Design Methodology for On-Chip Power Grid Interconnect: AI/ML Perspective

Sukanta Dey

Ph.D. Defense Seminar, 22nd March 2021

Supervisors

Prof. Sukumar Nandi

Dr. Gaurav Trivedi

Department of Computer Science and Engineering

Indian Institute of Technology Guwahati

Guwahati – 781039, Assam, India

Overview

- Motivation
- Introduction and Objective
- Contributions:
  - Power Grid Analysis using Probabilistic Approach
  - Design Space Exploration of PG Interconnect
  - ML Approach for PG Design
  - ML Approach for Aging Prediction
- Conclusion and Future works

Motivation

- Apple's new iPhone/iPad bionic processor in Sept 2019*
  - 8.5 billion transistors (A13) /10 billion transistors (A12x)
  - Die area: 98.48 mm^2 (A13) /122 mm^2 (A12x)
  - 0.7 V supply voltage (2^{nd} gen TSMC 7nm Tech)
- 20 billion or more transistors in future mobile SoC.
- **Challenge is distributing 0.7 V to 20 billion or more transistors**
- Better design methodology required.
- EDA/VLSI Design comes to rescue.
- To reduce design cycle time and manpower.

Courtesy: *https://en.wikichip.org/wiki/apple/ax/a13

VLSI Design objective

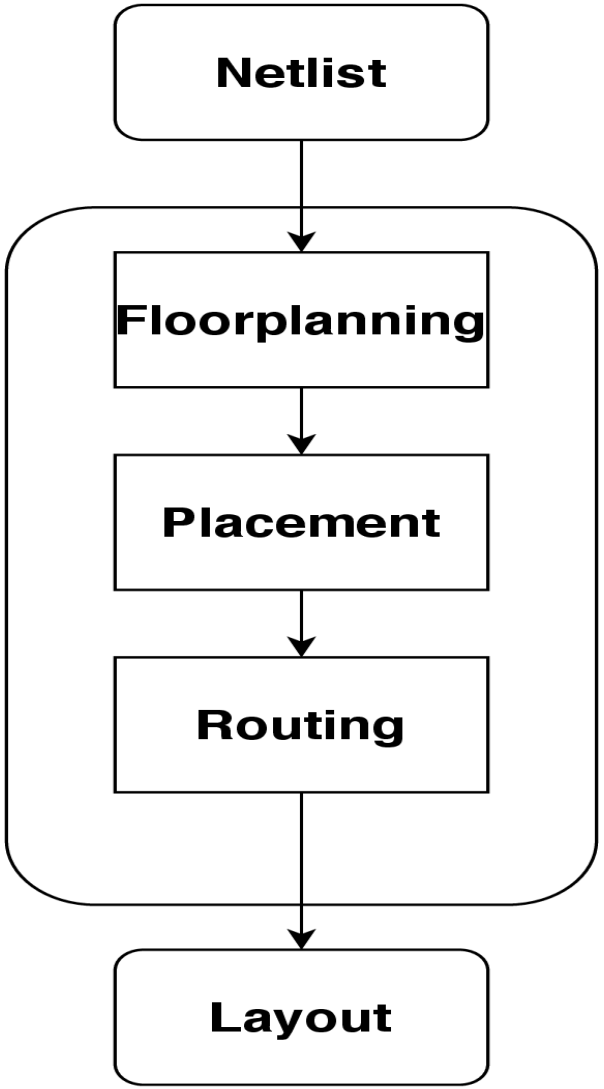
- **Objective:** To create the layout from the design specification.
- **Requirement**: To **decrease the human effort** in the design process and to **automate** and make the **design cycle faster**.

What is VLSI Physical Design?

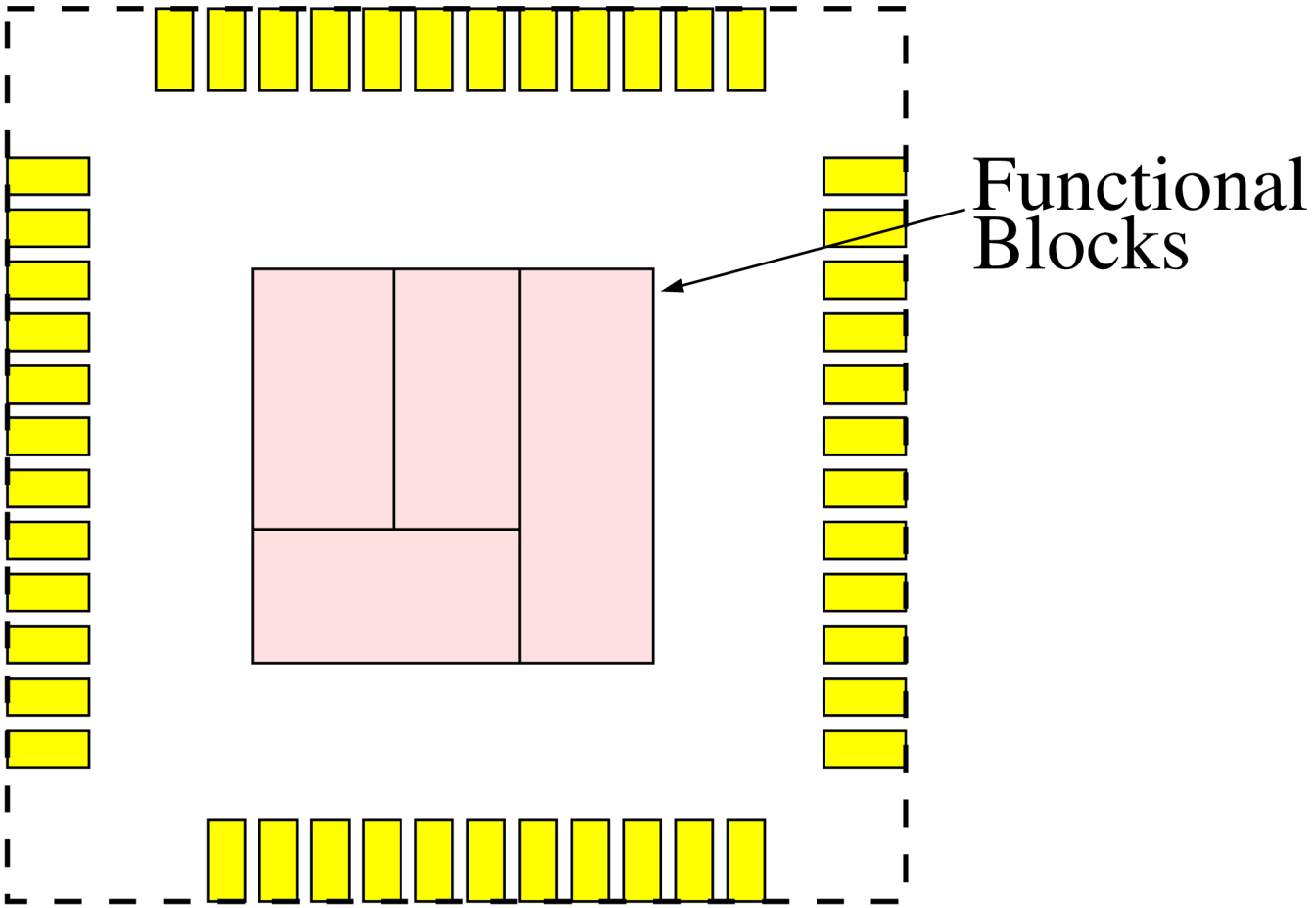
- The process of **converting** the **specification of an electrical circuit** called netlist into a **geometric representation** called layout.
- The layout must take optimum silicon area.
- Simultaneously, other issues must be minimized.

VLSI Physical Design Flow

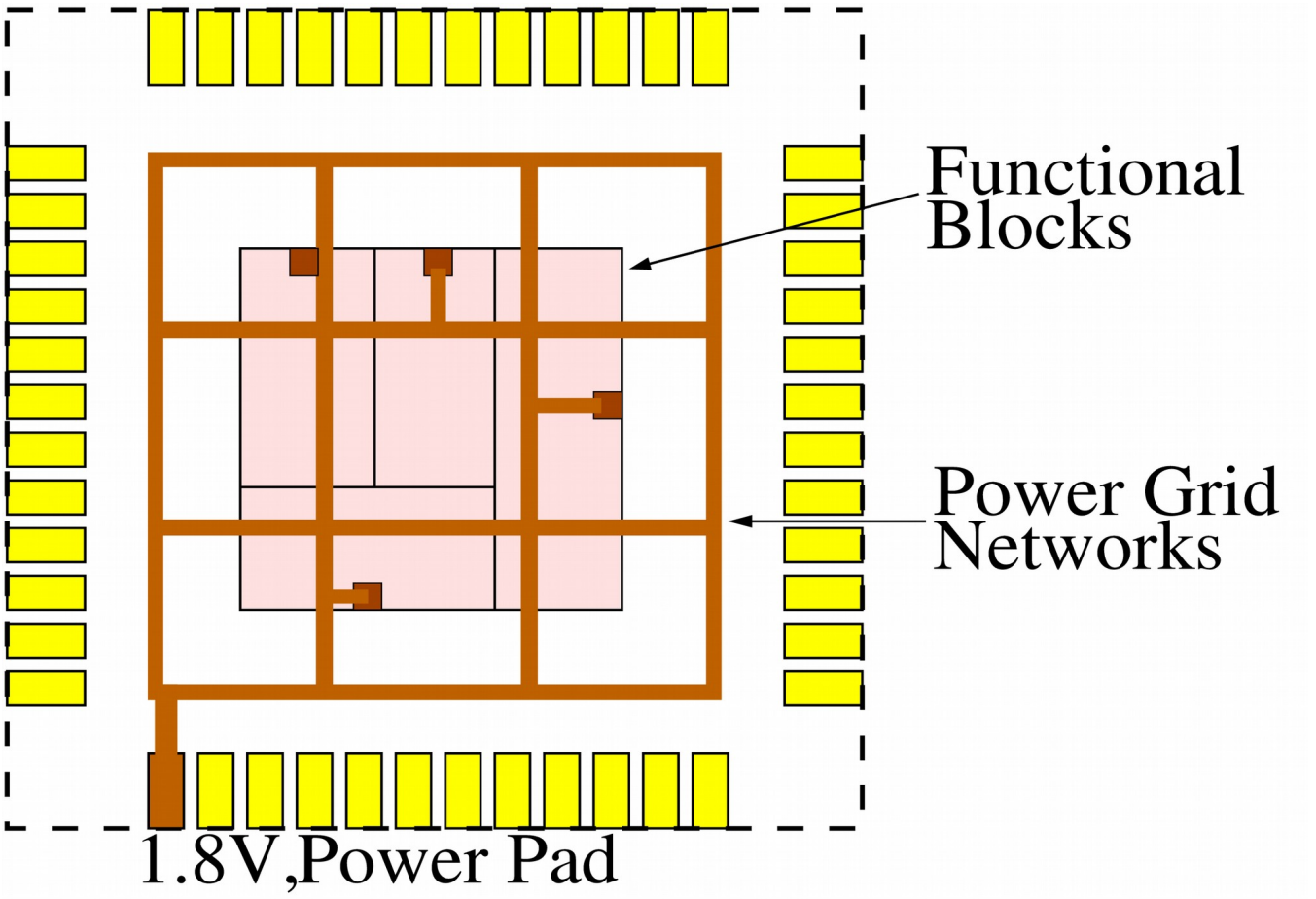
- Floorplanning.
- Placement.
- Routing.



