

BEARS

Berkeley Education Alliance
for Research in Singapore

SinBerBEST

Singapore-Berkeley Building Efficiency
and Sustainability in the Tropics

Sensing, Data Mining and Modeling

UC Berkeley: Costas Spanos (PI), Alex Bayen

*NTU/NUS: Yeng Chai Soh (PI), Xie Lihua, Hock Beng Lim,
Quek Ser Tong, Wang Qing-Guo, Kevin Kuang, Tham Kok Wai,
Keck Voon Ling, Weng Khuen Ho, Wen Jian Cai, Edmund Lo*



Thrust 1 mission statement

Mission statement of the sensing thrust:

“To design, develop, prototype, test and deploy novel sensing technologies, data analytics paradigms and actionable models for energy efficient buildings in the tropical climates”

Thrust 1 mission statement

What thrust 1 is:

- An effort to try “new” technology and approaches through SinBerBest, in particular off the shelf technology.
- Develop ourself the technology when needed
- An opportunity to walk out of the program with changed paradigms in sensing, data collection, data filtering and data analysis
- An effort which interfaces with other thrusts to support the sensing, modeling and data mining arm of their research agenda

What thrust 1 is not:

- An “IT” effort which implements other thrusts’ specs , the testbed for SinBerBest (as a hardware/software contribution, it will be part of THE testbed of thrust 6).

Strategic decision in light of NRF goals:

- Choice in the sensing work do be done: “humans, resources usage vs. traditional sensing”

Schedule, logistics, achievements of 2012

Schedule:

- Hiring of post docs and research staff / fellows / associates has started on both Singapore and UC Berkeley sides
- Hiring of students on the Singapore side was based on the NTU/NUS cycle (unless students were already here).
 - Staff working on the thrust on the Berkeley side:
 - Staff working on the thrust on the Singapore side:

Equipment purchase:

- Development of the sensors / motes, has started in Berkeley
- Back end computing infrastructure is to be purchased immediately to support motes (this is different from office desktop/laptops)

Schedule, logistics, achievements of 2012

Re-orgs:

- Professor Steven Glaser left the project, for health related reasons.
- As a consequence, the WP 1.3, “Embedded sensing and devices” was partially canceled, and reorganized into other workpackages. In particular, the following activities are now partially on hold, canceled or reorganized:
 - Sensor node hardware design: mote customization, sensing package finalization.
 - Prototyping and testing sensor nodes, signal processing, de-noising.
 - Preliminary deployment, building systems integration, and data collection.
 - Signal processing and data analytics tool box.
- The activity “Integration of CO₂ sensing in the sensing system” was canceled for now, as we have no expert on the Berkeley side for CO₂ sensing.
- Thrust 1.7 (first principle modeling) has also been canceled on the Berkeley side after the departure of Professor Glaser

Re-orgs:

- Professor Steven Glaser left the project, for health related reasons.
- As a consequence, the WP 1.3, “Embedded sensing and devices” was partially canceled, and reorganized into other workpackages. In particular, the following activities are now partially on hold, canceled or reorganized:
 - Sensor node hardware design: mote customization, sensing package finalization.
 - Prototyping and testing sensor nodes, signal processing, de-noising.
 - Preliminary deployment, building systems integration, and data collection.
 - Signal processing and data analytics tool box.
- The activity “Integration of CO₂ sensing in the sensing system” was canceled for now, as we have no expert on the Berkeley side for CO₂ sensing.
- Thrust 1.7 (first principle modeling) has also been canceled on the Berkeley side after the departure of Professor Glaser

WP1.1 - Modeling and data mining

Principal Investigator: Costas Spanos, Alex Bayen

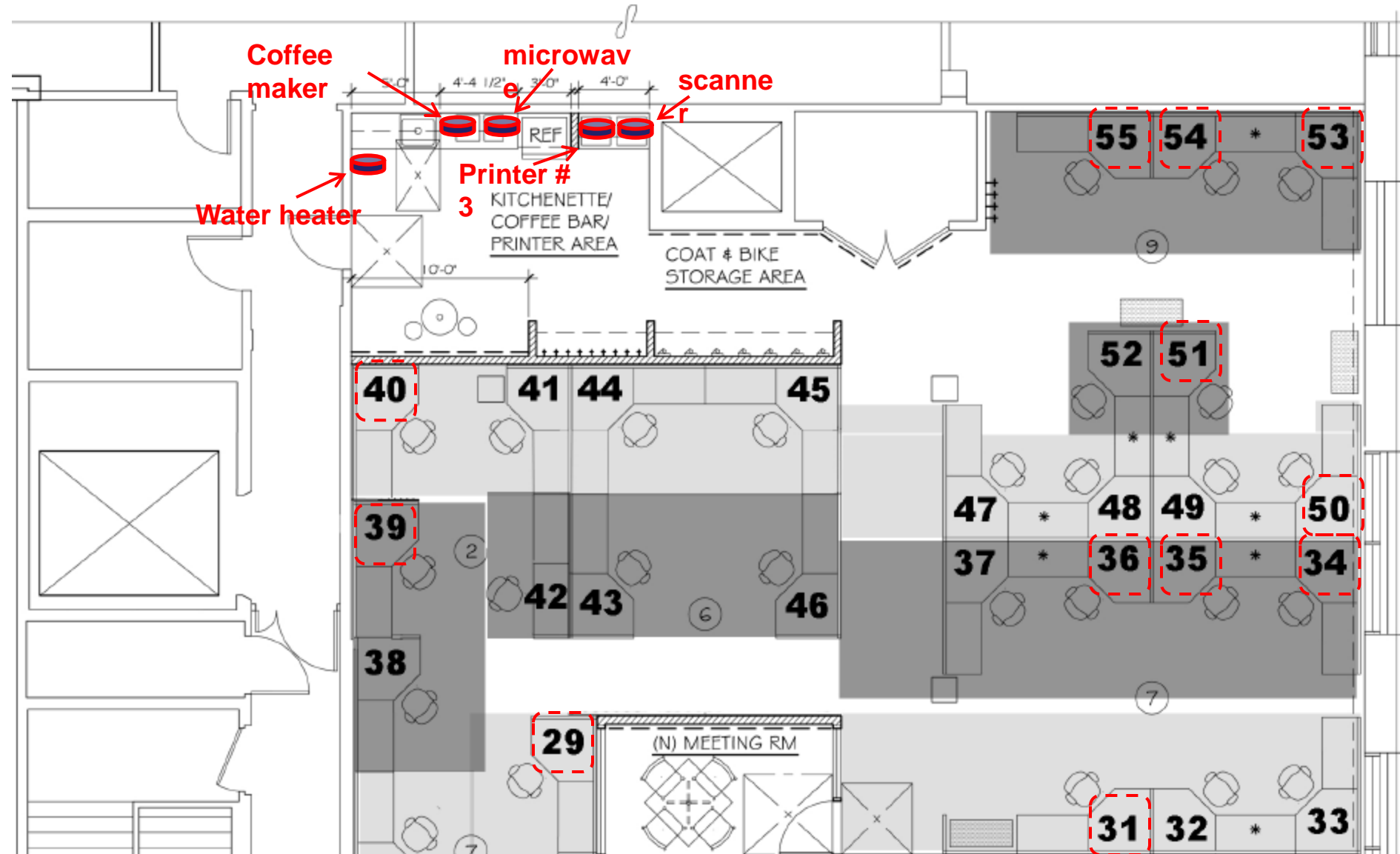
Other investigators: TBD

Research staff/students: 1 Ph.D Student, 1 undergraduate, 1 postdoc

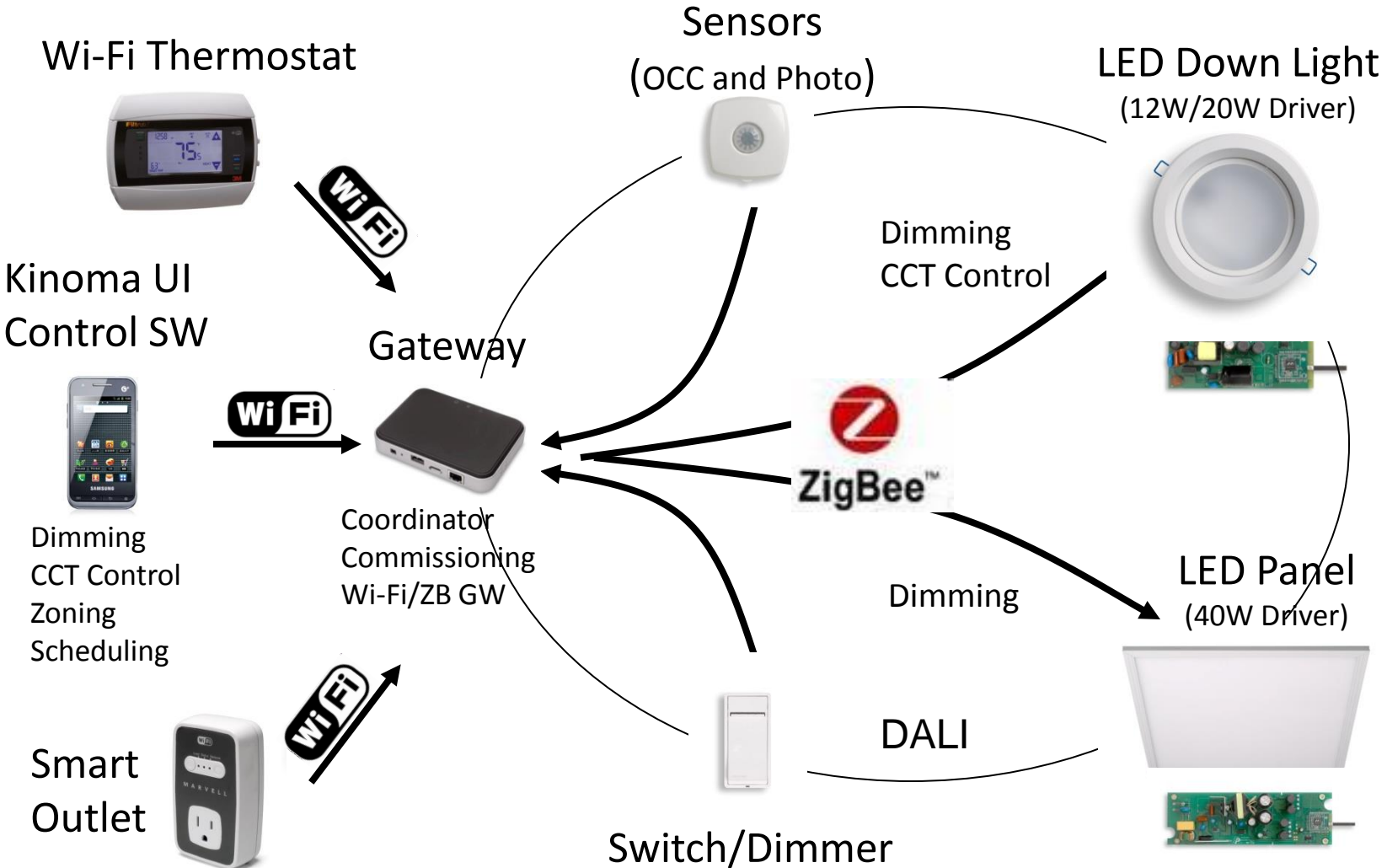
Milestones

- Machine learning for flagging and filtering outliers in energy consumption.
- Incorporation of exogenous variables from data collected in WP1-3.
- Mapping of outlier status to causal analysis.
- Design of EVOP experimentation sequences for focused parametric estimation of model variables.
- Optimization of the data collection based on the modeling and data mining (compressed sensing).
- Demonstration of automatically tunable Energy Plus model of building segment for SinBerBEST floor in the CREATE building.

Platform for individualized monitoring and social network gaming in a shared work space



Big Picture: Smart Platform



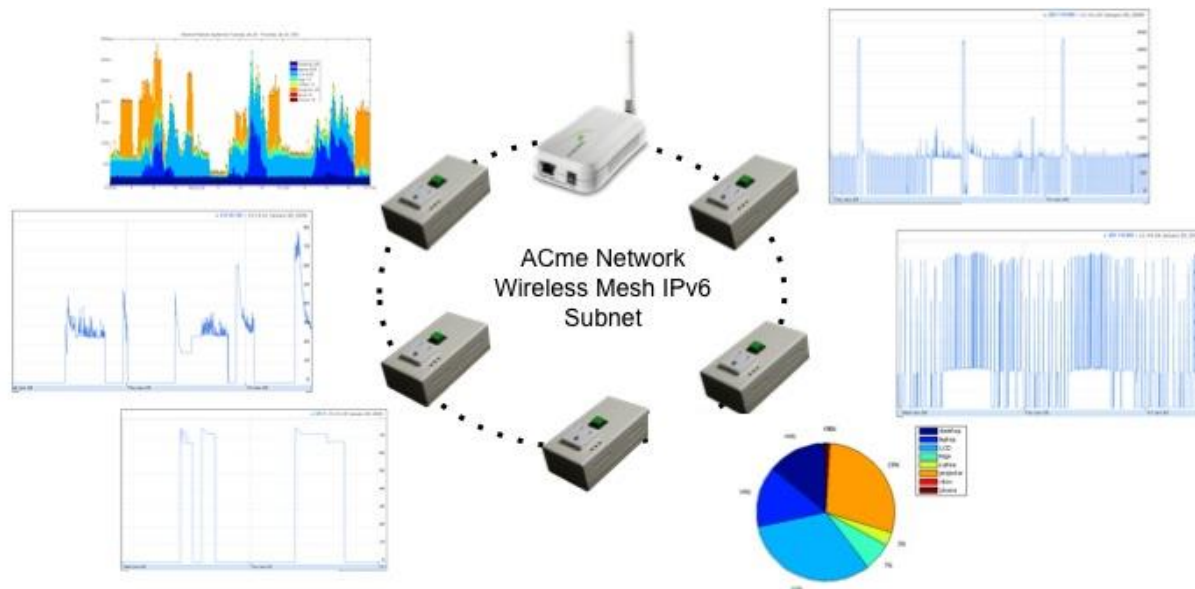
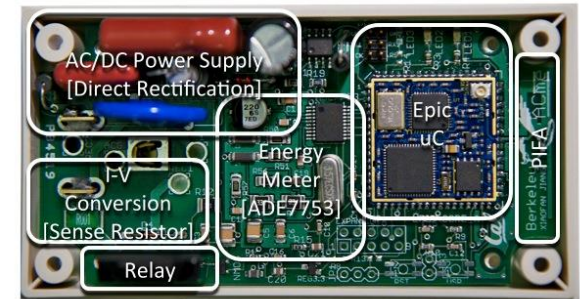
Zooming in - Kill-a-Watt & XBee

