DapuStor

PIPULS: Predicting I/O Patterns Using LSTM in Storage Systems

Dr. Qing Yang

Distinguished Engineering Professor

University of Rhode Island

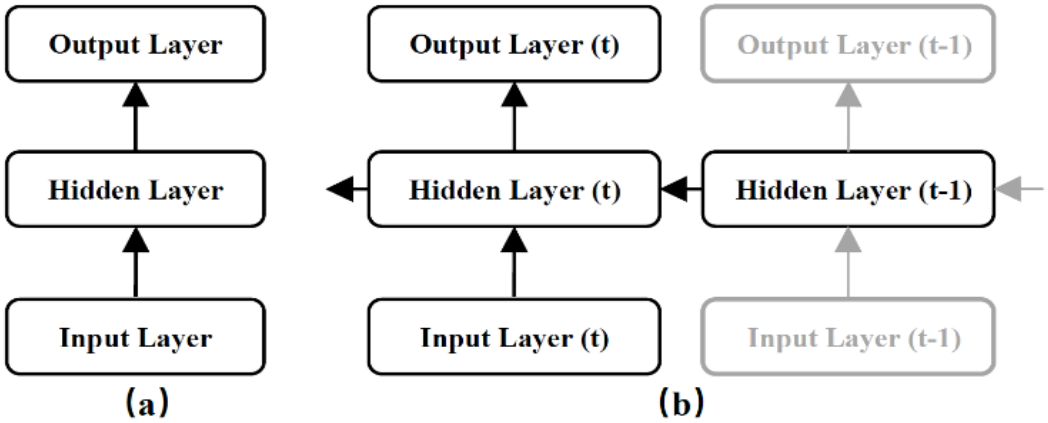
CSO, Shenzhen Dapu Microelectronics Ltd Co.

Storage I/O Patterns

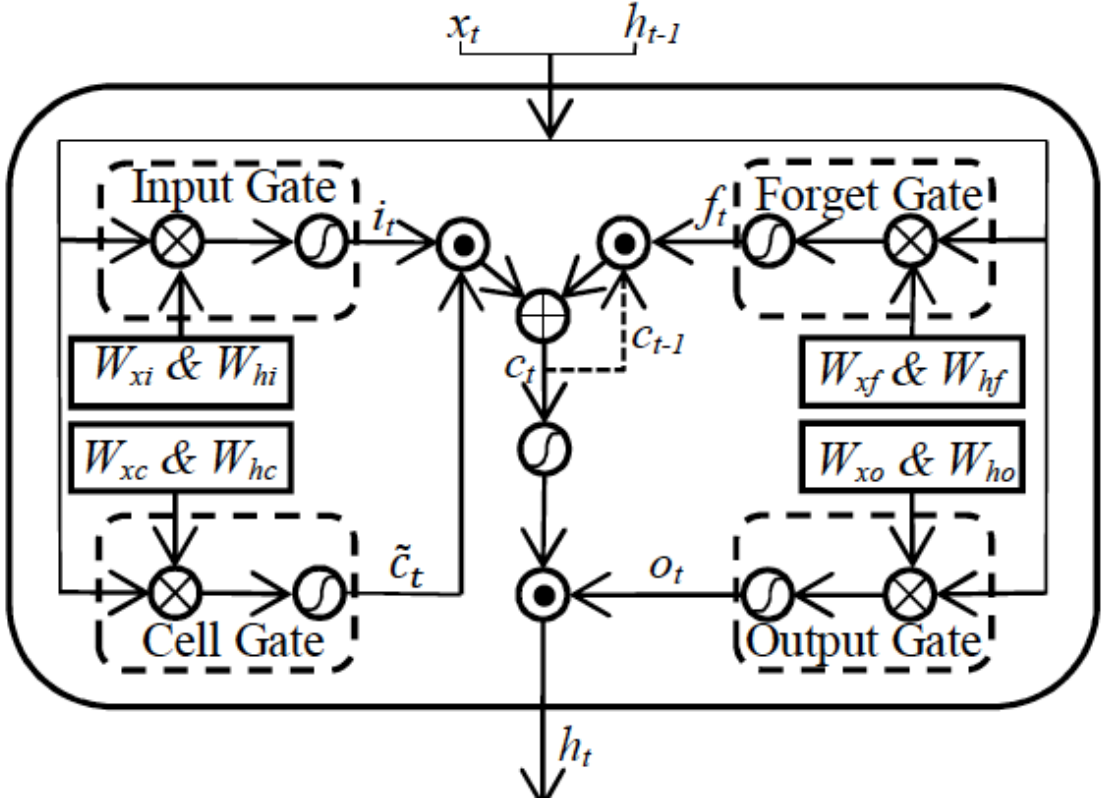
- Vary greatly across different applications
- Accurate prediction can help
  - Proper scheduling of SSD activities
  - Garbage collection while idle
  - Caching and prefetching etc.
  - For Energy Efficiency:
    - ✓ fully utilizing different power modes and fine tuning can result in energy saving

