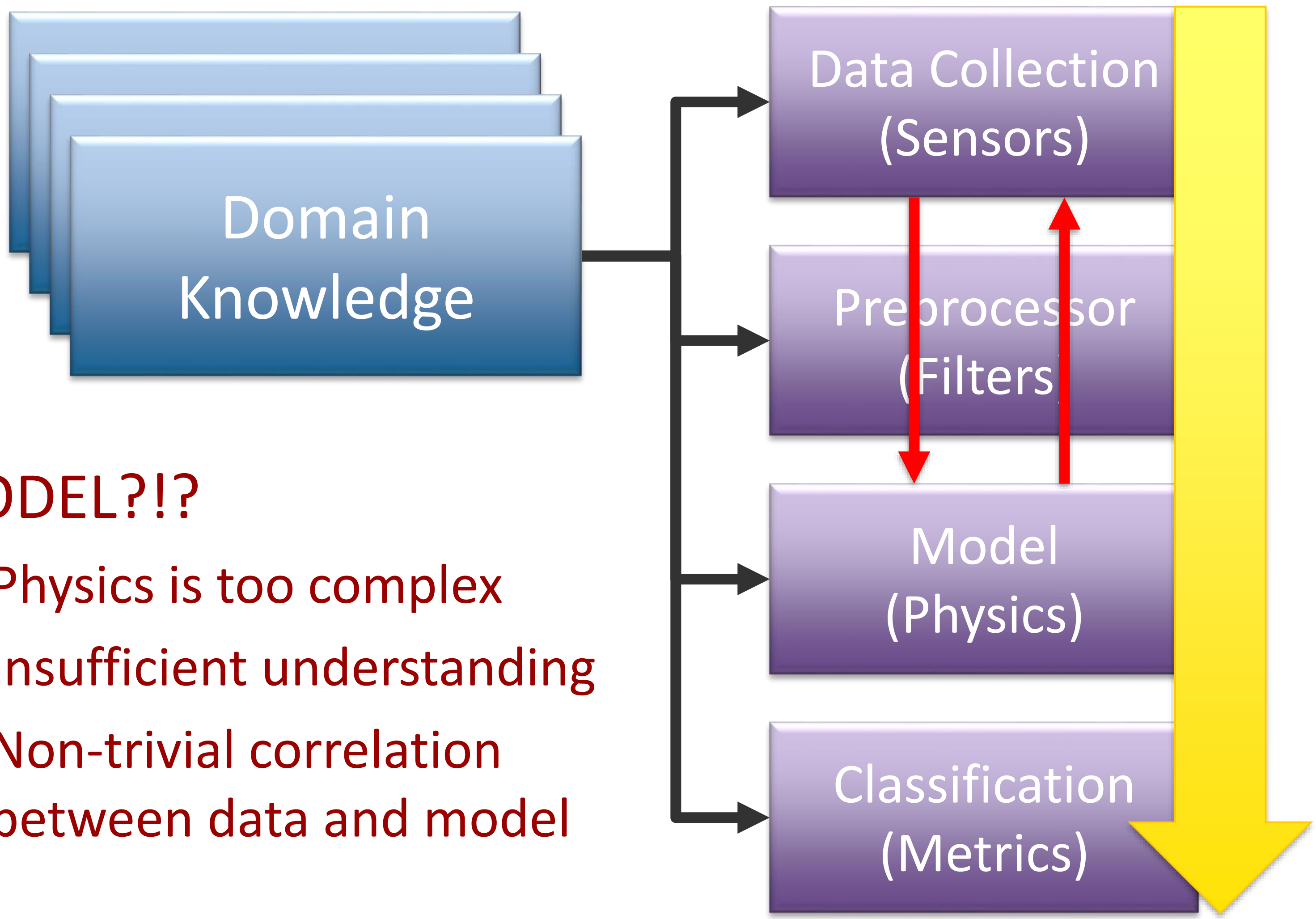
Young H. Cho, Ph.D.
University of Southern California

