

# Two-Stage Sequence Model for Maximum Throughput in Cluster Tools

January 21-23, 2021  
SAMI 2021

*Taehee Jeong<sup>1</sup>, Kunj Parikh<sup>1</sup>, Raymond Chau<sup>2</sup>, Chung Ho Huang<sup>2</sup>,  
Henry Chan<sup>2</sup>, and Hyeran Jeon<sup>3</sup>*

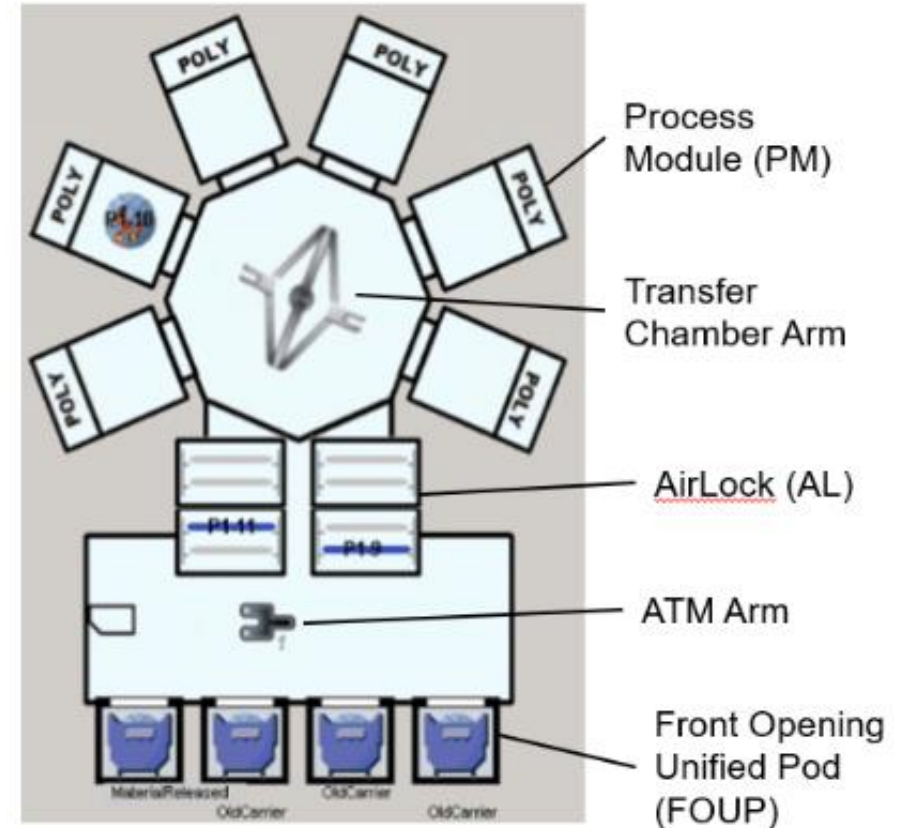
<sup>1</sup>San Jose State University

<sup>2</sup>Lam research

<sup>3</sup>University of California, Merced

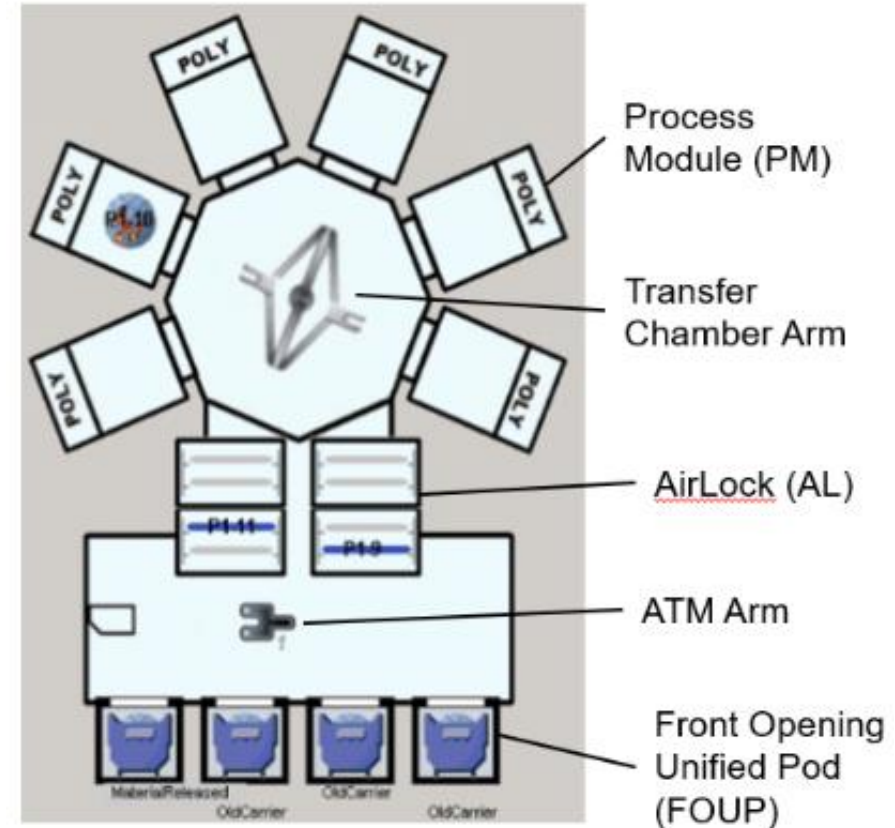
# Cluster tool

- Semiconductor wafer processing tool
- It processes a batch of wafers with a programmed scheduler
- It consists of multiple process modules (**PMs**) that process wafers in parallel
- To be processed in PMs, wafers are moved between wafer containers (**FOUP**) and vacuumed buffer (**AirLock**) or between AirLock and PM via robotic arms.



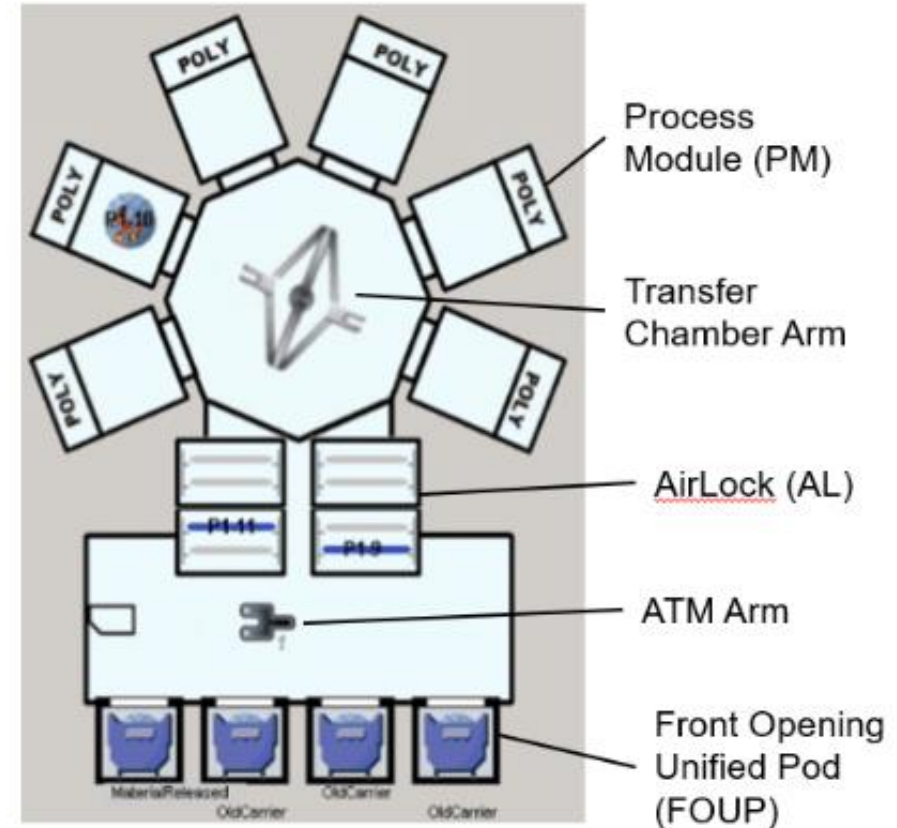
# Process module (PM)

