

# EDA-Schema

An Open Graph Datamodel Schema and Dataset for Design Automation

**Dr. Ioannis Savidis**

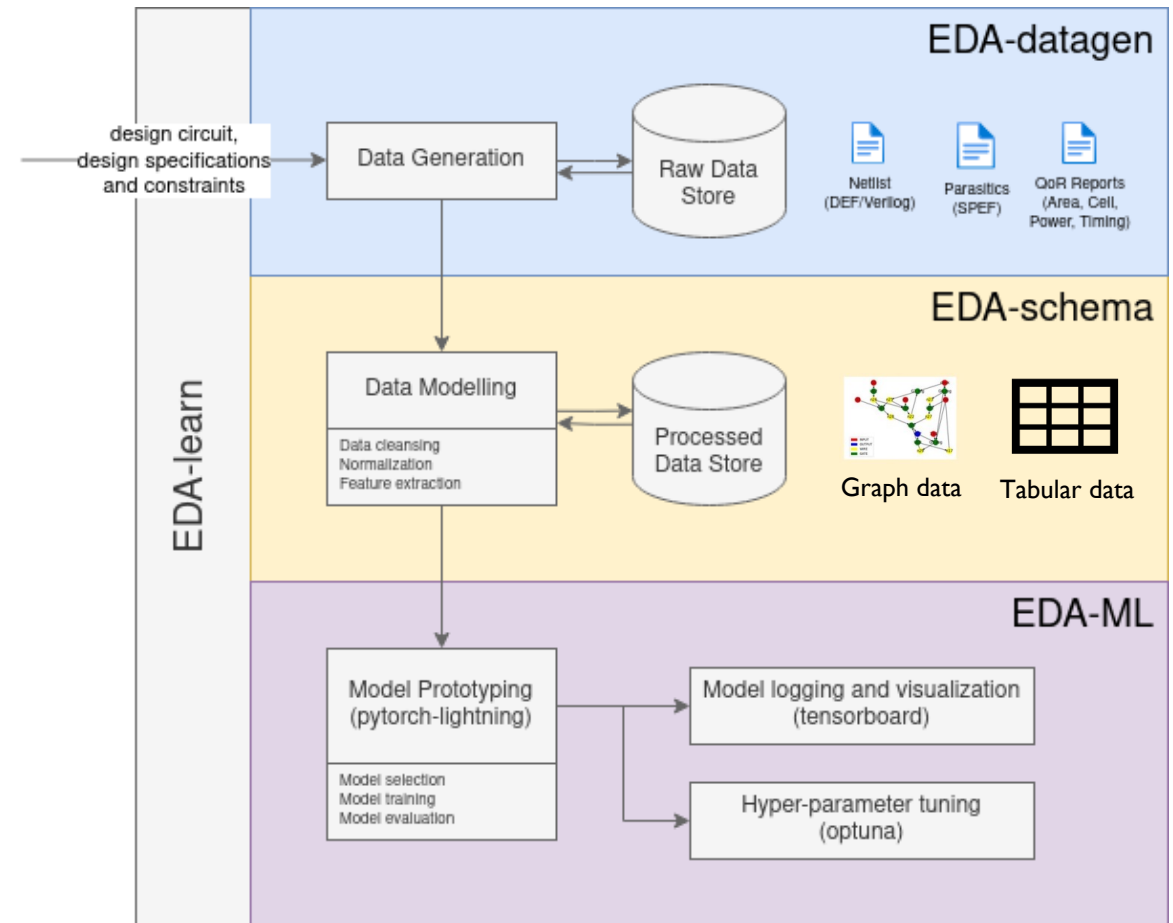
Department of Electrical and Computer Engineering  
Drexel University