- Commercially available power meter from P3 International
 - LCD Display, multiple unit display, energy monitoring
 - Predict power costs
-
- “Piggyback” off Op-Amp outputs
 - 10mF capacitor to support transmission
 - Two XBee ADC channels sampled
 - LED to indicate packet transmission



ACme Sensor

- Commercial Solution
 - ADE7753 core IC; shunt resistor I-V conversion
 - Open source control platform
 - Online public database (sMAP)



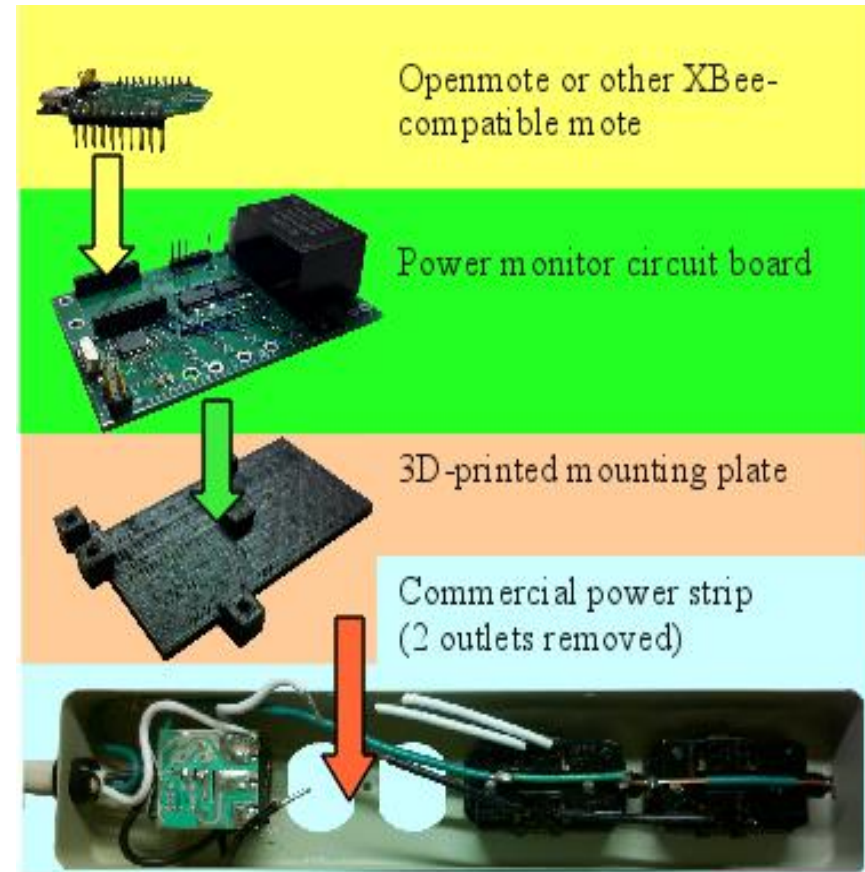
Sensing: Better Power Monitoring

Power Strip v1

- 4 inputs
- ZigBee
- Onboard Intelligence

Power Strip v2

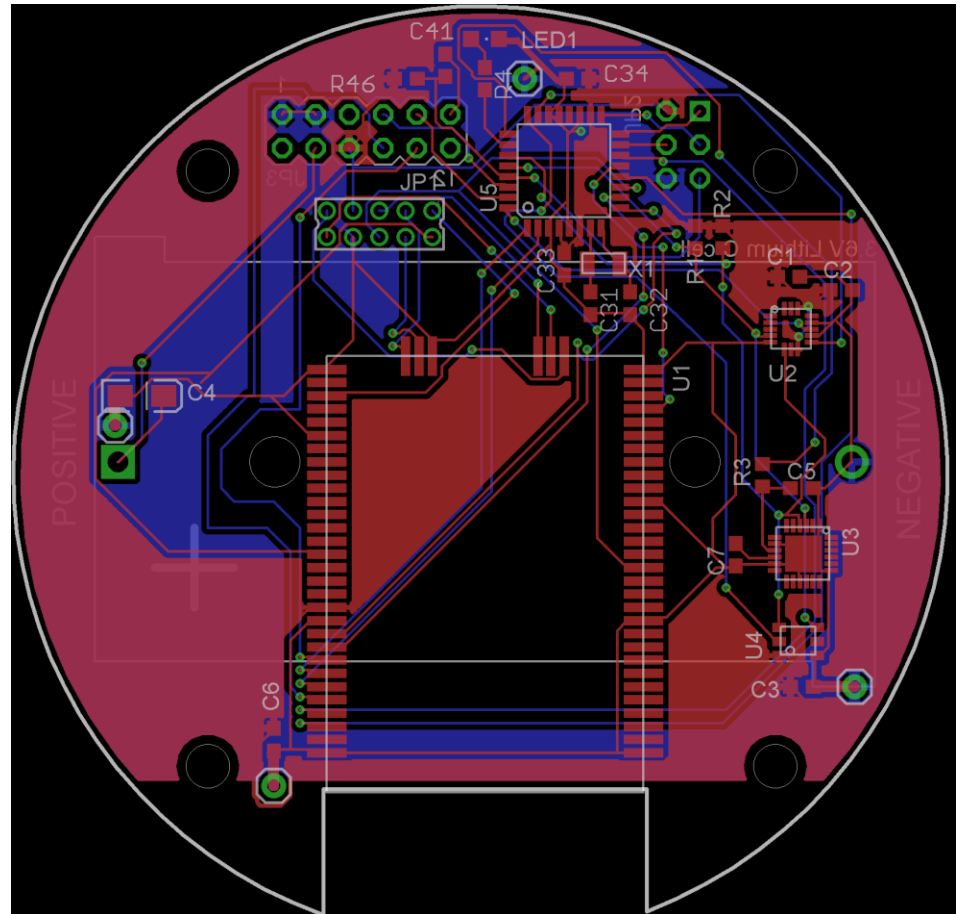
- 6+ outlets
- SmartMesh
- Control of Outlets



Sensing: Environmental

Sensor Puck

- Temperature, Ambient Light, Humidity, Vibration
- Arduino-compatible ATmega328
- SmartMesh IP Network
- 8 Year Lifetime using Lithium C cell



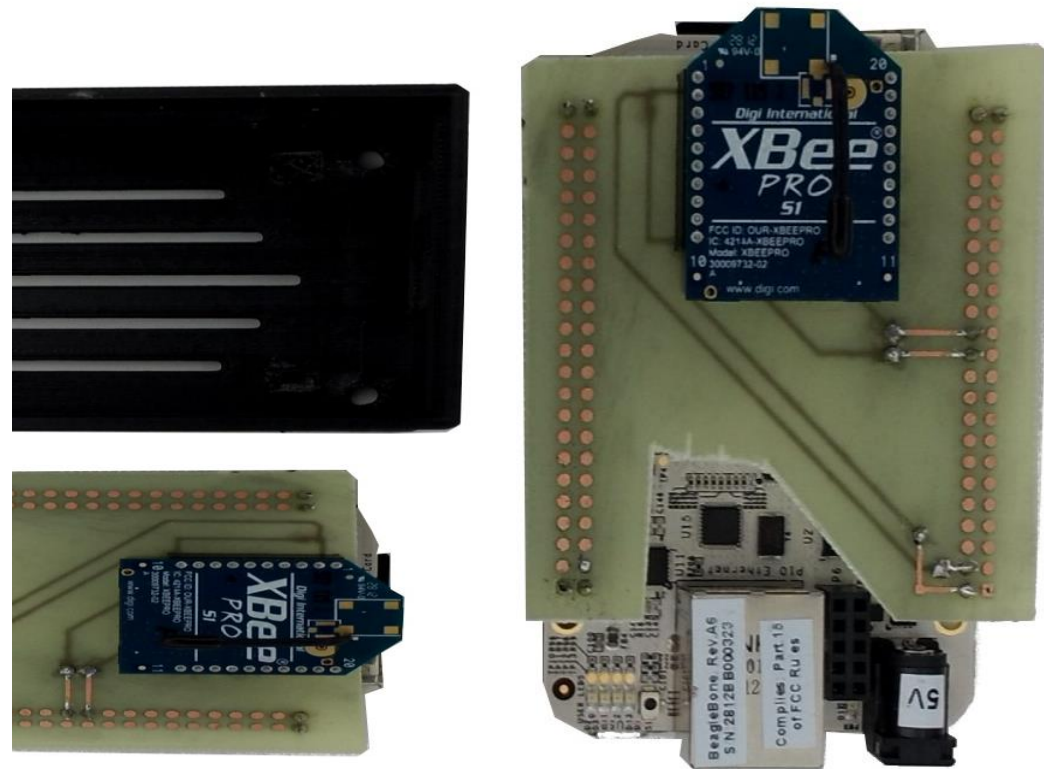
Custom Gateway Nodes

BeagleBone Linux computer.

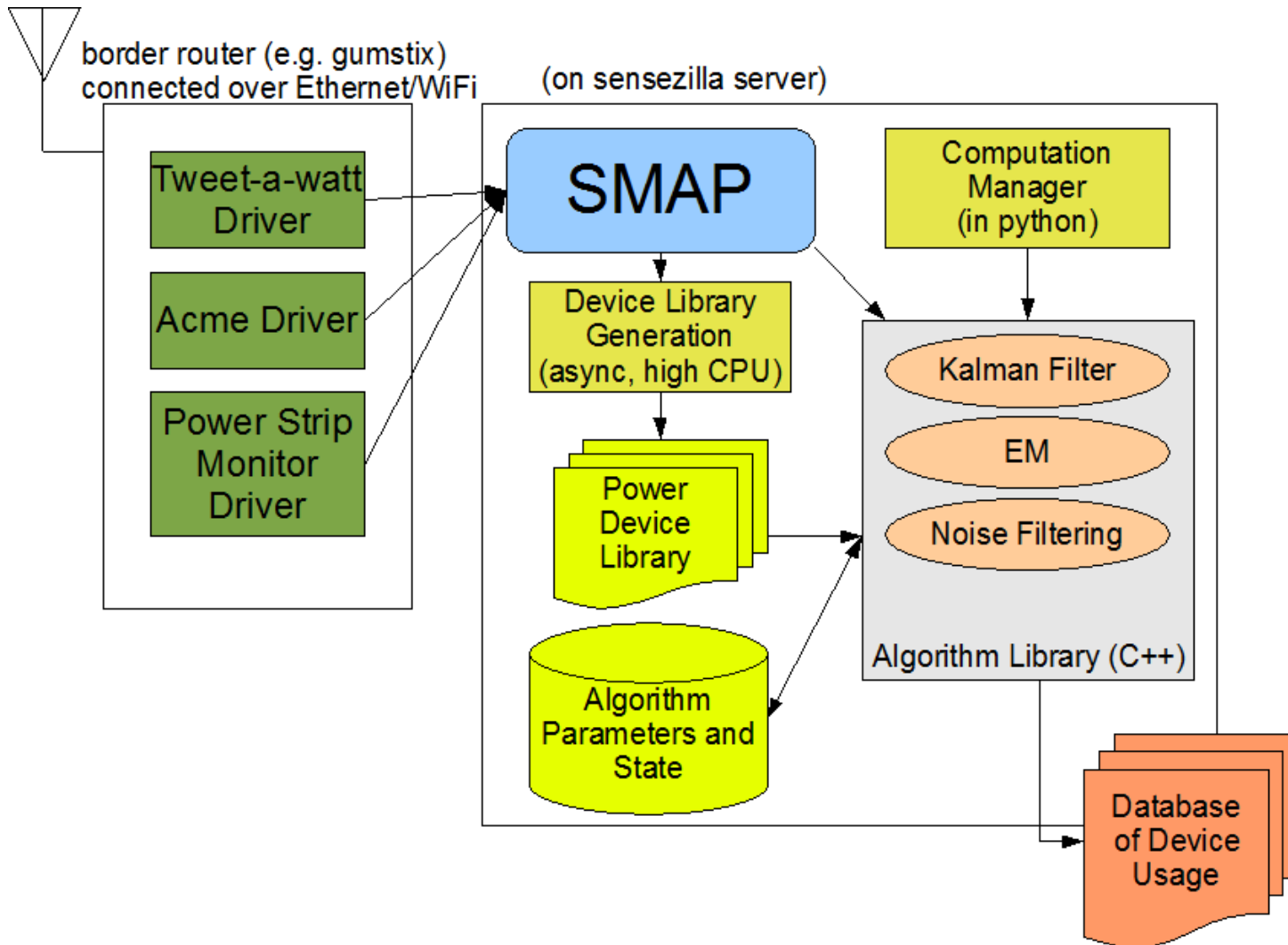
Custom shield for XBee module.

Custom 3D-printed enclosure.

Uses python software framework common with server.

