Lawrence Berkeley National Laboratory

Recent Work

Title

Linking energy-cyber-physical systems with occupancy prediction and interpretation through WiFi probe-based ensemble classification

Permalink

<https://escholarship.org/uc/item/9b52v37f>

Authors

Wang, W
Hong, T
Li, N
et al.

Publication Date

2019-02-15

DOI

10.1016/j.apenergy.2018.11.079

Peer reviewed

1Linking energy-cyber-physical systems with occupancy predication 2and interpretation through WiFi probe-based ensemble classification

3 Wei Wang ^a, Tianzhen Hong ^b, Nan Li ^c, Ryan Qi Wang ^d, and Jiayu Chen ^{e1}

4^a School of Architecture, Southeast University, 210018, Nanjing, Jiangsu, China

5^b Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory,
6Cyclotron Road, Berkeley, CA 94720, USA

7^c Department of Construction Management, Tsinghua University, 100084, Beijing, China

8^d Department of Civil and Environment Engineering, Northeastern University, 433 SN, 360
9Huntington Avenue, Boston, MA 02115, USA

10^e Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621,
11AC1, Tat Chee Ave, Kowloon, Hong Kong

12

13Abstract:

14With rapid advances in sensing and digital technologies, cyber-physical systems are
15regarded as the most prominent platforms to improve building design and
16management. Researchers investigated the possibility of integrating energy
17management system with cyber-physical systems as energy-cyber-physical systems to
18promote building energy management. However, minimizing energy consumption
19while fulfilling building functions for energy-cyber-physical systems is challenging
20due to the dynamics of building occupants. As occupant behavior is one major source
21of uncertainties for energy management, ignoring it often results in energy wastes
22caused by overheating and overcooling as well as discomfort due to insufficient
23thermal and ventilation services. To mitigate such uncertainties, this study proposed
24an occupancy linked energy-cyber-physical system that incorporates WiFi probe-
25based occupancy detection. The proposed framework utilized ensemble classification
26algorithms to extract three types of occupancy information. It creates a data interface
27to link energy management system and cyber-physical systems and allows automated
28occupancy detection and interpretation through assembling multiple weak classifiers
29for WiFi signals. A validation experiment in a large office room was conducted to

11*Corresponding author. tel.: +852 3442 4696; fax: +852 3442 0427.

2e-mail addresses: jiayuchen@cityu.edu.hk (Jiayu Chen)

3

4

30examine the performance of the proposed occupancy linked energy-cyber-physical
31systems. The experiment and simulation results suggest that, with a proper classifier
32and occupancy type, the proposed model can potentially save about 26.4% of energy
33consumption from the cooling and ventilation demands.

34

35**Keywords:** Energy-Cyber-Physical Systems, Building occupancy, Wi-Fi probe
36technology, ensemble algorithm

Nomenclatures

TPM	Transition probability matrix of one occupant x_k	G_{other}	Load from other potential sources
x_k^{i-o}	Probability that occupancy status transfers from “in” to “in” or “out”	Q_r	Load of room r
x_k^{i-i}			
N_{i-i}	Frequency that occupancy status transfers from “in” to “in” or “out”	E_r	Energy cost to satisfy the cooling load at room r
N_{i-o}			
x_k^{Mac}	MAC address of occupancy x_k	m_r	Total supply air flow rate
$X(t)$	Input feature vector at time t	T_s	Supply air temperature
Y	Actual occupancy vector	$m_{OA,r}$	Outdoor air flow rate of room r
$F(x)$	Ensemble occupancy algorithm function	R_p	Outdoor air requirement for each occupant
$f_m(x)$	Meta occupancy algorithm function	P_r	Total number of occupants
m			
w_m	Weight value of function m	R_a	Outdoor air requirement for per area
L	Loss function	A_r	Total floor area of room r
Q_{nor}	Non-occupant-related load	$E_{ven,r}$	Energy use for ventilation of room r
Q_i	Occupant-related load	$Q_{vent,i}$	Ventilation load of room r
$Q_{inf,r}$	Heat gains from infiltration of room r	h_{OA}, h	Enthalpy value of outdoor and room air
$Q_{surf,r}$	Heat gains from surface of room r	$p_{pred,A}^r$	Prediction value of occupancy type A
$m_{inf,r}$	Flow rate of the infiltration air	t_0	Time resolution of the occupancy
C_p	Specific heat capacity of air	T	Length of the averaging time window
$T_{i,r}$	Temperature of room r	TP	Number of true positives
T_{air}	Temperature of outdoor air	TN	Number of true negatives
$A_{surf,r}$	Surface area of room r	FP	Number of false positives
K_{surf}	Heat transfer coefficient of surface	FN	Number of false negatives
G_p	Heat gain from per occupant	BM	Baseline model
G_{eq}	Load from equipment	OLE	Occupancy-linked e-CPS model
		M	

37

38

391. INTRODUCTION

40Buildings consume more than 40% of primary energy among all energy-consuming
41sectors [1] and energy bills become the largest overhead in building maintenance and
42operation budget. An increasing number of building owners and decision makers
43recognize promoting building energy efficiency as the most cost-effective approach
44for conservation. In modern buildings, the majority of energy is consumed by the
45mechanical/facility systems, which consists of heating, ventilation, air-conditioning
46(HVAC), lighting, water, safety, and similar allied subsystems. However, promoting
47energy efficiency of these facility systems is extremely challenging, as they usually
48have to comply with complicated working conditions, comfort requirements, and
49dynamic energy demand. In recent years, researchers propose to integrate both
50physical building systems with engineered cyber models so that building systems can
51be monitored, coordinated, controlled, optimized with a computing and
52communication core [2]. The integrated system is able to model, visualize, and
53operate complex building systems with various computing tools, and such systems are
54called cyber-physical systems (CPSs). With advances in the sensors, sensor networks,
55and embedded computing systems, CPSs unlocked the potential of optimizing
56building energy systems, such consolidated system is called energy-cyber-physical
57systems (e-CPSs) [3]. The ideal e-CPSs are designed to reduce the power demand
58through computational optimization so that the demand can be satisfied by the
59available power with minimum waste [3]. In this context, strategies were developed to
60optimize building facility operation through frequency control, voltage control, or
61sleep state scheduling [4]. However, dynamic demand caused by occupants and
62distributed operation cause poor system coordination in the centralized control system
63[5]. The physical facility systems require the computational outcomes from cyber
64model to optimize their operation, but the biggest challenge is the unreliable and
65incorrect demand estimation, which often results in energy wastes or unsatisfied
66thermal comfort.