- In a PM, a wafer is processed according to a programmed **recipe time**
- When a PM finishes a wafer processing, it spends **WAC time** to remove residual chemicals and impurities formed inside the PM before processing another wafer



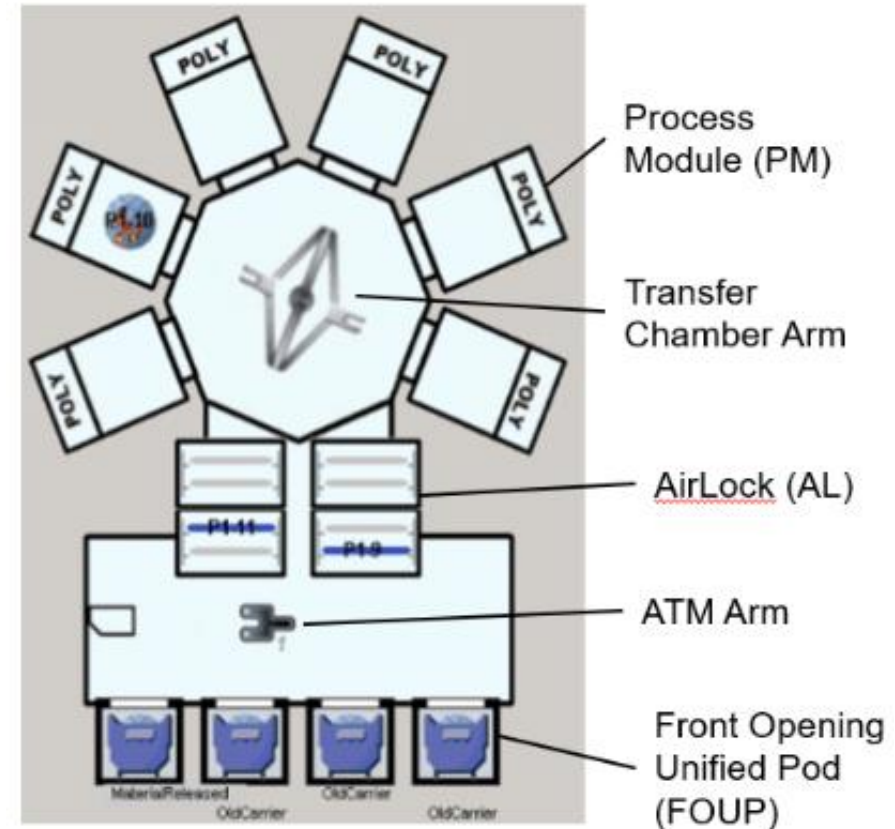
# Scheduling

- Scheduling
  - determines the timing and operations of robotic arms
- Factors
  - recipe time, WAC time, arm transport time, the availability of the PMs and AirLocks.
- Not deterministic
  - various recipe times, mechanical operational latency of robotic arms, unexpected runtime issues



# look-ahead timing (JIT1)

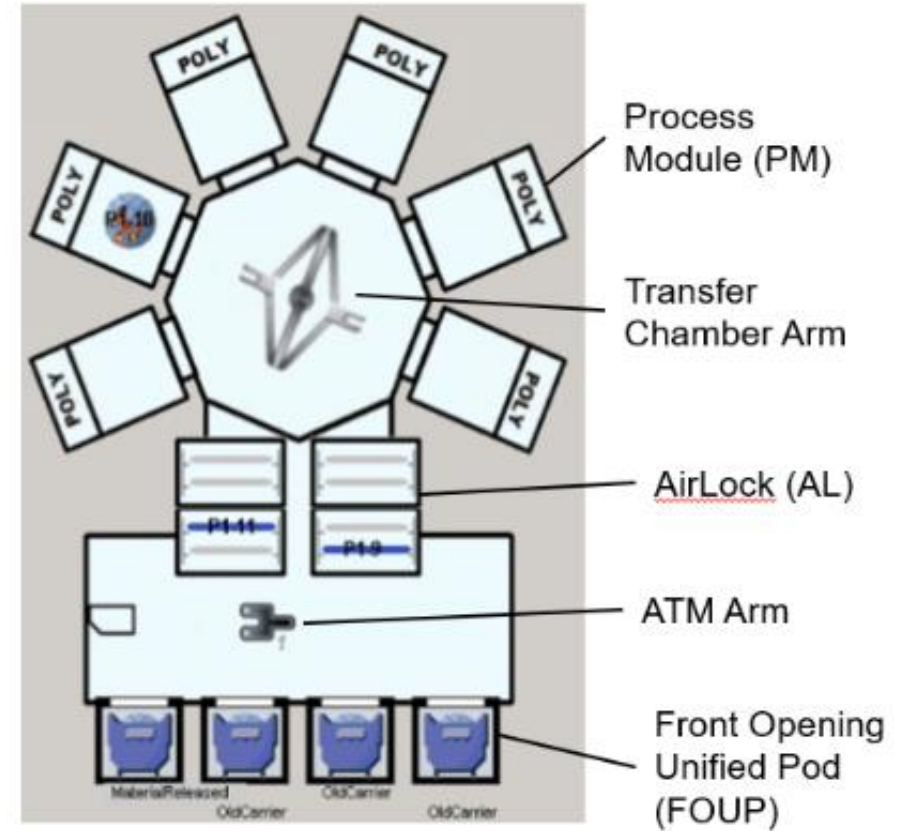
- Scheduling parameter
- It determine scheduling look-ahead timing
- When JIT1 is N, the scheduler initiates the picking of a wafer from AirLocks N seconds before the end of WAC time
- It reduces the overhead of the transportation time of the transfer chamber arm





# Throughput

- defined as number of wafers processed in an hour
- critical factor that determine the productivity of cluster tool