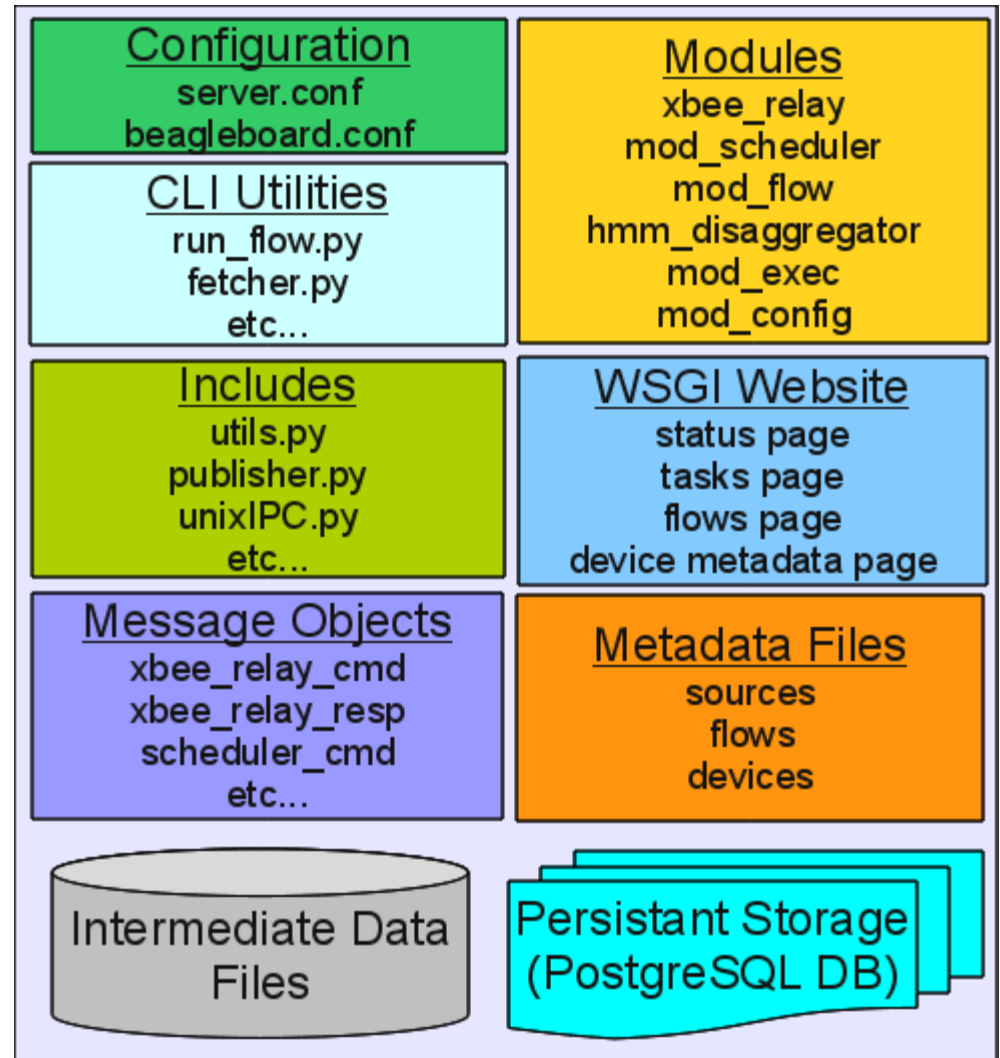


Software: Overview



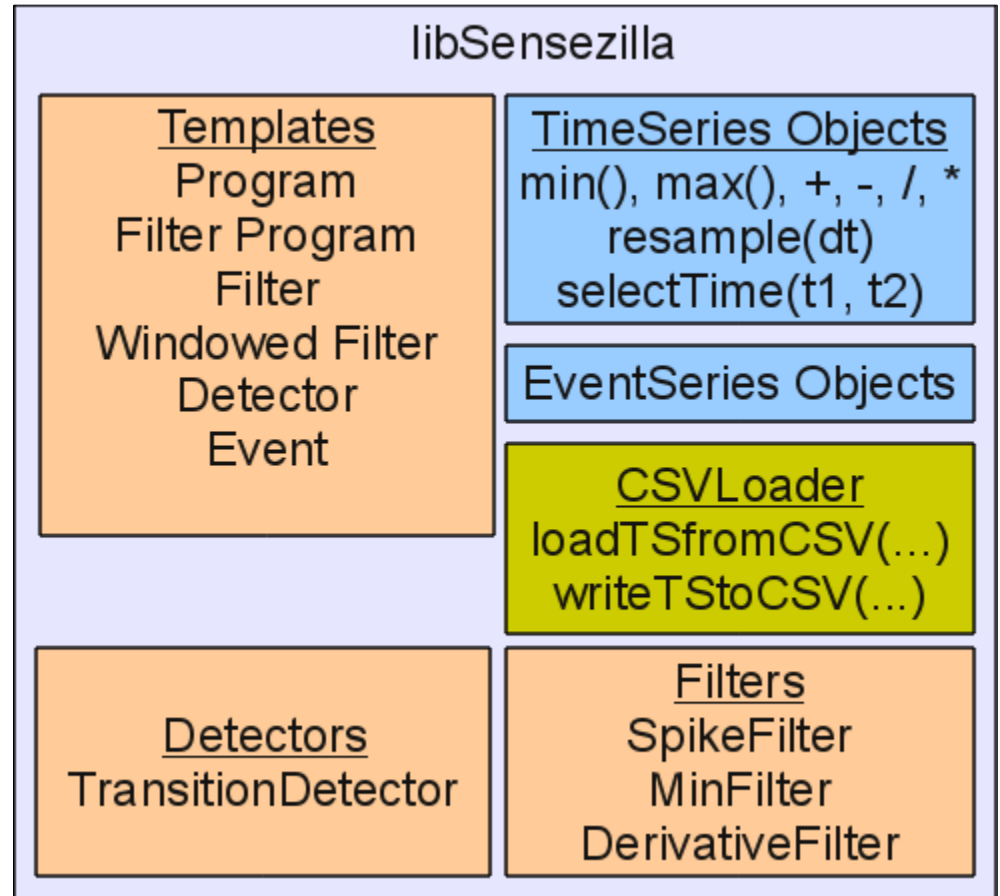
Software: python framework

- Focus on modularity, consistency, and rapid development.
- Segmentation of functionality into separate processes.
- Centralized and parameterized configuration files.
- Scheduler to run individual processing steps with dependency management.
- Flow-based computation manager using “Makefile-like” instructions.

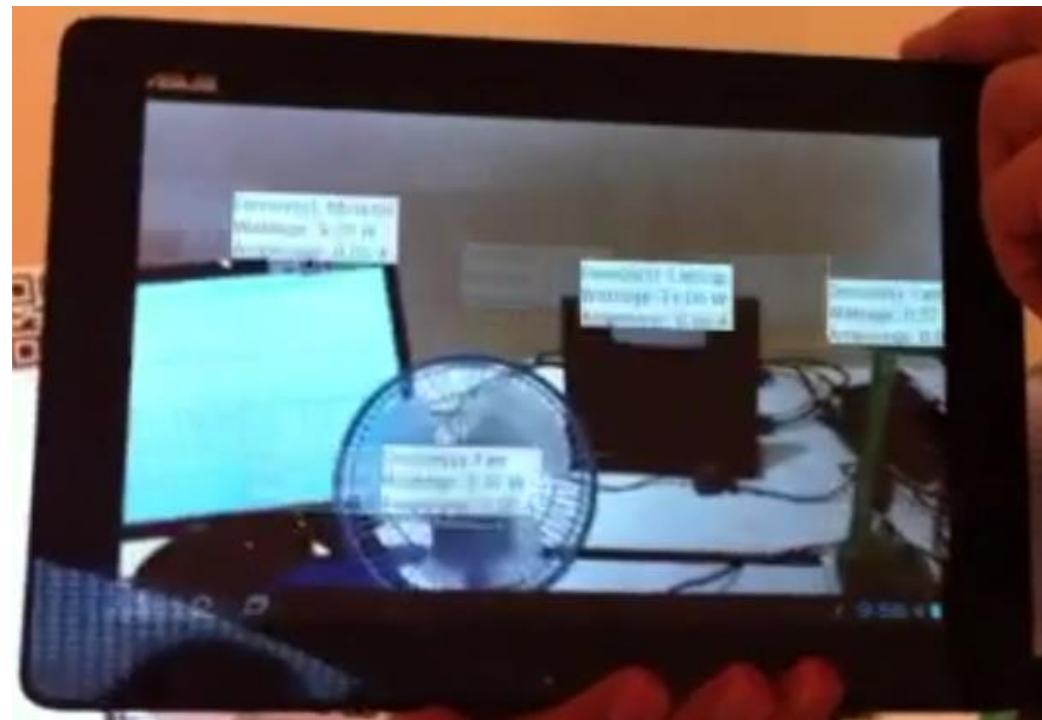


Software: C++ framework

- Focus on efficiency and consolidation of repeated code.
- Eclipse IDE development and example projects.
- MATLAB code ported to C++ saw 14x speed increase.

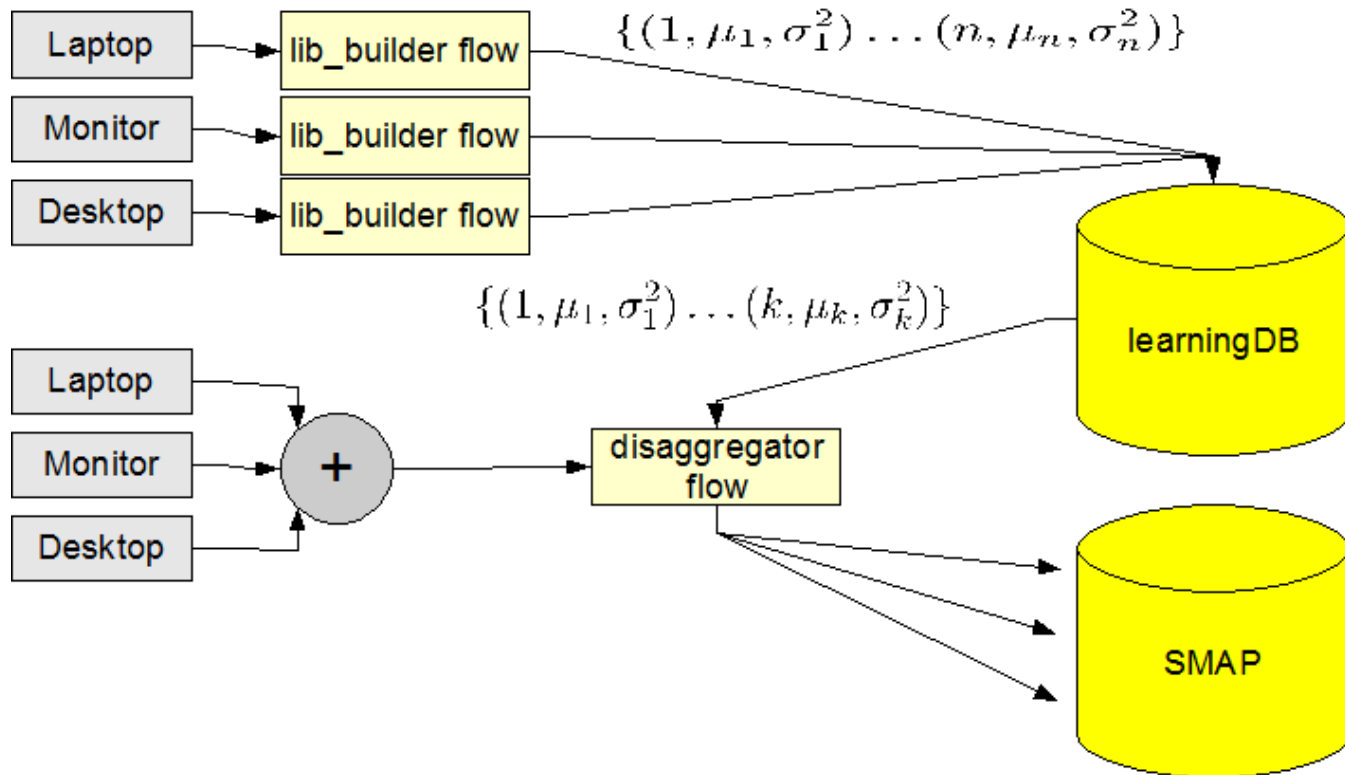


- Augmented Reality Android app locates QR-codes placed on power-consuming appliances and augments the video stream with consumption information.
- Can be used as a tool to find energy wasters.
- Showcases data flow from sensor to server to end-user.
- Future improvements:
 - Control devices / lighting
 - Heat maps of power use
 - Show other types of data



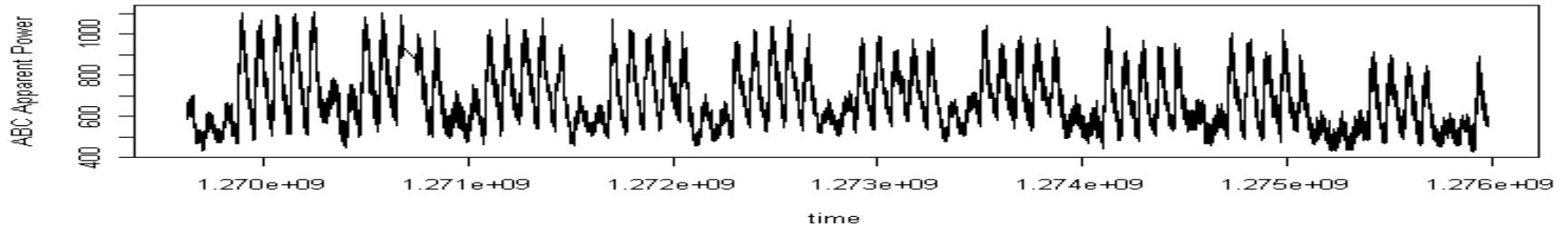
Application: power desagregation

- Autonomously estimates power usage of several devices plugged into one sensor using Hidden Markov Model.
- Automated learning from individually monitored devices.