67Therefore, a well-integrated e-CPSs should ensure reliability of demand information,
68which is usually captured by the physical system. With accurate and meaningful data
69inputs, the cyber model can provide effective operation suggestions. However, a
70building's energy demand is mainly generated by occupants' thermal, lighting, and

functional requirements, which are extremely dynamic and difficult to be captured by the physical building system. Conventional e-CPSs can synchronize physical mechanical and energy management systems with digital models, but they lack the ability to response to uncertain demand of occupants. Due to this constraint, in practice, conventional e-CPSs usually are rigid and static systems that based on certain assumed operation schedules. To fill this research gap, this study proposes to implement structured occupancy information to bridge the cyber and physical systems and form a new occupancy linked e-CPSs. Such system incorporates WiFi probe technology and interpreters that are based on ensemble Wi-Fi signals classifiers. The WiFi probe infrastructure on the physical model side and the ensemble signal classifiers on the cyber model side can be integrated and bridged by the accurate and reliable occupancy estimation. With such occupancy information, accurate demand can be estimated and the facility operation can be optimized for the energy saving purpose.

The rest of the paper is organized as follows. Section 2 reviews related works, including energy-cyber-physical systems (e-CPSs) studies and buildings. Section 3 introduces the framework and quantitative occupancy linked e-CPSs. Section 4 describes the validation experiment. Section 5 presents the results of experiment and simulation. Section 6 discusses the implication and limitation of this study, and Section 7 concludes this study.

91

92 **2. BACKGROUND**

93 **2.1 Energy management and cyber-physical system**

With the increased capability and decreased cost of wireless sensors, CPSs are capable of capture various building information through efficient networks and abundant computing powers. Thus, researches proposed to develop CPSs for building energy management systems in future smart buildings [6]. Kleissl and Agarwal looked at modern smart buildings entirely as a cyber-physical energy systems and examined the opportunities with joint optimization of energy use by occupants and information processing equipment [7]. Balaji et al. explored two case studies on smart buildings and electric vehicles to examine the feasibility of implementation of CPSs for energy

102management [8]. Zhao et al. developed a conceptual scheme for CPSs based energy
103management in buildings that combines the building energy information system, net-
104zero energy system, and demand-driven system [9]. Paridari et al. proposed a cyber-
105physical-security framework that also includes building energy management system
106(BEMS) with resilient policy and security analytics [10]. Based on upon these efforts,
107researchers concluded that e-CPSs is one of most prominent platforms in promoting
108building efficiency by introducing energy management into the cyber-physical
109interaction loop.

110Current research on e-CPSs mainly focuses on framework design and data-driven
111control. For the framework design studies, researchers integrate building information
112models (BIM) [11] and energy simulation programs [12], such as Modelica [13] or
113EnergyPlus [14], with physical sensor networks. For example, Delwati et al.
114compared the design features of the demand-controlled-ventilation methods with
115Modelica and proposed guidelines for building ventilation designers [15]. Hong et al
116simulated variable refrigerant flow systems with EnergyPlus and tested the model
117with typical houses in California [16]. Grigore et al. studied a case of deploying an e-
118CPSs for thermal optimization through electrical load monitoring, forecasting, HVAC
119control, and smart grid integration [17]. Behl et al. proposed an open source e-CPSs,
120DR-Advisor, which also allows data-driven modeling and control with rule-based
121algorithms. Based on a comparison with DOE commercial reference buildings, their
122system showed a 17% energy saving [18]. For the data-driven thermal control studies,
123researchers focus on converting physically captured data to system operation schedule
124and settings. For example, Ferreira et al. utilized neural network to implement
125predictict control to imporve thermal comfort in public buildings [19]. Costanzo et
126al.employed reinforcement learning tool to develop data-driven control for heating
127systems [20].

128As the premise of effective e-CPSs is to ensure human-centric services (e.g. thermal
129comfort, visual comfort) while saving as much as possible energy, researchers
130recognized that occupancy information played a central role to guarantee the e-CPSs'
131performance in smart buildings [21]. Latest studies suggest that accurate occupancy
132information not only links the physical building systems and cyber models but also
133mitigates the discrepancies between the designed/simulated and the actual building

operation performance [22]. Menezes et al. conducted a comprehensive study on the non-domestic buildings and concluded that occupancy information is significant to building energy and occupancy comfort benchmarking [23]. Liang et al. also stated occupancy data should be included to improve accuracy of building energy use predicting since occupancy is highly correlated with energy use and thermal comfort [24]. Wang et al. applied neural networks and WiFi technology to predict occupancy and integrate it to efficient building HVAC control and save 20% energy through avoiding overheating and overcooling [25]. Barbeito et al. assessed occupant thermal comfort and energy efficiency in buildings using statistical quality control (SQC) with integrated big data web energy platform [26]. Zhang et al optimized ventilation systems to satisfy occupant thermal comfort and saved 7.8% of total energy consumption [27]. Korkas et al. proposed a study of matching energy generation and consumption with occupant behavior to guarantee occupant thermal comfort and developing demand response in microgrids with renewable energy sources [28]. Chen et al. applied occupant feedback based model predictive control (MPC) for thermal comfort and energy optimization and proposed a novel dynamic thermal sensation model, saving 25% of energy use while maintaining thermal comfort level [29]. Lim et al. discussed occupant visual comfort in office spaces based on occupants' behaviors and reported 33.39% of lighting energy saving [30]. Shen et al. integrated lighting control strategies with occupancy state to guarantee visual comfort and resulted in a 48.8% saving [31].

155

1562.2 e-CPS and occupancy information

Usable and efficient building cyber models require a good understanding of occupants' energy demand and meaningful inputs from physical building systems [32,33]. Many studies suggested that the actual energy consumption of physical buildings severely deviates from the estimations of cyber models due to incorrect estimation of occupancy behavior [34]. Significant discrepancies between actual and estimated energy performance have been observed due to the complicated interrelationship between occupancy and building facility operation and the uncertainty of human behavior [35]. Oldewirtel et al. investigates the potential of using occupancy information to realize a more energy efficient building climate

control and in the simulations with alternating occupancy, the savings are in the range of 50% of the savings with homogeneous occupancy [36]. Hong et al. discussed ten questions concerning occupant behavior and building energy performance [37]. The International Energy Agency (IEA) Energy in Building and Community (EBC) Programme Annex 66 also highlighted and concluded that occupancy and occupants' behaviors are the most significant role for various research of enhancing building performance and human-centric services [38]. However, both physical building and cyber model are seldom changed in CPSs after the building has been built and the system uncertainties mainly arise from dynamic occupants' behavior and weather conditions. Many studies concluded that the occupancy information is one of the most significant considerations in energy conservation or low energy building design [39,40]. Therefore, as occupancy is the most critical data sources in energy demand estimation, e-CPSs should allow accurate and reliable occupancy information exchange between the physical system and cyber model.