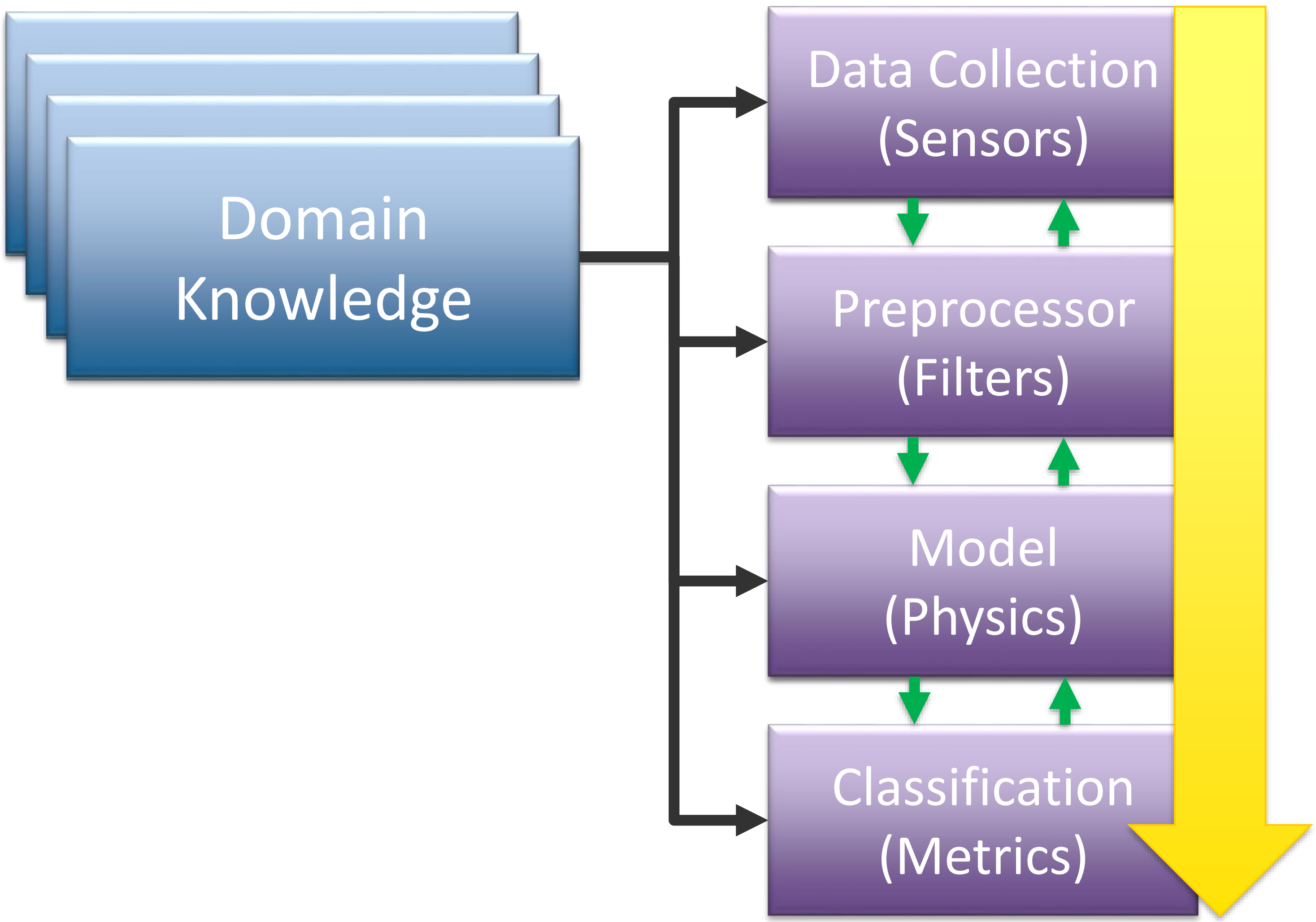
EE 542
Lecture 20: Machine Learning 2

Traditional Data Analytics



- MODEL?!?
  - Physics is too complex
  - Insufficient understanding
  - Non-trivial correlation between data and model

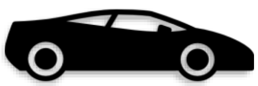
My Thoughts


Domain
Knowledge
Data Collection
(Sensors)
Preprocessor
(Filters)
Model
(Physics)
Classification
(Metrics)

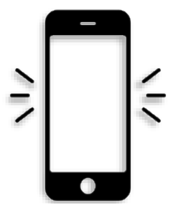
Machine Learning Approaches

	Unsupervised Learning	Supervised Learning
Discrete	Clustering	Classification
Continuous	Dimensionality Reduction	Regression

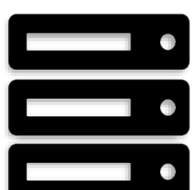
Automobile Application



- Data is collected from car's OBD2 port

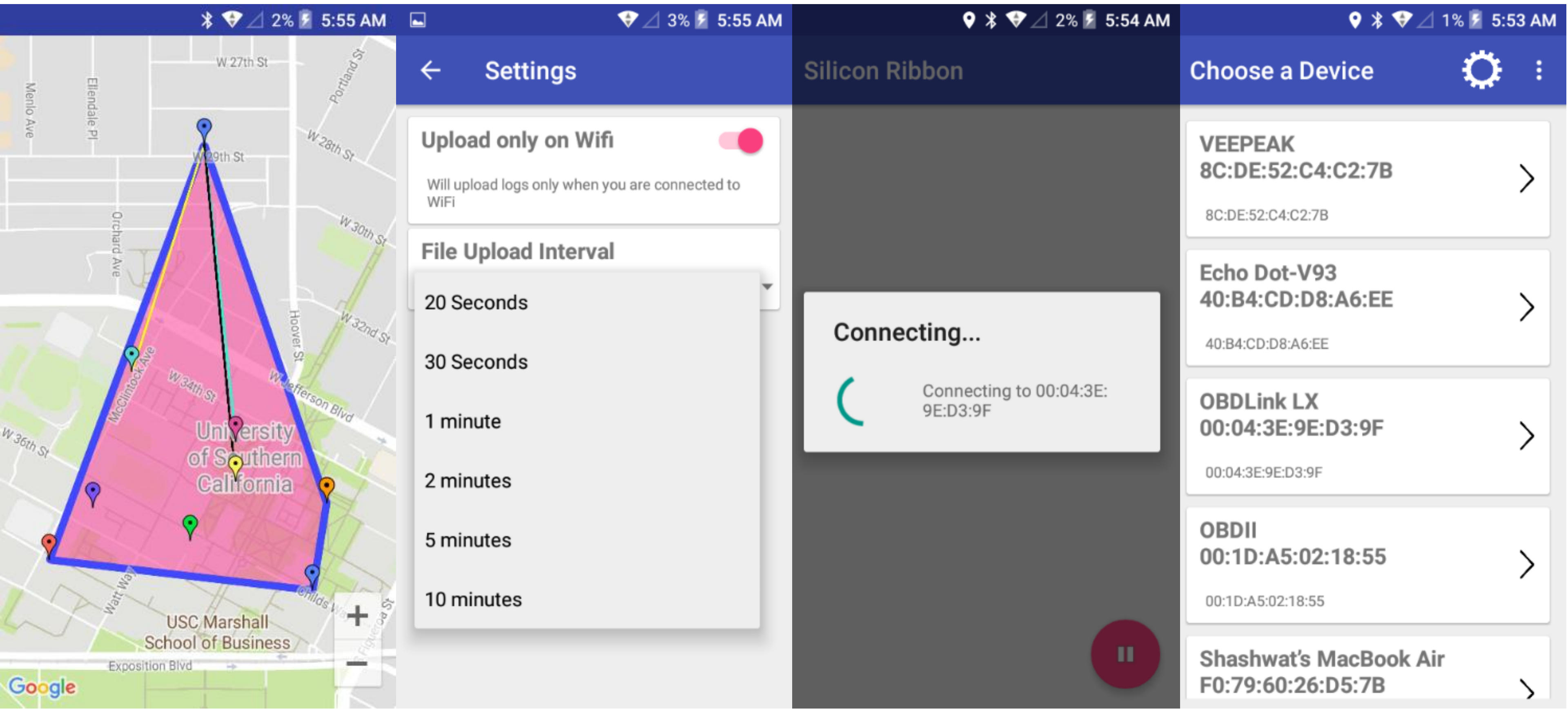


- The app receives the data over Bluetooth and creates chunks of it.
- The app sends these chunks to the server periodically.