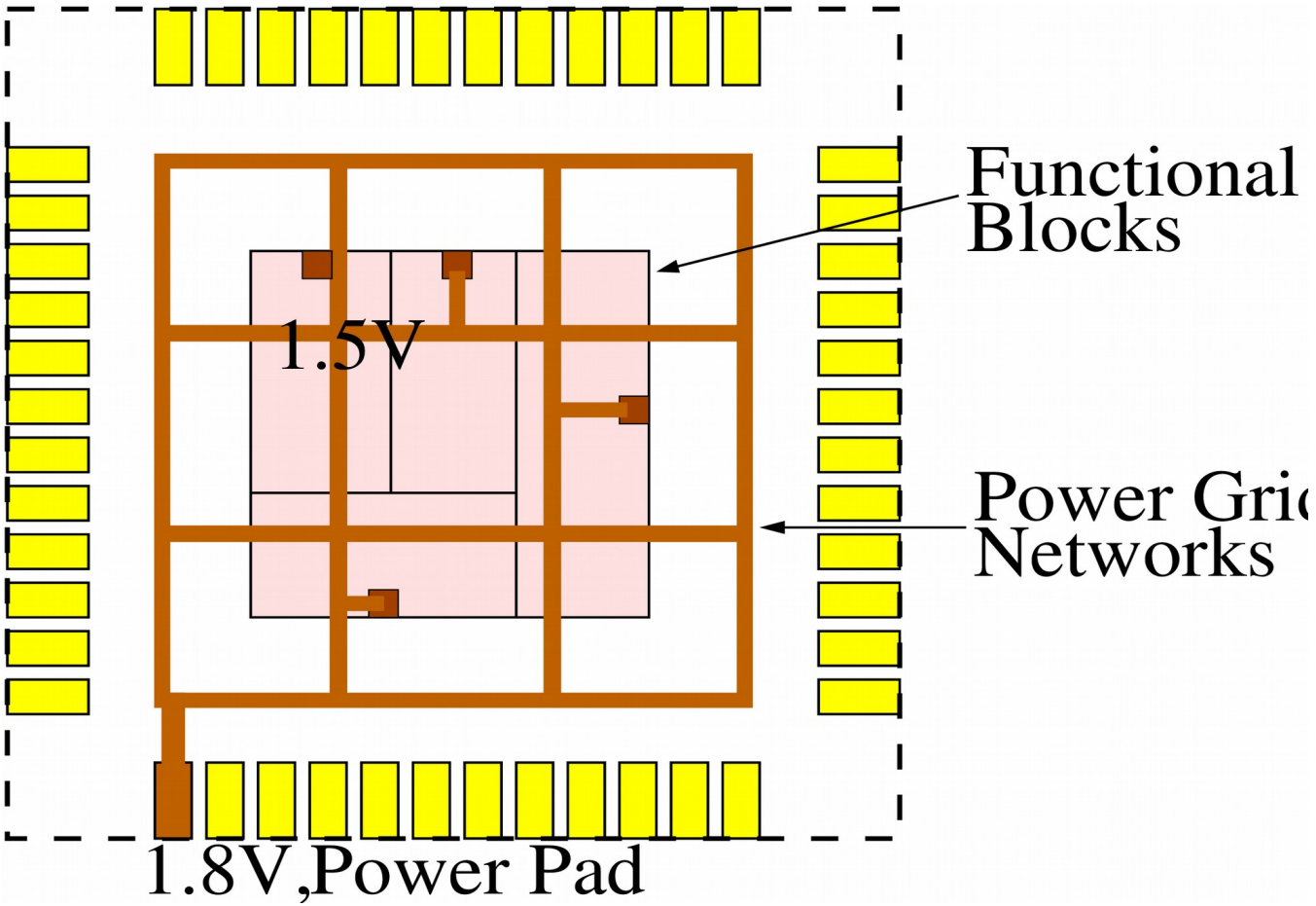
Floorplan



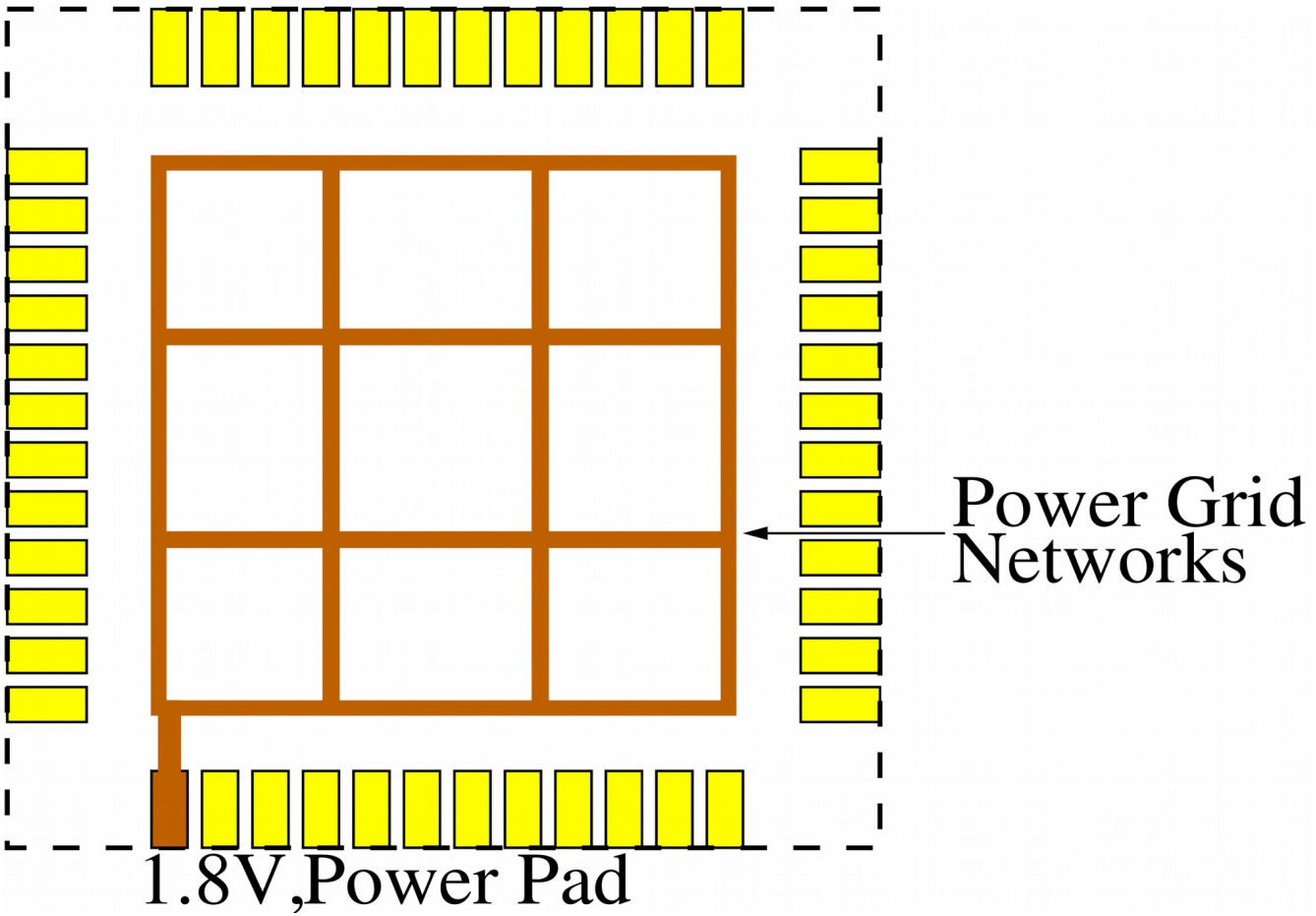
Power Grid Network Connections



Power Grid Network Connections



Power Grid Network Connections



Power Grid Modeling

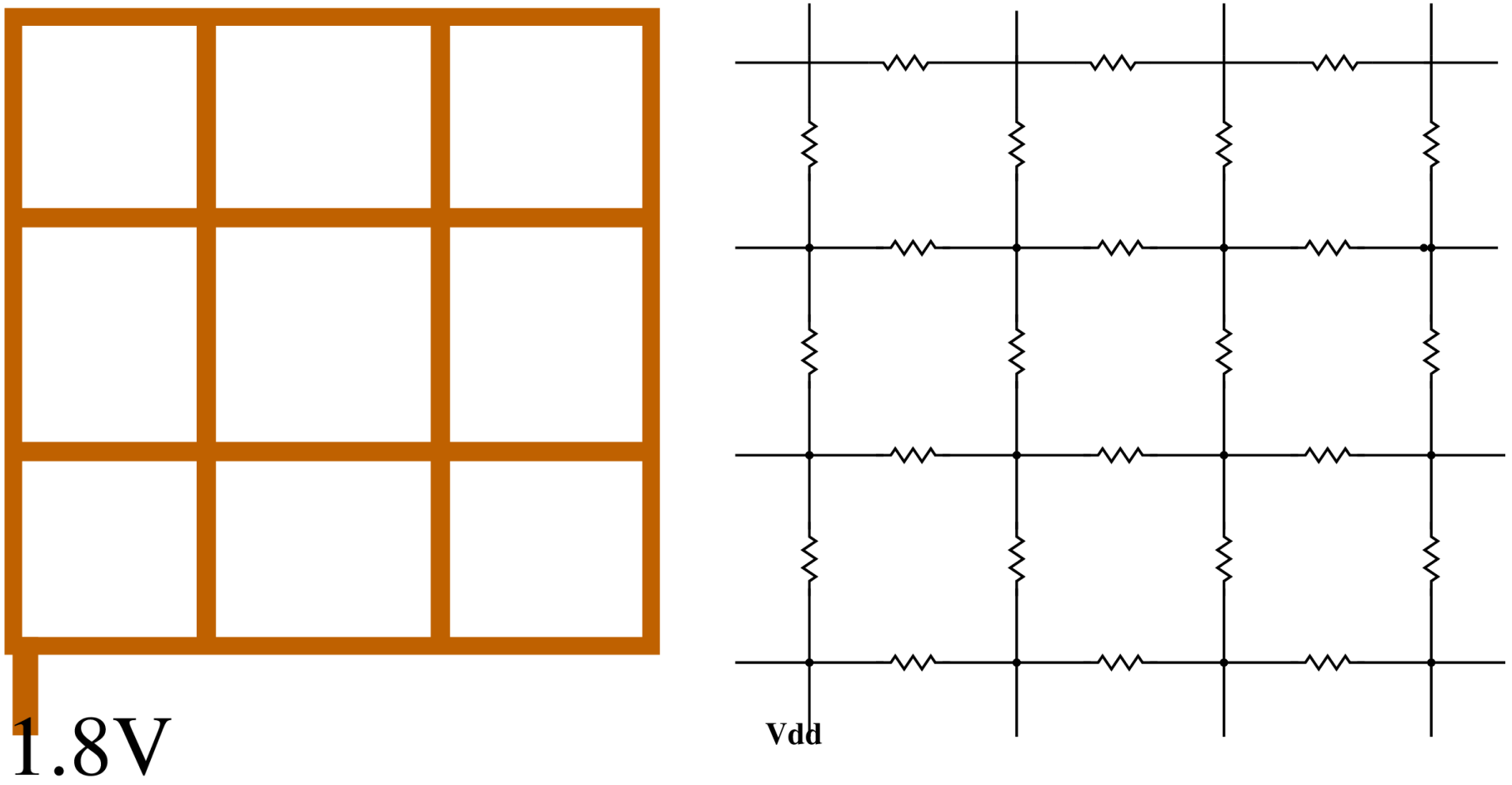
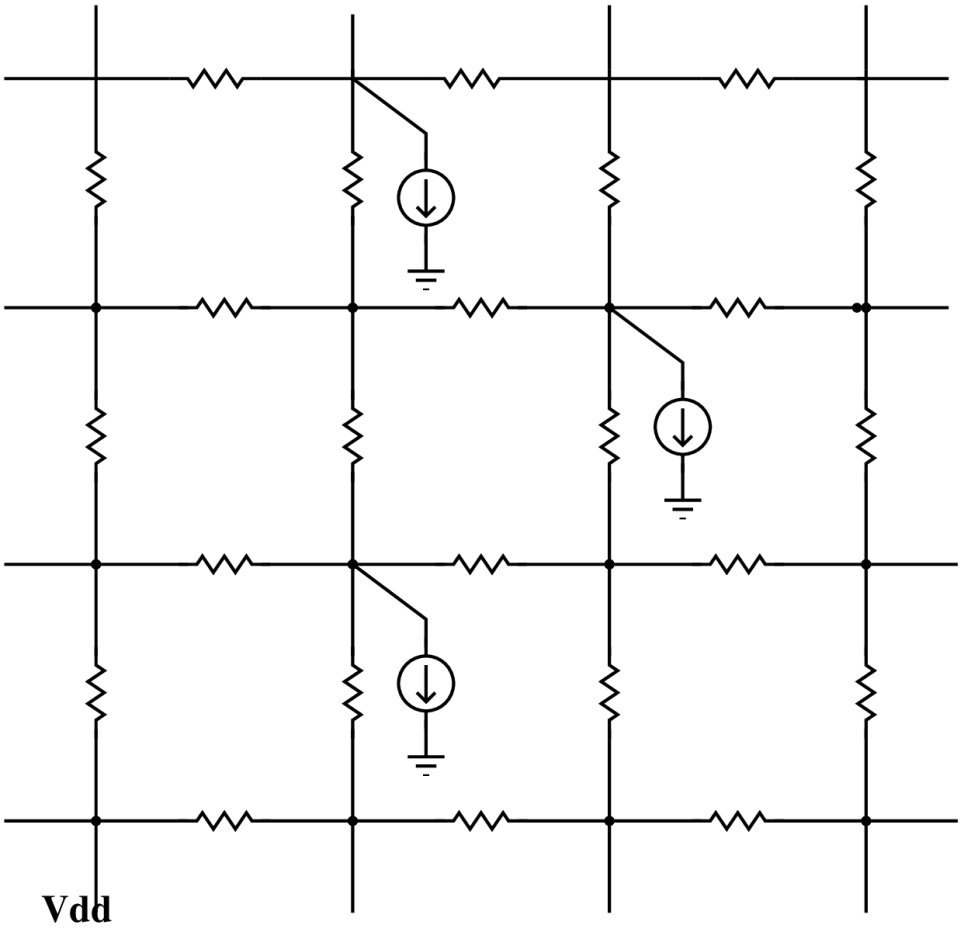