Real opportunities for improving current e-CPSs exist where sensors, Information and Communication Technology (ICT), and data analytics can provide real-time occupant-related energy demand to guide building operation. Due to the complicated interrelationship of the energy consumption in building facilities and occupant behaviors [35,36,41], implementing occupancy information to improve building energy efficiency has been proven a feasible and cost-effective approach. For example, Kim et al. employed occupancy in simulation models and significantly reduced the deviated plug-load estimation [42]. Yang et al. investigated energy consumption of three institutional building in Singapore with the variability of daily occupancy and additional occupancy due to visitors [43]. Yang and Becerik-Gerber reported in their studies that the occupancy profiles-based operation schedule and room assignment can reduce 8% of HVAC energy use [44]. Pisello et al. suggested human-based energy retrofits can effectively promote energy efficiency in residential buildings with simulated post-occupancy information [45]. Chen et al. utilized occupancy information to visualize and validate the impact of occupants' behavior on commercial buildings [46].

To acquire occupancy information, researchers have proposed various methods. Jin et al. detected occupancy information through environmental sensing based on proxy

198measurements, such as temperature and CO₂ concentrations, and achieved 0.6044
199mean squared error and 55% ventilation cost reduction [47]. Other researchers
200focused on using smart meters to infer occupancy presence when no data or limited
201data is available and reported a detection accuracy of 93% for residences and 90% for
202offices, respectively [48]. On the other hand, Radio frequency identification (RFID)
203can be applied for indoor occupant positioning, e.g. Weekly applied RFID based
204sampling importance resampling particle filtering algorithm for occupant positioning
205in a real office and achieved an accuracy of 50% estimates within 3 m range and 90%
206estimates within 5 m range [49]. WiFi networks are the most preferable infrastructure
207in existing buildings, since they are efficient, affordable, and convenient [50]. In
208addition, WiFi access points are usually pre-installed in most modern buildings and
209multiple networks can cross-reference each other. The occupants' smartphones can
210serve as signal receivers or tags by measuring the signal strength indicators (RSSI)
211and hardware addresses. Thus, with these considerations, researchers developed
212various WiFi-based occupancy approaches to optimize HVAC operation [51]. For
213example, Chen et al. showed the number of Wi-Fi connections have a positive
214relationship with building energy consumption [52]. Balaji utilized WiFi networks
215and smartphones to adjust HVAC operation setting and achieved a 17.8% electricity
216saving [53]. Jin et al. proposed a PresenceSense research with data collection through
217multiple sensing sources, including ultrasonic sensors, acceleration sensors, and WiFi
218[54]. Zou et al. proposed a non-intrusive occupancy sensing system, called WinOSS,
219to count WiFi-enabled mobile devices, which can achieve 98.85% occupancy
220detection accuracy when occupants stay stationary [55]. Zou et al. claimed
221implementing Internet of Things (IoT) technologies the counting accuracy can be as
222high as 99.1% [56].

223Although many researchers recognized that the key of e-CPSs to promote building
224energy efficiency is integrating occupancy information, the interface to bridge sensing
225outcomes and e-CPS platform remains unfeasible. Inspired by previous researches,
226this study intends to develop a quantitative framework to interpret dynamic WiFi
227signals as useful occupancy schedules and profiles for cyber energy models. To
228achieve this goal, this study proposed an occupancy linked e-CPSs model (OLEM) to
229take advantage of existing Wi-Fi infrastructure in buildings and to incorporate

ensemble classification algorithm for occupancy detection and predication. The proposed OLEM utilized three occupancy data formats as interface and WiFi probe technology toolset to bridge energy management system and CPSs.

233

234**3. METHODOLOGY**

235**3.1 Occupancy linked e-CPSs**

A fundamental e-CPSs framework includes at least a physical building system and cyber model for energy management and optimization. The physical building model reflects the actual conditions and performance of a building while the cyber model is a digital twin that can be used for various computational processes. The physical buildings usually have sensors and sensor network installed which allows acquiring the various types of environmental information, such as temperatures, CO₂ concentration, and relative humidity (RH), and system operation information, such as supply/outdoor air flow rate and temperature, pump efficiency, and instantaneous energy load. The building information model is the key to associate both components and to create a dependable digital twin for the actual building. The building information model contains static features and dynamic operation settings. The static features include building materials, geometry, location, system type, and etc., while the dynamic operation settings include the operation schedule, efficiency, and settings of HVAC, lighting, and security systems.

To extend conventional e-CPSs, this study proposes to integrate dynamic occupancy information to enable data exchange between the physical building and cyber model. As the physical infrastructure of the building system, Wi-Fi networks were utilized to obtain the signal strength of occupants' device/tag. The obtained occupancy information serves as the inputs for a cyber model for data analysis and system optimization. To connect both components of e-CPSs, this study also developed an occupancy interpreter based on ensemble algorithms to convert Wi-Fi signal strengths to occupant number and schedule. Once detailed occupancy information is captured, the cyber model can conduct energy simulation with the building information model and suggest proper operational settings for the facility/mechanical systems. The Figure 1 shows the structure of the proposed occupancy linked e-CPSs.

* Conventional e-CPSs

261 * Occupancy linked e-CPSs

Fig. 1. The scheme of the occupancy linked e-CPSs.

2643.2 Wi-Fi Probe-based ensemble learning algorithm for occupancy prediction

265 This study proposes to utilize Wi-Fi probes as the active detector for occupants
 266 (occupants are assumed to have a smartphone or tag with the capacity of Wi-Fi
 267 connection) and the proposed prediction algorithm implements a set of ensemble
 268 algorithms. The algorithm serves as the occupancy interpreter to convert received Wi-
 269 Fi signal strengths to the number and residency patterns of occupants and send the
 270 results as the inputs for energy simulator. The process of data interpretation includes
 271 three steps: (1) Feature extraction; (2) Ensemble learning; and (3) Occupancy pattern
 272 matching. Figure 2 shows a simplified process of the proposed algorithm.

273

274 Fig. 2. The process of the Wi-Fi Probe-based ensemble learning for occupancy.

275

2763.2.1 Feature extraction

277The appearance of occupants in a building space shows a strong stochastic
 278characteristic [57], thus, the occupancy prediction is usually modeled as a Markov
 279process [58,59], in which current occupancy status depends on previous occupancy
 280status. For example, the probability of an occupant leaves a space only feasible when
 281he/she is already in the space. Therefore, the feature extraction step models an
 282occupant status in a given space as “in” or “out” and the transfer probability and
 283transition matrix of the Markov process can be modeled as

$$TPM|_{x_k} = \begin{bmatrix} x_k^{i-o} & x_k^{i-i} \\ x_k^{o-o} & x_k^{o-i} \end{bmatrix} \quad (1)$$

Where $TPM|_{x_k}$ represents the transition probability matrix of one occupant x_k . In the transition matrix, x_k^{i-o} and x_k^{i-i} denote the observed probability that one occupant whose status is “in” at the current time would be “out” or still “in” at the next time. x_k^{o-o} and x_k^{o-i} denote the observed probability that one occupant whose status is “out” at the current time would be “out” or “in” in the next time interval. The probability can be computed with an observed conditional probability based on Bayesian models.