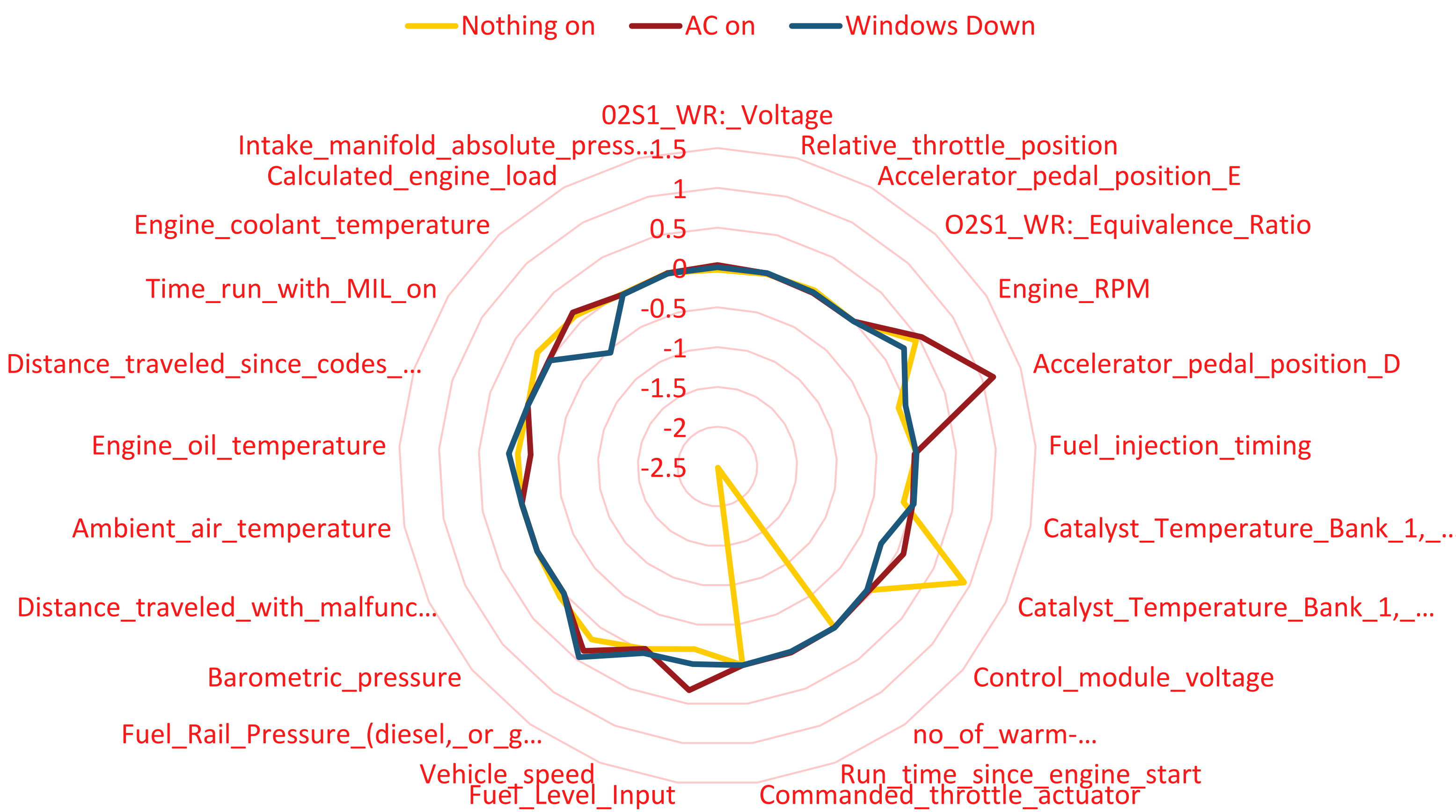
- The server receives the data in the form of binary files.
- It is then preprocessed and converted to CSV.
- The csv is then used for machine learning.

Collection and Transfer



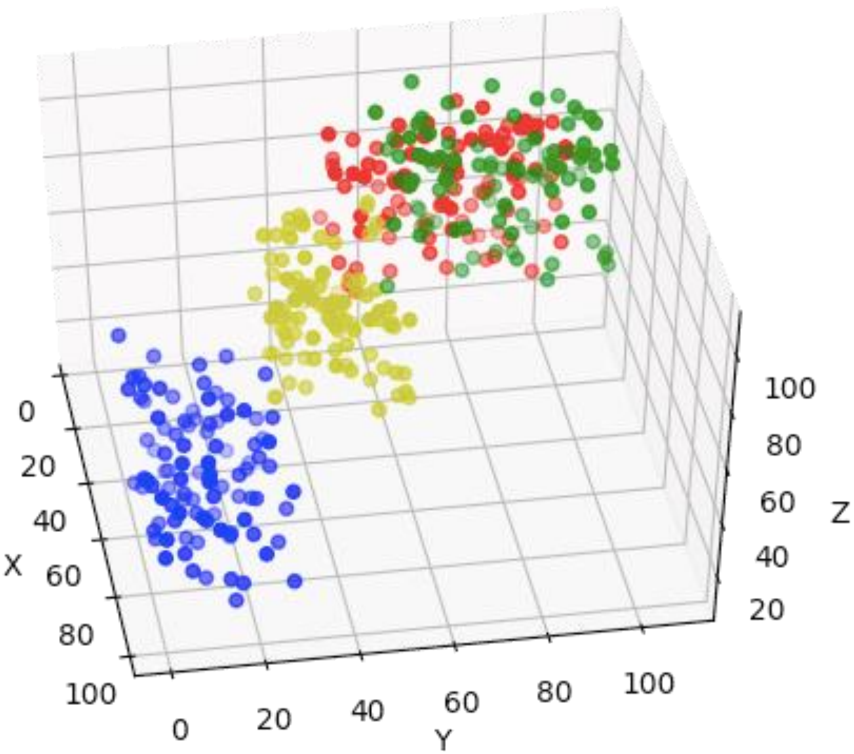
Cloud Storage

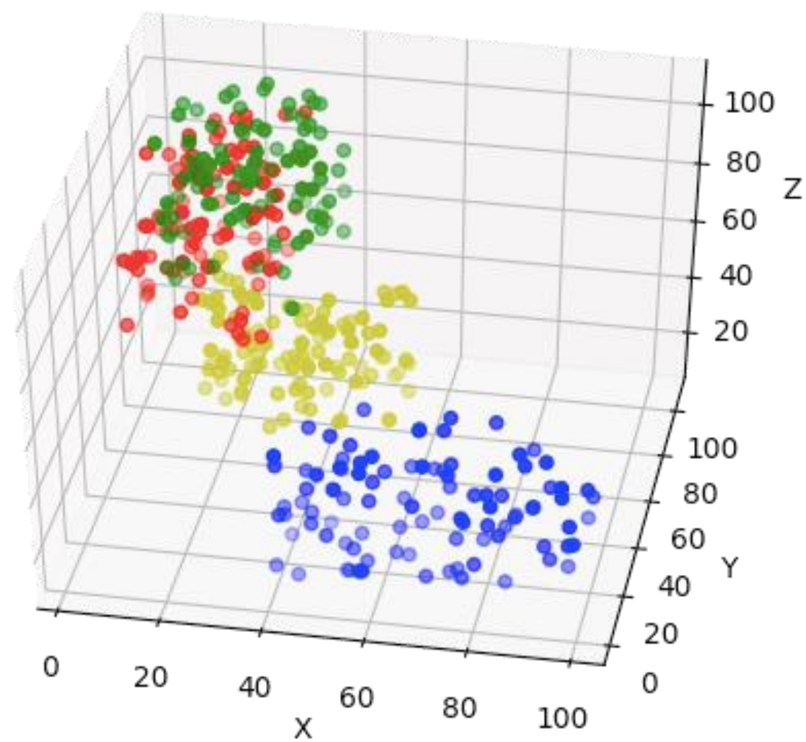
Manually Classifying Data



Supervised ML

A/C OFF/Window UP 1
A/C OFF/Window UP 2
A/C ON/Window DOWN
A/C OFF/Window DOWN run 1