DAC Birds-of-a-Feather 2024

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

## Main Components

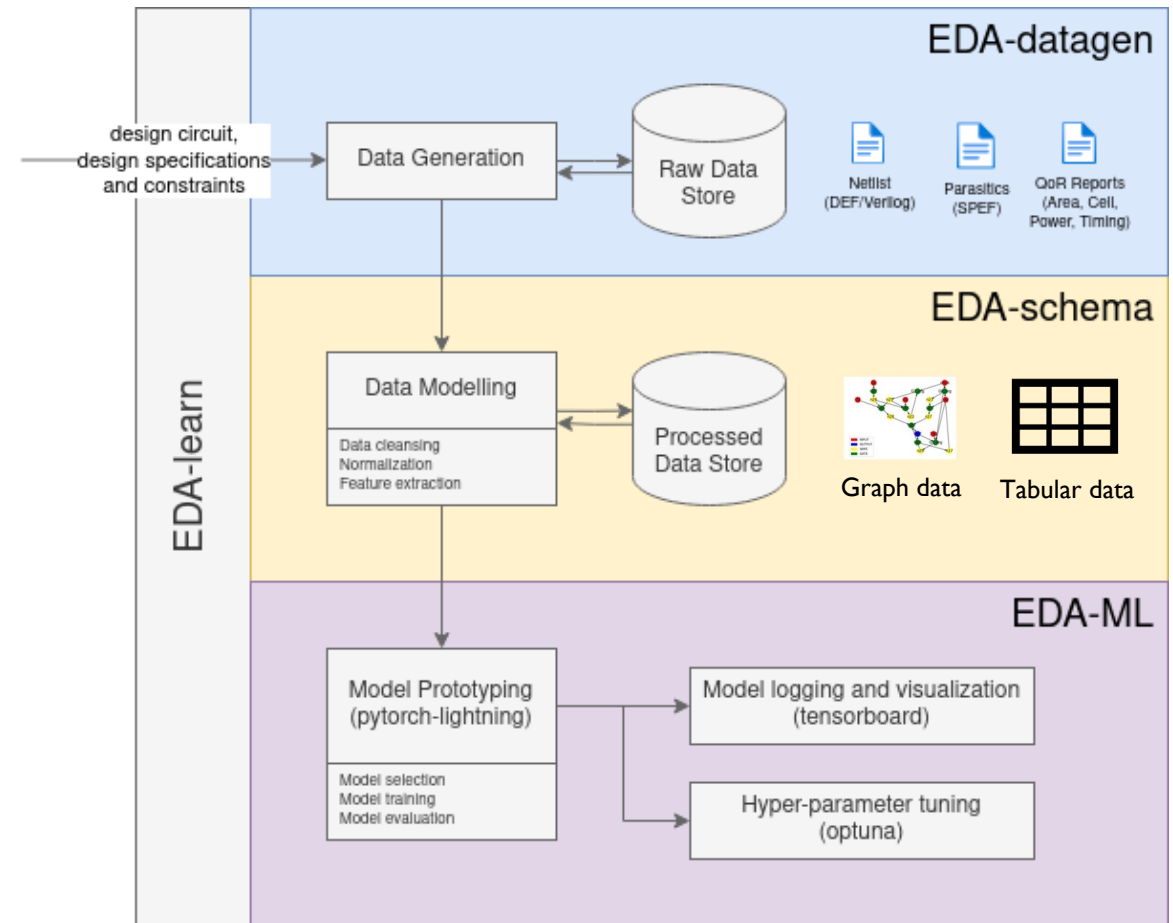
- EDA-datagen
  - Physical design flow automation for parameterized large scale dataset generation
- EDA-schema
  - Property graph data-model schema for circuit data representation
- EDA-ML
  - Rapid prototyping and evaluation for EDA based machine learning models