$$x_k^{i-i} = P(\text{observed state} = i | \text{observed state} = i) \quad (2)$$

Therefore, the occupied probability of one media access control (MAC) address is

$$x_k^{i-i} = \frac{\sum N_{i-i}}{\sum N_{i-i} + \sum N_{i-o}} x_k^{o-o} = \frac{\sum N_{o-o}}{\sum N_{o-o} + \sum N_{o-i}} \quad (3)$$

Where N_{i-i} is the frequency in which the occupancy status transfers from “in” to “in”. N_{i-o} is the frequency in which the occupancy status transfers from “in” to “out”. Similarly, N_{o-o} and N_{o-i} represent the frequencies in which the occupancy status transitioned from “out” to “out” and from “out” to “in”, respectively. With an assigned probability for MAC addresses in the room. Each MAC address is formatted as

$$x_k = \{x_k^{Mac}, x_k^{o-i}, x_k^{i-i}\} \quad (4)$$

Then, suppose there are n occupants at one time spot t, then input feature vector at time can be as

$$X(t) = \{x_1^{Mac}, x_1^{o-i}, x_1^{i-i}, \dots, x_k^{Mac}, x_k^{o-i}, x_k^{i-i}, \dots, x_n^{Mac}, x_n^{o-i}, x_n^{i-i}\} \quad (5)$$

300

3013.2.2 Ensemble learning algorithms

There are main families of ensemble methods. The first method is averaging, which builds several estimators independently and then average predictions through minimizing their prediction variance, such as Bagging methods and Forests of

Randomized Trees. The second method is boosting, which builds sequential estimators to reduce the bias by combining several weak models, such as AdaBoost and Gradient Tree Boosting. The ensemble learning algorithm in this study integrates multiple meta-estimators through boosting method.

Through feature extraction, raw data can be interpreted as an input vector of $(X(t), Y)$. Where Y is actual occupancy (label) as the learning object and $X(t)$ are extracted features in previous section. The ensemble learning is built upon numbers of multiple meta-estimators, which are usually simple and weak models, such as a decision tree. Decision tree uses a tree structure to create a model that predicts the value of a target variable based on several input variables. The tree can be learned by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset until the splitting no longer adds value to the predicting model. Figure 3 shows the structure of the ensemble learning for occupancy prediction. $X = \{x_1, x_2, \dots, x_N\}$ is defined as a set of N observations of Wi-Fi dataset inputs with associated output $Y = \{y_1, y_2, \dots, y_N\}$.

320

Fig. 3. The ensemble learning algorithm for occupancy prediction.

323

Suppose the ensemble outputs can be estimated from the aggregated results from multiple meta-estimators as:

$$F(x) = \sum_{m=1}^M w_m f_m(x) \quad (6)$$

Where $f_m(x)$ are the basis functions of meta-estimators. n is the index of meta-estimators and w_m is the weight parameter assigned to one meta-estimator. The

iterative form of above equation can be represented as:

$$F_m(x) = F_{m-1}(x) + w_m f_m(x) \quad (7)$$

w_m is the weight of the estimators. In each iteration, the decision tree $f_m(x)$ is chosen to minimize the loss function L given the current model $F_{m-1}(x_i)$.

$$F_m(x) = F_{m-1}(x) + \arg \min_f \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + f(x)) \quad (8)$$

Other than the regular decision tree, the meta-estimators can be substituted with other more complicated classifiers. This study also embedded three other ensemble algorithms (Gradient Tree Boosting classifier, Random Forest classifier, and Adaptive Boosting classifier) in the occupancy prediction model.

(1) Gradient Tree Boosting (GTB)

Gradient Tree Boosting (GTB) classifier is a generalization of boosting to arbitrary differentiable loss functions. GTB classifier can easily handle the mixed type of data and is robust to outliers with improved loss functions. GTB attempts to solve the minimization problem numerically via steepest descent, the direction of which is the negative gradient of the loss function.

The GTB algorithm generates a model, which combines multiple simple trees in sequence. The minimum error is achieved by searching the best split of trees. The simple process of GTB can be illustrated as:

- Initial predicted value is assumed for all observation in the datasets. Error is calculated using the assumed predictions and actual datasets.
- A decision tree model is created using the errors. Split the tree branches to search the minimal error.
- Model should be updated and be used to generate new predictions. New errors can be calculated with new predictions and actual datasets.
- Repeat this process till maximum number of iterations is reached or error converges.

354(2) *Random Forests (RF)*

355Random Forests (RF) is another ensemble machine learning algorithm that follows
356the bagging technique. The base estimators in random forest are decision trees. Unlike
357bagging meta estimator, RF classifier randomly selects a set of features which are
358used to decide the best split from the training set. By doing this, the sample bias can
359be eliminated and the best split among trees can be selected. With averaging, the
360variance of meta-estimators can be minimized, hence yielding a better model.

361The RF model create multiple trees for subsets of the whole dataset. Each tree is much
362smaller than that of GTB. The final classification is the aggregated results based on all
363trees. The minimum error is achieved by properly selecting trees for subsets. The
364process of a random forest algorithm can be summarized as:

- 365 • Random subsets are created from the original dataset (as bootstrapping).
- 366 • Formulate decision trees for subsets. At each node in the decision tree, only a
367 random set of features are considered for the best split.
- 368 • An optimized decision tree model is fitted for each subset for all features.
- 369 • The final predictions of the outputs are averaged from the predictions of all
370 decision trees.

371

372(3) *Adaptive Boosting (AdaBoost)*

373Adaptive Boosting (AdaBoost) classifier, one of the simplest boosting algorithms,
374implements multiple sequential rules (weak classifiers) on the meta-estimators. The
375predictions from all of the estimators are combined through a weighted majority vote
376(or sum) to produce the final prediction. For each successive iteration, the weights are
377individually modified and the learning algorithm is reapplied to the reweighted data.

378The AdaBoost uses rules to classify the inputs, and the final classification is the
379aggregated results based on all rules. Different from RF, AdaBoost assigns unequal
380weights to subsets. The minimum error is achieved by properly selecting rules and
381subset weights. Below is a brief summary of the process of performing the AdaBoost
382algorithm:

- 383 • Assign equal weights to all observations in the dataset.

- 384 • Rule models are built for subsets and compute the predictions for the whole
- 385 data set.
- 386 • Compute errors by comparing the predictions and actual data. Update the rule
- 387 models and assign higher weights for incorrectly predicted observations.
- 388 • Repeat above steps until errors are minimized.