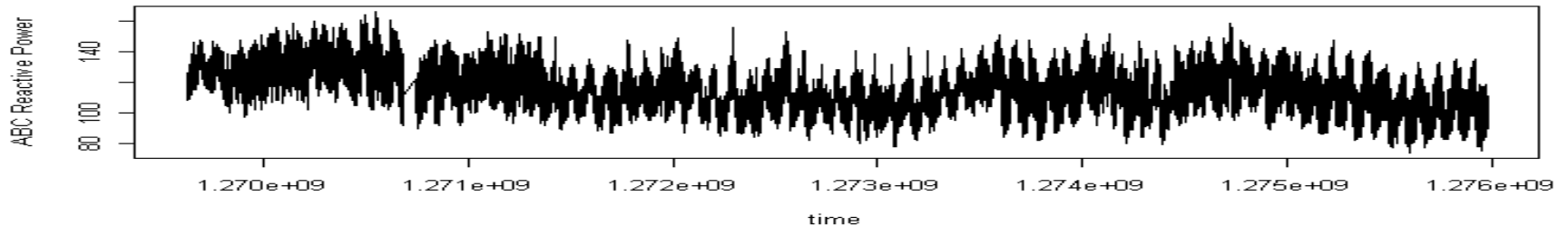


Example of High Level Monitoring

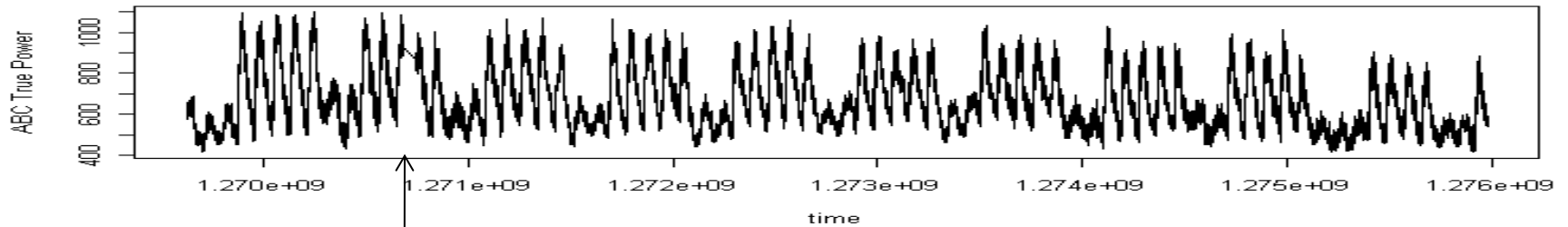
ABC apparent power against time, Device 1A, March 26th - June 7th



ABC reactive power against time, Device 1A, March 26th - June 7th



ABC true power against time, Device 1A, March 26th - June 7th

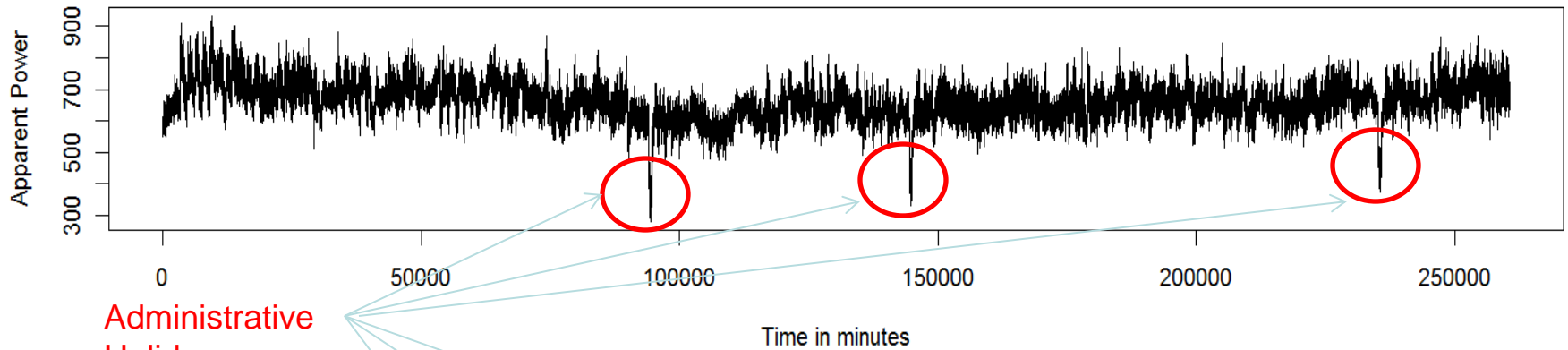


Main lighting transformer

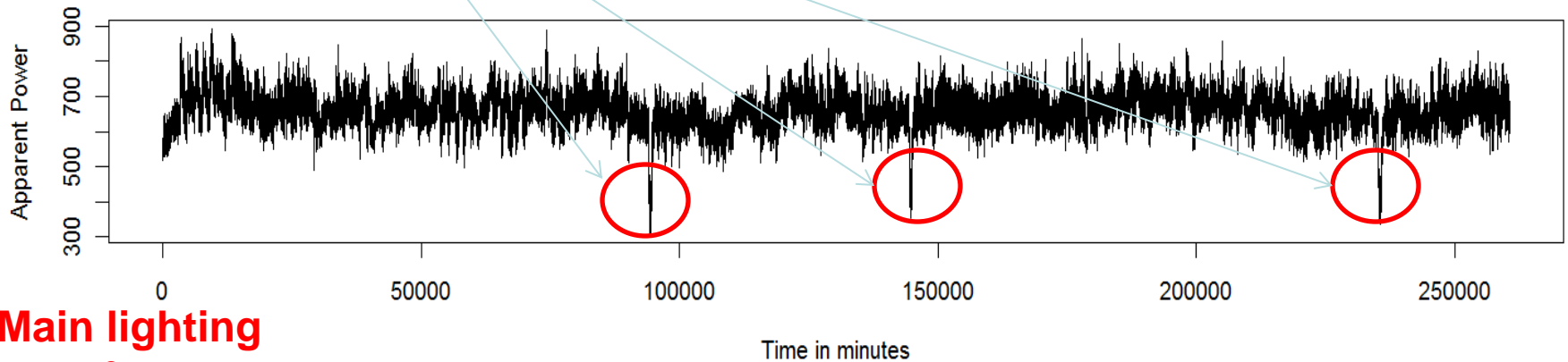
Missing Data for all devices (Apr 7th – 8th) – same for all other devices

Filtering, minutes average, apparent power device 1A

Graph of the minutes average of ABC apparent power for Device 1A
daily and weekly patterns removed (27th March - 23rd September)



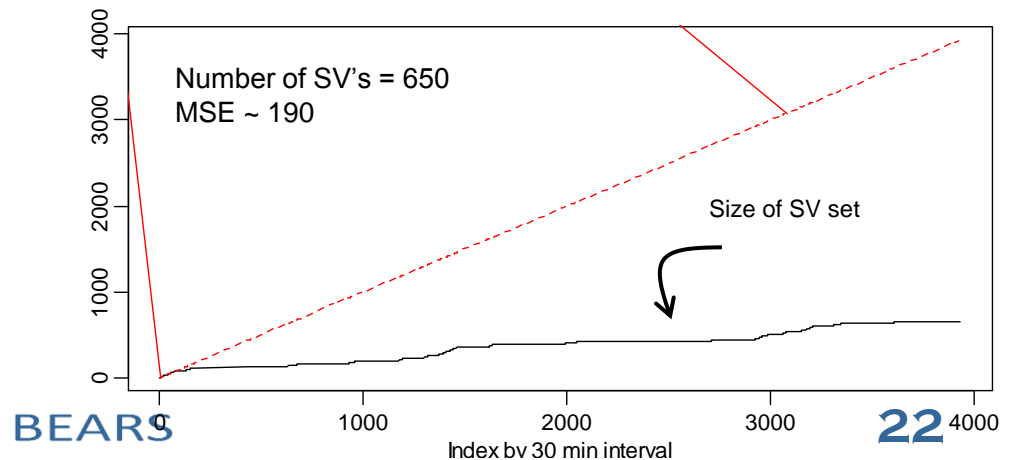
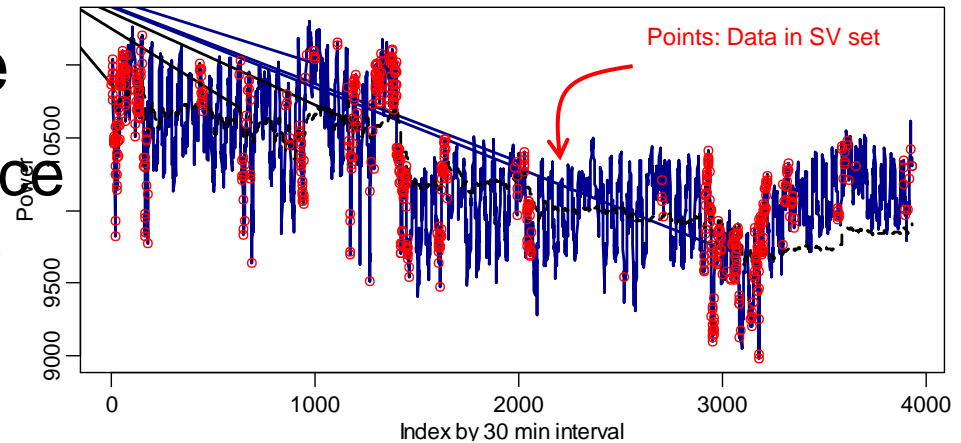
Graph of the minutes average of ABC apparent power for Device 1A
daily and weekly patterns, semester/non-semester differentials removed (27th March - 23rd September)



**Main lighting
transformer**

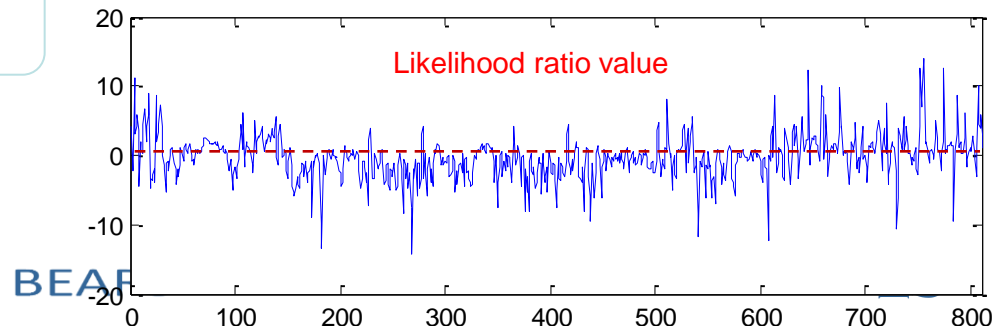
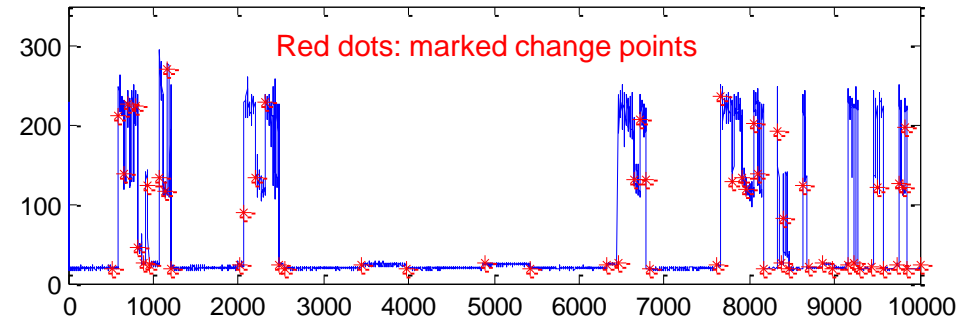
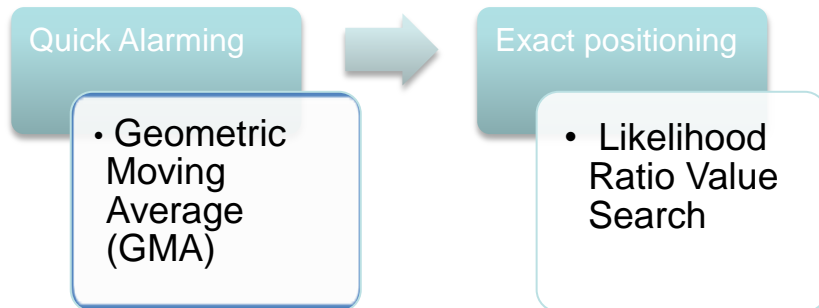
Real Time Anomaly Detection

- Gaussian model not good for periodic patterns;
- S-ARMA model not good for nonlinear pattern
- Support Vector Machine
 - Regression in kernel space
 - Example: SVM model for nonlinear time series



Real Time Anomaly Detection

- Detection Criterion:
 - Geometric Moving Average (GMA): efficiently catches the change in mean/variance;
 - Likelihood Ratio (LR) Search: exact change positioning
- Example: office desk power profile



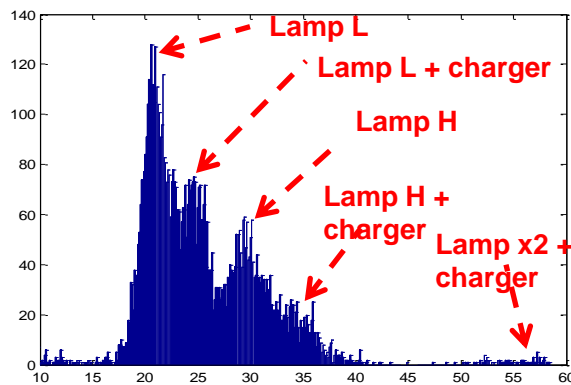
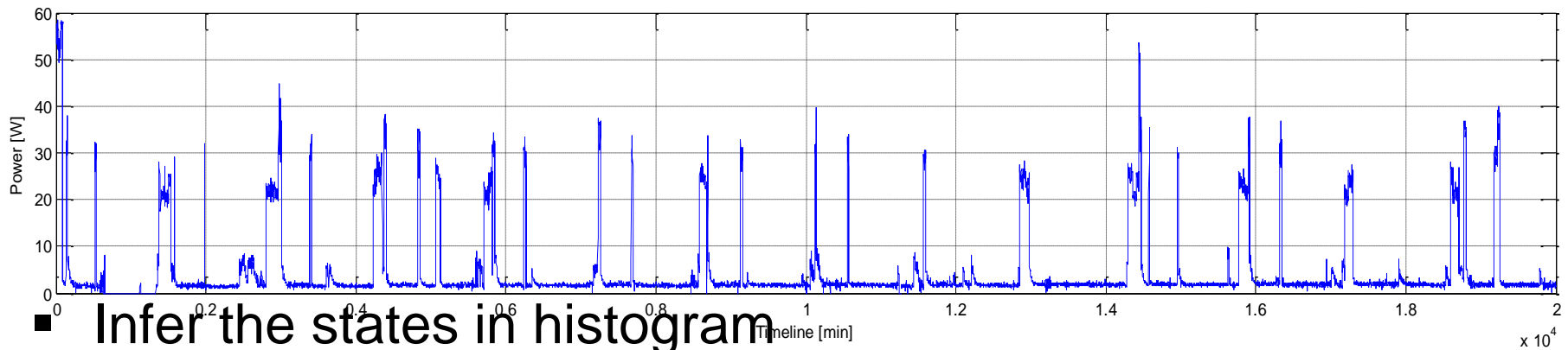
Device-State Model