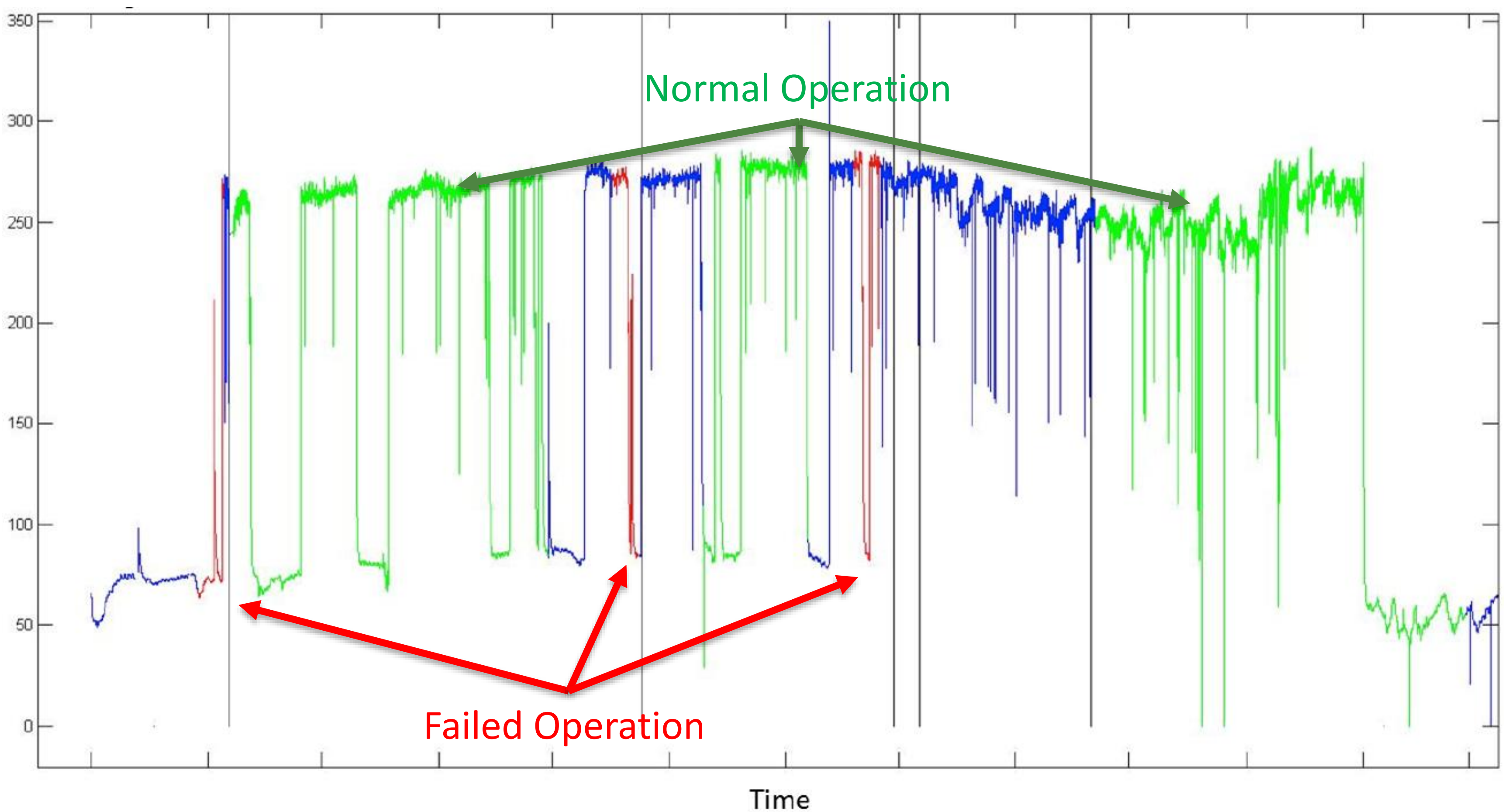
Lessons Learned

- Careful Data Collection/Handling
  - Repeatable Experiments
  - Minimal Differences (Date/Time/Temperature)
  - Experiment One Parameter at a Time
- Results are Still Difficult to Interpret
  - Data Spread Range without Explanation
  - **Opportunity to Hypothesize/Test Resulting Model/Weights**

Oilfield Applications

- Oilfield Production Pipeline
- Industrial Internet of Things
- Sensors Placed at the Pipelines
- Can We Use ML on the Data to Predict Failures?
- Can We Use ML on the Data to Optimize Production?