389

3903.2.3 Occupancy pattern matching

391 Buildings consume energy to ensure the thermal comfort and indoor air quality for
 392 occupants. The energy load of a building can be categorized as non-occupant-related

393 load ($Q_{nor,i}$) and occupant-related load ($Q_{i,i,i}$) . The non-occupant-related load

394 comes from the heat transfer across the building envelope and outside environment,
 395 which highly depends on weather conditions. The total energy load can be roughly
 396 estimated as

$$Q_{nor,r} = Q_{inf,r} + Q_{surf,r} \quad (9)$$

$$Q_{inf,r} = m_{inf,r} * C_p * (T_{i,r} - T_{air}) \quad (10)$$

$$Q_{surf,r} = A_{surf,r} * K_{surf} * (T_{i,r} - T_{air}) \quad (11)$$

397 Where $Q_{inf,r}$, $Q_{surf,r}$ are the heat gains from infiltration and surface,
 398 respectively. $m_{inf,r}$ is the flow rate of the infiltration air; C_p is the specific heat
 399 capacity of air; $T_{i,r}$ and T_{air} are the temperature of a room and outdoor air,
 400 respectively; $A_{surf,r}$ is the surface area of a room; K_{surf} is the heat transfer
 401 coefficient.

402 The occupant-related load includes internal gain from occupants and equipment
 403 operated by occupants.

$$Q_{i,r} = \sum_{P_r} G_p + \sum_{P_{eq}} G_{eq} + \sum G_{other} \quad (12)$$

$$Q_r = Q_{nor,r} + Q_{i,r} \quad (13)$$

404 Where P_r is the number of occupants and G_p is the heat gain from per
 405 occupant. G_{eq} contains the load from computers, water heaters, lights etc.; P_{eq}

406 is the index of equipment; Q_r is the total cooling load of a room. At room level,
 407 the ventilation and air conditioning system should provide enough conditioned air to
 408 maintain proper indoor temperature and the air handling system should supply
 409 sufficient fresh air.

$$E_r = Q_r = m_r * C_p * (T_{i,r} - T_{s,r}) \quad (14)$$

410 Where E_r is the energy cost to satisfy the cooling load at room level. m_r is the
 411 total supply air flow rate. T_s is the supply air temperature.

412 In practice, American Society of Heating, Refrigerating and Air-Conditioning
 413 Engineers (ASHRAE) standards recommends minimum ventilation approach, which
 414 requires a rough estimation on the number of occupants. The suggested ventilation
 415 amount includes both a people component (to dilute contaminants from people and
 416 their activities) and an area component (to dilute contaminants from non-occupant-
 417 related sources that are more related to floor area than occupants) [60]. Outdoor
 418 airflow required in the breathing zone of the occupied space or spaces in a zone
 419 should be computed first.

$$m_{OA,r} \geq R_p * P_r + R_a * A_r \quad (15)$$

420 Then,

$$E_{ven,r} = Q_{vent,r} = m_{OA,r} * (h_{OA} - h_i) = m_{OA,r} * (f(T_{air}, H_{air}) - f(T_{i,r}, H_{i,r})) \quad (16)$$

421 Where $m_{OA,r}$ is the outdoor air flow rate of a room. R_p is the outdoor air flow
 422 rate requirement for each occupant. R_a and A_r are the outdoor air flow rate
 423 requirement for per area and the total floor area of room, respectively. $E_{ven,r}$ and
 424 $Q_{vent,r}$ are the energy consumption for cooling of ventilation. h_{OA} and h_i are
 425 the enthalpy value of outdoor air and room air, respectively. H_{air} and $H_{i,r}$ are
 426 the humidity of outdoor air and indoor air, respectively.

427 Based on above itemized energy loads, to match the system operation and energy
 428 simulation model, this study utilized three operation schedules based on different
 429 occupancy types. Figure 4 illustrates a typical occupancy schedule of each occupancy
 430 type. In the baseline simulation model, all other system operation settings, such as the
 431 supply air flow rate and outdoor air flow rate, are either set by facility managers or
 432 captured by sensors.

Fig. 4. Sample occupancy schedules for three occupancy types.

436(1) Type A occupancy

437Type A occupancy reports the continuous and exact occupancy information (number
438of occupants in a space) that estimated by the ensemble algorithm. The operative
439temperature and relative humidity settings are computed with ASHRAE standard
44062.1-2013 recommended thermal comfort based on the number of occupants. Then
441the minimum outdoor air flow rate can be computed accordingly.

$$m_{OA} = m_{pred.min}^{OA} = R_p * p_{pred.A}^r + R_a * A_r \quad (17)$$

442Where $T_{i,r} = T_{setting}$ and $H_{i,r} = H_{setting}$ are the temperature and humidity settings.
443 $p_{pred.A}^r$ is the predicted results of type A occupancy. $m_{pred.min}^{OA}$ is the minimum
444outdoor air flow rate based on such data type.

446(2) Type B occupancy

As the detected occupancy is often contaminated by random noise and the optimization for system operation is periodical, discrete occupant number with suitable time interval is preferable in many cyber energy models. In addition, fluctuations in occupancy could result in excessive adjustments. Therefore, Type B occupancy applies time window to average occupancy within its length.

$$p_r = p_{pred.B}^r = \frac{t_0}{T} * \sum_{i=0}^{T/t_0} x_i \quad (18)$$

Where $p_{pred.B}^r$ is the predicted occupancy. t_0 is the time resolution of the occupancy. T is the length of the averaging time window.

(3) Type C occupancy

Type C is a simplified categorical scale occupancy for the ease of system operation. In type C occupancy, the predicted results are divided into four levels, including zero, low, medium, and high. The mechanical system can switch between setting scenarios based on the building occupancy level.

In summary, the entire process of occupancy prediction with the ensemble algorithm is illustrated in Figure 2.

-
1. Feature abstraction from Wi-Fi dataset
 2. Define occupancy patterns
 3. Define Input $X = \{x_1, x_2, \dots, x_n\}$, Output $y = \{y_1, y_2, \dots, y_n\}$, a set of base estimators $F = \{f_1(x), f_2(x), \dots, f_M(x)\}$. Loss function L
 4. Select parameter in parameters tuning set
 - For $i = 1$ to M number of iterations):
 - (a). Compute residuals
 - (b). Fit pseudo-residuals using base estimator
i.e. set f_m to minimize $L(y, f_m(x))$
 - (c). Find multiplier, $w_m = \operatorname{argmin}_f \{L(y_i, F_{m-1}(x) + f_m(x))\}$
 - (d). Update $F_{m-1}(x) + w_m f_m(x)$
 - Output: occupancy model $F_m(x)$
 - Calculate assessment metric (MAE, RMSE)
 5. Output occupancy model to minimize assessment metric for occupancy patterns
 6. Output occupancy pattern file
-