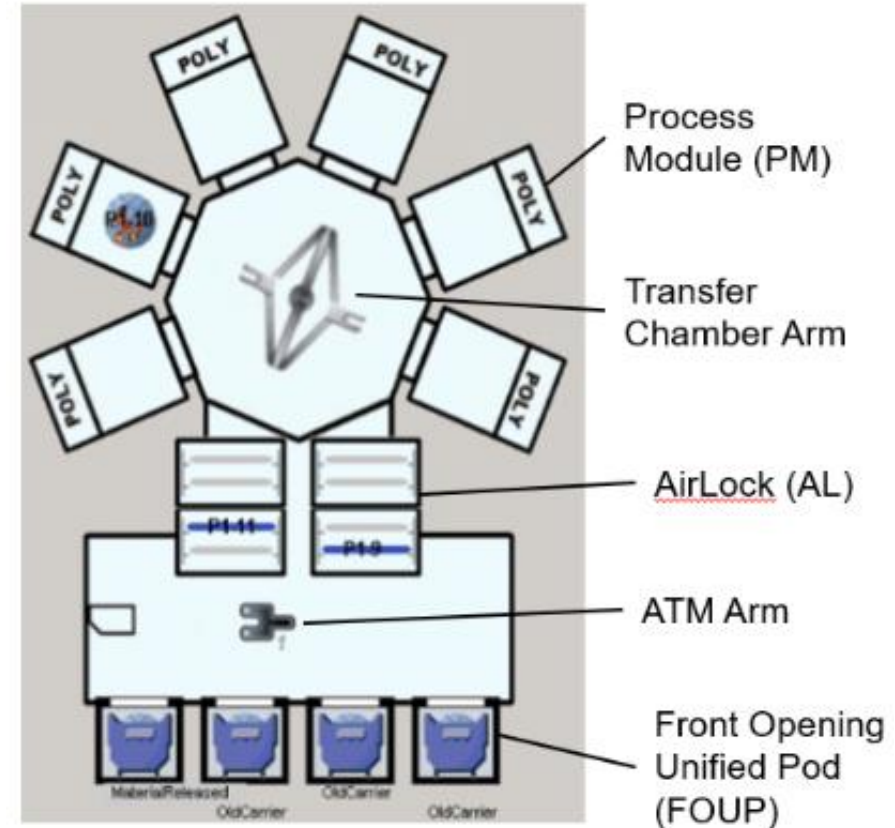


# Goal

Predict optimum look-ahead timing values (JIT1) to produce maximum throughput with a given configuration

# Data Collection of Throughput

- After completing the operation with the 25 wafers, one throughput data point is generated.
- Cluster tool simulator is used, which is developed and used by Lam Research.
- Simulator runs the same number of steps that the real cluster tool may take and generates the estimated throughput with the given input configurations



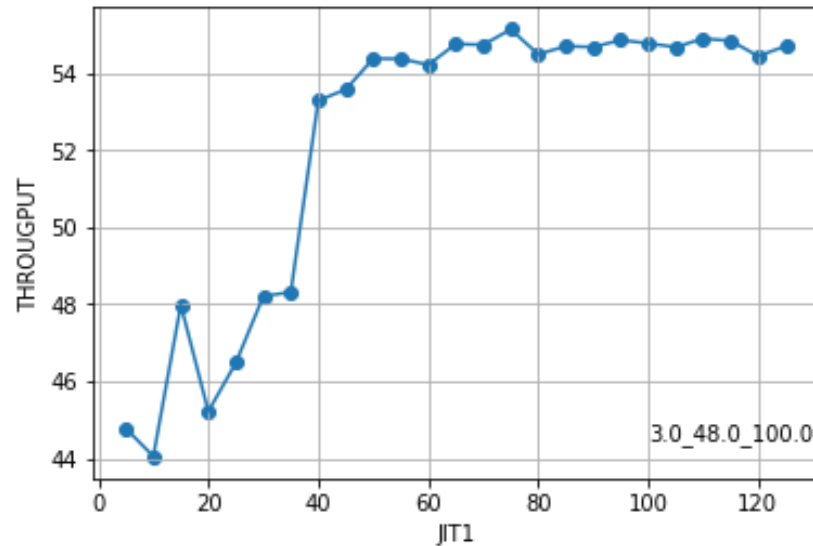


# Input Data Configuration

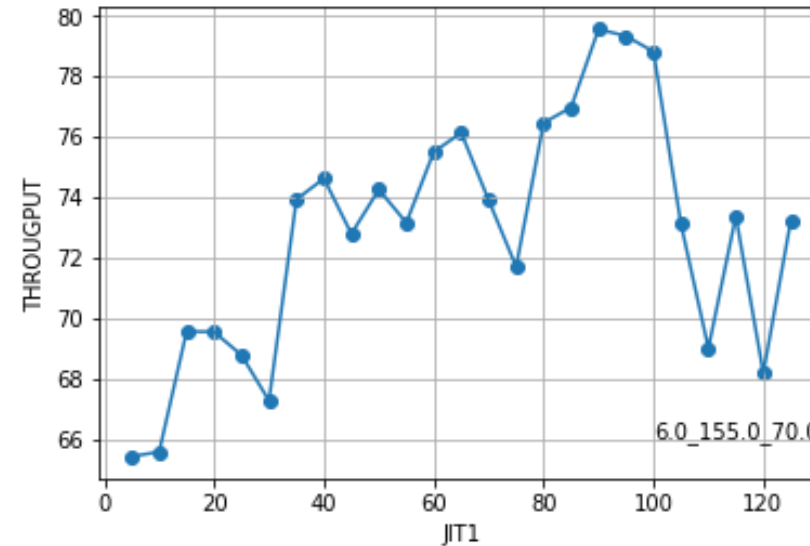
- The number of PMs: 3 ~ 6
- Recipe times: 44 ~ 218
- WAC times: 20 ~ 100
- 780 combinations → most popular 216
- JIT1 values: 5 ~ 125 with interval of 5

Parameters	Values
Number of PMs	3, 4, 5, 6
Recipe Time	44, 48, 85, 95, 105, 115, 125, 130, 135, 145, 155, 162, 218
$WAC^{TM}$ Time	20, 30, 31, 34, 40, 45, 48, 50, 53, 55, 60, 70, 71, 94, 100
JIT1	25 values from 5 to 125 with interval of 5
Total Data Points	5400