Supervised Learning Example



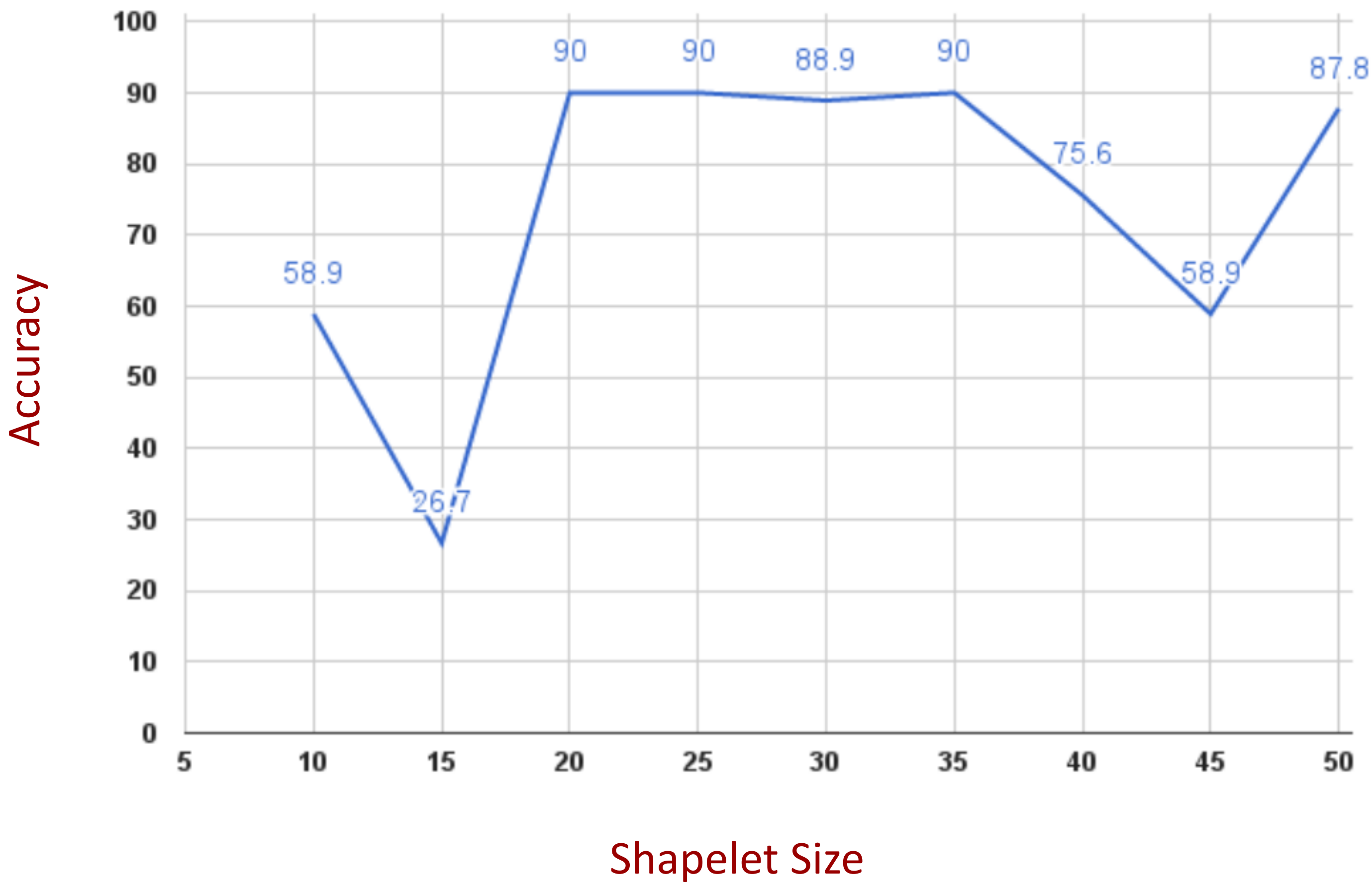
Dataset for One of the Sensors for Gas Compressor Valve

USC Viterbi School of Engineering

Time Series Shapelet

- Portion of Timed Sequence of Sensor Data (A Shape)
- Training: Given training dataset for the sensor, search and extract a window of fixed sequence of data (shapelet) that differentiate failed operation to normal operation
- Classifier: Function that quantifies the similarity between the shapelet and a sequence of test data enclosed in the moving window of the same size
- Modification: Shapelet was modified based on the observed data sequence to increase effectiveness

Result

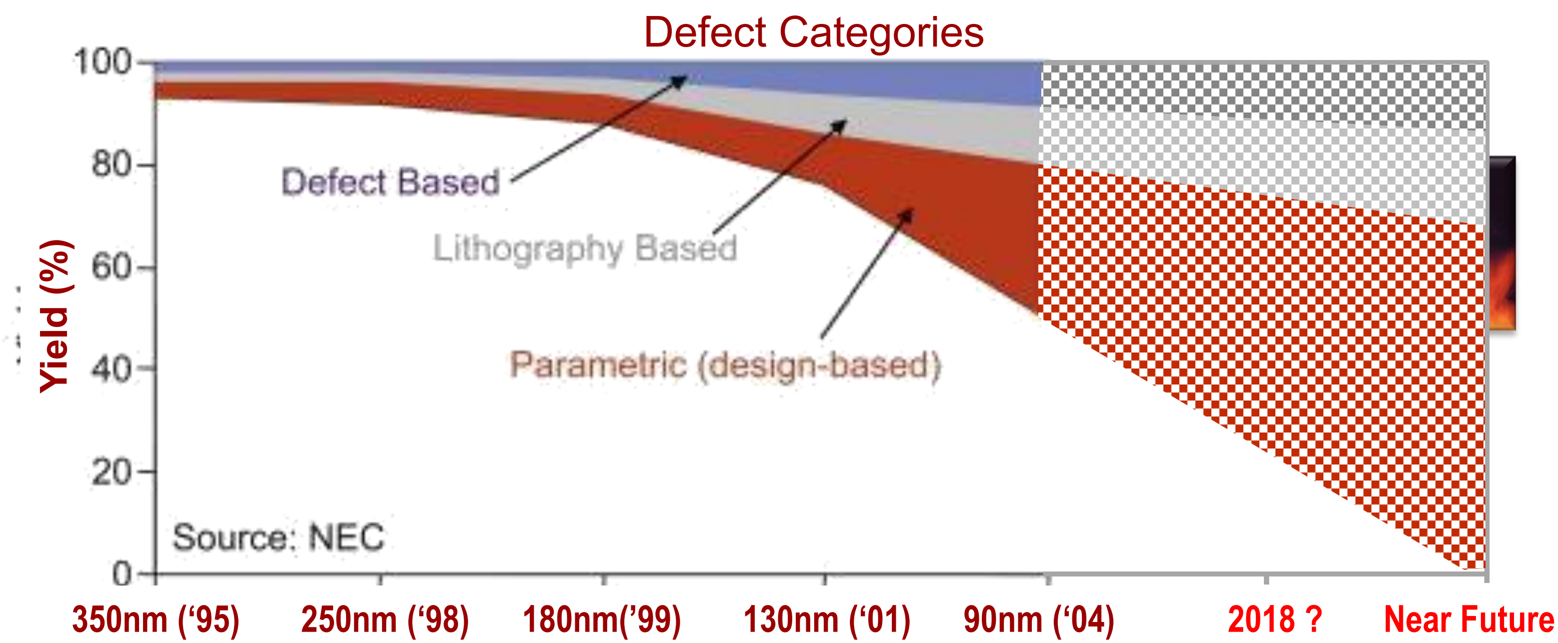

Accuracy
100
90
80
70
60
50
40
30
20
10
0
5
10
15
20
25
30
35
40
45
50
58.9
26.7
90
90
88.9
90
75.6
58.9
87.8
Shapelet Size

Lessons Learned

- Traditional Classifiers were NOT useful
  - This is a common method used in industry
  - Sometimes they work, but cannot determine why
  - Limits for the classifiers are not the absolute
- Need Application Specific Classifiers
  - Data abstraction may be modified to use classifiers
  - May be better to come up with a new classifier

Power Problem in Chips

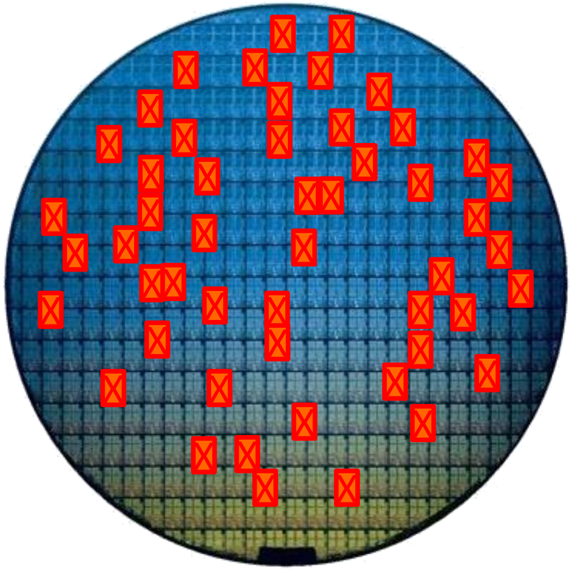
Even yield is now largely affected by power



Parts of the chip run too HOT
even though nothing is wrong structurally!