463 Fig. 5. The pseudocode of the ensemble algorithm for occupancy prediction

464

465 **4. VALIDATION EXPERIMENT**

466 **4.1 Physical conditions of the experiment testbed**

467 To examine the proposed occupancy linked e-CPSs, this study also conducted a
468 validation experiment in a large office space. The testbed has an area of about 200
469 square meters and 20 long-term residents during the experiment period. Figure 6
470 shows the space layout and sensors setup. The room equipped with a dedicated
471 outdoor air system to bring outdoor air into indoor areas without air handling process.
472 The indoor air is conditioned by the fan coil unit with the variable refrigerant flow
473 and the indoor air circulation is driven by positive pressure. The entire room has Wi-
474 Fi coverage with three Wi-Fi probes. During the experiment, TA465-X sensor system
475 (produced by TSI Co.) was utilized to monitor the indoor air temperature, relative
476 humidity, and airflow rate. The CO₂ concentration of return air of the fan coil unit
477 was used to approximate the CO₂ concentration of the indoor air after air mixing. To
478 eliminate the uneven air mixing, three environmental sensors were evenly installed at
479 the ceiling (3m). Air flow meters were installed near outdoor inlets to monitor the air
480 flow rate of the ventilation system. Two overhead cameras were installed to record the
481 entrance and exit events of occupants. During the experiment, the occupants aware of
482 the Wi-Fi experiment and were instructed to switch on their Wi-Fi signal on their
483 mobile devices. Table 1 shows the specifications of the installed sensors, including
484 data storage types, sensing intervals, range, accuracy, and resolution. The experiment
485 lasted for nine days.

Fig. 6. Space layout and equipment setup.

Table 1. Sensors used in the experiment.

Sensors	Camera	Wi-Fi Probe	Environmental Sensors			
			Air flow rate	Temperature Sensors	Humidity Sensors	Other Sensors
Recorded Variables	Time, Actual occupancy	Time, MAC address, RSSIs	Time, Temperature, Relative humidity, Air flow rate, Air pressure			
Data Storage	Online	Online	Local			
Sensing interval		30s	1min	1min	1min	
Range			0 - 9999 ft/min	14 - 140 °F -10 – 60 °C	0 to 95%	
Accuracy			±3% or ±3 ft/min	±0.5°F (±0.3°C)	< 3%	
Resolution			1 ft/min	0.1°F (0.1°C)	0.10%	

4914.2 Cyber model for energy management and simulation

492Figure 7 shows the energy cyber model applied in this study. The model was
 493developed with EnergyPlus and DOE2 to optimize facility operation. Based on BIM
 494models, the energy cyber model is able to incorporate construction materials, building
 495geometries, and schedule of operation to estimate the energy consumption of the
 496building. With co-simulation with other programming languages, such as Matlab or
 497Python, the model is capable of tuning system settings to minimize energy
 498consumption. This study employed Eppy, a Python package that can manipulate
 499EnergyPlus IDF files [61], to search for the optimal system settings. It takes full
 500advantage of the rich data structure and idioms that are available in Python and
 501provide availability of designing expected energy model and algorithm to integrate
 502physical and cyber models. Eppy can help programmatically navigate, search, and
 503modify EnergyPlus IDF files. Users can use Eppy to create one or multi new IDF
 504files, make changes to original IDF files, change occupancy schedule in all the
 505interior zones, and read data from the output files after EnergyPlus simulation run.
 506Related to occupancy linked e-CPSs, Eppy provides an interface to link occupancy
 507results from ensemble models as the input to cyber energy model with Python.

Fig. 7. The energy cyber model for the experiment tested.

510The cyber model matches the physical room with a size of 20 m (length) x 10 m
 511(width) x 3 m (height) and 20 occupants. Internal heat sources were set as 75W for per
 512person, 150W for per computer, and 35W for per lamp. The light schedule followed
 513the on/off schedule and the schedule for computers was assumed to same as the

occupancy schedule. Hong Kong has a subtropical climate and high-density highrise urban form. According to statistics [62], the typical mean, minimum, maximum values of monthly average temperature are around 23.4°C, 13.3°C, and 29.8°C, respectively. Also, relative humidity (RH) of Hong Kong is high and minimum, maximum values of monthly average RH are 78.2%, 60%, and 90%, respectively. The typical Hong Kong weather condition was used and the heat transfers from wall, floor, and ceiling were ignored since the experiment was conducted in one inner zone adjacent to conditioned zones. The cooling temperature setpoint is 24°C and there was no heating.

523

4.3 Data processing

4.3.1 Actual occupancy information

To collect the ground truth for training the ensemble learning algorithms and assessing the model errors, two cameras were installed above the two entrances of the experiment testbed. The number of occupants was counted through video analysis based on the camera records. The counted numbers were synchronized with the internet timestamp with a five-minute interval. To match the Type C occupancy data, the number also was also translated to categorical occupancy levels as specified in Table 2.

Table 2. The threshold setting for categorical occupancy levels

Occupancy level	Number of people
Zero (0)	0
Low (25%)	1-6
Medium (50%)	7-14
High (75%)	15-20

534

4.3.2 Model parameters tuning

To improve the facility operation with reliable occupancy information, it is necessary to identify, compare, and optimize the ensemble model through parameter tuning. The training model implemented n-fold cross-validation method. In this study, the raw dataset has total 882 samples and about 70% of dataset was used for model training

and 30% for model validation and test. Table 3 shows the search space for the parameters tuning. The multi-variable comparison in the exhaustive grid search is applied to identify the best assembly of model parameters. For the RF classifier, the number of estimators determines the results precision and training time, while the number of features affects the accuracy and the diversity of results. For GTB and AdaBoost classifiers, learning rate affects the boosting step length of the gradient descent procedure.

Table 3. Parameters search space for the occupancy ensembled model

Algorithm	Parameter	Range
GTB	Number of estimators	[100; 150; 200; 250; 300; 400; 500; 600; 800; 1000; 1200]
	Learning rate	[0.01; 0.02; 0.05; 0.1; 0.2; 0.25; 0.3; 0.4; 0.5]
	Min_samples_split	[2; 3; 4; 5; 6; 8; 10; 15]
	Max_tree_depth	[3; 4; 5; 6; 7; 8; 9; 10; 12; 15]
AdaBoost	Number of estimators	[100; 150; 200; 250; 300; 400; 500; 600; 800; 1000; 1200]
	Learning rate	[0.01; 0.02; 0.05; 0.1; 0.2; 0.25; 0.3; 0.4; 0.5]
Random Forest	Number of estimators	[100; 150; 200; 250; 300; 400; 500; 600; 800; 1000; 1200]
	Max_features	['all'; 'sqrt'; 'log2']
	Min_samples_leaf	[1; 2; 3; 4; 5; 6; 7; 8; 9; 10]

548

4.3.3 Error assessment

To evaluate the effectiveness and accuracy of the model, both the mean average error (MAE) and root mean squared error (RMSE) metrics were used for Type A and Type B occupancy. For discrete Type C occupancy, the Accuracy (ACC) is defined with true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of the confusion matrix.

$$TPR = \frac{TP}{TP + FP} \quad (19)$$

$$TNR = \frac{TN}{TN + FN} \quad (20)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

Meanwhile, the value of the area under curve-receiver operating characteristic curve (AUC-ROC) is applied, which is created by the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. For the unbalanced dataset, Balanced Accuracy (bACC) can be used to average the TPR and TNR, which can be presented in the following formula:

$$bACC = \frac{TPR + TNR}{2} \quad (22)$$

According to ASHRAE standard 62.1-2013 [60], the fresh air volume of the ventilation system and the occupant-related thermal load of the air conditioning system are determined by the number of occupants. The errors in the occupancy assessment could directly affect the energy usage of the building. Therefore, the e-CPSs can be significantly improved with the occupancy information incorporated.

