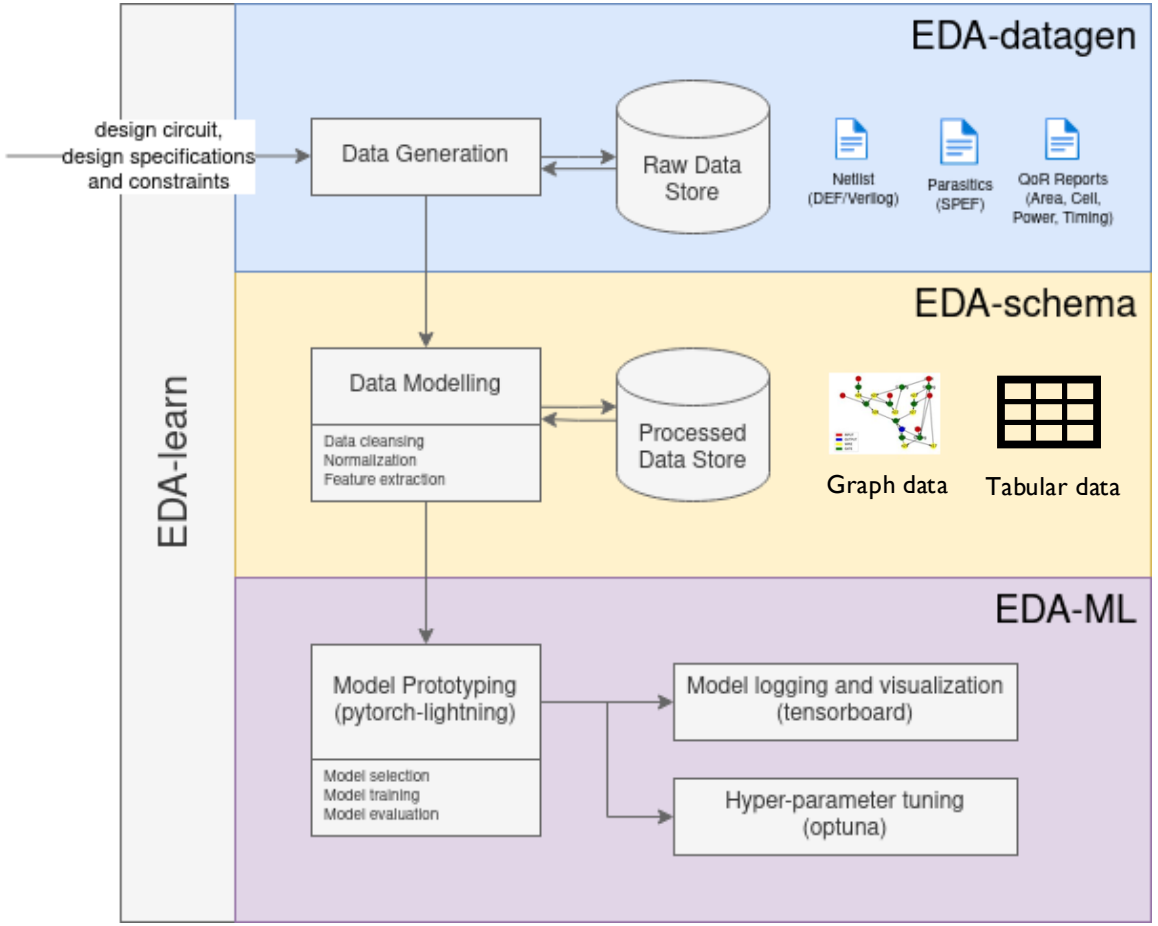
P. Shrestha, S. Phatharodom, and I. Savidis, "Graph representation learning for gate arrival time prediction," *Proceedings of the ACM/IEEE Workshop on Machine Learning for CAD (MLCAD)*, pp. 127–133, Sept. 2022.
P. Shrestha and I. Savidis, "Graph representation learning for parasitic impedance prediction of the interconnect," *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, May 2023