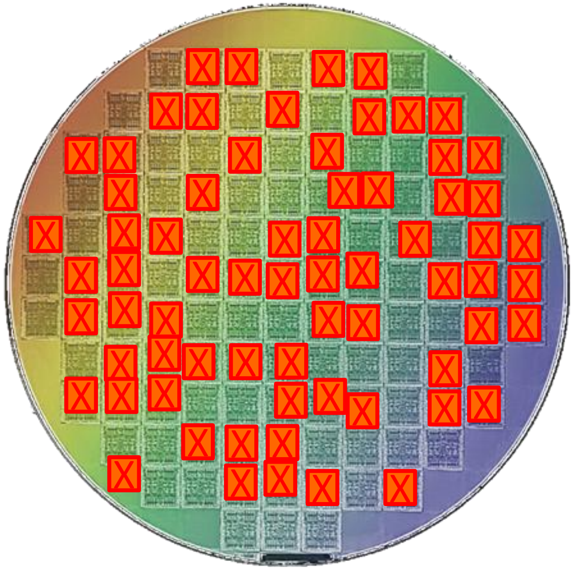
Power Problem in Chips

130nm Pentium III
April 2001



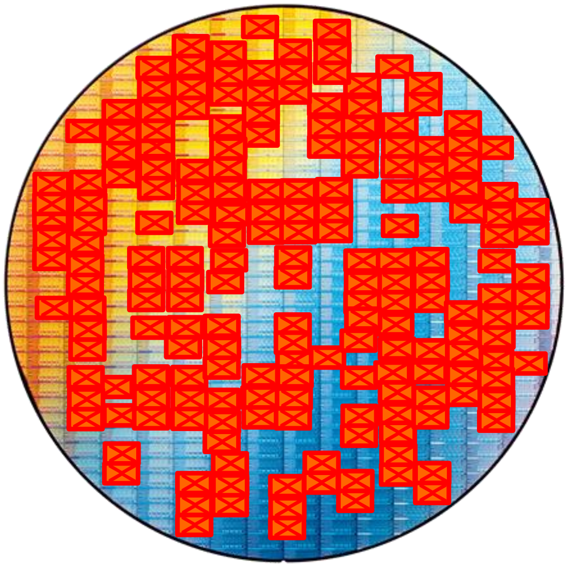
~25% non-determinism

90nm Nvidia G80
November 2006



~50% non-determinism

22nm Processors
Today!



> 75% non-determinism?!?