565

566 5. RESULTS

567 5.1 Environmental conditions

In the experiment field, dedicated outdoor air system and fan coil unit is under operation. The former system delivers the outdoor air to inner space directly without cooling and the latter cools indoor circulating air. Figures 8 show the environmental conditions during the experiment period. In Figure 8 (a), the outdoor air supply flow rate is 180 cfm (cubic feet per minute) for each outdoor air inlet consistently and the supply air flow rate for each supply air inlet is over 300 cfm but less than 400 cfm most of the time. The outdoor air was supplied uninterrupted during the night even if the cooling services from supply air terminals were closed. Figure 8 (b) shows that the measured supply air temperature varies periodically from 15°C to 25°C, which is caused by the periodical cycling operation of the fan coil system. During the experiment, the outdoor air temperature ranged from 30°C to 35°C, which is a typical summer day in Hong Kong. Figure 8 (c) reports the relative humidity.

580

581 Fig. 8. Environmental conditions of a typical experiment day (a) Air flow rate (top)

582

(b) temperature (middle) (c) relative humidity (bottom).

583

5845.2 Predicted occupancy

585 This study performed a grid search to determine optimal values for the parameters of
 586 the tree-based ensembles. The features of Wi-Fi dataset described in Eq. 5 were
 587 considered as the input variables. The GTB classifier consists of 150 estimators with a
 588 learning rate of 0.01. To split an internal node, the model requires a minimum 8
 589 samples and a maximum tree depth of 15. The AdaBoost classifier has 100 estimators

with a learning rate of 0.2. The RF classifier has 250 estimators and 10 minimum sample leaf. Table 4 summaries the averaged errors of all three type of classifiers after tuning. Among all three types of classifiers, the AdaBoost classifier shows the highest accuracy.

Table 4. Averaged errors for the three ensemble learning algorithms.

	RFs			GTB			AdaBoost		
	MAE	RMSE	Accu.	MAE	RMSE	Accu.	MAE	RMSE	Accu.
Type A	2.66	3.31		2.89	3.58		2.54	3.30	
Type B	2.63	3.32		2.81	3.53		2.41	3.06	
Type C			71.0%			66.0%			72.7%

Figure 9 presents the predicted results for all three occupancy types with the AdaBoost classifier. Type B occupancy used a 30 minutes sliding time window to smooth the predicted occupancy. Type C occupancy levels are categorized as zero, low, medium, high. The detailed error comparison by days is listed in Table 5 and Table 6 shows the normalized confuse matrix of AdaBoost classifier for Type C occupancy. From detailed assessment results, it shows Day 5 and 7 have the almost best accuracies for type A occupancy with 1.88 and 1.91 of MAE and 2.40 and 2.30 of RMSE respectively. For type B occupancy, Day 3 shows the best accuracy with 1.48 of MAE and 2.48 of RMSE. For the detailed accuracy of Type C occupancy, it can be found that Day 2, 4, 6, and 7 have no “Zero” level occupancy, while Day 3, 5, and 7 have no “High” level occupancy. The best accuracy is shown on Day 7, where accuracies are 61.1% for “Low” and “Medium” levels occupancy, respectively. The total accuracy for Type C occupancy is 72.7% and AUC-ROC value is 0.82. According to Eq. 22, bACC in this study is 70%. The results suggest that although variance there is no significant differences or outlier are observed cross days for MAE and RMSE. Results of Type C occupancy indicate that the classifiers are more suitable for partial occupancy since the overall accuracy of medium occupancy level is much higher than the other levels.

Fig. 9. The predicted occupancy (a) Type A Occupancy (top), (b) Type B Occupancy (middle), (c) Type C Occupancy (bottom).

Table 5. Averaged errors and accuracy of three occupancy types

	Type A		Type B		Type C				
	Occupancy		Occupancy		Occupancy				
	<i>MA</i>	<i>RMS</i>	<i>MA</i>	<i>RMS</i>	<i>Zer</i>	<i>Low</i>	<i>Mediu</i>	<i>Hig</i>	<i>Total</i>
	<i>E</i>	<i>E</i>	<i>E</i>	<i>E</i>	<i>o</i>		<i>m</i>	<i>h</i>	
Day 1	2.69	3.38	1.73	2.22	0	85.7%	96.8%	0	76.2%
Day 2	2.15	2.89	1.93	2.45	-	35.5%	93.3%	0	77.8%
Day 3	2.16	3.05	1.48	2.21	0	76.3%	64.2%	-	70.6%
Day 4	3.75	4.40	3.56	4.12	-	50.0%	98.2%	0	50.7%
Day 5	1.88	2.40	1.85	2.14	0	63.6%	95.0%	-	86.5%
Day 6	3.23	4.01	3.12	3.77	-	36.8%	100.0%	0	56.3%
Day 7	1.91	2.30	3.13	3.71	-	61.1%	95.4%	-	90.4%
Total	2.54	3.30	2.41	3.06	0	60.0%	95.0%	0	72.7%

619

Table 6. The normalized confusion matrix of Type C occupancy results

	Zero	Low	Medium	High
Zero	0.00	1.00	0.00	0.00
Low	0.00	0.60	0.40	0.00
Medium	0.00	0.05	0.95	0.00
High	0.00	0.00	1.00	0.00

620

6215.3 Energy performance and analysis of the occupancy linked e-CPSs

622To access the potential energy savings using occupancy-linked e-CPSs, this study
 623simulated three scenarios of energy consumption for both the proposed model and
 624traditional e-CPSs. The baseline model (BM1) is the traditional e-CPSs that use
 625ASHRAE recommended occupancy (ASHRAE Standard 62.1-2013) schedule for
 626energy management and facility operation. The occupancy-linked e-CPSs model
 627(OLEM) implemented the three types of predicated occupancy as modeling input and
 628updated the system operation with new optimized setting parameters. Another
 629benchmarking model (BM2) implemented the actual occupancy information (captured
 630by cameras) as the inputs for the occupancy linked e-CPSs model to estimate its
 631energy saving potential and track the errors.