Supervised Learning Model

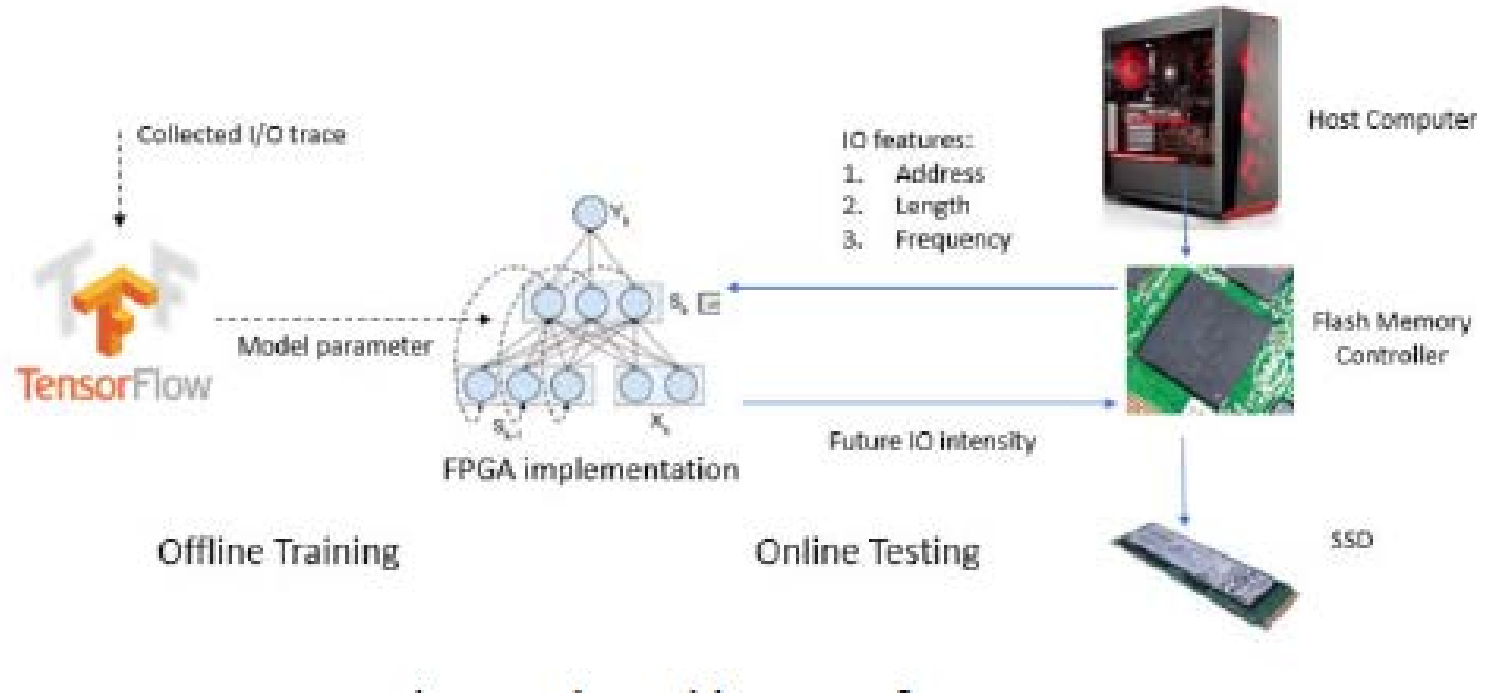
- LSTM neural network: recurrent neural network
  - Time series data: e.g.weather forecast, language modeling and speech recognition