Fig: Power Grid Network and its equivalent resistive networks

Power Grid Modeling



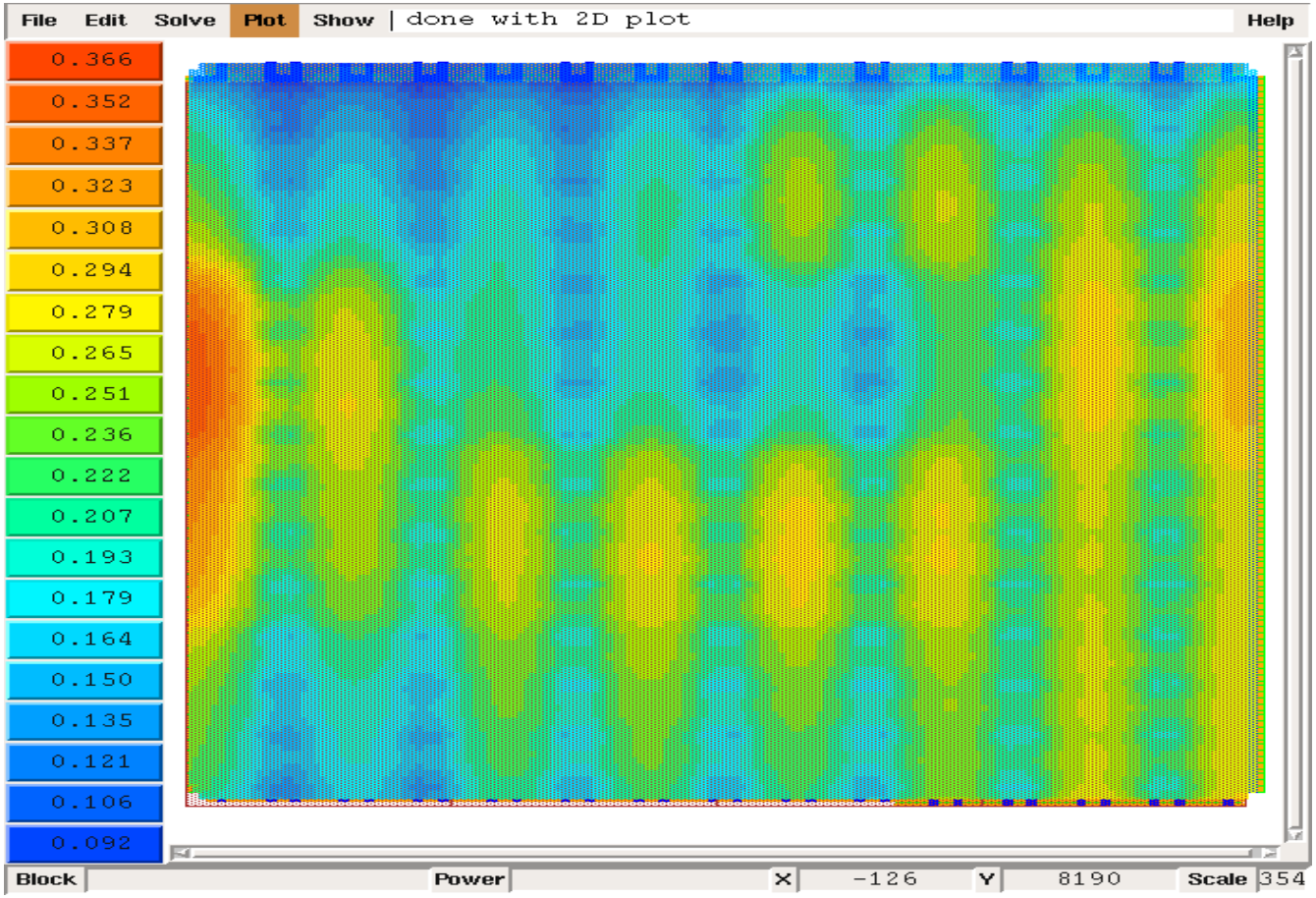
Using KCL & KVL, System of Linear Equations are formed: [G][V] = [I] formed

Challenge is to solve for large power grid networks. For eg: think of power grid design of Apple A12x SoC with 10 billion transistors

Design Challenges in Power Grid

- **IR drop**
- **Electromigration**

Figure: IR drop map of IBM processor



Courtesy: IBM

Design Challenges in Power Grid: Electromigration

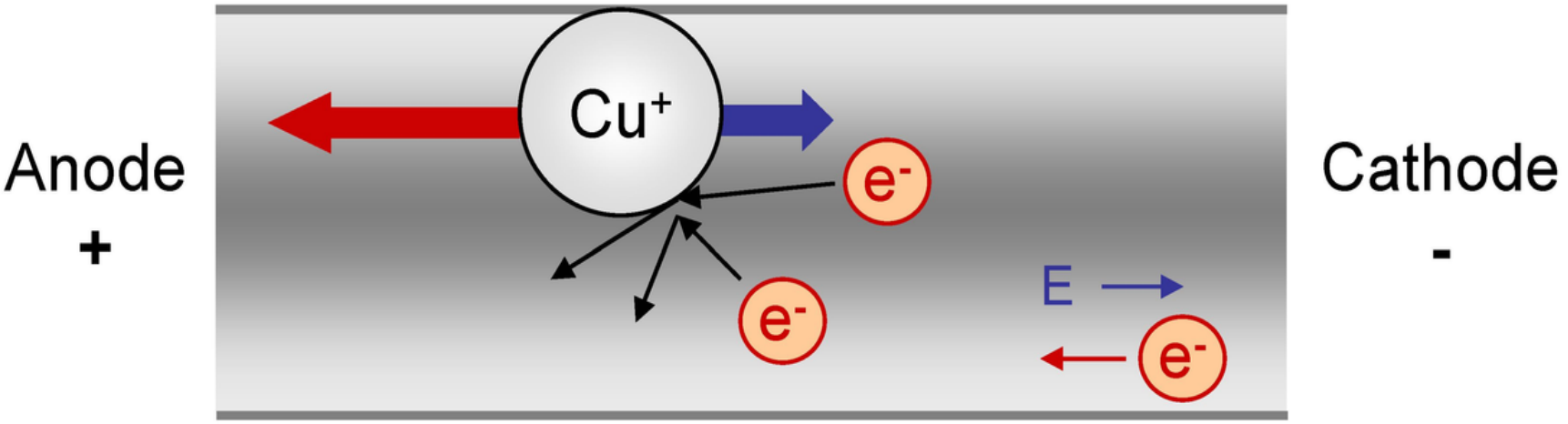
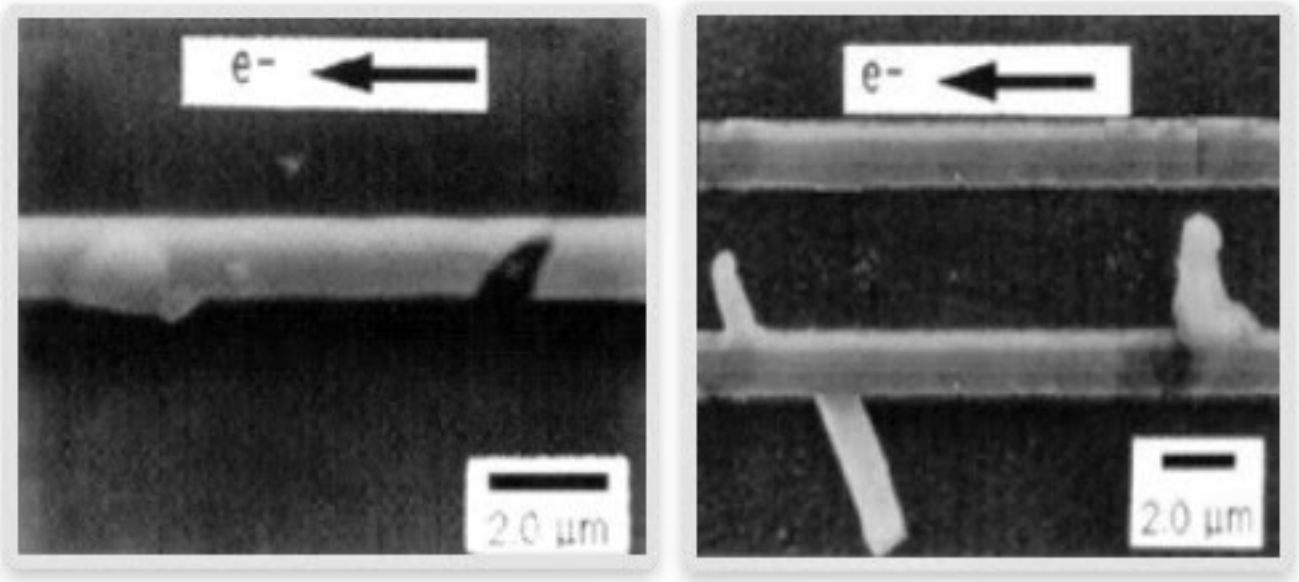


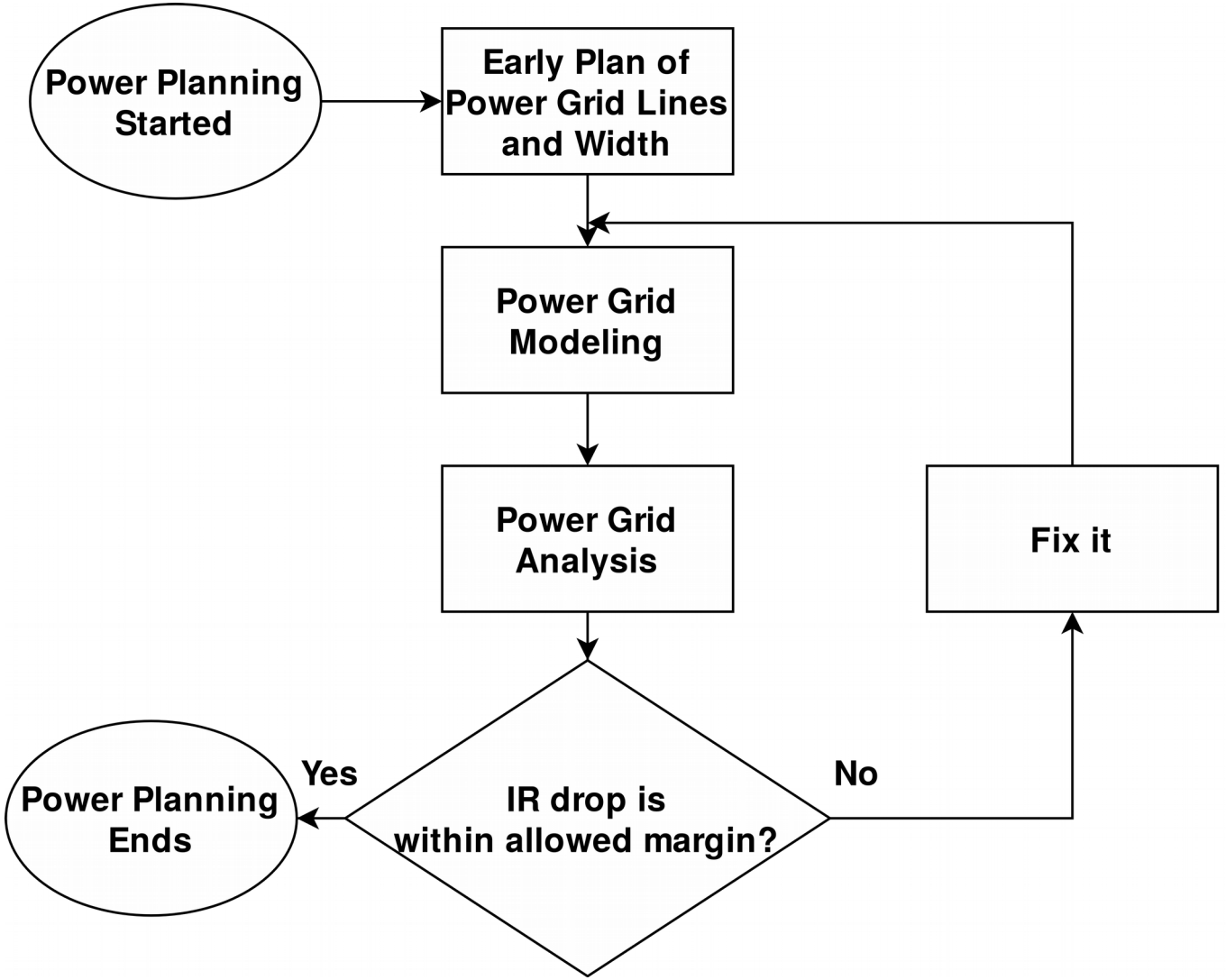
Figure: Metal wire cross section



Courtesy: IEEE

Figure: SEM Image of Void (open circuit) and hillock (short circuit)