# Behavior of Throughput verse JIT1

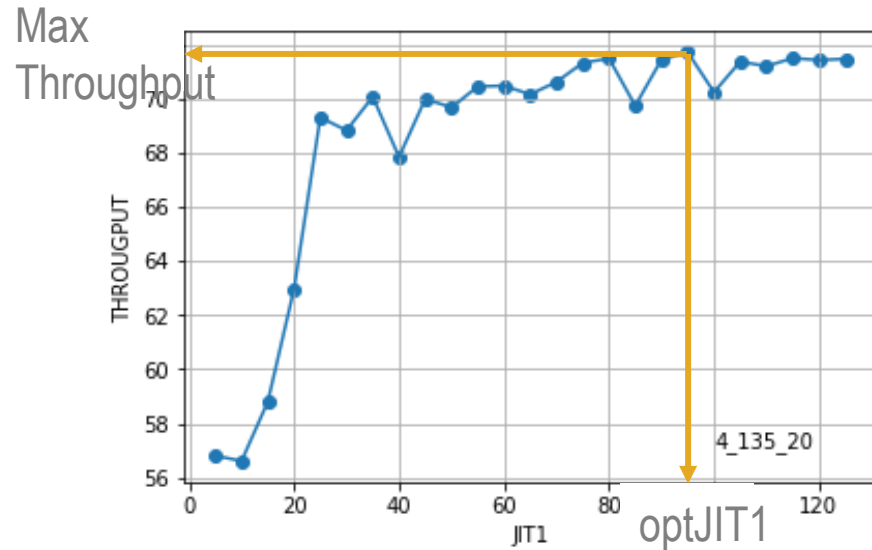


Number of PMs: 3  
Recipe time: 48  
WAC time: 100

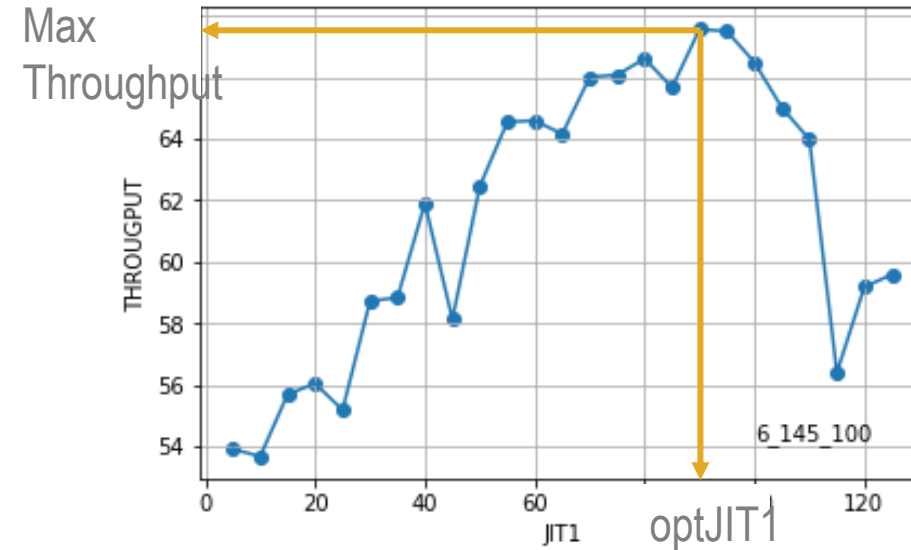


Number of PMs: 6  
Recipe time: 155  
WAC time: 70

# Fining Max Throughput & optimum JIT1



Number of PMs: 4  
Recipe time: 135  
WAC time: 20



Number of PMs: 6  
Recipe time: 145  
WAC time: 100

# metrics

- Mean Squared Error (MSE)

$$MSE = \frac{1}{m} \sum_{i=1}^m (y - \hat{y})^2$$

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y - \hat{y})^2}$$

- Pearson correlation coefficient

$$pc = \frac{\sum_{i=1}^m (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^m (x - \bar{x})^2} \sqrt{\sum_{i=1}^m (y - \bar{y})^2}}$$

- Intersection-Over-Union (IOU)

$$IOU = \frac{\text{true positive}}{\text{true positive} + \text{false positive} + \text{false negative}}$$

# Hyperparameter optimization for Deep Neural Network

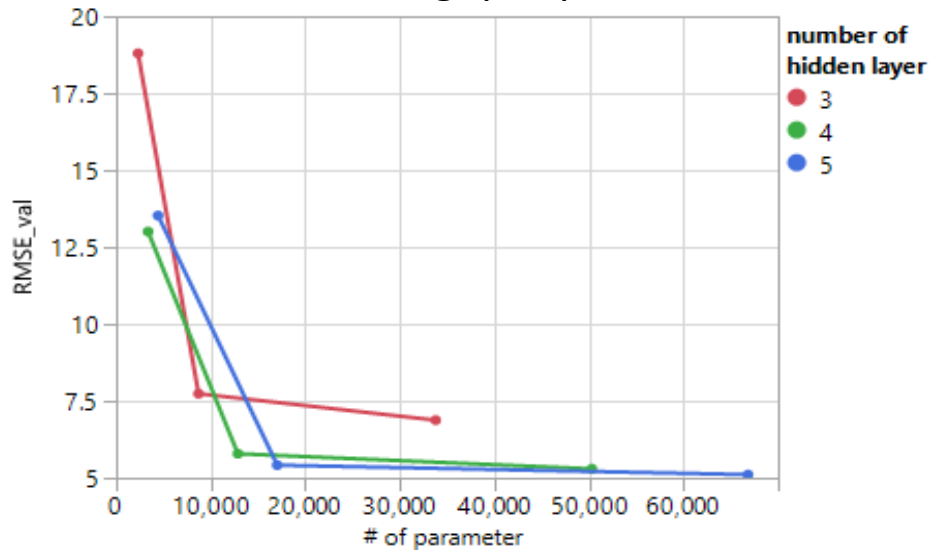
- Input variables:
  - PM count
  - RECIPE time
  - WAC<sup>TM</sup> time

- Output prediction:
  - Maximum throughput
  - Optimum JIT1 (look-ahead time)

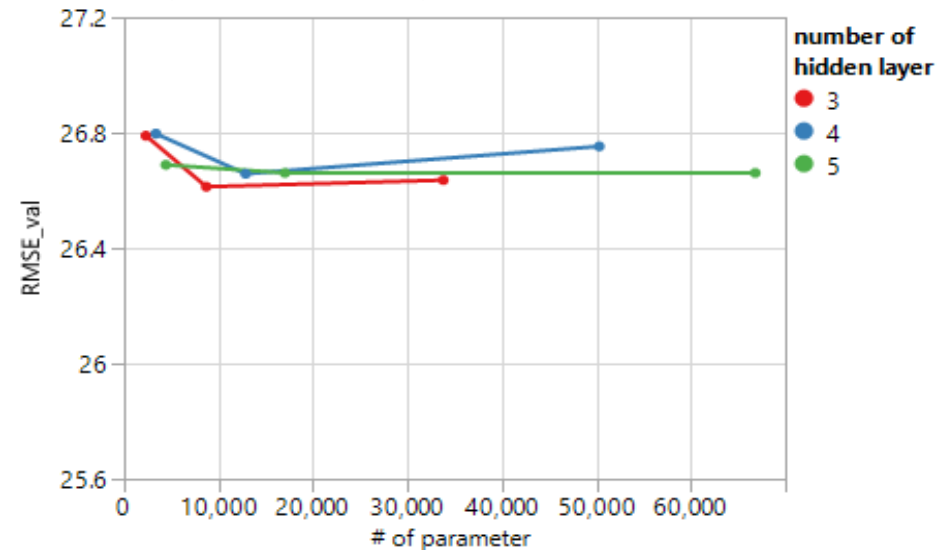
- Training conditions:
  - number of epochs : 100
  - Optimizer: Adam
  - learning rate: 0.0001
  - activation function: ReLU
  - 5fold-cross validation

- Hyperparameter optimization
  - Number of hidden layer
  - Number of neurons per layer