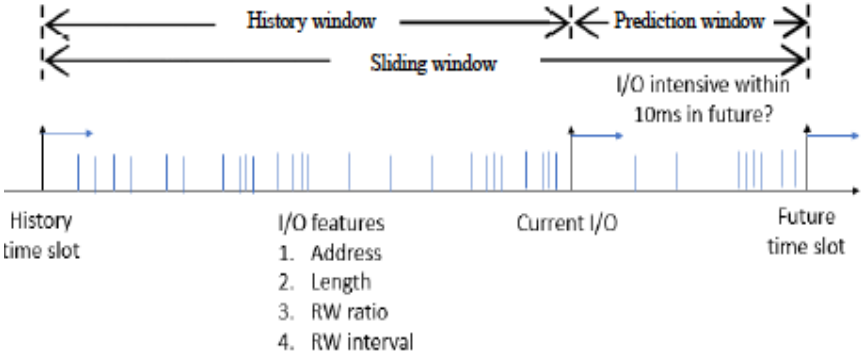
Architecture of an LSTM Cell

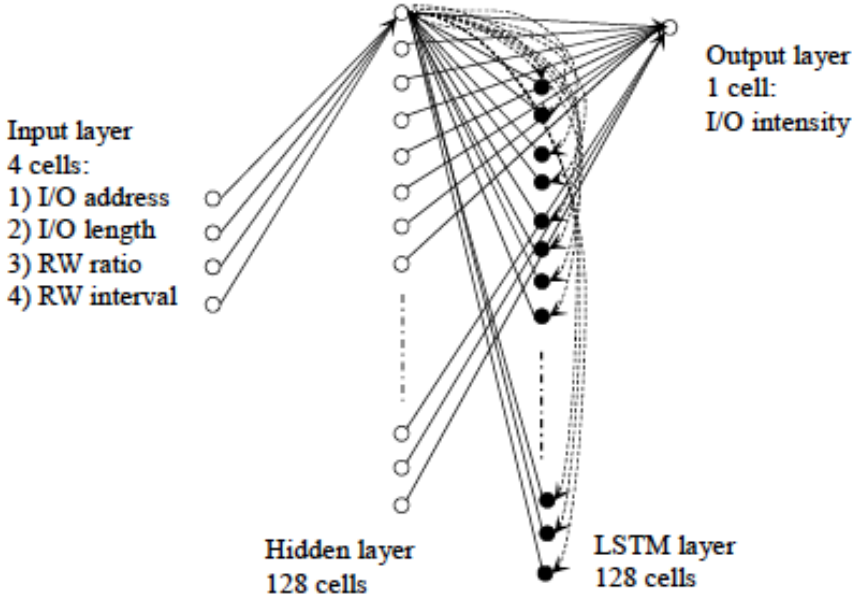


PIPLULS Architecture



Input & Output of the Model

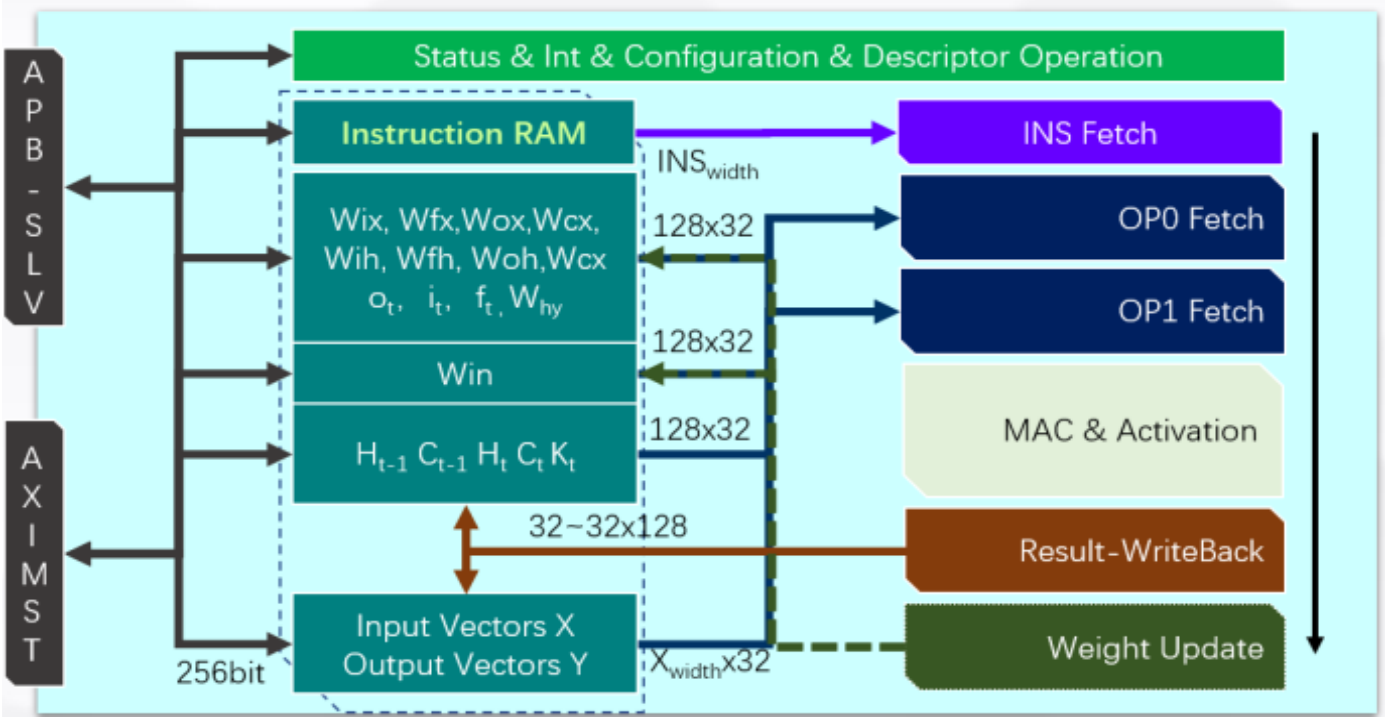




PARAMETER SIZE OF THE WEIGHT/BIAS MATRIX

	Weight matrix size	Numbers of weight matrix	Bias matrix size	Numbers of bias matrix
Input layer	128 × 4	1	128 × 4	1
Hidden layer	128 × 128	8	128 × 128	8
Output layer	128 × 1	1	128 × 1	1
Total size	264K parameters			