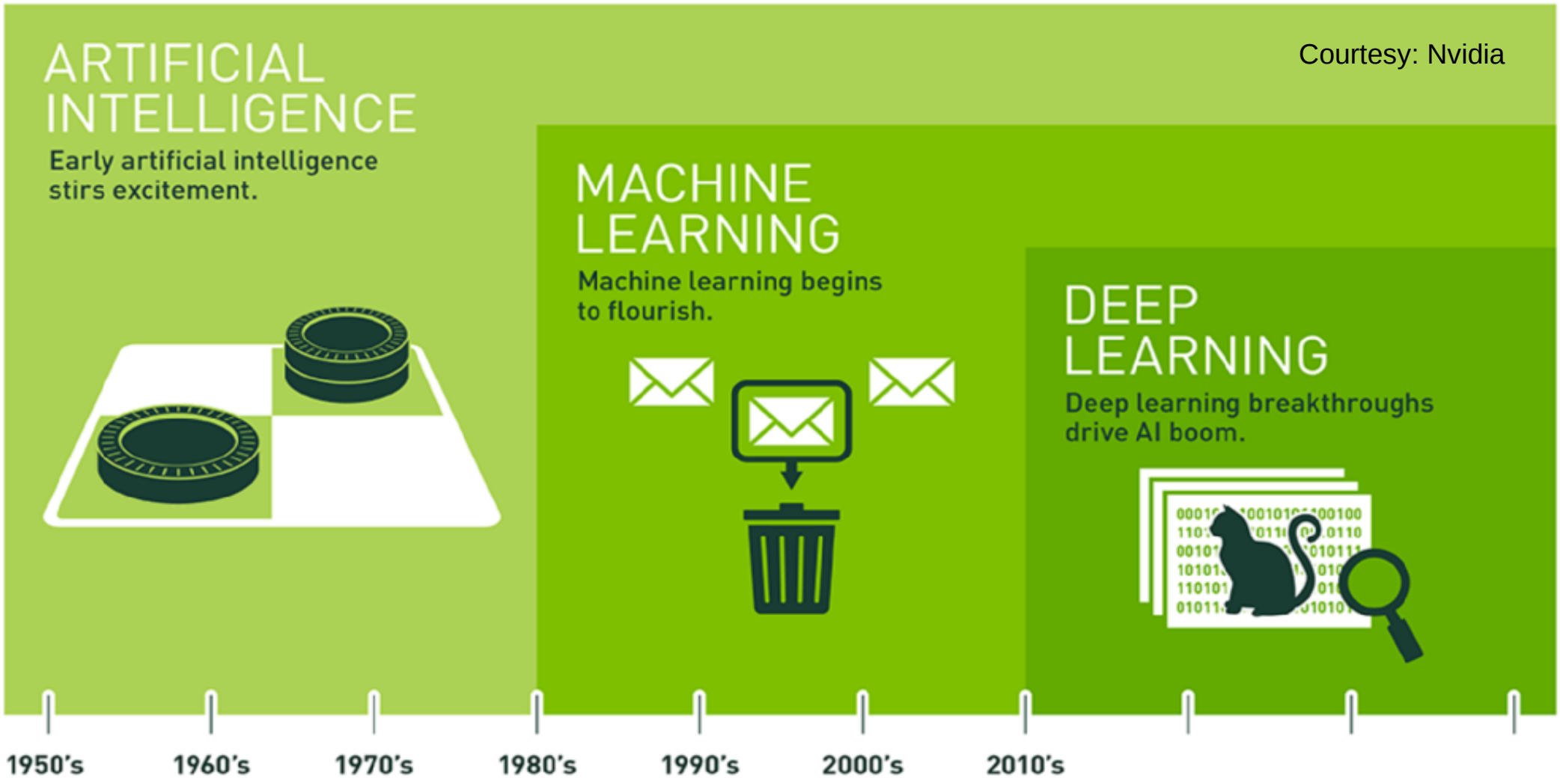
On-Chip Power Planning



Previous Works

- **Early 2000s**: Many works on Power Grid Analysis, Few works on Area Optimization of Power Grid. [DAC'00/01/02/03, ICCAD'05/08, DATE'06/08]
- **Early 2010s**: Works mostly focused on Fast Power Grid Analysis/Verification using Linear Algebric Methods.[DATE'11/13, ICCAD'10/12/13, DAC'11/12].
- **Since 2013**: Electromigration Issues in Power Grid [DAC'13/14/16/17, ICCAD'13/16/17/19, DATE'17]
- **Recently, 2018/19/20**: Learning-based approaches in different issues in Power Grid Design. [ICCAD'19, DATE'19]
- **So Why Learning Approaches now?**

AI/ML History



Courtesy: Nvidia

Since an early flush of optimism in the 1950s, smaller subsets of artificial intellgigence – first machin learning, then deep learning, a subset of machine learning – have created ever large disruptions

What is AI/ML?

- **Artificial Intelligence:** Intelligence demonstrated by machines.
- **Machine Learning:** Inlelligence demonstrated by machines using its previous experiences.
- Supervised Learning: Labelled data.
- Unsupervised Learning: Unlabelled data.

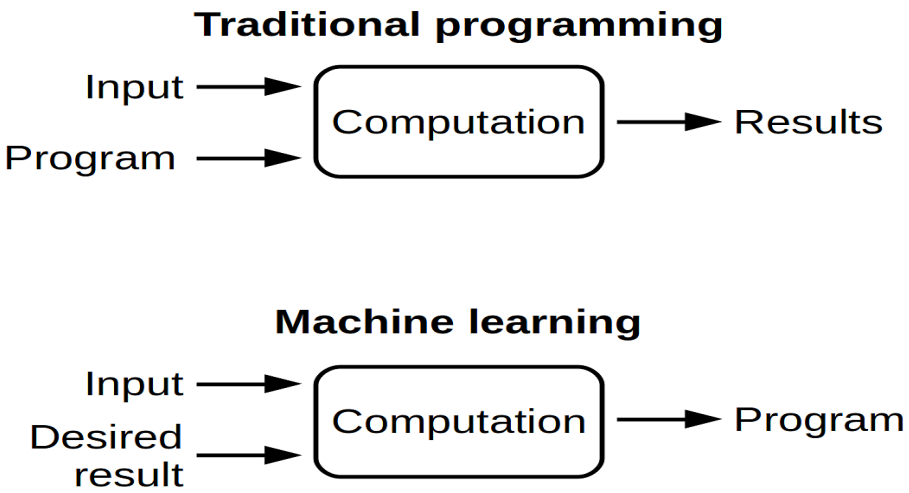


Figure: Difference between Traditional programming and Machine Learning

Why AI/ML for PG Design? Why Now?

- Designs are getting complex and large.
- Big data is associated with each design.
- Time-consuming design cycle.
- Involves vast human resource.
- New technology nodes evolve in every alternate year.
  - 14nm in 2014, 10nm in 2016, 7nm in 2018, 5/3nm in 2020.
- AI/ML can help in automation/semi-automation.
- Recently, Placement, Routing problems are also solved using AI/ML.

Why ML for EDA? Why Now?

- Intelligent Design of Electronic Assets ***(IDEA)*** program sponsored by DARPA in 2018.
- It's under an initiative of $1.5 billion.
- Involve many US universities and industry.
- *IDEA aims to create a* ***"no human in the loop" 24 hour turnaround*** *layout generator for System-On-Chips, System-In-Packages, and Printed Circuit Boards.*

https://spectrum.ieee.org/tech-talk/semiconductors/design/darpa-picks-its-first-set-of-winners-in-electronics-resurgence-initiative

Thesis Objective & Problem Statement

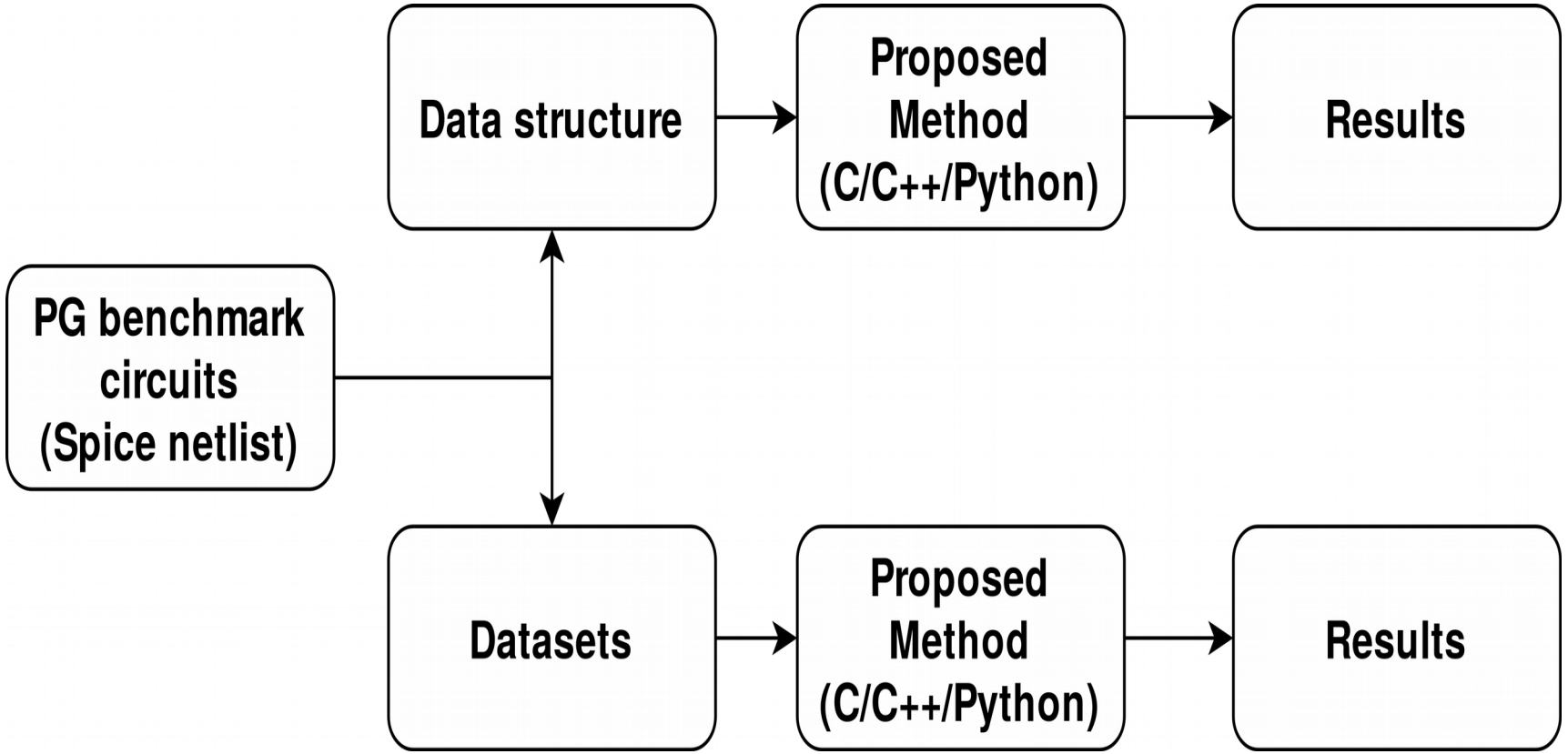
- **Objective:** Designing Relibale On-Chip Power Grid.
- Finding the hotspots for a large scale power grid network is challenging task.
- **Problem 1**: Solving power grid network in efficient way with better runtime.
- **Problem 2**: Obtaining optimum power grid design parameters.
- **Problem 3**: Calculating the Electromigration Aging time of the power grid network during the design time.
- **Our Main focus is AI & ML-based approaches.**

Thesis Contributions

- **First Contribution**: Power Grid Analysis using Probabilitisc Approach
- **Second Contribution**: Design Space Exploration of PG Interconnects
- **Third Contribution**: Machine Learning Approach in PG Design
- **Fourth Contribution**: Machine Learning Approach in Aging Prediction of PG

Simulations and Experimental setup

- All the methods are implemented in C/C++, Python.
- OS : Ubuntu 14.04/16.04/CentOS 6.
- Machine: Intel E5-2650/i5.
- Memory: 32/64 GB

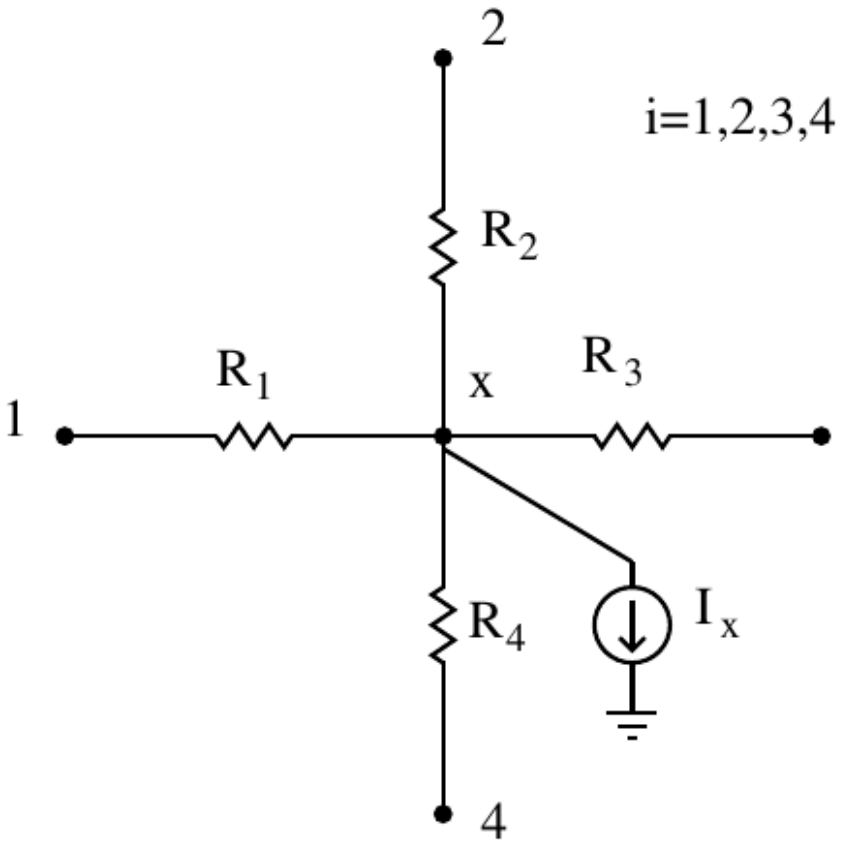


First Contribution: Power Grid Analysis using Probabilitisc Approach

- Power grid network is converted to a unweigthed graph.
- Objective is to speedup power grid analysis process.
- Levy Flight-based Probabilistic Method is used for traversing the graph.
- Results.