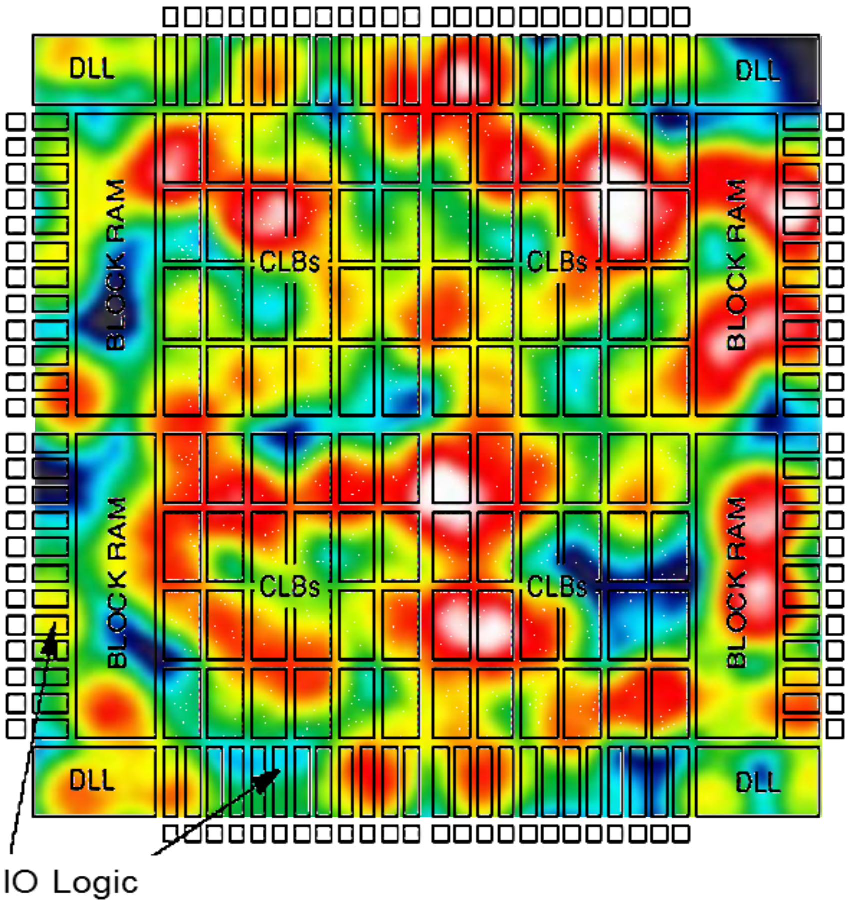
~$300B per Year Industry

USC Viterbi
School of Engineering

Our Test: Nvidia GTX1070

Total power consumption for 3 best cores is the same as 2 worst cores

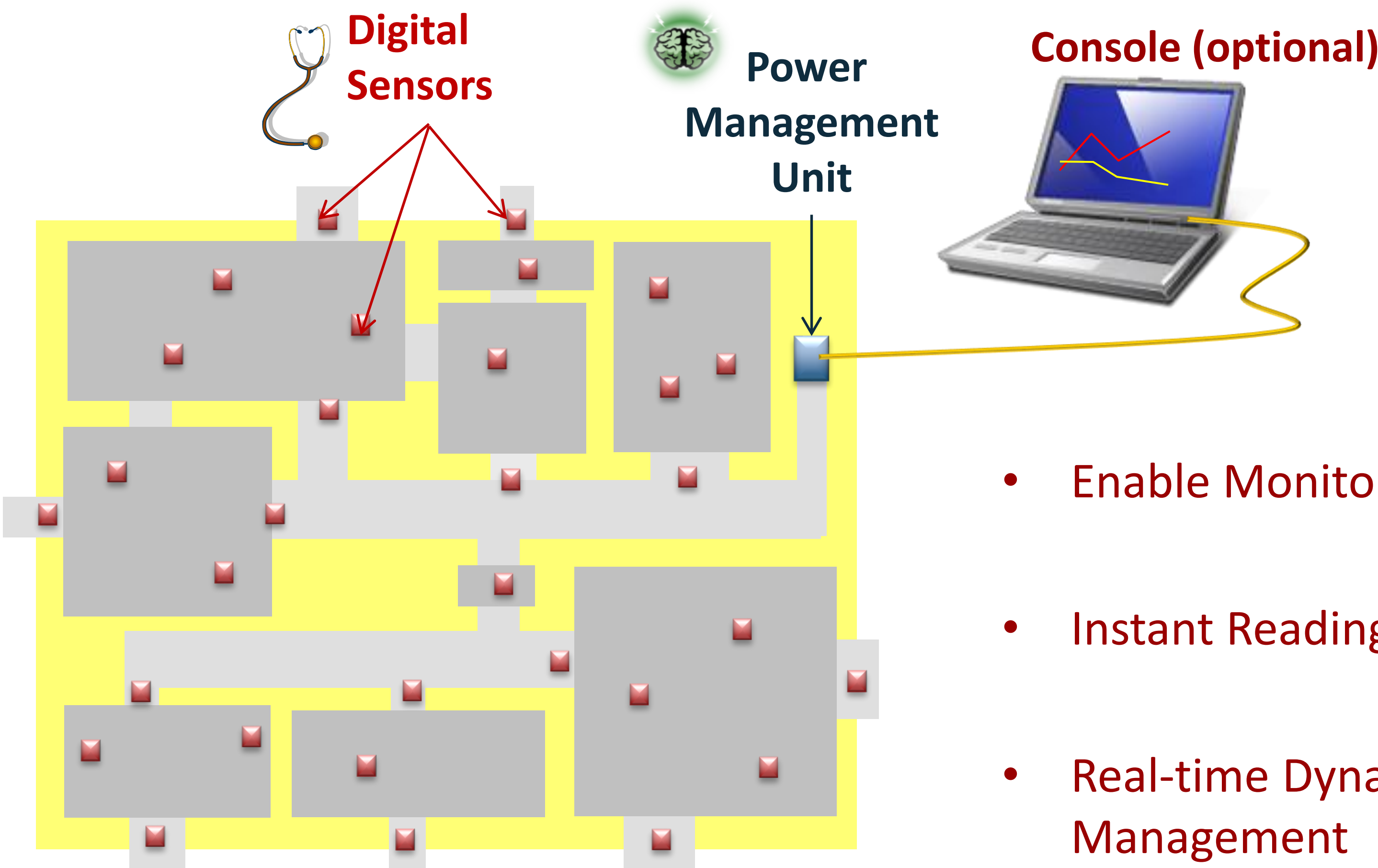
Field Programmable Gate Arrays



Measurement Experiments

- **Experiment Circuit**
  - Regular structure
  - 200MHz design
  - Entire FPGA mapped
- **Xilinx Spartan 3E**
  - 90nm CMOS technology
  - **3-5% power variation on the same IC**
  - **~14% average power variations across two different ICs**
- **Xilinx Spartan 6**
  - 45nm CMOS technology
  - **~9% power variation across the same IC**
- **7 Series and beyond**
  - < 45nm CMOS technology
  - **Higher variations expected**

On-line Power Monitoring



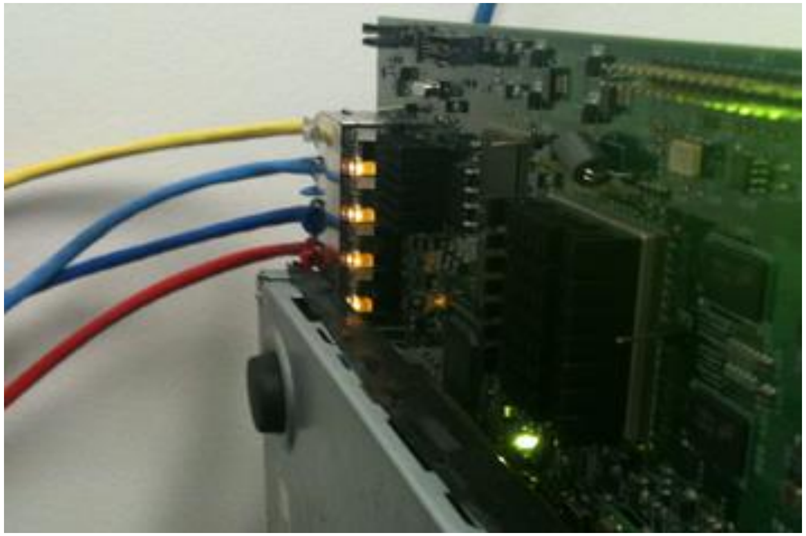
Internal layout of a Microprocessor

- Enable Monitoring
- Instant Readings
- Real-time Dynamic Management

On-line Power Monitoring

- **Hardware Instrumentation**
  - External total chip power sensor (a low-cost ADC at power source)
  - Digital logic based probes within chip
  - Negligible chip overhead and resilient to noise
  - Can leverage existing digital instrumentations
  - 100% scalable to all future technology node
- **Purely Software Instrumentation (Alternative to HW)**
  - Software only solution can be used in place of hardware probes
  - 2-5% software overhead dependent on program structure
- **Software Analytics**
  - **Big Data** created in real-time through digital monitoring
  - **Machine learning** applied to the Big Data to extract chip-specific parameter
  - Built-in self-calibration and real-time measurements
  - Measurements at any granularity (down to per cycle)
  - Determine sub-component level dynamic power and leakage
  - Predict dynamic power and leakage for tasks/operations

Working Prototypes



Internet Protocol Router prototype

OpenRISC based SoC prototype

- Combinational Logic to SoC Design on FPGAs
- On chip as well as on board Power Measurements
- 3 Patents held through USC