HARDWARE ARCHITECTURE



PREDICTION RESULTS

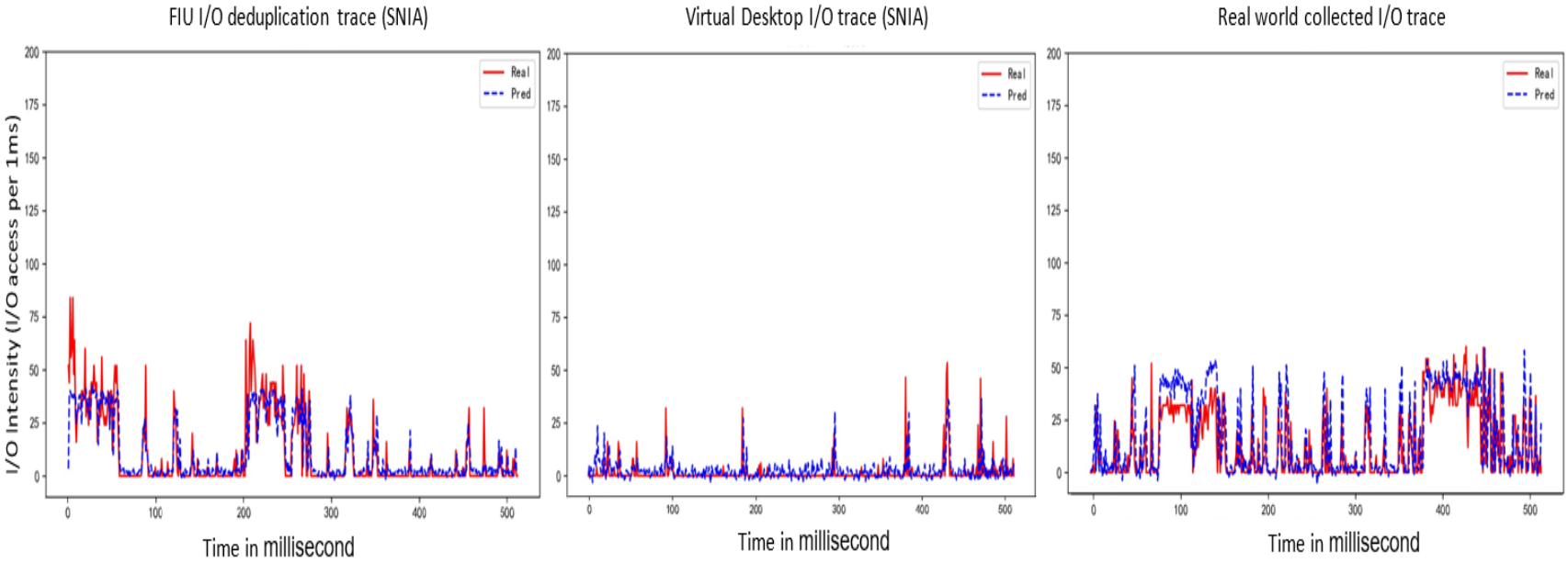
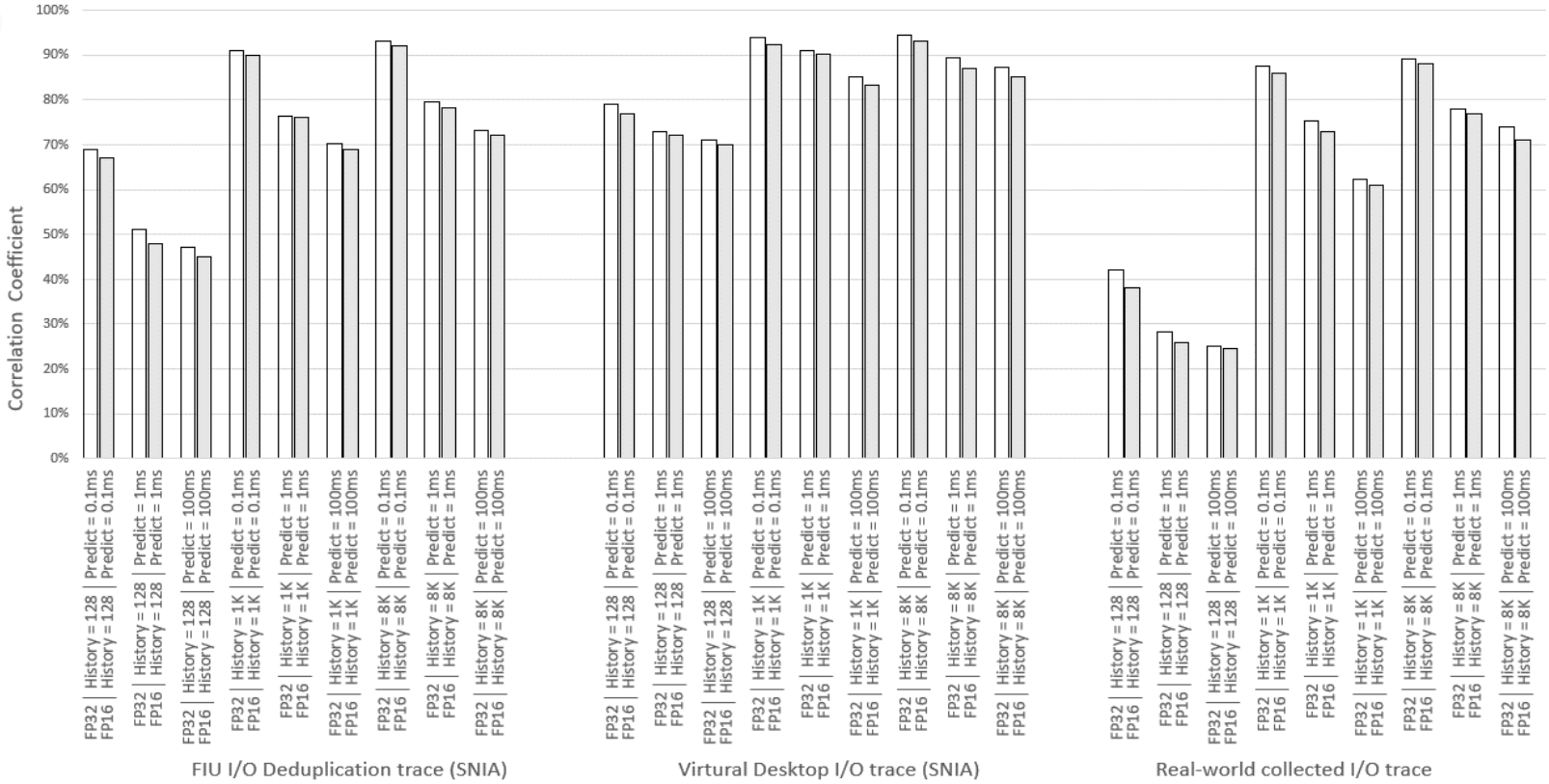


Figure 9 the prediction result of three storage I/O traces (configuration: history=*1K*, predict period=*1ms*, precision=*float 16 bit*)

PREDICTION ERROR RATES



Hardware Resource Usage

	Xilinx VU9P	PIPULS FP32	PIPULS FP16
Logic cells	2586K	2.1K	1.8K
LUT	1182K	4.1K	1.3K
Flip flop	2364K	9.5K	6.5K
DSP block	6,840	256	128
Block RAM	28MB	1MB	0.5MB

Summary and Conclusions

- A New LSTM Model
  - Predicting I/O Patterns
    - Highly Accurate
    - High Speed
- Working Prototype
  - Hardware inference
    - FPGA implementation
    - Low resource usage
  - Software Training