EDA-learn

- Framework developed within Drexel-ICE to pursue ML problems in EDA

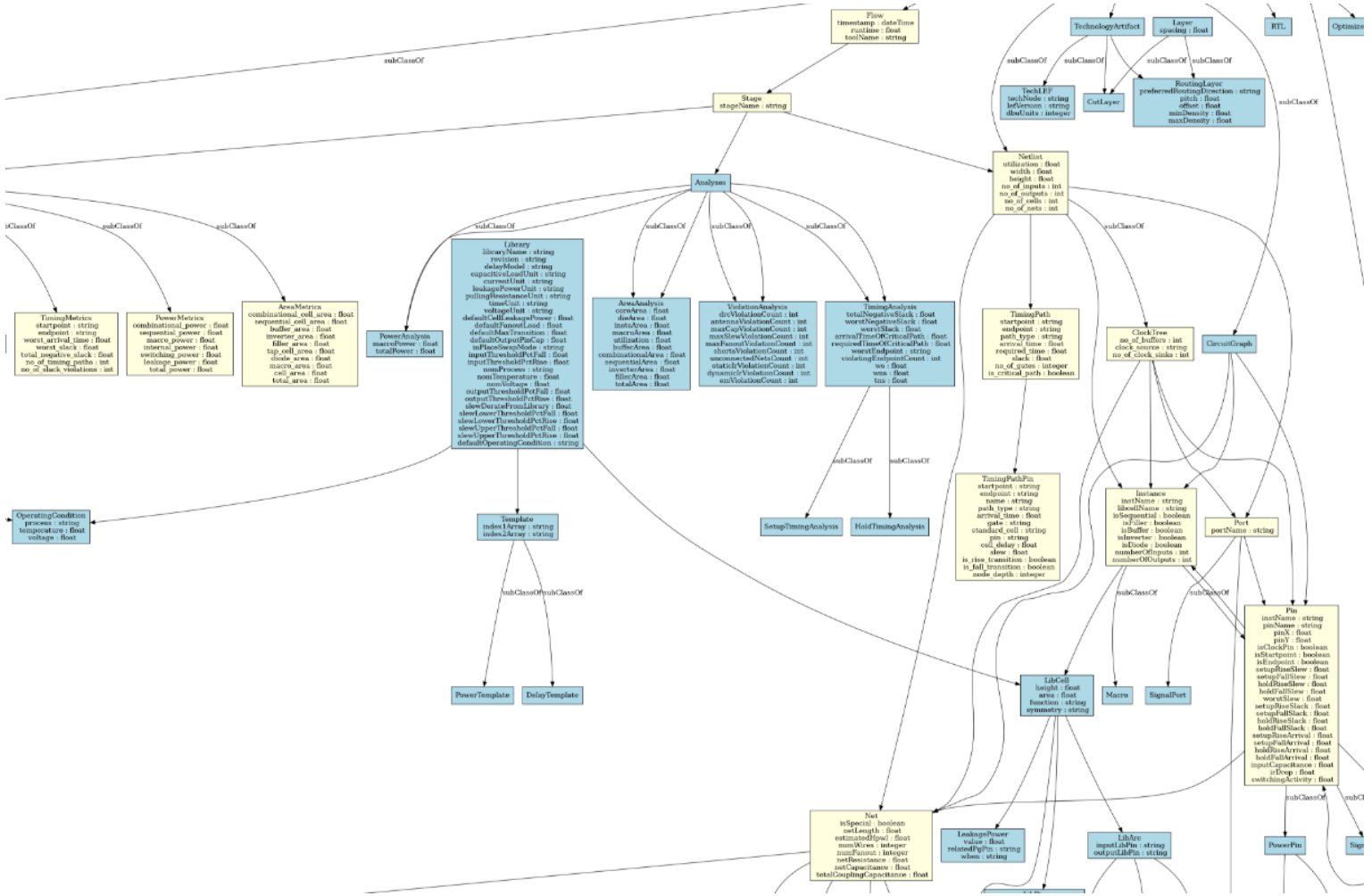
Main Components

- EDA-datagen
  - Physical design flow automation for parameterized large scale dataset generation
- **EDA-schema**
  - **Property graph data-model schema for circuit data representation**
- EDA-ML
  - Rapid prototyping and evaluation for EDA based machine learning models