USC Viterbi School of Engineering

Comparison

Measurement Technology	Deviation from ADC	Method of Measurement	No. of Channels	Sampling Rate	Workload Dependent	Target Power Monitoring Level
Direct ADC Measurement	-	Dedicated Current Sensors	6	4KS/s	No	Board level
Runtime Power Monitoring	12.5% (Average)	Power Model Table Look Up	22	1K S/s	Yes	Architectural components
Modeling w/Event Counters	10% (Average)	Power Model Event Counters	12	NA	Yes	Architectural components
USC OASYS RESEARCH	**1.25 % (Max)**	**Weighted Power Measurements**	**350**	**250 KS/s**	**No**	**Board, Architectural, and Sub-circuit**

Comparison between type of power estimation techniques

Comparison

Measurement Technology	Method of Measurement	No. of Channels	Sampling Rate	Overhead
Altera PowerPlay	Estimated power through simulation	4	-	USB, software based simulation
Xilinx XPower	Estimated power through simulation	14	-	USB, software based simulation
Performance-driven Clustering	Estimated Power derived through simulation	-	-	Inaccurate due to manufacturing variations
Intel Itanium	Model based power estimation	120	125kS/s	Counters, Voltage regulators, PLL and freqcontrollers.
AMD On-chip power estimation	Model based power estimation	95	-	Counters, Voltage regulators, PLL and freq controllers.
USC OASYS RESEARCH	**Weighted Counter Values**	**350**	**250 KS/s**	**Digital Sensors**

Comparison between other implementations

USC Viterbi School of Engineering

Lessons Learned

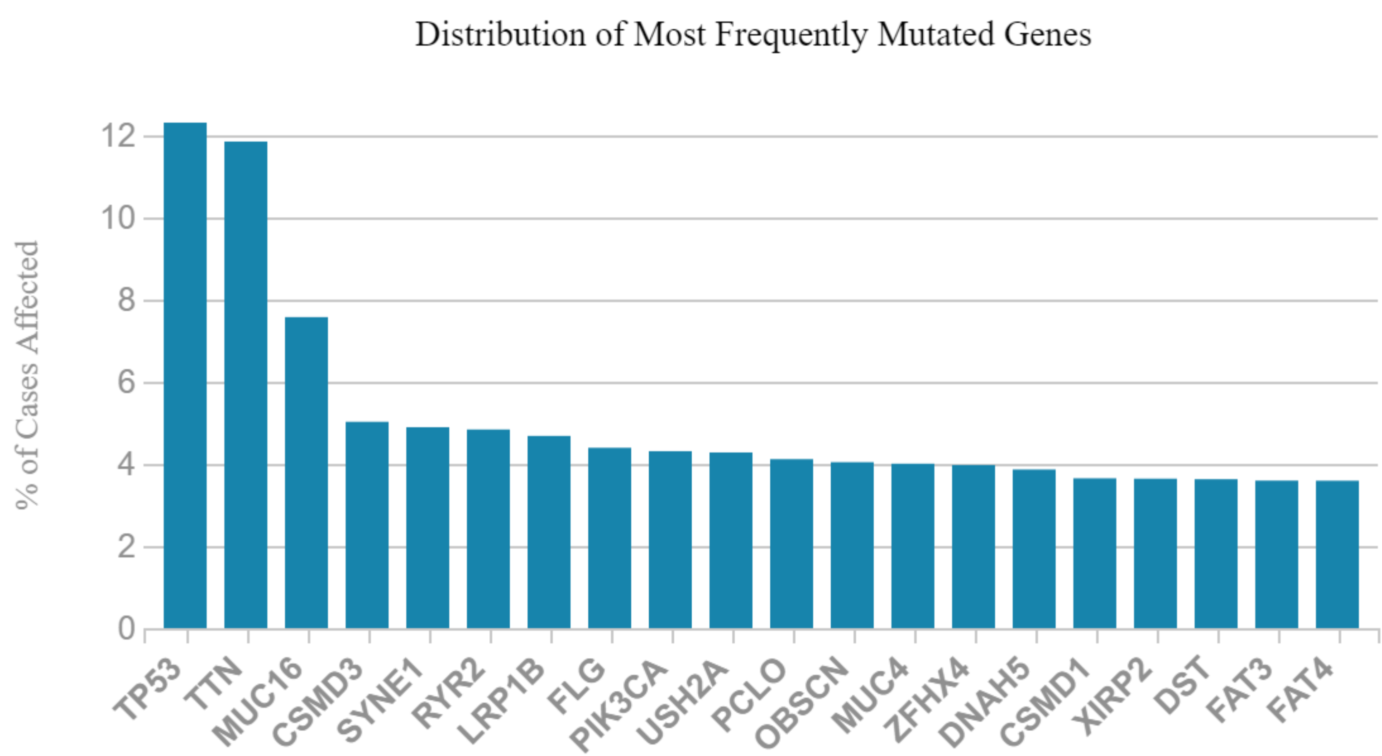
- Power Consumption is going to be (IF NOT ALREADY) the Biggest Problem for Cloud Computing and Data Center
  - Fine-grained monitoring with management is necessary
  - Big wins in chip yield and life, precise health monitoring, energy savings, and highest performance
- Machine Learning can be very powerful if the fundamental models can be integrated into the algorithm
  - **Top-down training** should be combined with **bottom-up data representation and fundamental model**

Summary

- Quality of Training Data Set
  - Emphasis of abstraction (retain the most relevant features)
  - Correct data labels
  - Filter out wrong/bad data from training set
  - Small set of high quality training is better than large set of bad training data
  - Manual prefiltering by human can drastically improve performance
- Custom Data Representation and Classifiers
  - Classifier that best fits the domain
  - Data context based modification
  - Default set of classifiers may never produce effective results
  - Application/Domain specific model integration may yield the best result
- Cloud Computing and Sensing Platform
  - Major problems with power, heat, and growing concerns for security
  - Need for fine-grained in-situ power and heat monitoring
  - Need for intelligent management

Current: Medical Research

- Predicting Cancer based on Biomarkers



Research Challenges

USC Viterbi School of Engineering

- High Complexity Data
  - Traditional methods may have limited usefulness
  - Should still evaluate their effectiveness
  - A method of reducing complexity or isolating parameters
  - Hierarchically cluster data to potentially isolate parameters
  - Recursively correlate and remove largest parameter to amplify smaller
- Data Quality and Quantity
  - Collect sufficient data from the healthy population for training
  - Possibly integrate data before and after cancer
  - A method of differentiating good versus bad data
  - Select training data from cluster that test data belongs
  - Collecting large amount of data to match the problem
  - Integrate additional body sensor and contextual data for each person
- Application-specific Knowledge
  - Integration of evidence support findings and/or models
  - Select biological models and integrate prediction function
- Platform Related Research
  - Power/Heat problems in both High-Performance and Mobile Systems
  - Minimally invasive fine-power/heat monitor and smart management
  - Security problems with Cloud computing and Mobile Systems
  - Non-invasive security monitor with signature (refer to presentation)