- 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



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

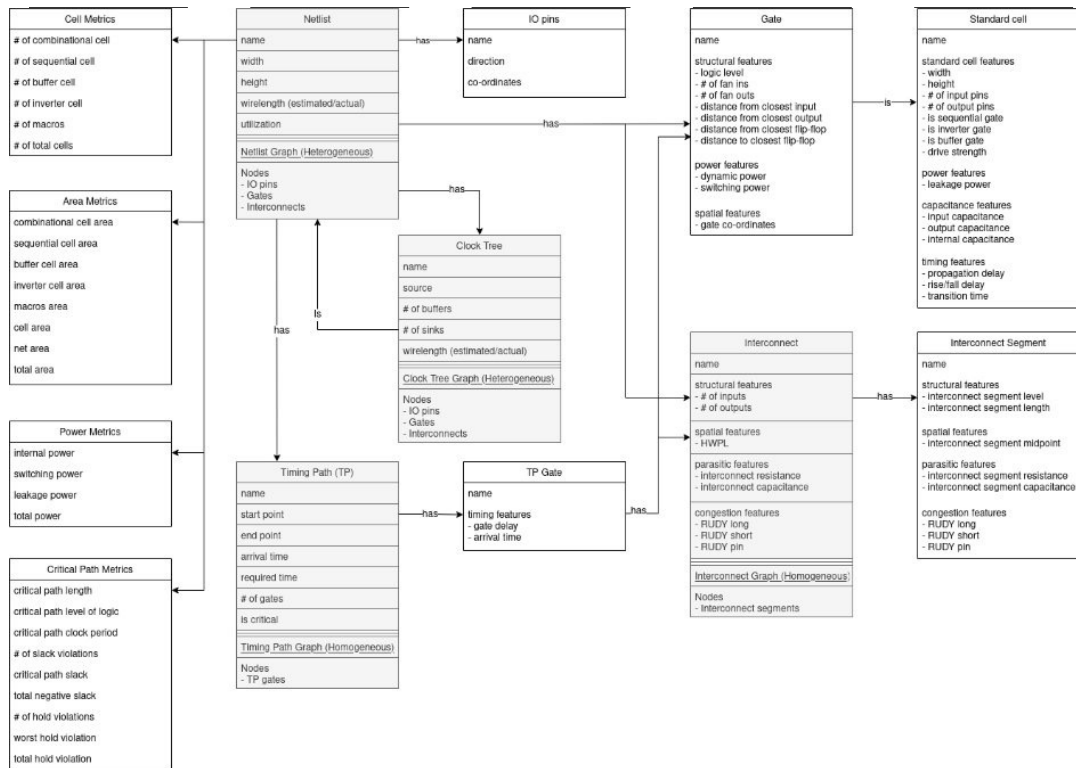


Fig. Entity Relationship Diagram

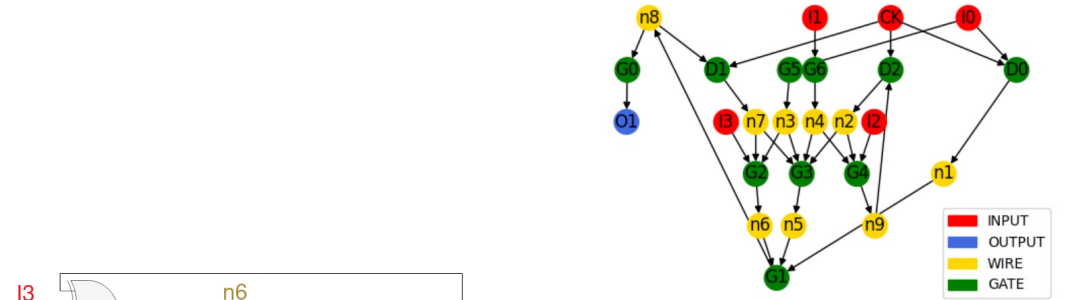


Fig. Netlist Graph

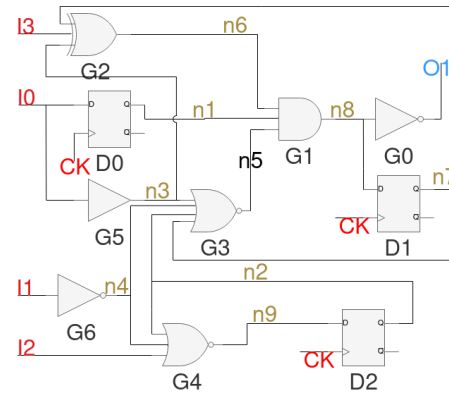


Fig. Circuit

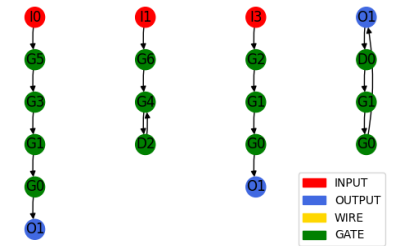
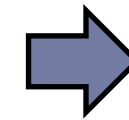