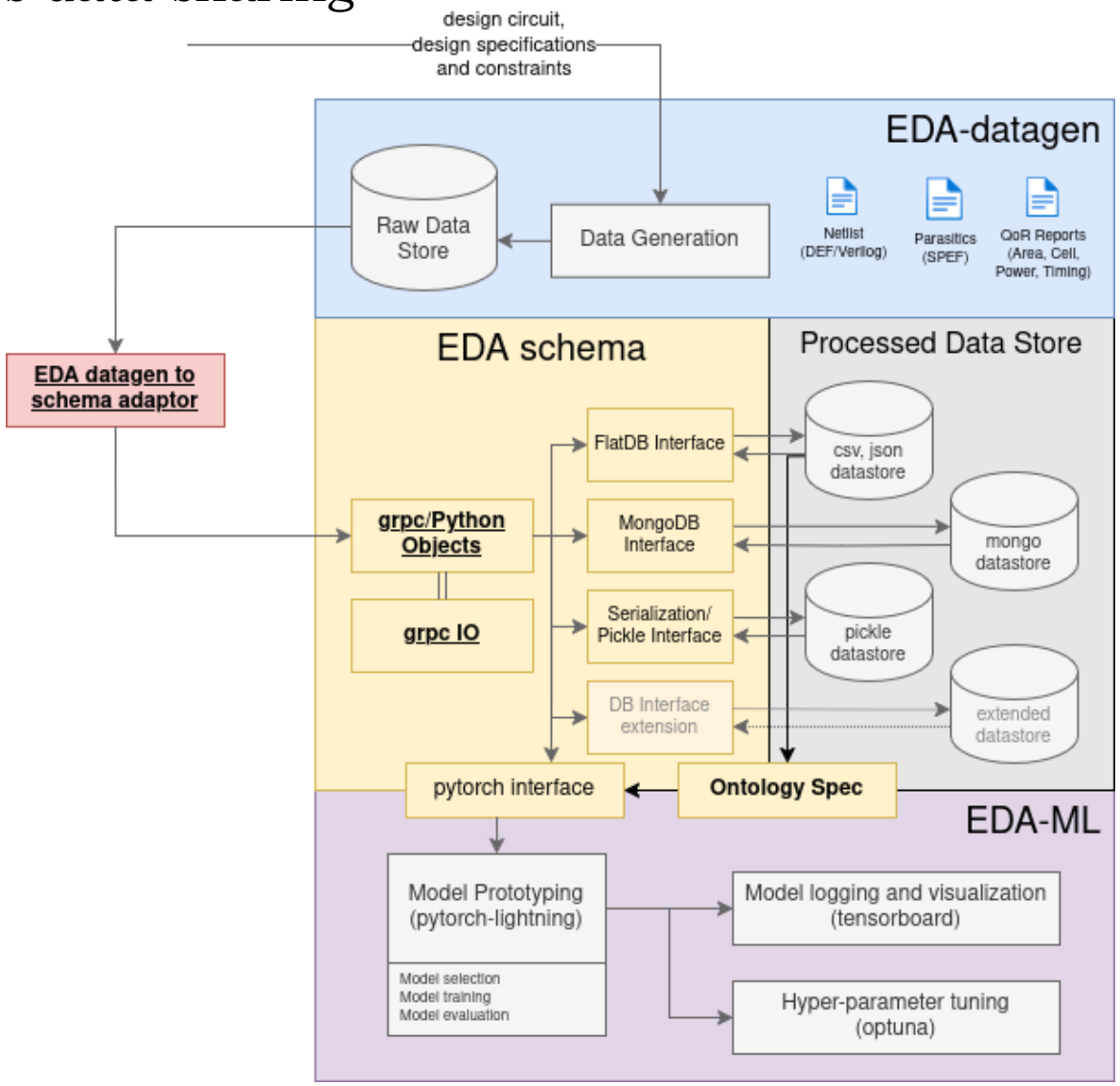
Si2 Schema Working Group

- Develop a standard EDA ontology for seamless data sharing
- Resource Description Framework (RDF) of physical design data
- **Participants**



Si2 Schema Working Group

- Develop a standard EDA ontology for seamless data sharing
- Resource Description Framework (RDF) of physical design data



- **Participants**

Drexel University · ASU Arizona State University · NC State University · Siemens · NXP Semiconductors · Silvaco · IBM · Ansys · Synopsys · Si2

Thank You!

Check out our open dataset at
github.com/drexel-ice/EDA-schema