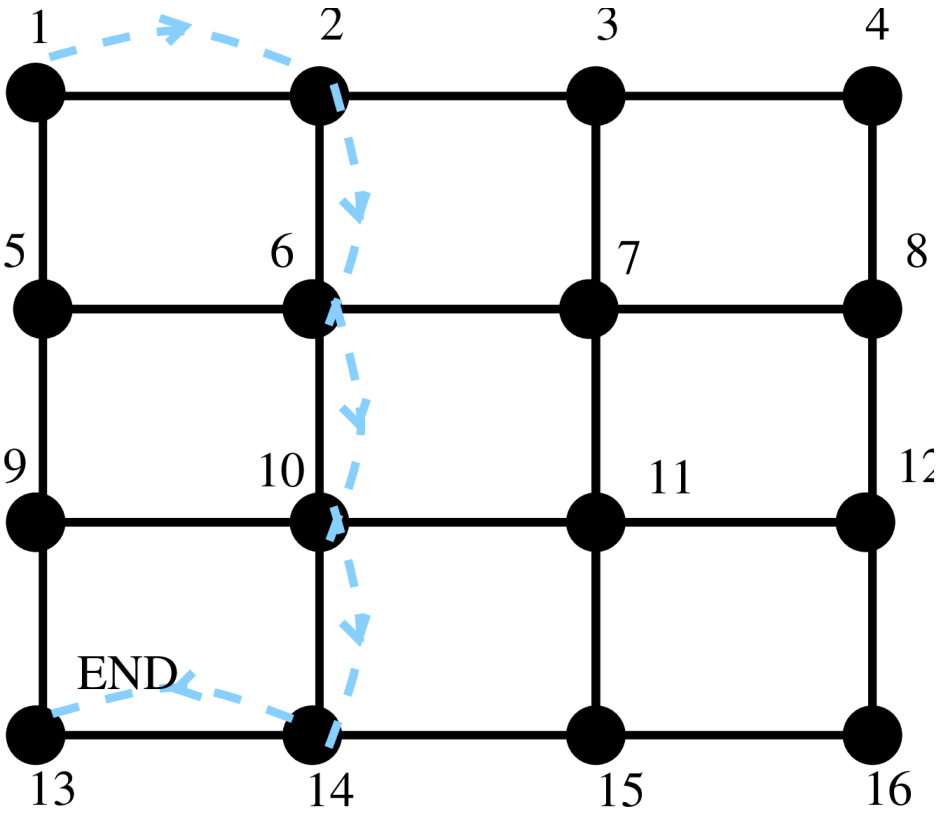
* Parts of work is published by Dey et al. “**Markov Chain Model using Levy Flight for VLSI Power Grid Analysis**” in **VLSID’17**.

Single Node of PG Network



$$V_x = \sum_{i=1}^{degree(x)} \frac{g_i}{\sum_{i=1}^{degree(x)} g_i} - \frac{I_x}{\sum_{i=1}^{degree(x)} g_i}$$

Previous Work*



- Circuit Equation

$$V_x = \sum_{i=1}^{degree(x)} \frac{g_i}{\sum_{i=1}^{degree(x)} g_i} - \frac{I_x}{\sum_{i=1}^{degree(x)} g_i}$$

- RW Eq. for gain f()

$$f(x) = E[total\ money\ earned |\ walk\ starts\ at\ node\ x]$$

$$f(x) = p_{x,1}f(1) + p_{x,2}f(2) + \cdots + p_{x,degree(x)}f(degree(x)) - m_x$$

- Compare both

$$p_{x,i} = \frac{g_i}{\sum_{i=1}^{degree(x)} g_i}, i = 1, 2, \cdots, degree(x)$$

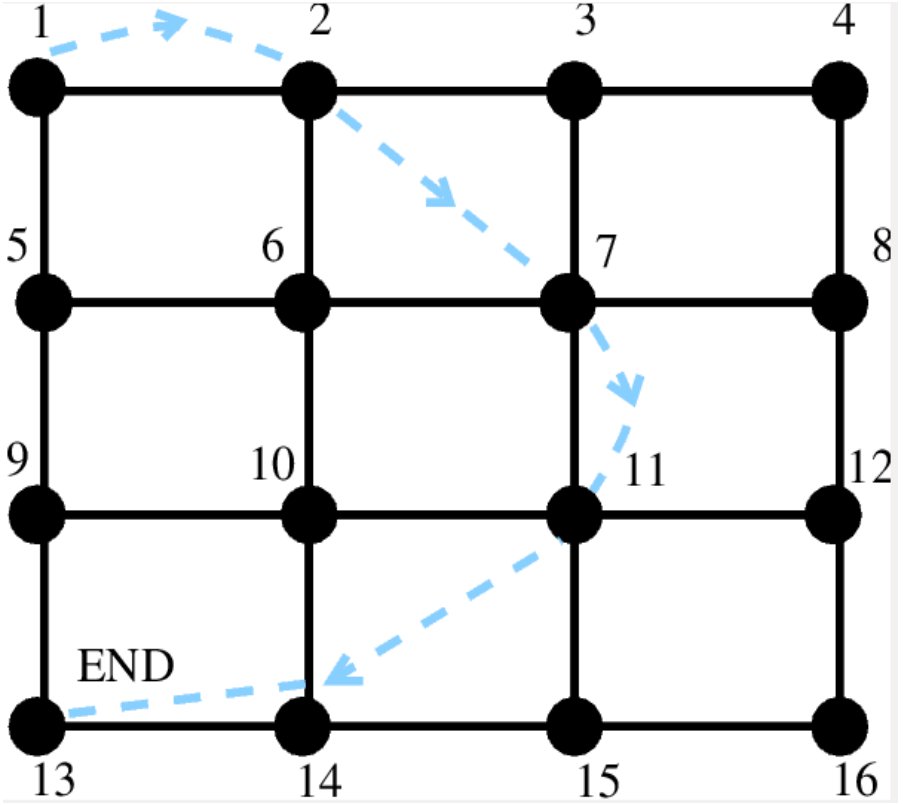
$$m_x = \frac{I_x}{\sum_{i=1}^{degree(x)} g_i}, m_0 = V_{DD}, f(x) = V_x$$

* Qian et al. "***Random Walks in Supply Network***" in **DAC'03**.

Our Contributions

- We try to remove the self-loops of Random Walks.
- Further, we adapted jumping strategy to speedup the convergence.
- Levy Flight is used to incorporate the jumping strategy.
- The reward is modified by calculating the effective resistance.
- To validate the method the result is tested for 49M nodes power grid network.

Our Contributions



$$\mathrm{m}_x = \frac{I_x}{\sum_{i=1}^{degree(x-1)} g_i + \delta g_i}, m_0 = V_{DD}, f(x) = V_x$$

Experimental Results*

TABLE I
SPEEDUP ANALYSIS OF PROPOSED METHOD USING LÉVY FLIGHT ON CPU

PG Circuits	t_{RW} (s)	t_{GS} (s)	t_{HLS} (s)	t_{levy} (s)	Speedup (t_{RW}/t_{levy})	Speedup (t_{GS}/t_{levy})	Speedup (t_{HLS}/t_{levy})
pgckt_40K	0.30	0.36	0.82	0.17	1.76×	2.11×	4.82×
pgckt_90K	0.65	1.62	1.84	0.30	2.16×	5.40×	6.13×
pgckt_250K	1.92	6.78	6.06	0.86	2.23×	7.88×	7.04×
pgckt_640K	7.98	19.31	20.00	2.19	3.64×	8.81×	9.13×
pgckt_1M	18.95	27.85	39.55	3.51	5.39×	7.93×	11.26×
pgckt_4M	297.21	117.76	154.78	14.61	20.34×	8.06×	10.59×
pgckt_9M	1513.4	272.74	349.66	33.88	44.66×	8.05×	10.32×
pgckt_16M	3326.44	486.03	651.15	61.49	54.09×	7.90×	10.58×
pgckt_25M	6263.80	760.10	1034.36	112.85	55.50×	6.73×	9.16×
pgckt_36M	9800.35	1094.01	1562.71	167.63	58.46×	6.52×	9.32×
pgckt_49M	14065.90	1498.20	2430.38	232.91	60.39×	6.43×	10.43×

Max speedup ~60X over RW method

First Contribution Summary

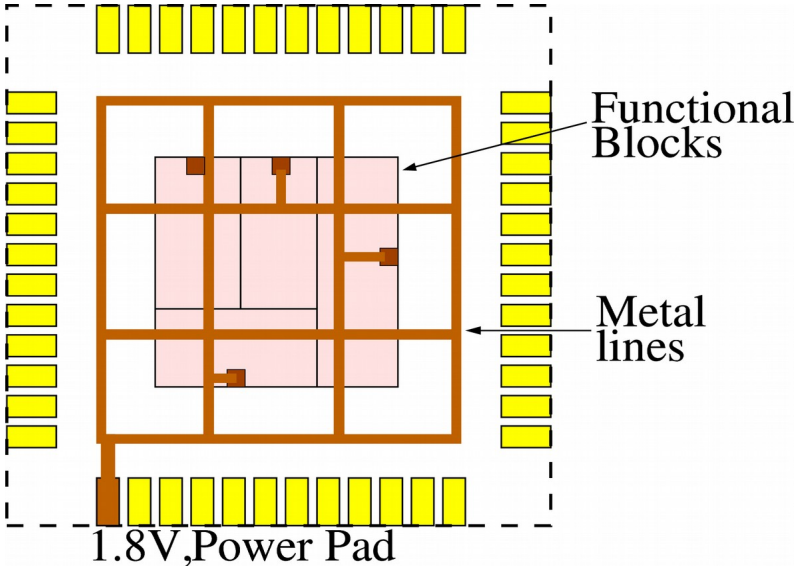
- Fast power grid method is proposed.
- We have achieved max. ~60X speedup over RW method for 49M node PG circuit.
- Solutions of the proposed method is similar to RW method (3-4% error).

Second Contribution: Design Space Exploration of PG Interconnects*

- First IR drop minimization problem using metaheuristics.
- First Multiobjective DSE for PG Design
- Problem Formulation
- Proposed DSE Framework
- Results.

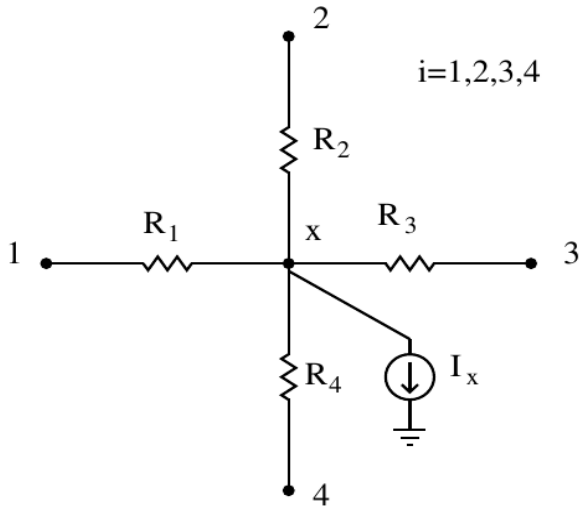
* Parts of work is published by Dey et al. **ISVLSI'18**. & Elsevier **MICPRO 2021**

Problem Formulation: IR Drop Minimization



$$\upsilon = \sum_{i=1}^{b} |I_i| R_i$$

$$= \sum_{i=1}^{b} |I_i| \frac{\rho l_i}{w_i},$$



$$\underset{I_i, w_i}{\text{minimize}} \quad \upsilon(I_i, w_i)$$

$$\text{subject to} \quad |I_{i\in E}| R_{i\in E} \le \xi, \sum_{i=1}^{b} l_i w_i \le \mathscr{A}_{max}$$

$$\frac{I_{i\in E}}{w_{i\in E}} \le I_m, w_{i\in E} \ge w_{min}$$

$$\sum_{i=1}^{K} I_{j_i} + I_x = 0 \; \forall j \in V$$

Problem Formulation: IR Drop-Area Minimization

- *P* represents the set of nodes and *Q* represents the set of branches

$$\underset{I_i, w_i}{\text{minimize}} \quad \begin{cases} v_{\text{IR total}} = \sum_{i=1}^{b} |I_i| \, \frac{\rho l_i}{w_i} \\ A = \sum_{i=1}^{b} l_i w_i \end{cases}$$