Maximum throughput prediction



Optimum JIT1 prediction

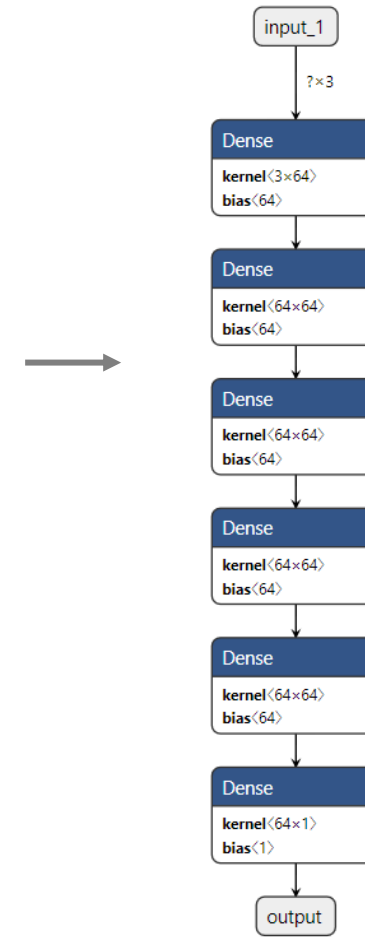


# Hyperparameter optimization for Deep Neural Network

5-cross-validation

result			Max. Throughput		Opt. JIT1	
Hidden layers	Neurons per layer	Num of params	Train RMSE	Valid RMSE	Train RMSE	Valid RMSE
3	32	2,305	19.47	18.80	26.19	26.79
4	32	3,361	12.40	13.01	26.26	26.80
5	32	4,417	13.40	13.53	26.13	26.69
3	64	8,705	7.22	7.74	26.03	26.61
4	64	12,865	6.39	5.79	26.03	26.66
5	64	17,025	5.92	5.42	26.02	26.66
3	128	33,793	6.48	6.88	25.99	26.64
4	128	50,305	5.89	5.30	26.02	26.75
5	128	66,817	5.61	5.11	26.01	26.66