- Filter out individual devices profile from total load profile
 - Lower cost & network requirement, since we only need one sensor for each seat/cubicle;
 - Some devices are indicator of occupancy, e.g. Laptop/Monitor, etc.
- Framework: Device-State Model



Device-State Model

- a power strip data with several devices

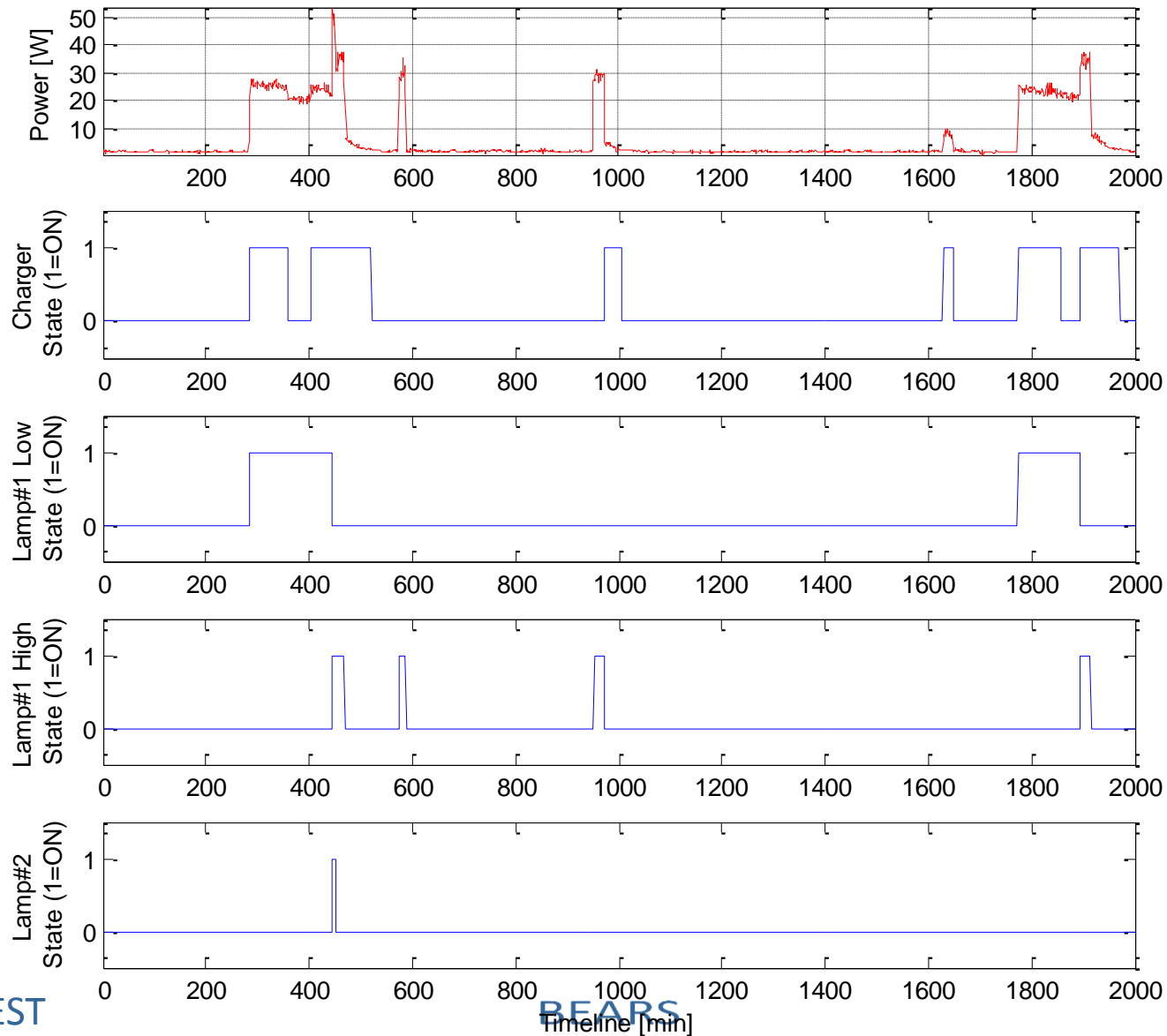


OFF state: 1~2w
Charger 4~8w
Lamp #1 Low: 20w
Lamp #1 High:
26~28w
Lamp #2: 20w

Device-State Model

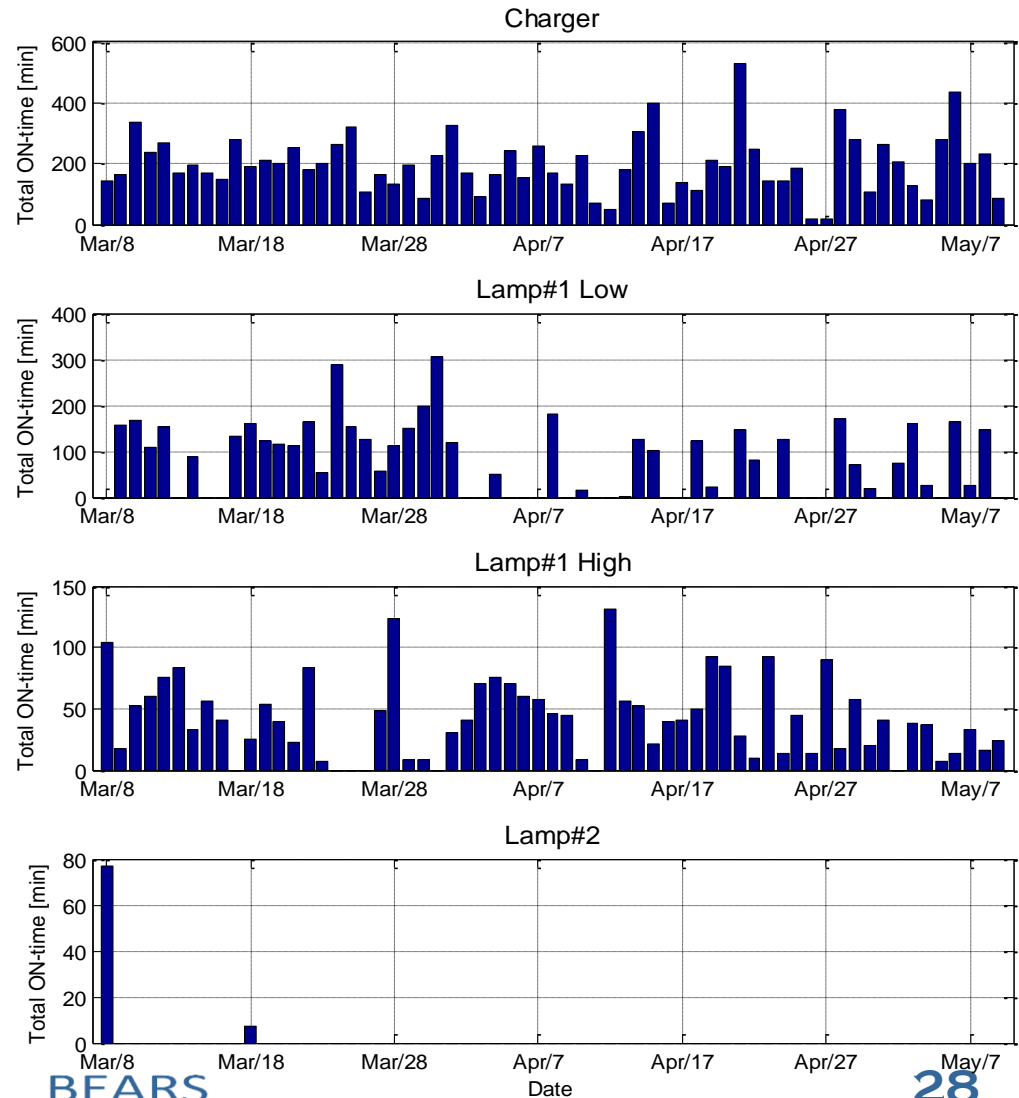
- Hidden Markov Model
 - Assume device state as hidden variable
 - Assume device holds at constant mean, and follows Gaussian distribution
- EM algorithm
 - Parameter updating
- Kalman filter for noise reduction
 - Filter out observation noise

Device-State Model



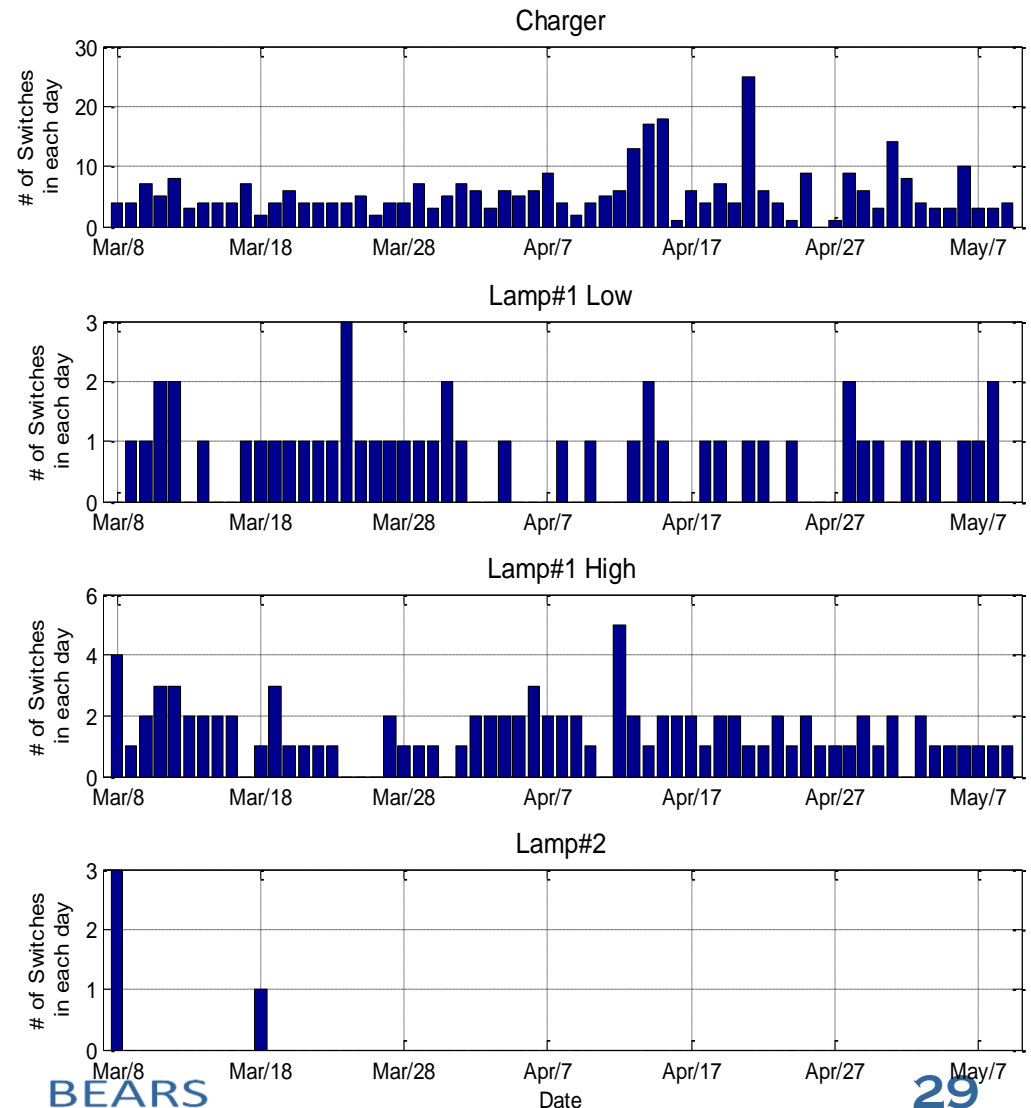
Device-State Model

- *Dynamics* of the device-usage can also be extracted
- Total ON-time of each device in each day
- Intense usage of *Lamp#1-Low* from Mar/17 ~ Apr/1 observed



Device-State Model

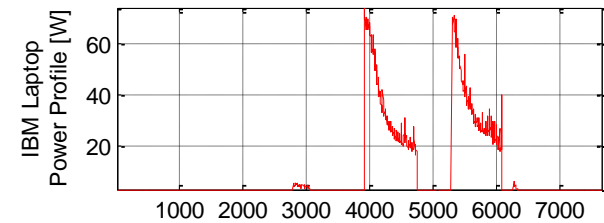
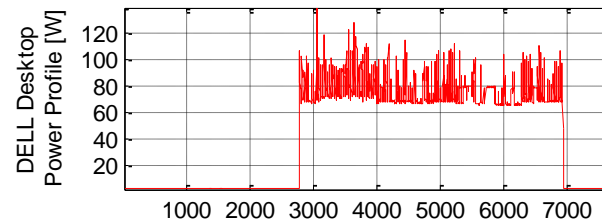
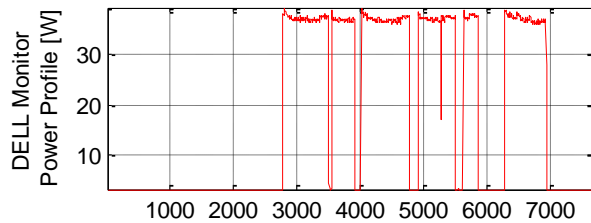
- # of switching of the devices in each day
- The dynamics of device-usage forms a more advanced framework for user profile study



Device-State Model

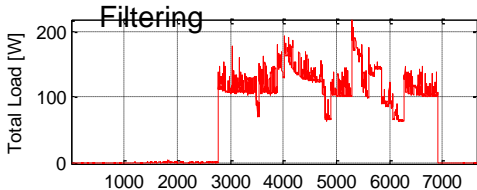
- For more complicated patterns, e.g. Laptop & iPhone, we can use exponential curve fitting.
- Example: a load with a desktop, a monitor and a laptop

True Device Power Profiles: For Reference

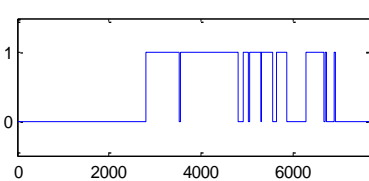


Total Load Curve: We use this to do Device-state

Filtering



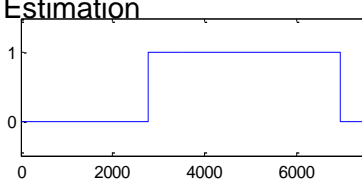
DELL Monitor State (1=ON,0=OFF)



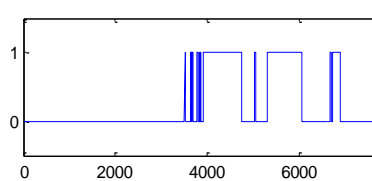
Device-state Model

Estimation

DELL Desktop State (1=ON,0=OFF)



IBM Laptop State (1=ON,0=OFF)



- Capture most important ON-OFF switching
- Some error: device power pattern non-Gaussian distributed

WP 1.2 - Smartphone based participatory sensing

Principal Investigator: Alex Bayen (UCB), Hock Beng Lim (NTU)

Other investigators: N.A.