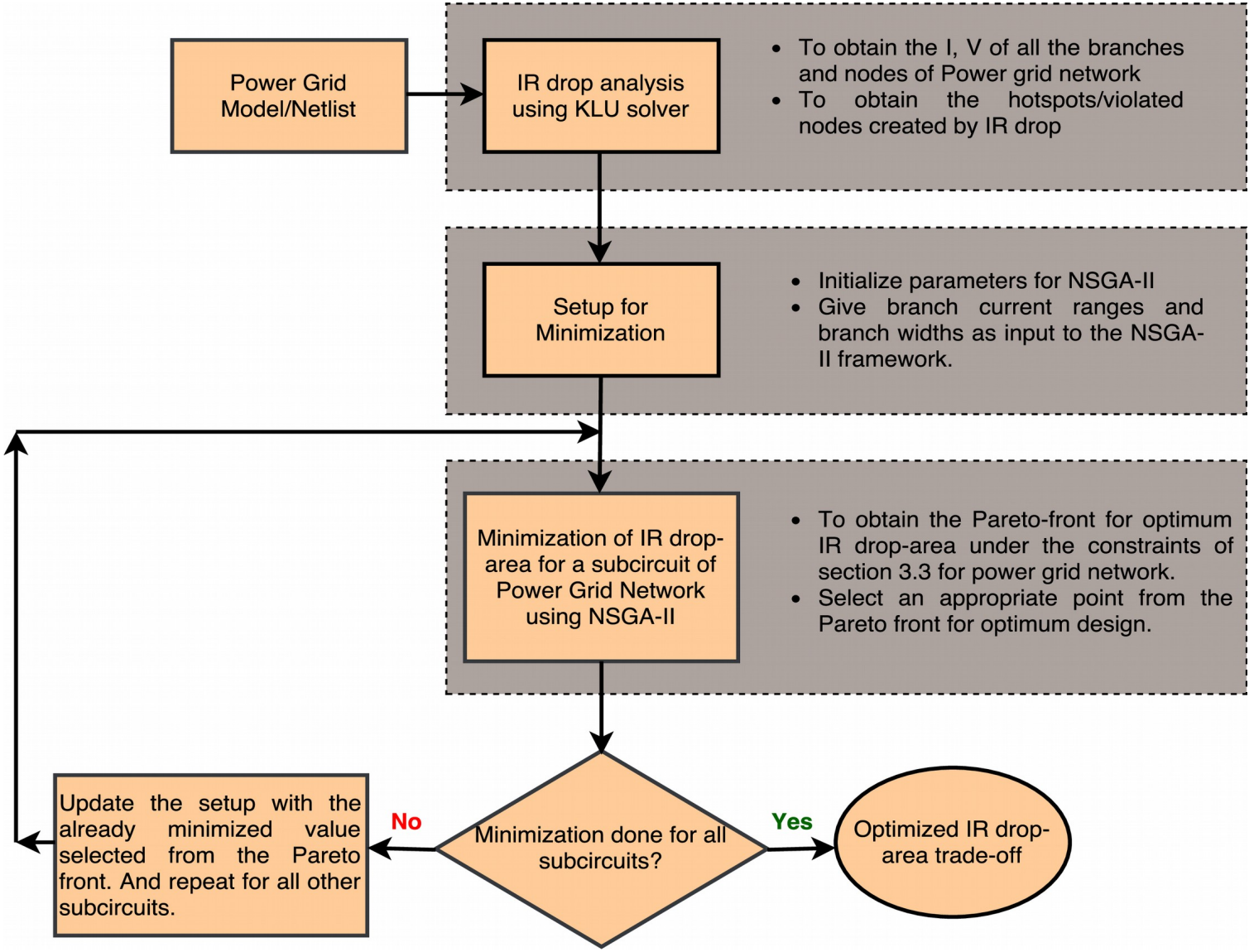
$$\text{subject to} \quad |I_i| \, \frac{\rho l_i}{w_i} \leq \xi \; \forall i \in Q,$$

$$\sum_{i=1}^{b} A_i \leq \mathcal{A}_{max},$$

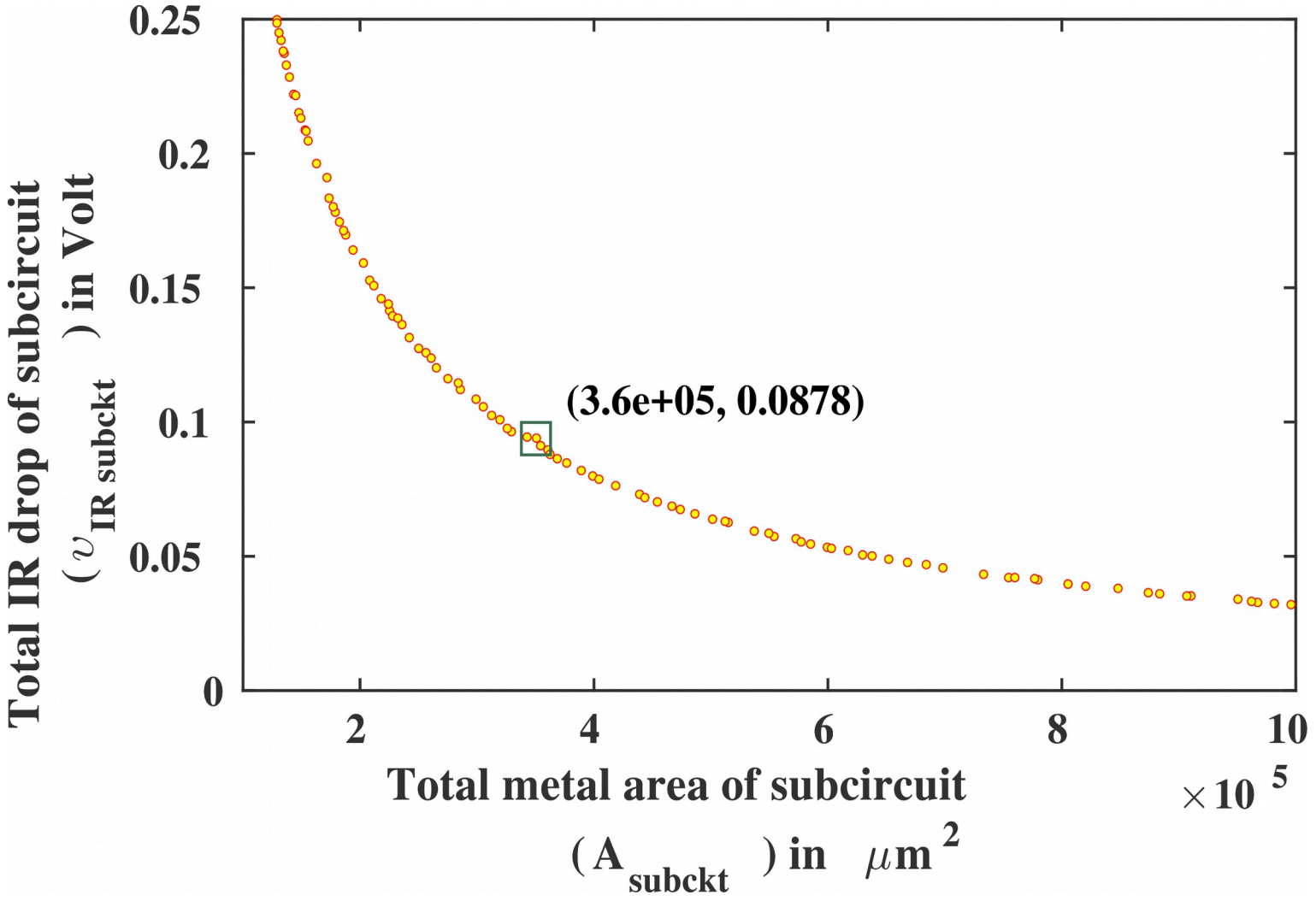
$$\sum_{i=1}^{K} I_{j_i} + I_x = 0 \; \forall j \in P \;\; (\text{KCL}),$$

$$I_i \leq I_{max}.$$

Proposed Framework



Pareto Front of a Subcircuit of ibmpg1



Note: Knee Point is selected as the suitable point from the Pareto front

Results

~7.31% reduction in worst-case IR drop

~8.51% reduction in metal routing area

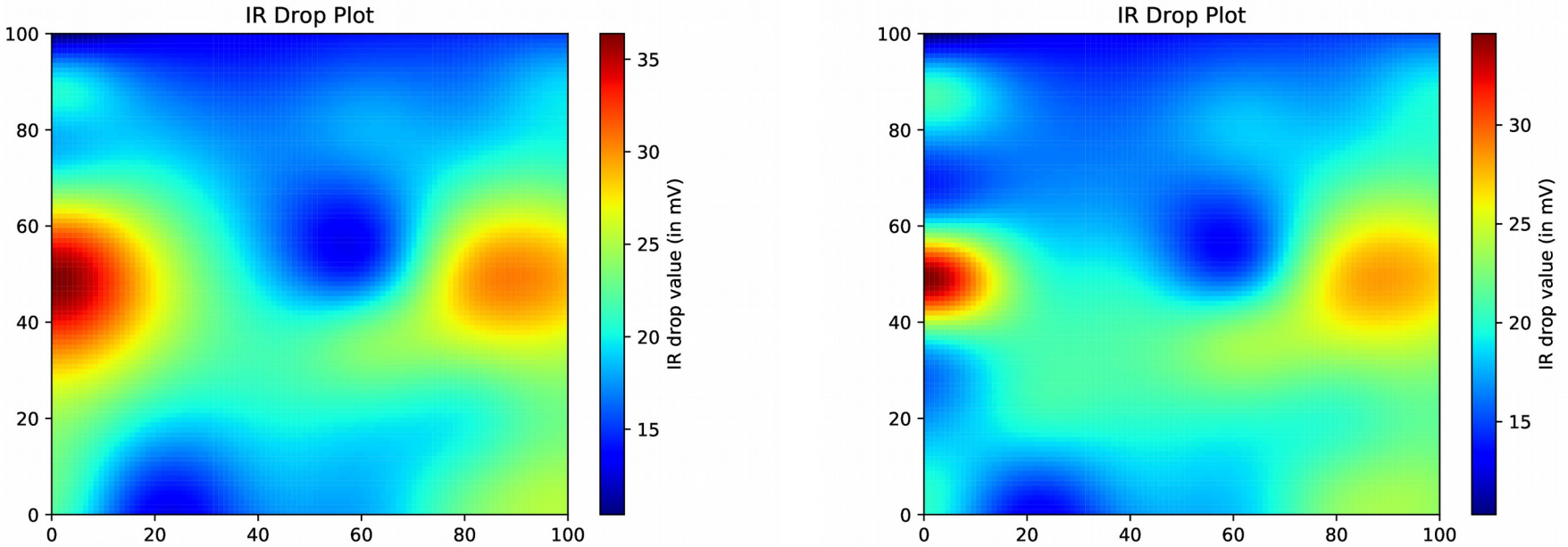


Figure: Two instances of overall IR drop map for ibmpg2 benchmark circuit. (a) Before optimization. (b) After using the proposed framework.

Second Contribution Summary

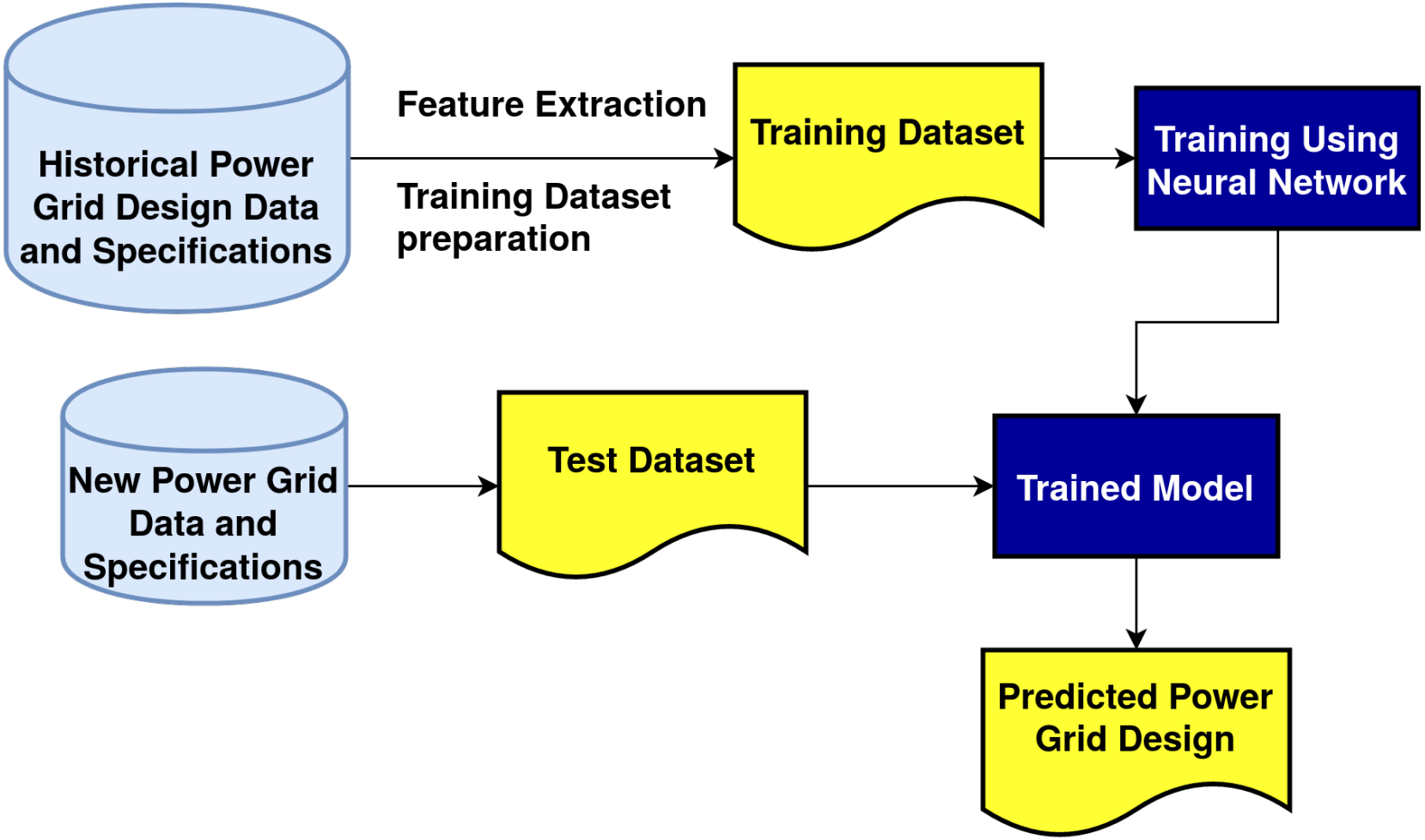
- DSE framework helps in obtaining IR drop and Area minimization for PG designs.
- These estimations help designer to obtain initial idea about the design parameters.