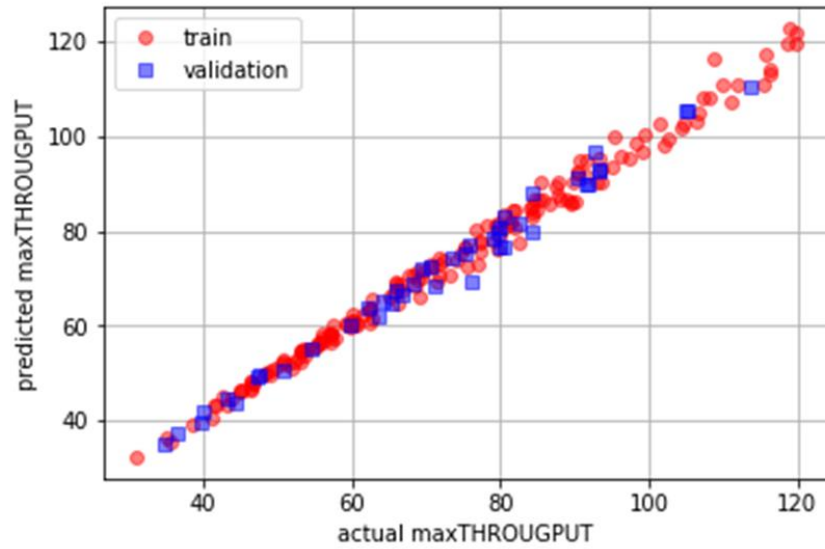
Hidden layer: 5  
N of neuron: 64





# Prediction using DNN models

Max Throughput prediction

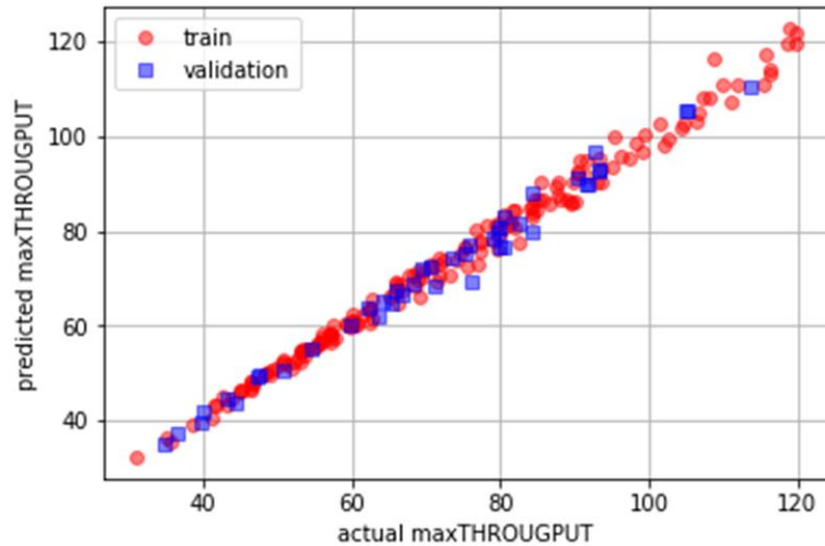


Pearson correlation coefficient

- 0.9956 for train data
- 0.9940 for validation data

# Prediction using Deep Neural Network (DNN) models

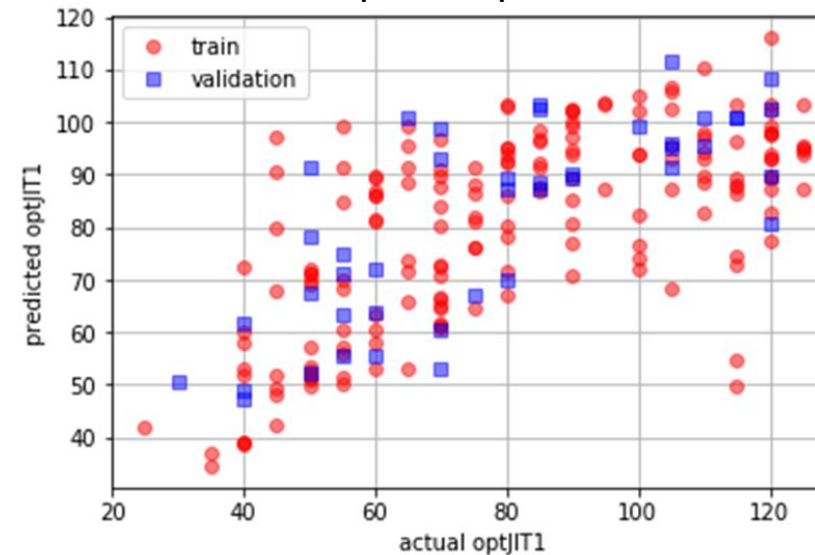
Max Throughput prediction



Pearson correlation coefficient

- 0.9956 for train data
- 0.9940 for validation data

opt JIT1 prediction



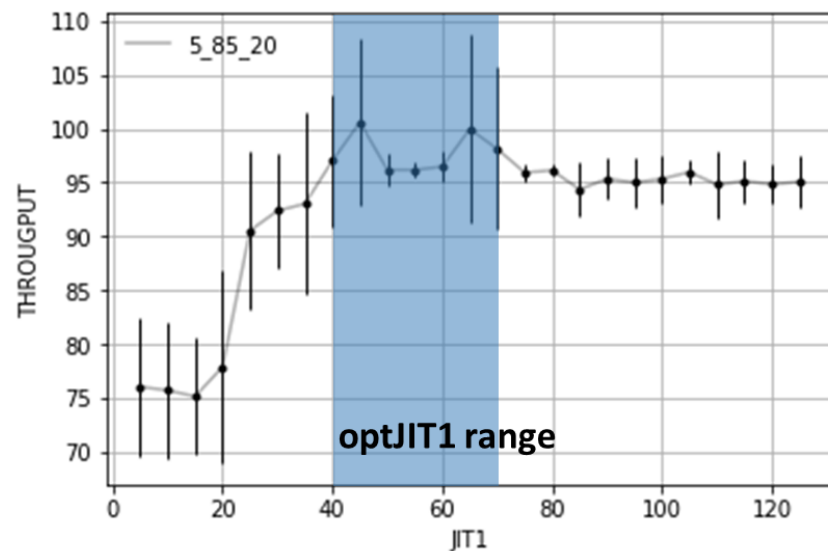
Pearson correlation coefficient

- 0.6473 for train data
- 0.7598 for validation data

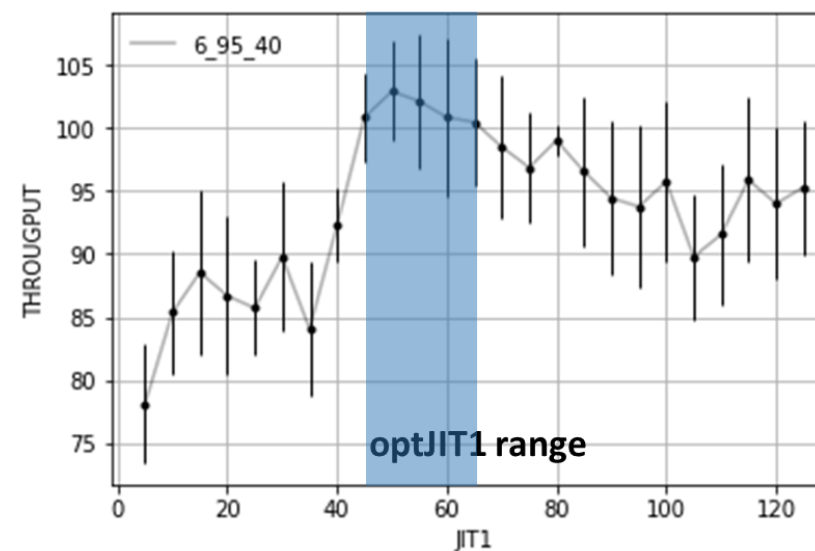
# Optimum JIT1 range

The poor accuracy of JIT1 prediction is sourced from the unclear distinct opt JIT1 value

Optimum JIT1 range based on highest throughput value and its neighboring data points within 5% span

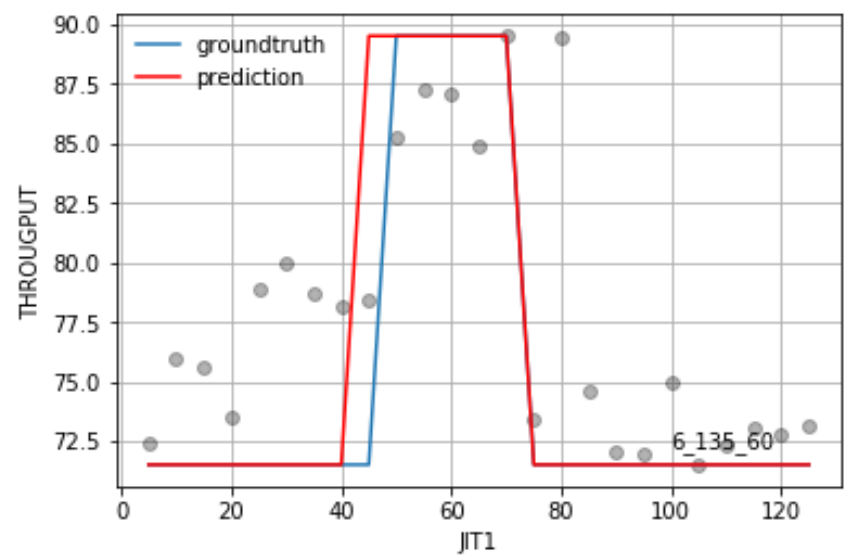


Number of PMs: 5  
Recipe time: 85  
WAC time: 20



Number of PMs: 6  
Recipe time: 95  
WAC time: 40

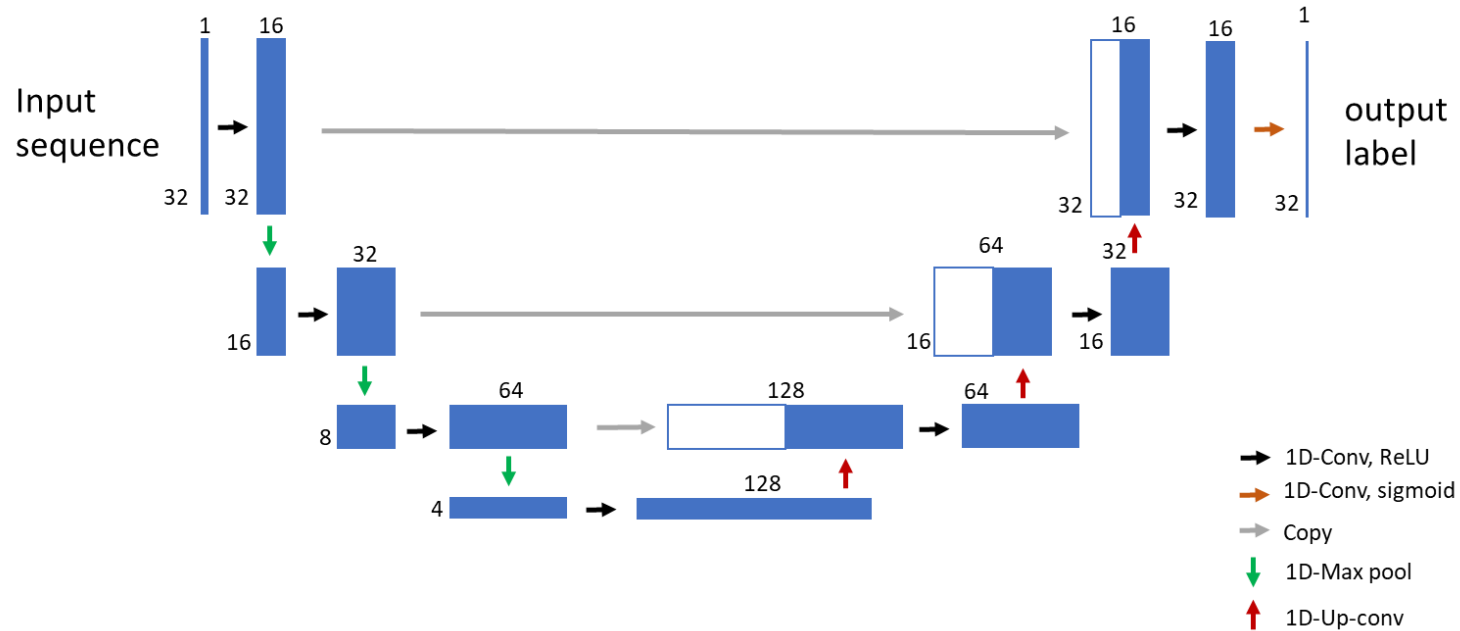
# Prediction of optimum JIT1 range



INPUT	Throughput	0	0	...	0	72	76	76	...	78	85	87	87	85	90	73	89	75	72	...	73	73
(OUTPUT of 1st Stage)																						
OUTPUT	JIT1 Label	0	0	...	0	0	0	0	...	0	1	1	1	1	1	0	0	0	0	...	0	0