Fig. Timing Path Graph

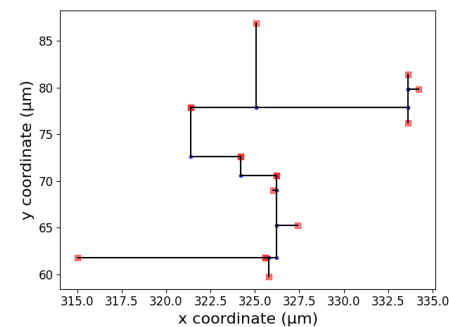


Fig. Interconnect Graph

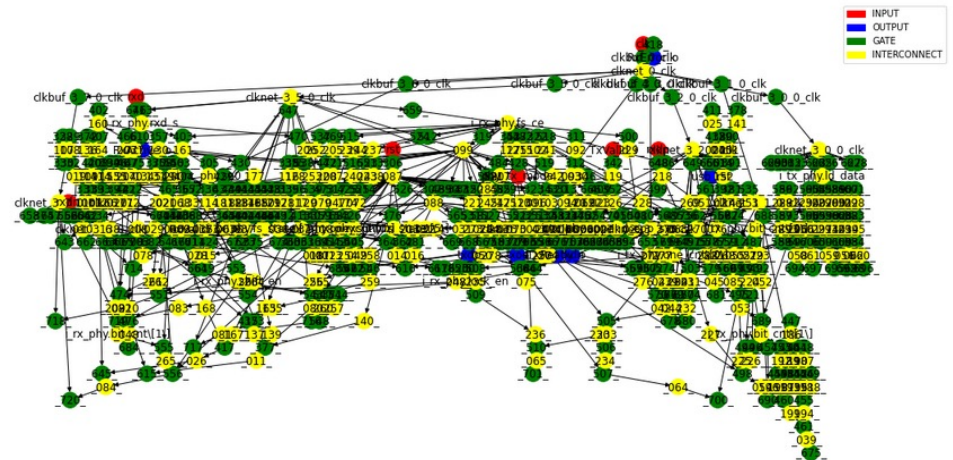
- Open Dataset
  - Designs: 20 IWLS'05 benchmark circuits
  - PDK: Skywater 130nm
  - Design Toolset: OpenROAD
- Dataset details
  - Overall circuits in dataset:  $49 * 20 = 960$
  - Number of gates in dataset: 7,468,228
  - Number of nets in dataset: 7,726,920
  - Number of timing paths in dataset: 1,561,975
  - Total dataset size: 82.836G

- The dataset is available publicly
- EDA-schema python library allows easy access to the dataset
  - <https://github.com/drexel-ice/EDA-schema>

### Loading a Netlist

A netlist, identified by attributes such as the circuit name, netlist ID, and design phase, is loaded from the database into an EDA-schema netlist graph object to extract specific details about the digital circuit.

```
netlist = dataset.load_netlist({'circuit': 'usb_phy', 'netlist_id': 'id-000020', 'phase': 'route'})
```



Dataset constraints and parameters

Parameters	Values or Ranges	# of Samples
Clock periods (ns)	{0.5, 1, 2, 5}	4
Aspect ratio	{0.5, 0.75, 1}	3
Max utilization	{0.3, 0.5}	2
Max skew (ns)	[0.01 - 0.2]	2
Max fanout	[50 - 250]	2
Max clock network capacitance (pF)	[0.05 - 0.3]	2
Max latency (ns)	[0 - 1]	2
<b>Total circuits per design</b>		<b>48</b>