Third Contribution: Machine Learning Approach in PG Design*

- First ML Model for PG Design
- Problem Formulation
- Feature Extraction
- Training Data Generation
- Proposed Learning Model
- Results

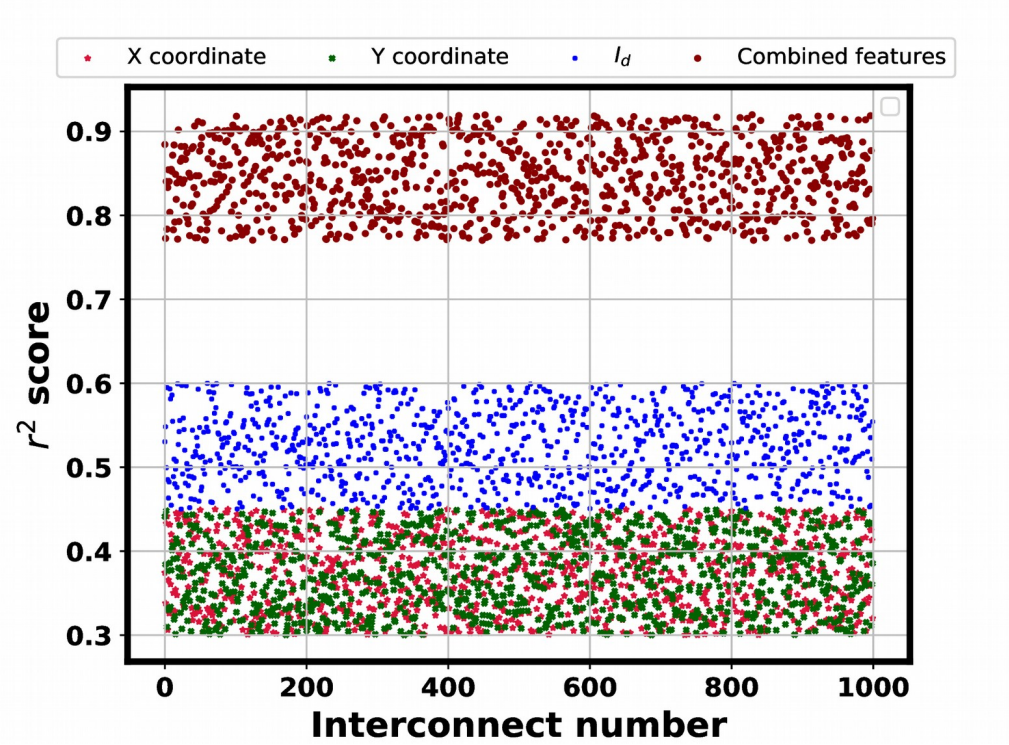
* Published by Dey et al. "**PowerPlanningDL: Reliability-Aware Framework for On-Chip Power Grid Design using Deep Learning**" in **DATE'20**.

Overview of ML Approach



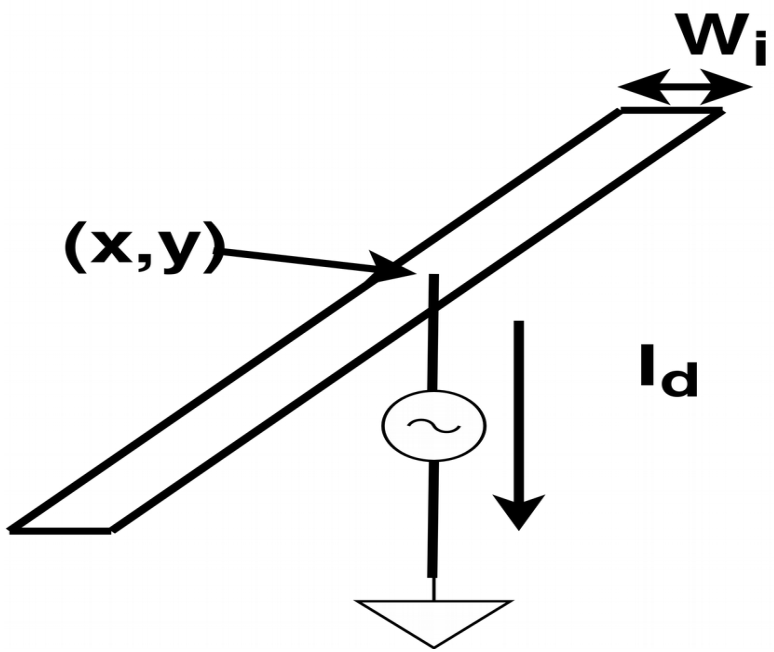
Feature Selection & Training Data

- r^2 score is evaluated
- Power Grid Designs Extracted from the IBM processors are employed.
- Label dataset is prepared for supervised learning.
- Input features: (x,y), I_d
- Output feature: w_i

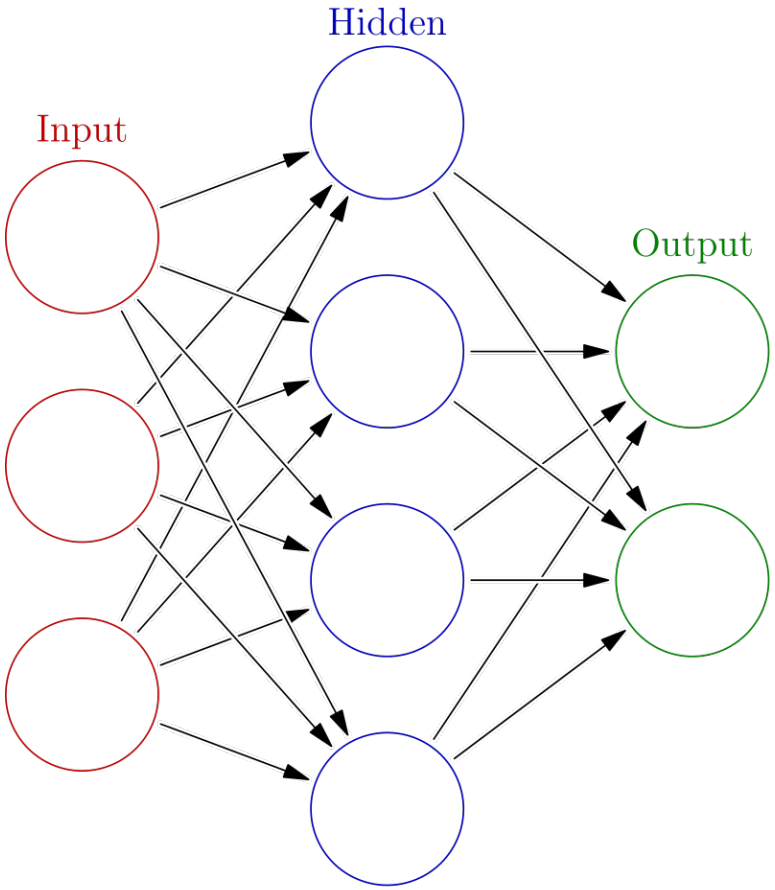


Problem Formulation

- **Problem 1**: Given (x,y) coordinate and I_d, then predict metal width required which satisfy IR and EM margin.
- **Problem 2**: Given the width and I_d, predict IR drop of the PG interconnect.

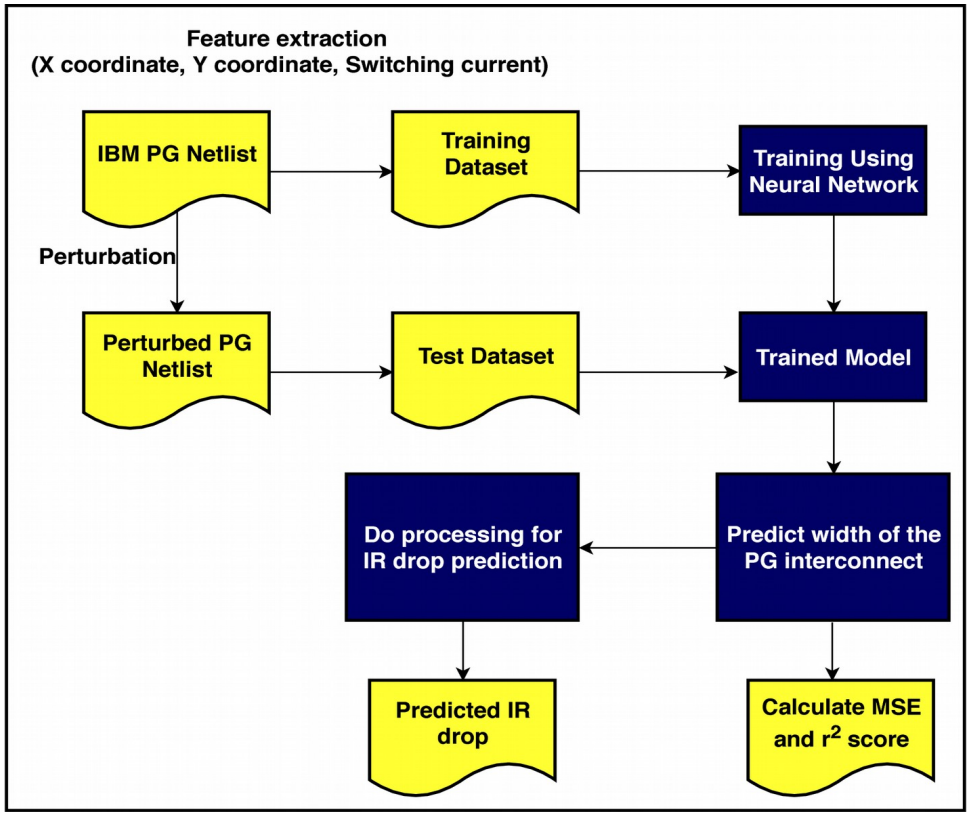


ML Model



- 10 hidden layers.
- Adam optimizer.
- Trained with the generated training data.
- Testdata set is created by perturbing training dataset and performance is tested.

ML Model



Algorithm 1: Wire width prediction by NN

Input: Training Set
Output: w_i and gradient

```
ForwardPropagation(X coordinate, Y coordinate, I_d, w_i)
{
    return loss function f
}
BackwardPropagation()
{
    return gradient
}
```

Algorithm 2: IR drop prediction

Input: Predicted width w_i
Output: Predicted IR drop

1 From switching current I_d and w_i;
2 Use kirchoff's law to predict IR drop.

Experimental Results:

	Time (sec)		Speedup
PG circuits	**Conventional**	**PowerPlanningDL**	$\frac{\textbf{Time}_{Conventional}}{\textbf{Time}_{PowerPlanningDL}}$
ibmpg1	6.85	3.56	1.92×
ibmpg2	23.46	11.88	1.97×
ibmpg3	29.50	8.07	3.59×
ibmpg4	52.4	11.83	4.42×
ibmpg5	74.80	12.74	5.87×
ibmpg6	97.5	17.41	5.60×
ibmpgnew1	102.58	21.50	4.77×
ibmpgnew2	48.60	10.86	4.47×

~5-6X max. speedup

Experimental Results: IR drop map

Same quality of results using ML!