↓ Highest Throughput Region Detection

# Semantic Segmentation network to predict opt JIT1 range



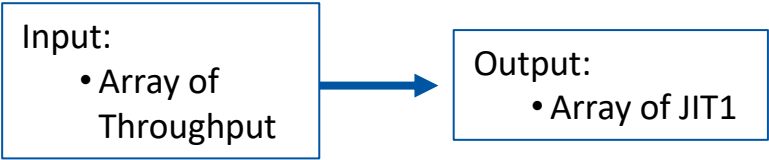
# 2<sup>nd</sup> stage network

Need a network to provide throughput

2<sup>nd</sup> stage

INPUT	Throughput	0	0	...	0	72	76	76	...	78	85	87	87	85	90	73	89	75	72	...	73	73	
(OUTPUT of 1st Stage)																							
		<div>Highest Throughput Region Detection</div>																					
OUTPUT	JIT1 Label	0	0	...	0	0	0	0	...	0	1	1	1	1	1	0	0	0	0	...	0	0	

Highest Throughput Region Detection





# 1<sup>st</sup> stage network

1<sup>st</sup> stage

		Zero Padding																																	
INPUT	Index	0	1	...	6	7	8	9	...	15	16	17	18	19	20	21	22	23	24	...	30	31													
	JIT1	0	0	...	0	5	10	15	...	45	50	55	60	65	70	75	80	85	90	...	120	125													
	N of PM	0	0	...	0	6	6	6	...	6	6	6	6	6	6	6	6	6	6	...	6	6													
	Recipe Time	0	0	...	0	135	135	135	...	135	135	135	135	135	135	135	135	135	135	...	135	135													
	WAC Time	0	0	...	0	60	60	60	...	60	60	60	60	60	60	60	60	60	60	...	60	60													
OUTPUT	Throughput	0	0	...	0	72	76	76	...	78	85	87	87	85	90	73	89	75	72	...	73	73	<div><div></div></div> <div>HighLow</div>												

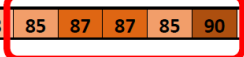

Input:

- PM count
- RECIPE time
- WAC<sup>TM</sup> time
- JIT1

Output:

- Array of Throughput

2<sup>nd</sup> stage

INPUT (OUTPUT of 1st Stage)	Throughput	0	0	...	0	72	76	76	...	78	85	87	87	85	90	73	89	75	72	...	73	73
																						
OUTPUT	JIT1 Label	0	0	...	0	0	0	0	...	0	1	1	1	1	1	0	0	0	0	...	0	0
																						

Input:

- Array of Throughput

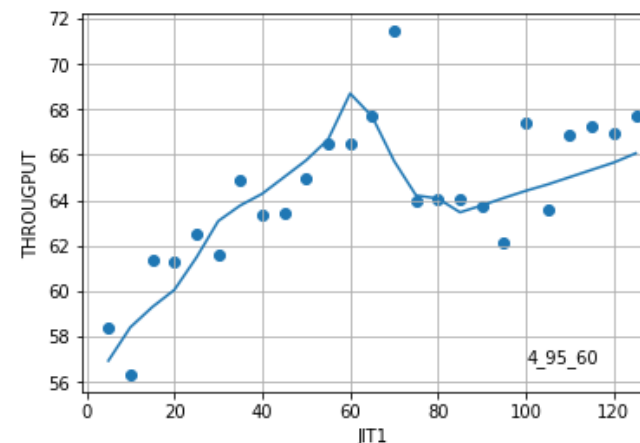
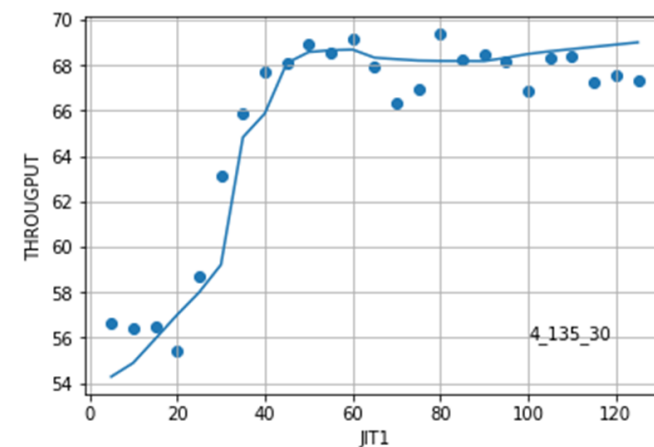
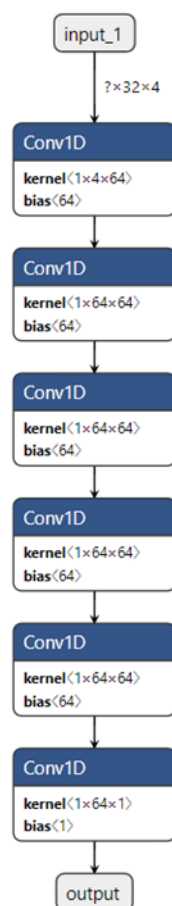
Output:

- Array of JIT1

# 1D Convolutional Neural Network to predict Throughput

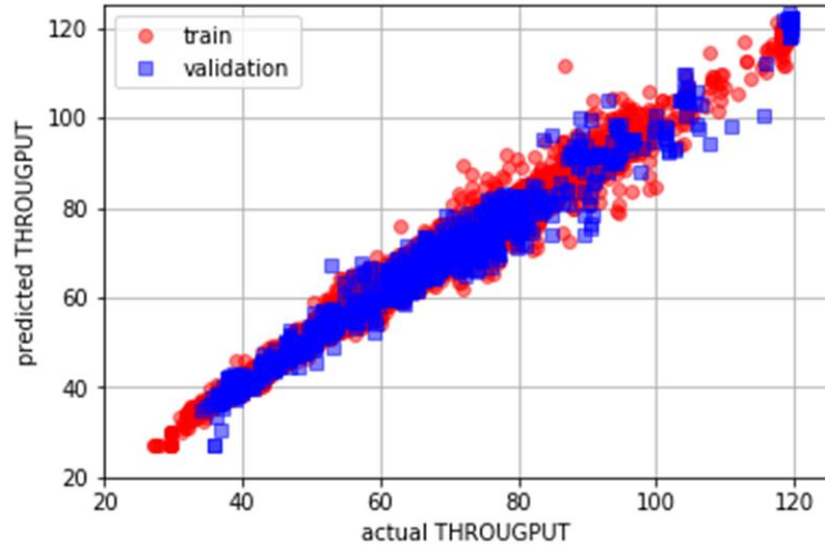
Predict an array of throughput instead of one data point