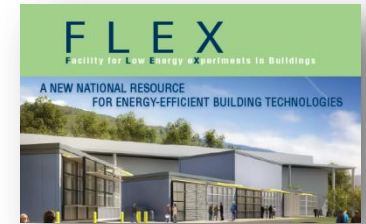
Research staff/students: 1 PhD student*, 1 post doc, 1 undergraduate

PI Alex Bayen was on paternity leave for part of 2012. This workpackage has been postponed for now. Hire of a systems staff engineer needed to start the WP. No hire has been done on this WP.

- Start: June 2013. Development of an iPhone and Android client
 - Motion, location
 - Microclimate control, user feedback
 - Activity survey
- Jan. 2014: start of integration with building infrastructure (PhD hire)
- June 2014: data collection and modeling work.
- Sep. 2014: Data assimilation into modeling work

Motivation

- Building energy monitoring/optimization is active area of research
 - Dynamic building-occupant interaction
 - New materials/systems verification testing
- It is promising to link multiple physical smart spaces in real-time.
 - Acquire and apply the knowledge to design, monitor, and manage smart spaces .
 - Improve building design, environmental modeling, energy resource optimization, and building control.



Our Solution - Ecosense



EcoSense: Cyberinfrastructure to Support Smart Spaces

- Wireless sensor networks provide the means to monitor the physical world in an unobtrusive manner.
- Pervasive middleware technologies provide mechanisms for interpreting spatial variability thus enabling classification
- High performance computing technologies such as grid and cloud computing provide the infrastructure for data management, processing and analysis.

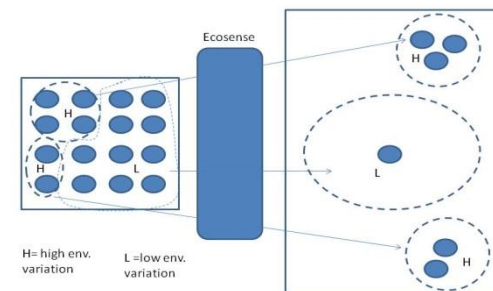
Context Modeling

- Based on the observations from an intensively instrumented smart space, to model and identify the areas with different environmental variability, which are referred to as *context zones*.
- Two dimensions in identifying environmental variability:
 - *Temporal variability*: the changing of the monitored characteristics over time (e.g. different temperatures at a specific location over time).
 - *Spatial variability*: the difference of the monitored characteristics over space (e.g. different temperatures in different locations at the same time).

Context Zones

- Based on temperature variability, identified locations requiring greater sensor density and vice versa
- Able to reduce sensors deployed

| <i>Smart Space</i> | <i>Node</i> | <i>Near airbox/vent</i> |
|--------------------|-------------|-------------------------|
| BubbleZero | 1-1 | No |
| BubbleZero | 2-1 | Yes |
| BubbleZero | 3-1 | Yes |
| BubbleZero | 4-1 | No |
| BubbleZero | 1-2 | No |
| BubbleZero | 2-2 | Yes |
| BubbleZero | 3-2 | Yes |
| BubbleZero | 4-2 | No |
| Intellisys | 1-1 | No |
| Intellisys | 1-2 | No |
| Intellisys | 1-3 | No |
| Intellisys | 2-1 | Yes |
| Intellisys | 2-2 | Yes |
| Intellisys | 2-3 | Yes |



Accomplishments

- A demo paper acceptance for presentation at the IEEE Consumer Communications and Networking Conference (CCNC 2013).
- Development and deployment of testbeds at:
 - BCA test chambers
 - ETH BubbleZERO
 - SinBerBest
- Demo of testbeds were shown during CREATE and SinBerBEST Open House.

WP 1.3 - Distributed sensing and cooperative control

Principal Investigator: Xie Lihua (Singapore)

Research staff/students: 1 Res Eng., 1 Postdoc, 1 Ph.D. student

This WP is 50% supported by control, 50% by sensing

Original planed milestones

- Set up of motes and network for human activities and IEQ monitoring
- Sensor and actuator placement
- Data collection, data assimilation
- Distributed localization and tracking by time of flight
- Human activity modeling
- Micro-environment modeling
- Cooperative distributed localized cooling control
- Experimental verification and performance evaluation

WP 1.3 - Distributed sensing and cooperative control

Summary of activity to this day:

- Literature review and survey on wireless indoor localization algorithms and techniques for human activity sensing.
 - For localization algorithms, both trilateration approach and fingerprinting approach have been well studied.
 - Literature review on various localization techniques, such as infrared (IR), ultrasound, RFID, UWB and WLAN/WiFi.
- Deployment of a test indoor WiFi-based localization system in CREATE SinBerBEST office which is cooperated with Y-find Technologies.
- Currently testing the localization accuracy of this system.
- Manpower hired: PhD student, Zou Han, joined Aug. 2012.

WP 1.3 - Distributed sensing and cooperative control

Proposal for this workpackage for 2013:

- Deploy at least one well-functional indoor localization system in SinBerBEST Office (base on WiFi or RFID) in order to capture occupant presence and behaviors by Q2 2013
- Start data collection and data analysis stage to further understand human activity patterns
- Develop human activity forecasting algorithms and occupancy dynamic models in buildings
- We also plan to integrate the localization system with other environmental sensors such as temperature, humidity and lighting sensors in our NTU lab first

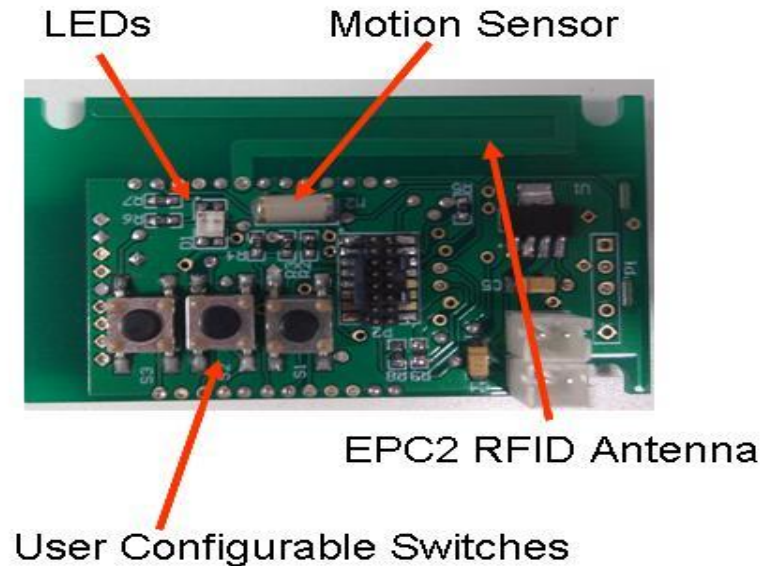
Purchase planned: Active RFID localization system in CREATE
SinBerBEST office: \$10000

WP 1.3 - Distributed sensing and cooperative control

Planned (and ongoing) partnerships

- Prof. Jia Qingshan's research team from Tsinghua University on human activity sensing, modeling and control. Prof. Jia and his research team have four years research experiences on related topics.

Wi-Fi Localization Tags

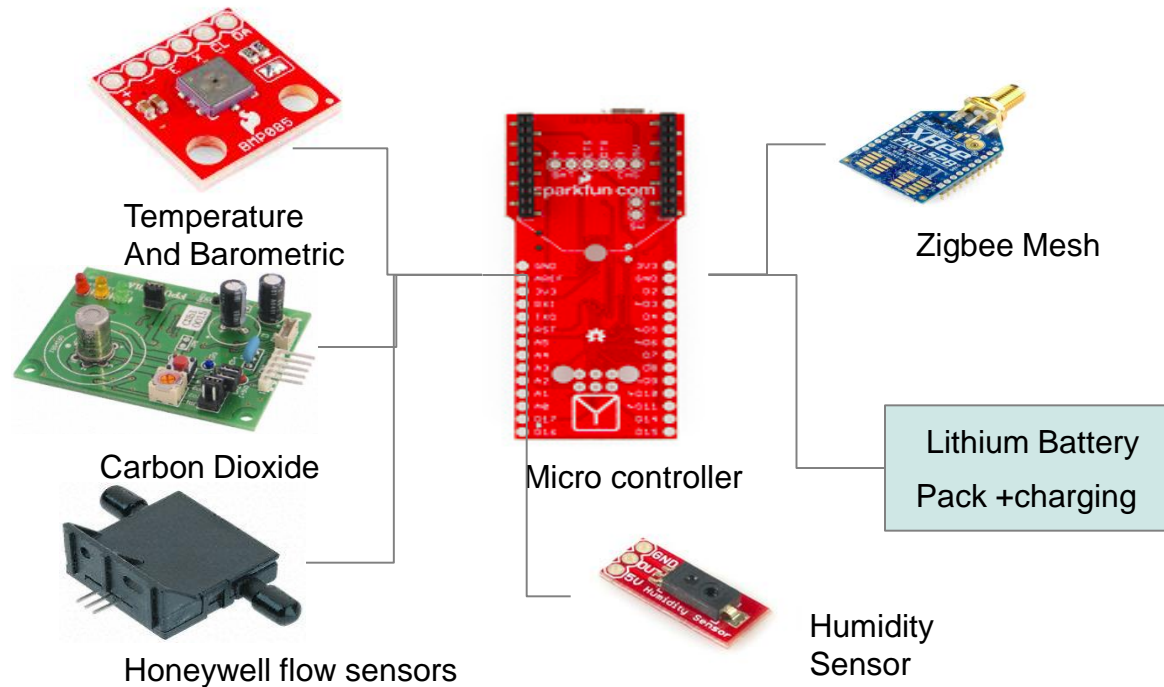


Tags PCB for illustrative purposes only, to be redesigned in 2013

Features:

- ⌘ Use existing Wi-Fi APs for localization
- ⌘ Sensors include 3 axis accelerometers MEMS, motion sensors
- ⌘ Support EPC2 RFID

Indoor Air Quality Sensor Node



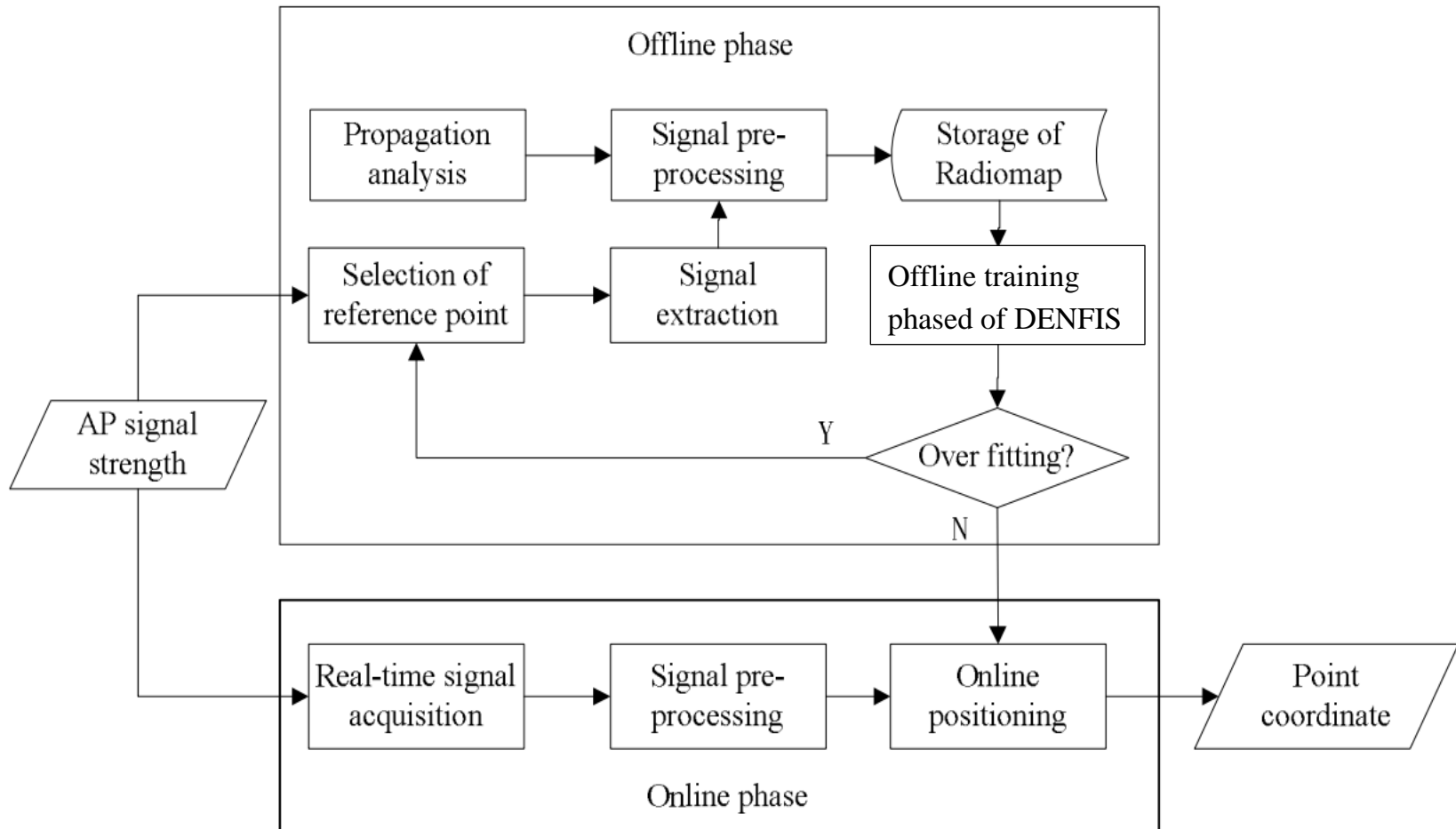
Features:

- Monitoring parameters: Temperature, Carbon Dioxide, Atmospheric Pressure, Air flow ,and Humidity
- Zigbee Wireless Mesh Nodes

DENFIS-based WLAN Indoor Positioning Algorithm

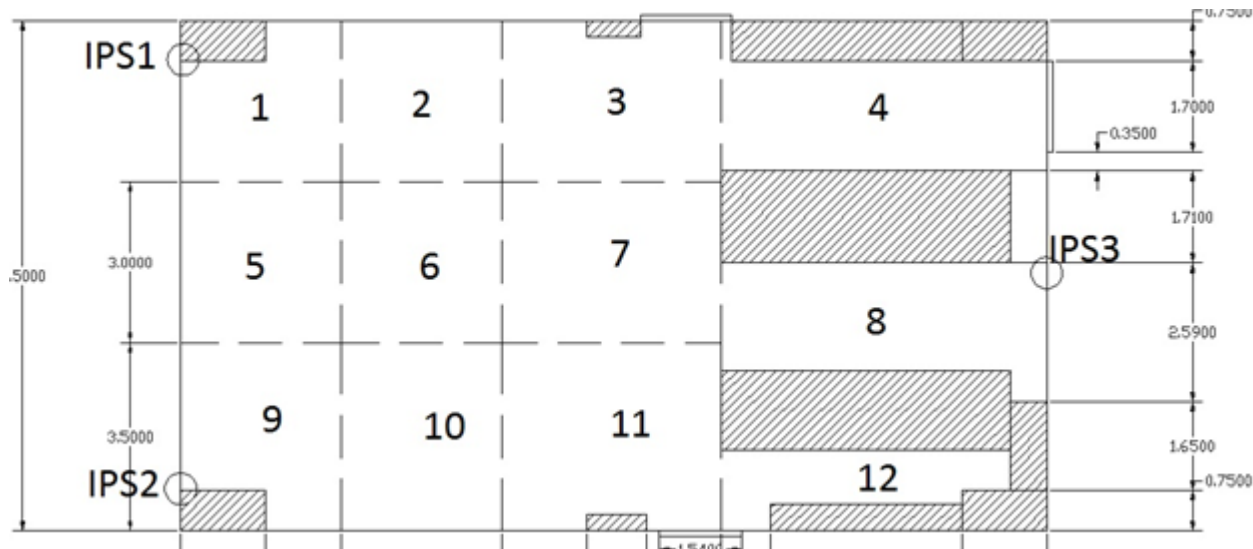
1. Dynamic evolving neural-fuzzy inference system (DENFIS) is good at adaptive offline learning. Based on this property, it is suitable for the training, searching and matching in fingerprinting method.
2. Through the acquisition and pre-processing of RSS signal characters, offline training phase of DENFIS could help us to build up the radio map of the indoor testing environment. During the online phase, DENFIS is used as the searching and matching algorithm.

DENFIS-based WLAN Indoor Positioning Algorithm



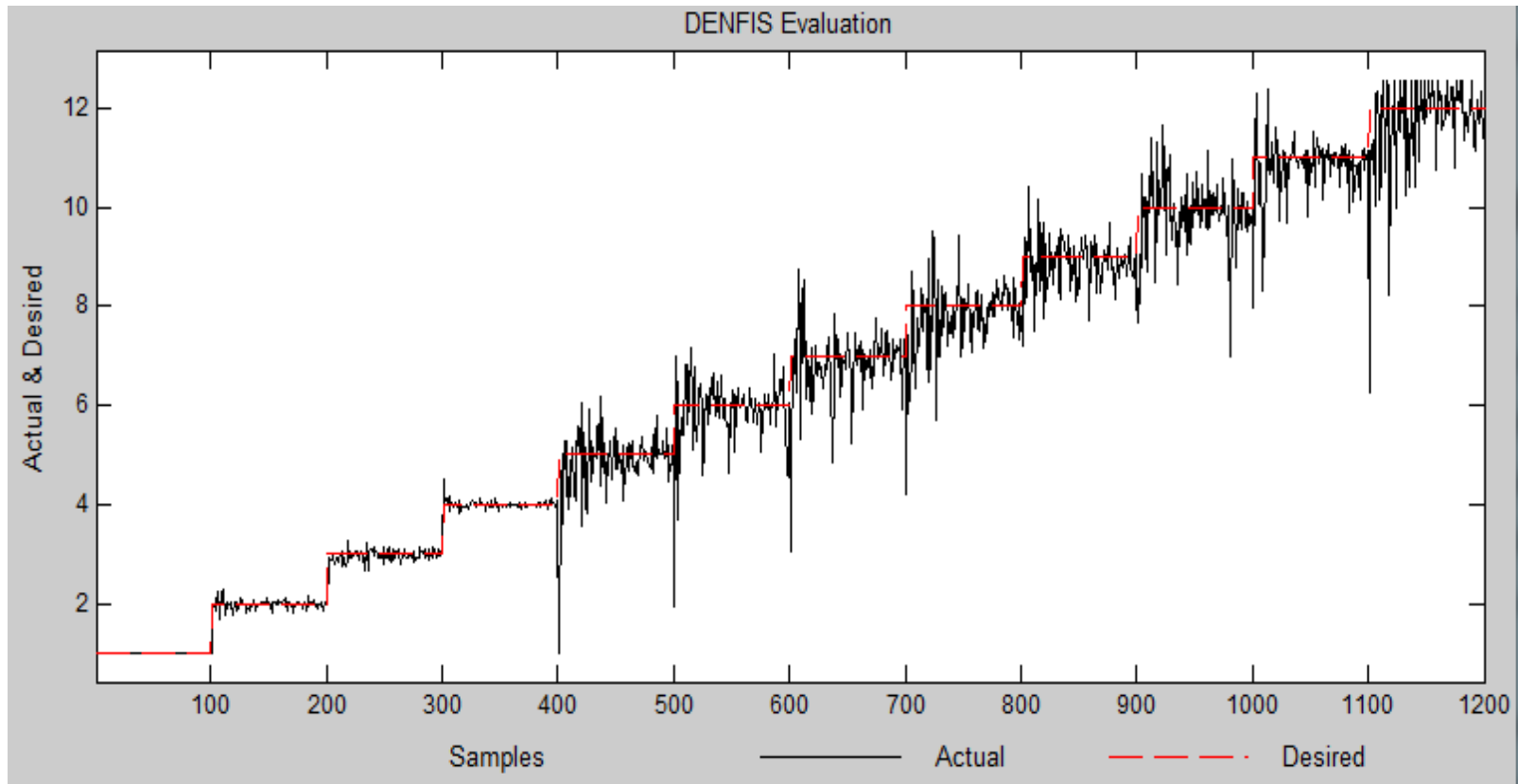
Fingerprinting method module of WLAN indoor positioning system

DENFIS Simulation



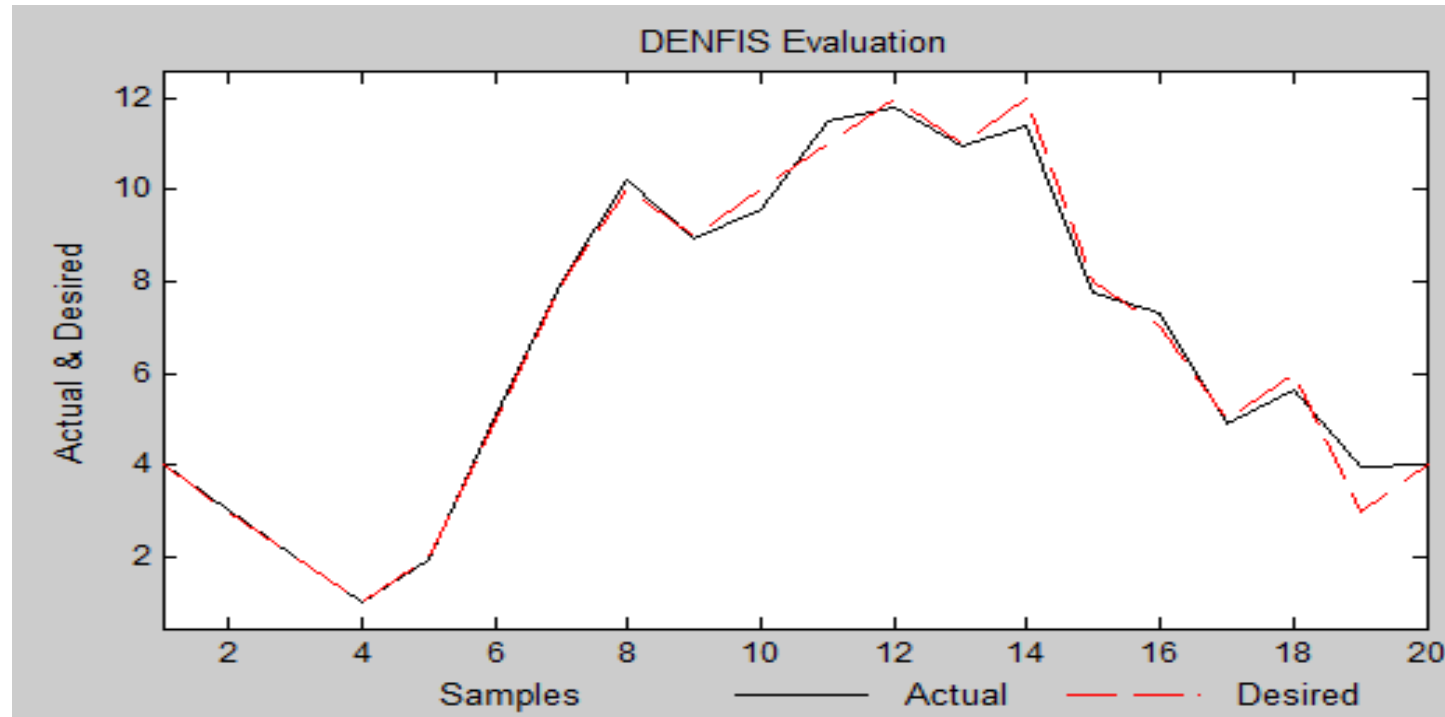
- Experimental environment (Area= $16.2 \times 9.5 \text{ m}^2$, divided into 12 small regions, 3 WiFi routers)
- 100 reference points (RPs) with their RRSs for each small region are picked up during the offline training phase
- 1200 sets of RRSs data for all the 12 regions are stored in the database

DENFIS offline training result



- 523 fuzzy rules are generated in the training process
- The average Root-Mean-Square Error (RMSE) is 0.594
- Total training time is 18.5s

DENFIS online evaluation result



- 20 sample points with their corresponding RRSs are tested based on the training result which is obtained in the offline phase.
- The estimated location regions (black line) based on DENFIS match quite well with the real location regions (red line) in general.

WP 1.4 Sensing, Modeling, estimation of occupancy, human activity

Principal Investigator: YC Soh (Singapore)

Other investigators: WJ Cai, Xie Lihua, Weng Khuen HO, Keck Voon LING,
Qing-Guo WANG, Jinming Xu, Zhenghua

Goal of the WP: to leverage sensing advances for energy monitoring,
and activity modeling.

- Collect more data points and more often
- Collect feedback from occupants

Algorithmic Advances

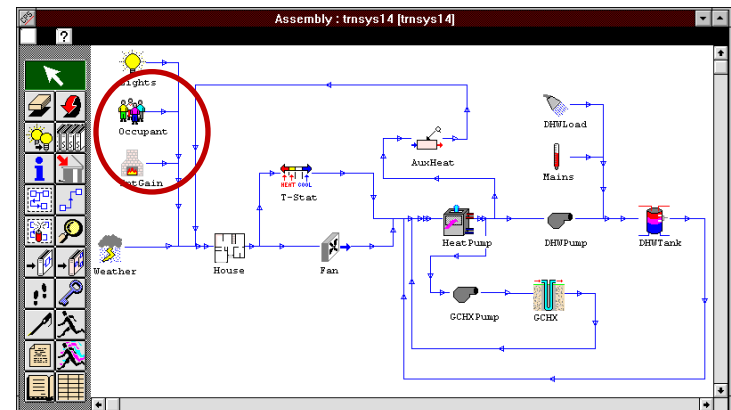
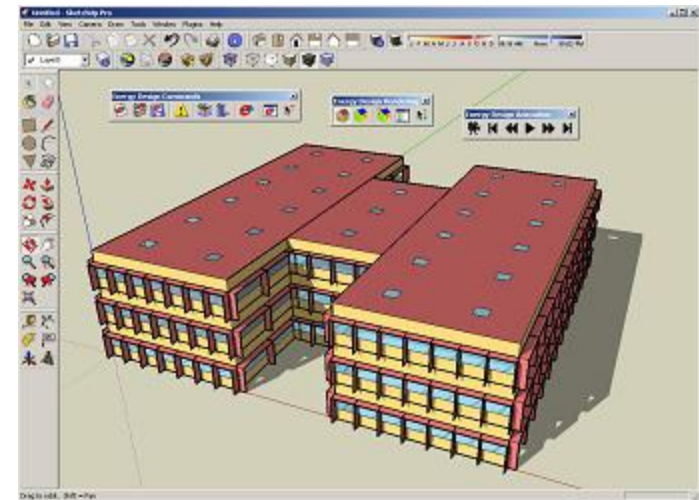
- Turn collected data into useful signals.