# Challenges in ML for EDA: Lack of standardized dataset

- Absence of standardized, open datasets in physical design space is a significant barrier to the advancement and validation of machine learning models

- Key Issues

<b>Non-Reproducible Datasets</b>
Researchers rely on custom datasets that are often not reproducible by others, leading to isolated findings and limited scientific validation

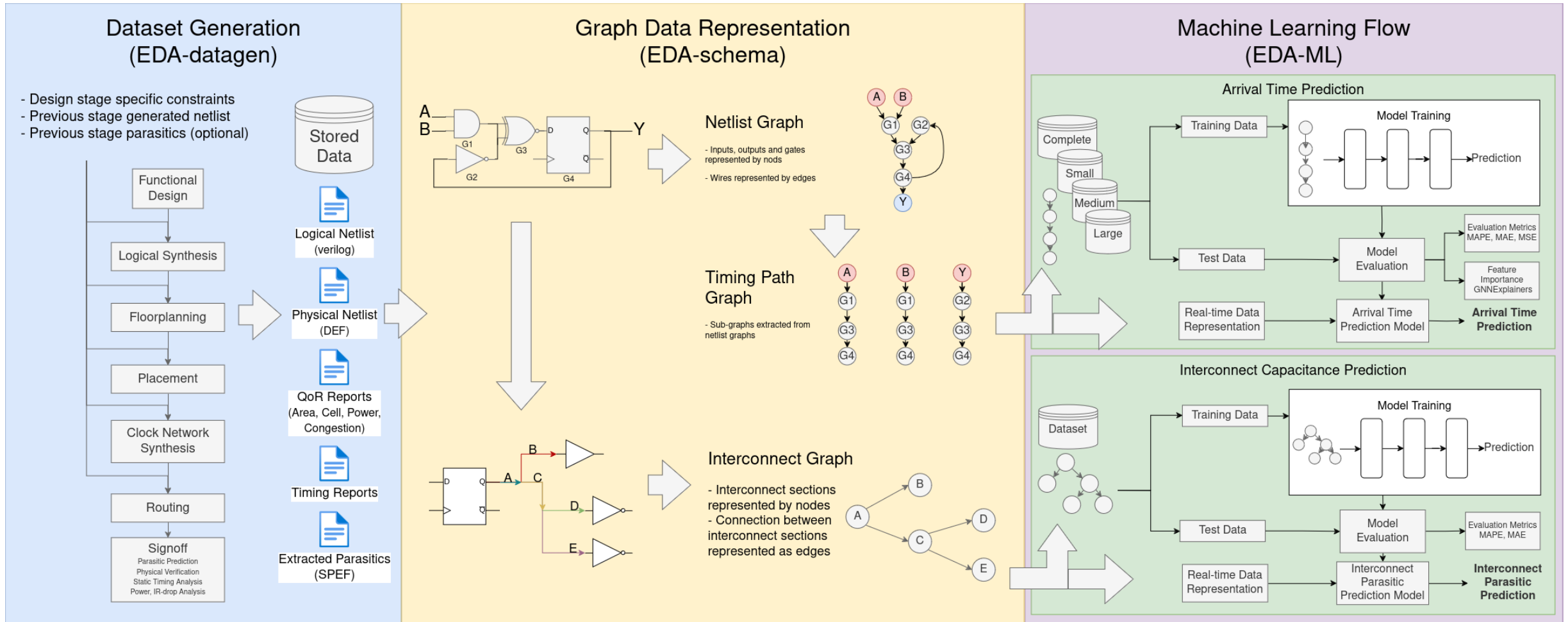
<b>Inconsistent Data Handling</b>
Variability in data preprocessing methods across different studies undermines the ability to compare and validate outcomes effectively

<b>Challenges in Data Sharing</b>
Creating common frameworks for model architecture and training streamlines development efforts and enhance reproducibility across studies

- Current research based on this flow are

a) arrival time prediction

b) interconnect capacitance prediction



# Thank You!

Check out our open dataset at  
[github.com/drexel-ice/EDA-schema](https://github.com/drexel-ice/EDA-schema)