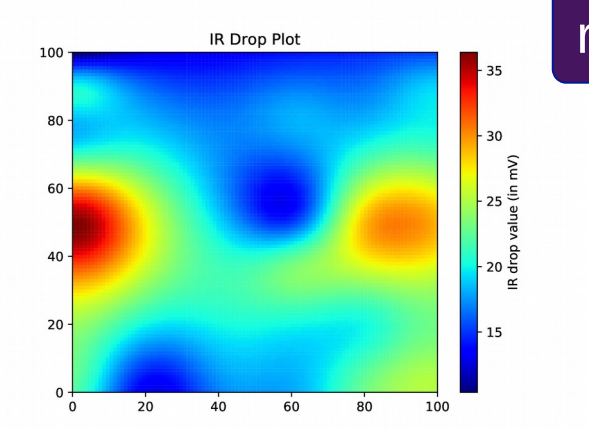


Figure: IR Drop Map of IBMPG2 using Conventional

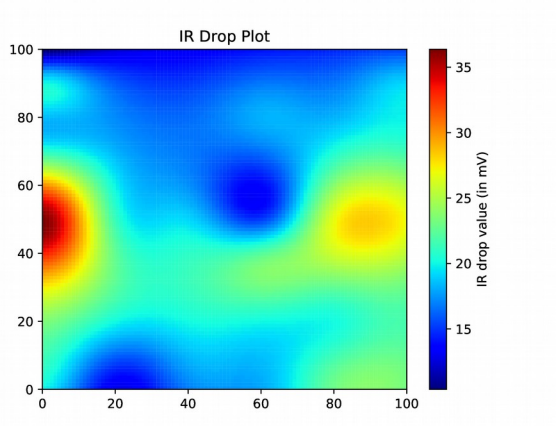


Figure: Predicted IR Drop Map IBMPG2

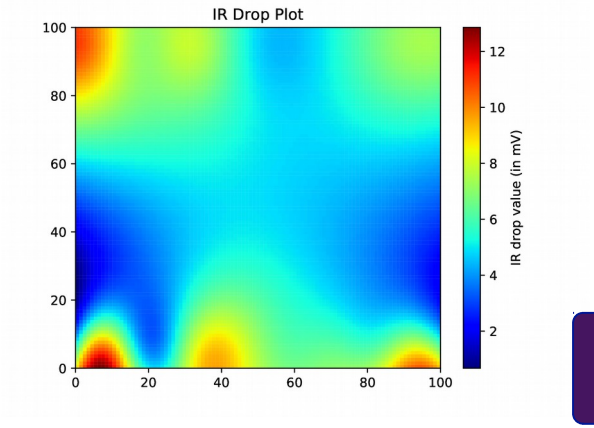


Figure: IRDrop Map of IBMPG6 using Conventional

Error ~2%

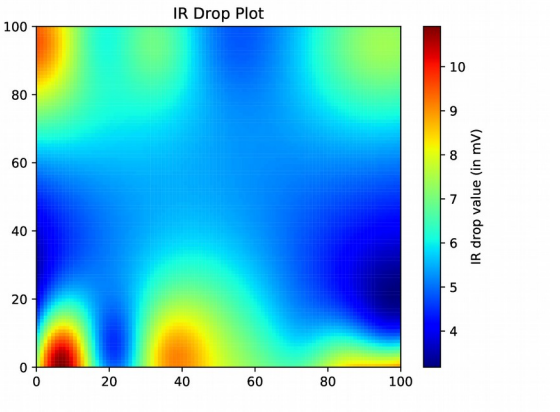


Figure: Predicted IR Drop Map IBMPG6

Third Contribution Summary

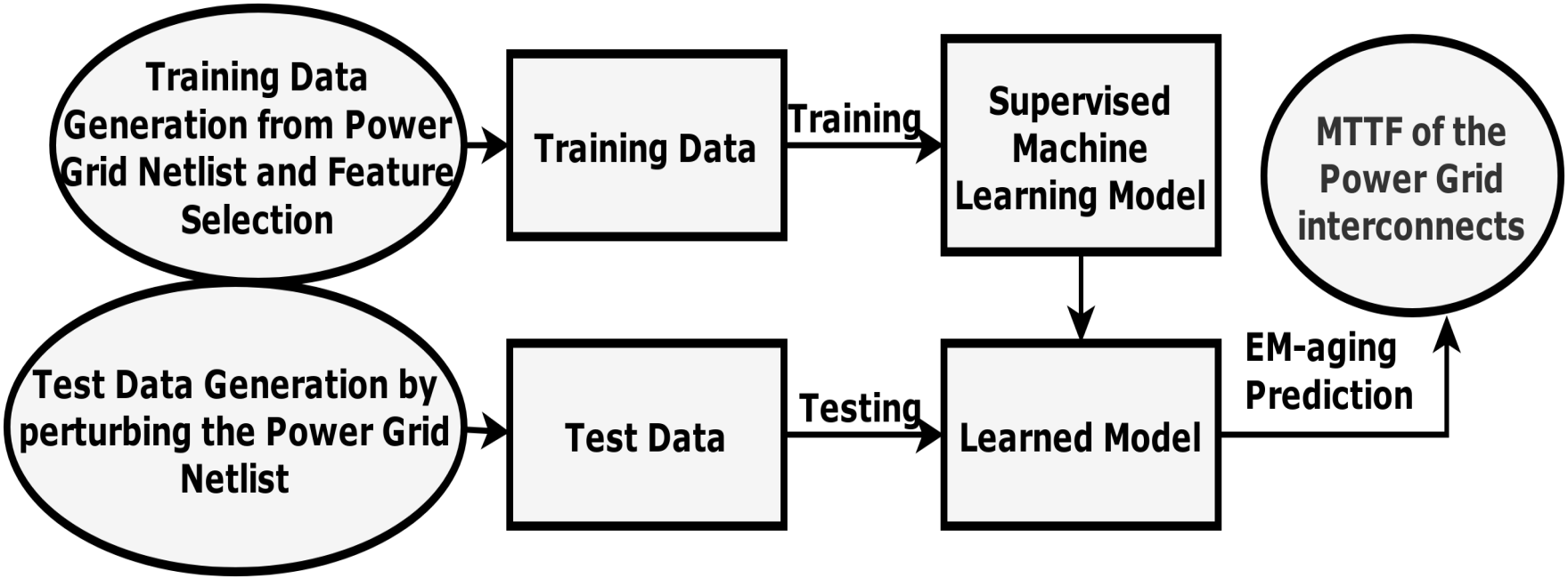
- Same quality of results using ML.
- Fast convergence (~5-6X max. Speedup)
- Low overhead (~2% error)

Fourth Contribution: Machine Learning Approach in Aging Prediction of PG*

- First ML Model for Aging Prediction of PG
- Problem Formulation
- Feature Extraction
- Training Data Generation
- Proposed Learning Model
- Proposed Failure Criterion
- Results show improvement than SOTA

* Published by Dey et al. "**Machine Learning Approach for Fast Electromigration Aware Aging Prediction in Incremental Design of Large Scale On-Chip Power Grid Network**" in **ACM TODAES**.

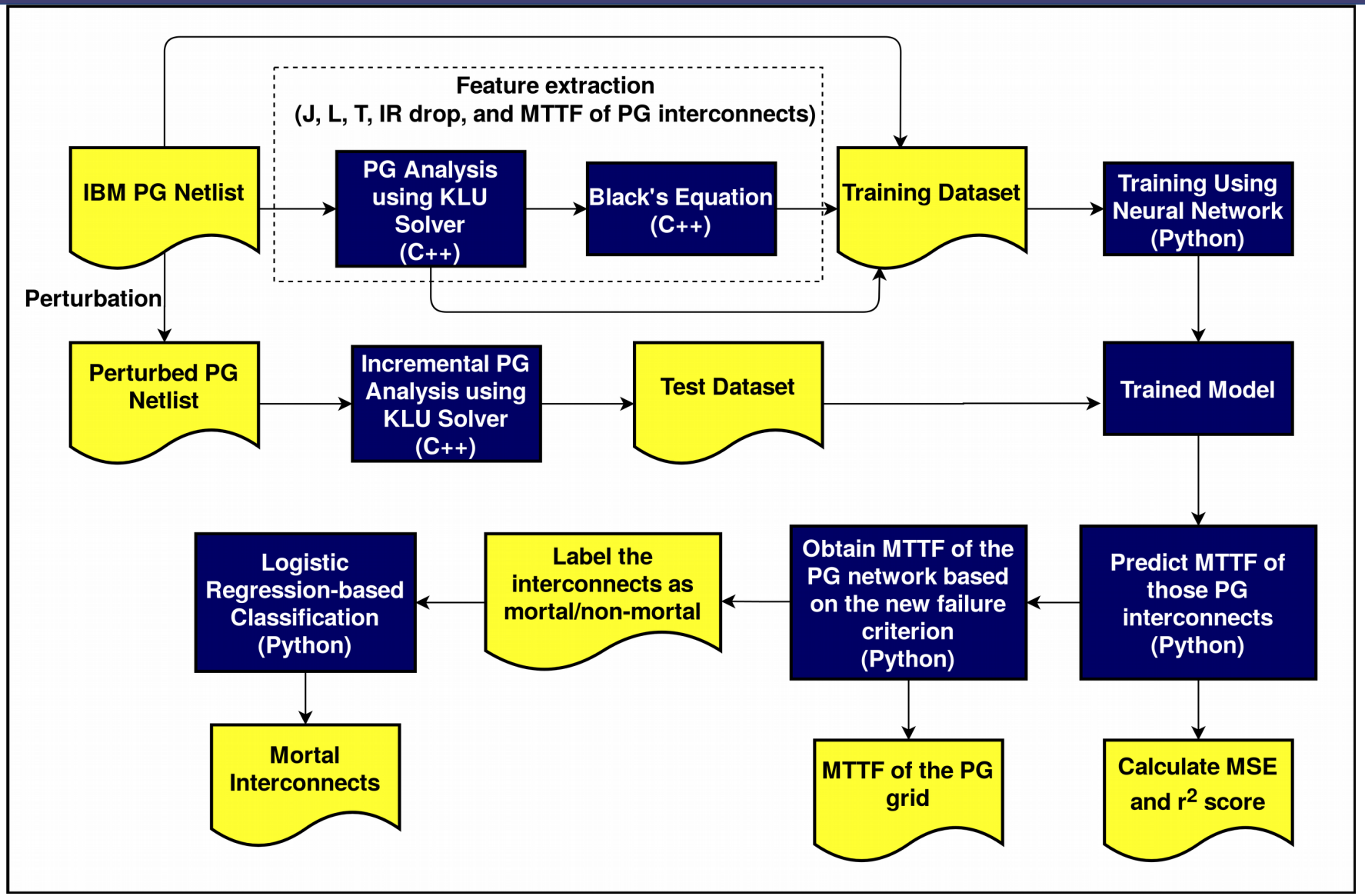
Overview of the ML Model



Feature Selection & Training Data

- r^2 score is evaluated
- Power Grid Designs Extracted from the IBM processors are employed.
- Label dataset is prepared for supervised learning.
- Input features: J, L, T, IR drop of the interconnect
- Output feature: MTTF or Mean Lifetime.

Proposed Method Flow



Experimental Results : CPU Runtime

Methods	CPU Runtime (t) (Hours)					Speedup			
	TCAD2016 [2] (t_H)	**ICCAD2017** [3] (t_{Ch})	**TCAD2018** [4] (t_C)	**IRPS2019** [5] (t_N)	**Proposed** (t_{ML})	$\frac{t_H}{t_{ML}}$	$\frac{t_{Ch}}{t_{ML}}$	$\frac{t_C}{t_{ML}}$	$\frac{t_N}{t_{ML}}$
PG Circuits									
*PG*1	0.02	0.02	0.001	0.000166	0.0001	200×	200×	10×	1.66×
*ibmpg*1	0.05	0.03	0.003	0.01000	0.0003	166.66×	100×	10×	33.33×
*ibmpg*2	0.11	0.31	0.04	0.02000	0.002	55×	155×	20×	10×
*ibmpg*3	5.83	4.27	0.41	0.07000	0.009	647.77×	610×	45.55×	7.77×
*ibmpg*4	14.71	6.81	2.31	0.11000	0.007	2101.42×	972.85×	330×	15.71×
*ibmpg*5	0.69	0.25	0.06	0.03000	0.006	115×	41.66×	10×	5×
*ibmpg*6	1.75	2.07	0.79	0.23330	0.009	194.44×	230×	87.77×	25.92×
*ibmpgnew*1	16.78	0.42	1.24	0.08000	0.013	1290.76×	32.06×	95.38×	6.15×
*ibmpgnew*2	15.32	2.60	0.43	0.06000	0.008	1915×	325×	53.75×	7.50×
*PG*2	10.94	1.12	1.06	0.10166	0.010	1094×	112×	106×	10.06×
*PG*3	-	-	-	0.13666	0.04200	-	-	-	3.25×
*PG*4	-	-	-	0.25666	0.10100	-	-	-	2.54×
Avg. Speedup						778×	277.85×	76.84×	10.74×

Our proposed ML-approach achieved significant speedup than all SOTA.

Experimental Results : MTTF

	MTTF (μ) (years)				
Methods	**TCAD2016 [2]** (μ_H)	**ICCAD2017 [3]** (μ_{Ch})	**TCAD2018** [4] (μ_C)	**IRPS2019** [5] (μ_N)	**Proposed** (μ_{ML})
PG Circuits					
*PG*1	14.01	6.10	8.51	6.5	13.25
*ibmpg*1	12.55	6.50	10.91	7.0	12.10
*ibmpg*2	18.75	6.78	10.11	12.1	12.55
*ibmpg*3	31.96	6.66	9.95	6.7	12.25
*ibmpg*4	33.39	9.83	11.95	16.7	17.48
*ibmpg*5	25.16	6.54	6.63	6.3	10.33
*ibmpg*6	19.87	9.53	11.96	11.2	12.41
*ibmpgnew*1	25.96	13.24	11.64	13.2	14.56
*ibmpgnew*2	21.80	5.72	6.72	7.3	13.24
*PG*2	17.85	8.32	9.32	10.3	11.21
*PG*3	-	-	-	7.2	10.51
*PG*4	-	-	-	6.8	8.47

Accuracy wise our ML approach is the closest to accurate method of TCAD2016, and better than all other SOTA results.

Fourth Contribution Summary

- ML approach can predict EM aging in PG design.
- MTTF value compared with accurate model, and compared to the other SOTA models.
- Proposed ML framework is faster than all SOTA models.
- Speeding up the MTTF prediction process helps in overall design sign-off time

Conclusions and Future works

- A fast and more effective power grid analysis technique is proposed, reducing the solving time of the circuit.
- We also attempt to design the power grid interconnects more efficiently by obtaining an optimum trade-off.
- This study includes machine learning techniques for power grid design and Electromigration-aware aging prediction of the power grid network.
- Works of the thesis will help the power grid designer to obtain an initial idea of different design metrics and to handle the reliability issues in the process of designing cost-effective as well as reliable chip.