Theoretical Advances

A complete model of the office space from input (energy use) to output
(occupant comfort)

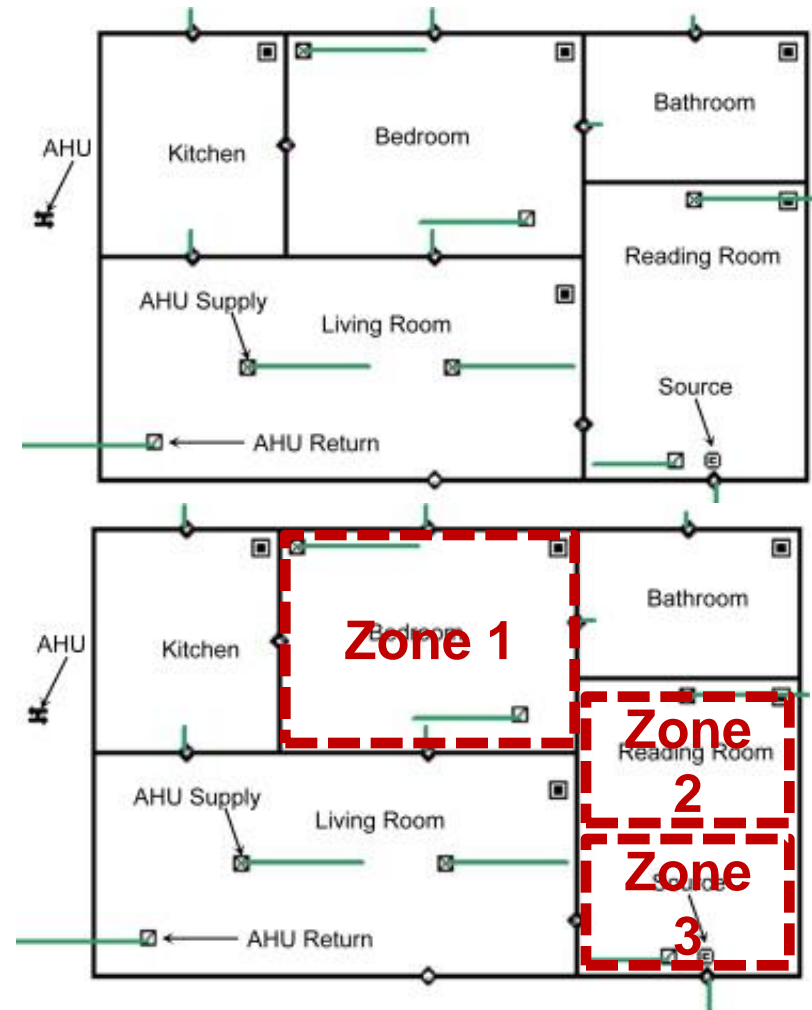
Motivations

- Serve as Inputs to some energy simulation tools(e.g., *EnergyPlus*, *ESP-r* and *TRANSYS*) for evaluation of building (and HVAC) systems in terms of energy efficiency.
- Facilitating demand-driven control strategies, e.g., *calculating the internal heat gain induced by human body and the associated appliance use, and thus the cooling load for each separated zone.*
- For occupancy estimation and prediction (e.g., *when using MPC strategies*)
- Crowd simulation, e.g., *emergency egress*



Occupancy Modeling Problems

- **ZONE:** Kitchen, bedroom and reading room - each is regarded as a separated zone. A Zone here refers to a general term that can also be used to represent an area that is not physically isolated, e.g., reading room can be further divided into two virtual zones.
- **AIM:** To construct a model that is capable of describing the real-time occupancy distribution of a building —the number of people in each separated zone over each time interval.



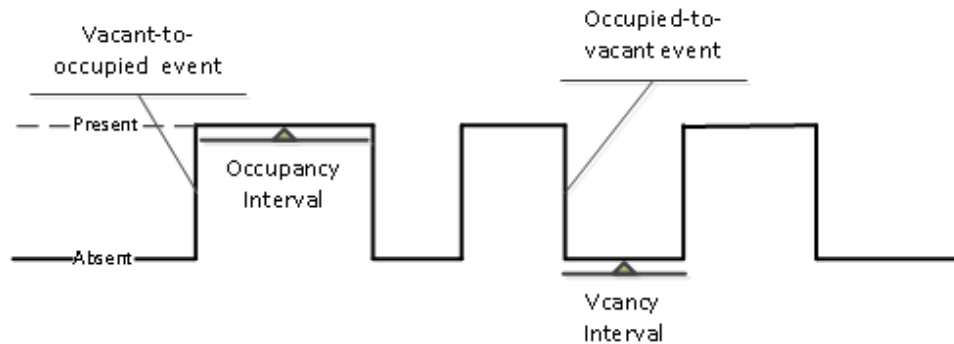
■ Single Person Single Zone

- **Probabilistic Model** (*Wang et al [3] - the clock-time information*) is not well treated.)
- **Markov Chain** (*Chuang et al [10] and Page, et al [9]*).
- Agent-based Model (*Liao, et al [6,7]*).
- Neural Network (*Dong, et al [4,5]*).

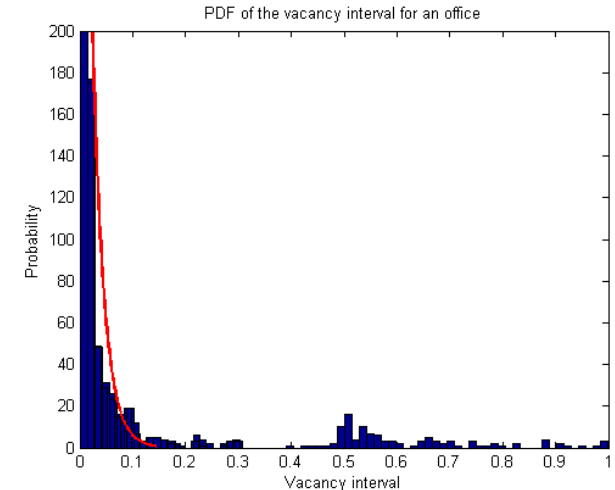
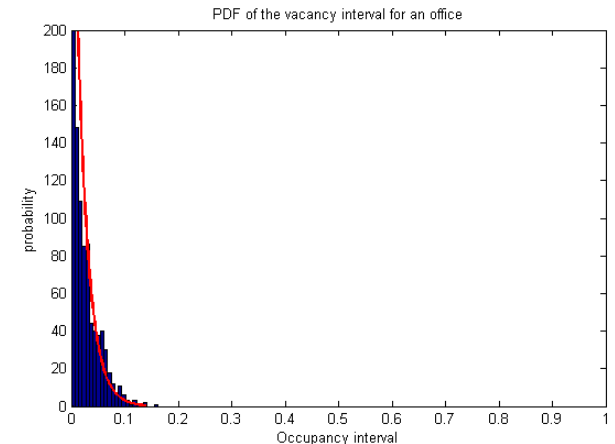
• Multi-Person Multi-Zone

- **Agent-based Model** (*Liao, et al [6,7] - performance is not good*)
- **Neural Network** (*Dong, et al [4,5] - much dependent on the training data and suffers from over-fitting problem*)
- **Time-varying Markov Chain** (*Page, et al [9] - computational prohibitive*)
- **Non-homogeneous Markov Chain** (*Chuang et al [10] - simple but does not fully account for the heavy-tail issue*)

Some Observations (*based on Robinson's data*)

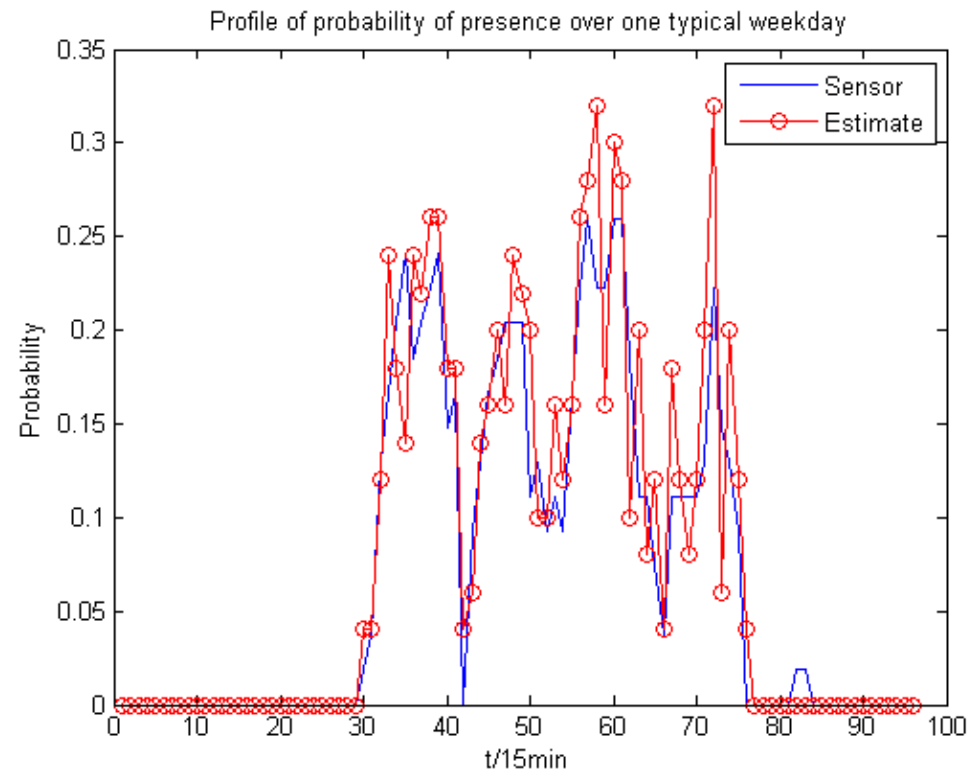


- *The horizontal axis is the time interval with 1 representing 24hs and the vertical axis represents the frequency of the interval falling within each bin.*
- *The histogram in the right, upon proper scaling, can be regarded as the probability distribution of the length of occupancy and vacancy interval over a recorded period.*



Observation: Either Occupied-to-Vacant or Vacant-to-Occupied event can, to some extent, be modeled as Poisson Process! Heavy-tail issue!

- Can reproduce key properties of occupants, such as
 - Times of arrival
 - Times of departure
 - Periods of absence
 - Periods of Presence
- Unable to deal with MPMZ Problems



Future Explorations

- **ARIMA Model**
 - **Adv:** provide good fit to the time series to be modeled
 - **Disadv:** difficult to guarantee the performance of prediction and no tools available to properly interpret certain phenomenon.
- **Event-driven Model**
 - Regard each primitive event as Poisson process
 - Time series of occupancy is triggered by nothing but the combination of all daily primitive events, e.g., sending an email.
- **Exploring other techniques**
 - reinforcement learning,
 - **Markov-modulated Non-homogeneous Poisson process** (Ihler et al[1]),
 - **Priority Queuing systems**(Albert-La et al, 2005 from Nature[2]); **Pareto** or Power-law Distribution.

Questions

Backup slides

- [1] A. **Ihler**, J. Hutchins, and P. Smyth, "Adaptive event detection with time-varying poisson processes," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 207-216.
- [2] A. L. **Barabasi**, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, pp. 207-211, 2005.
- [3] D. **Wang**, C. C. Federspiel, and F. Rubinstein, "Modeling occupancy in single person offices," *Energy and Buildings*, vol. 37, pp. 121-126, 2005.
- [4] B. **Dong**, C. Cao, and S. E. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy and Buildings*, vol. 37, pp. 545-553, 2005.
- [5] B. **Dong** and K. P. Lam, "Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network," *Journal of Building Performance Simulation*, vol. 4, pp. 359-369, 2011/12/01 2011.
- [6] C. **Liao**, Y. Lin, and P. Barooah, "Agent-based and graphical modelling of building occupancy," *Journal of Building Performance Simulation*, vol. 5, pp. 5-25, 2012/01/01 2011.
- [7] C. **Liao** and P. Barooah, "An integrated approach to occupancy modeling and estimation in commercial buildings," 2010, pp. 3130-3135.
- [8] D. **Robinson**, "Some trends and research needs in energy and comfort prediction," in *Windsor Conference*, 2006.
- [9] J. **Page**, D. Robinson, N. Morel, and J. L. Scartezzini, "A generalised stochastic model for the simulation of occupant presence," *Energy and Buildings*, vol. 40, pp. 83-98, 2008.
- [10] **Chuang** Wang, D. Yan, and Y. Jiang, "A novel approach for building occupancy simulation," *Building Simulation*, vol. 4, pp. 149-167, 2011/06/01 2011.

Appendix: Initial plan for budget (Singapore side)

| WP | Manpower resource requested | Manpower for year 1 | Equipment/devices/software |
|---|---|---------------------|----------------------------|
| Sensing, Data mining, and Modeling (Costas & Alex) | 2 RFs | | |
| Smartphone based participatory sensing | 1 PhD 1 RF | | \$40K |
| Embedded sensing and devices | 1 PhD 1 RF | | \$60K |
| Integration of CO2 sensing in the sensing system | 1 PhD 1 RF (50% from T4?) | | \$30K |
| Distributed sensing and cooperative control | 1 RA 1 PhD 1 RF (50% from T2?) | | \$60K |
| Modeling and estimation of occupancy & human activity | 3 PhDs 2 RFs | | \$30K |
| First principle modeling | 2 PhDs | | \$60K |
| TOTAL | 1 RA 9 PhDs 8 RFs | | \$250K |