Hidden layer: 5  
N of neuron: 64



dots: actual throughput  
solid line: predicted throughput

# 1D Convolutional Neural Network to predict Throughput

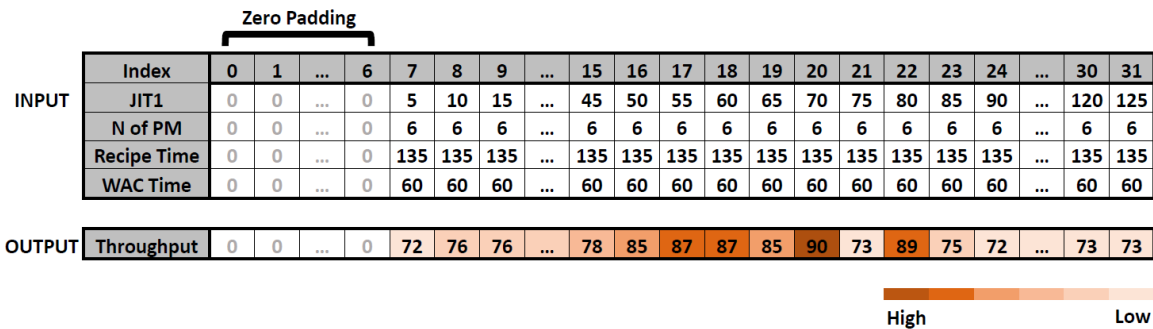


Pearson correlation coefficient

- 0.988 for train data
- 0.984 for validation data

# Two-stage model for optimum scheduling parameter prediction

1<sup>st</sup> stage



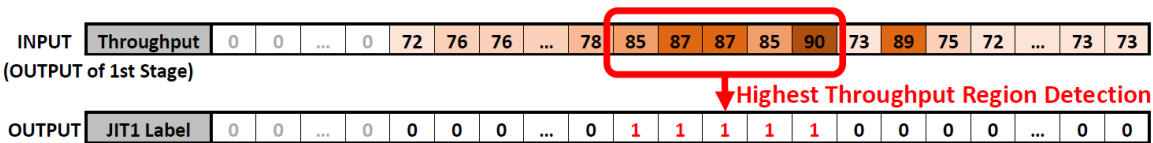
Input:

- PM count
- RECIPE time
- WAC<sup>TM</sup> time
- JIT1

Output:

- Array of Throughput

2<sup>nd</sup> stage



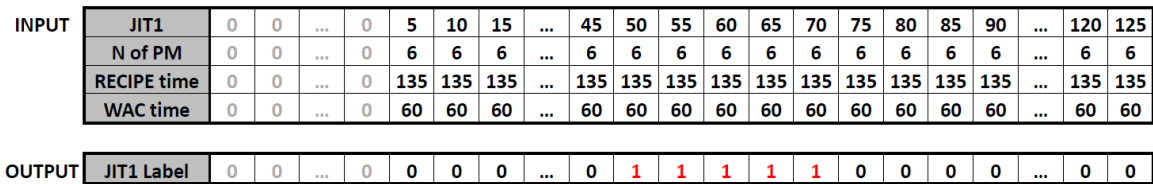
Input:

- Array of Throughput

Output:

- Array of JIT1

End-to-End



Input:

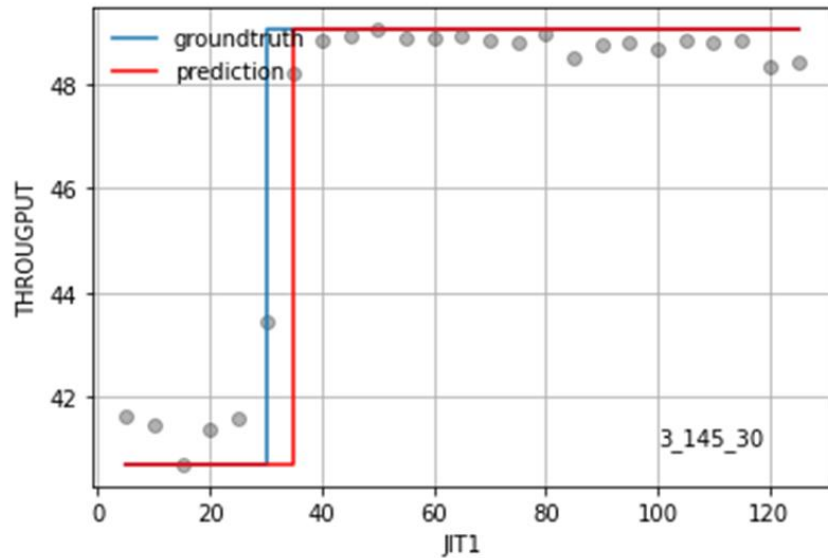
- PM count
- RECIPE time
- WAC<sup>TM</sup> time
- JIT1

Output:

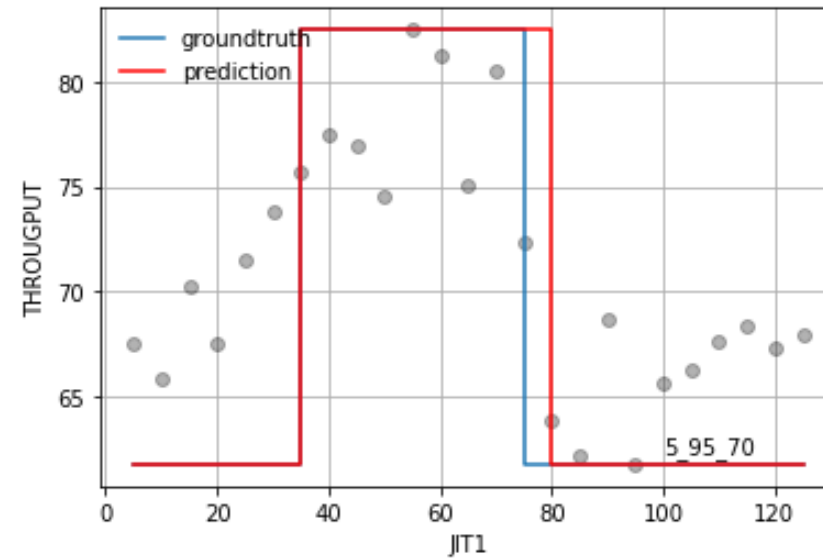
- Array of JIT1

# Opt JIT1 prediction using 2 stage network

The mean IOUs of train and validation data set are 0.783 and 0.778.



Number of PMs: 3  
Recipe time: 145  
WAC time: 30



Number of PMs: 5  
Recipe time: 95  
WAC time: 70

# Summary

- Optimizing scheduling of a cluster tool is important to produce maximum throughput.
- We propose a novel two-stage sequence model that consists of an one-dimensional convolution neural network and a semantic segmentation network.
- The proposed model effectively predicts the best scheduling parameter range for maximum throughput.

- 1<sup>st</sup> stage model: one-dimensional (1D) convolution neural network (CNN) that predicts throughput of the given series of parameters
- 2<sup>nd</sup> stage model: semantic segmentation model that identifies optimum JIT1 range for maximum throughput