632Figure 10 and 11 shows the simulated cooling load with different occupancy types. In
 633the simulation, the thermostat HVAC terminals in BM1 were set to default
 634temperature and the mechanical operation was mainly affected by the weather
 635condition. From both figures, it can be seen that the energy consumption for the
 636cooling load in BM1 is significantly higher than BM2 and OLEM, which included
 637occupancy as inputs for load estimation. In addition, all three occupancy types are
 638similar to each other and Type C seems closer to the actual demand.

Fig. 10. Simulated daily cooling load based on three occupancy types.

Fig. 11. Simulated hourly cooling load based on three occupancy types.

Another energy consumption component for the HVAC system is the fresh air amount. The mechanical drives and fans consume a large amount of energy when the air handling units deliver the outdoor air into indoor spaces. The physical building deploys on/off the system with a fixed flow rate about 1440 m³/h. However, according to ASHRAE standard, the flow amount is obviously insufficient given the number of occupants in the experiment office. Figure 12 and 13 show the simulated minimum outdoor air flow rate and amount. Both figures suggest that the outdoor air amount in BM1 is far less than the demand according to the number of occupants. Type A occupancy performs the worst among all three types, this could be caused by the tracking errors result from data fluctuation.

654

655 Fig. 12. Simulated daily outdoor air amount based on three occupancy types.

Fig. 13. Estimated hourly outdoor air flow rate based on three occupancy types.

Then the total energy consumption of air conditioning and ventilation was aggregated and compared for all three models. BM1 was used as the reference and potential savings are computed as a percentage less than the energy consumption of B1. Table 6 summarizes the aggregated results. The averaged savings vary from 24.71% to 26.31% and all three occupancy types have a close performance. The results indicate that the fixed flow rate of conditioned air could easily result in over-cooling and energy wastes.

Table 6. Energy saving potentials for different occupancy types (compared with BM1)

	Type A		Type B		Type C	
	BM2 vs. BM1	OLEM vs. BM1	BM2 vs. BM1	OLEM vs. BM1	BM2 vs. BM1	OLEM vs. BM1
Day 1	33.46%	39.27%	34.27%	34.61%	34.65%	36.17%

Day 2	43.16%	38.07%	41.50%	36.08%	41.39%	40.77%
Day 3	21.54%	8.55%	16.38%	22.37%	18.99%	18.12%
Day 4	22.60%	14.61%	23.79%	28.01%	19.06%	20.08%
Day 5	3.86%	16.47%	4.69%	8.94%	6.12%	10.90%
Day 6	26.00%	24.62%	26.62%	23.87%	24.68%	24.83%
Day 7	31.26%	29.52%	31.81%	29.14%	31.31%	31.87%
Total	26.29%	24.71%	25.91%	26.31%	25.47%	26.37%

667

6686. DISCUSSION

669With the rapid technological development of ICT and IoT, an increasing number of
670buildings are encouraged to install various sensors and sensor networks to facility
671smarter management and control. Combining these technologies, e-CPSs allow new
672advances such as data analytics, artificial intelligence to be utilized in optimizing
673building control for higher energy efficiency and human-centric services. This study
674extended conventional e-CPSs by introducing occupancy detection and prediction
675components so that the occupancy information can be included for better service and
676less energy waste. The detected occupancy can be used as dynamic information
677exchange between the physical building and cyber models so that the optimization
678boundary conditions can be updated timely. For existing buildings, since all building
679features have been determined, the major uncertainties in e-CPSs arise from weather
680conditions and occupancy variations. The occupancy-linked e-CPSs mitigated the
681occupant-related uncertainty by incorporating a reliable occupancy prediction
682mechanism. Accurate occupancy information allows building management system to
683turn off certain functions when occupants are absent to avoid waste. The validation
684experiment results suggest that the accuracy can reach 72.7% and reveal that when
685incorporating occupancy information, the e-CPSs is capable of implementing the
686demand-based facility management to promote building energy efficiency. For
687example, the validation experiment suggests 24% of energy saving potential and
68833.3% air amount compensation. With the proposed ensemble algorithm, e-CPSs can
689receive occupancy information with acceptable accuracy, especially when the
690occupancy was categorized. Also, it can be observed from the experiment that three
691types of occupancy information show no significant differences in the simulation and
692Type C occupancy is more suitable for practical implementation in e-CPSs control as

693it requires less computational power and is easier for practical deployment.

694One challenge in conventional e-CPSs is that many predefined human-centric control
695approaches conflict with the occupants' actual preferences and activities since
696occupancy is stochastic and changeable in different buildings. This study contributes
697to the research gap by proposing a theoretical framework for occupancy-linked e-
698CPSs model and a feasible ensemble algorithm to predict occupancy with proper data
699sources. As WiFi networks become a premise of all cloud-based platforms and cyber
700models, it is naturally compatible with e-CPSs without additional cost. The highly
701accessible WiFi technologies in modern buildings can help boost applicability of
702proposed OLEM. For existing buildings with Wi-Fi installation, through deploying
703fast and reliable artificial intelligence technologies, such as the proposed ensemble
704algorithms, the occupancy becomes accessible to e-CPSs and creates a significant
705synergy among all cyber models. In addition, with the cumulation of the detected and
706predicated occupancy, designers also can rethink and refine the building space design
707and mechanical system selection for new buildings. For example, it is possible to
708integrate WiFi-based occupancy-driven lighting control for smart buildings [63] and
709include the lighting system into the e-CPSs platform. Additionally, the unprecedented
710increase of human activities in buildings, infrastructures, and vehicles generates a
711complex and interdependent system in modern cities. The advances in the world wide
712web technologies allow an efficient information sharing through cloud among e-
713CPSs. Under such a context, the occupancy studies for e-CPSs can also be extended to
714urban scale. For example, the occupancy information can be associated with the
715human mobility between buildings and can be used for inter-building energy demand
716assessment. The information gathered from occupancy linked e-CPSs can be used for
717regional electricity grid design and human-centric urban planning. Another inspiring
718research direction is to integrate OLEM with smart grids for dynamically computed
719demand at the building side to achieve smart grids or microgrids optimization. In
720addition, such implementation also requires new technologies to protect the
721occupants' security and privacy during occupancy detection [64].

722This study also yields to limitations, which can be resolved in future studies. Firstly,
723the validation experiment constraint to small space (an office room). It is suggested to
724study a larger building space with multiple rooms so that the impact of indoor

commutes can be included. Also, rooms with different functions also have their unique occupancy patterns and mechanical system selection. Secondly, the energy consumption in this study mainly results from cooling load and ventilation due to the tropical climate condition and short experiment period. However, there are various energy consuming services systems in buildings, such as lighting, security, heating, and etc., which are also closely associated with human behaviors and inter-dependent with each other.

7. CONCLUSION

This study proposed a theoretical framework for implementing occupancy information as dynamic links for e-CPSs. The framework adopted WiFi Probe technology and ensemble classifiers to interpret WiFi connections as reliable and usable occupancy information. Three occupancy types (Type A, B, and C) have been compared in a validation experiment to examine the accuracy and feasibility of the proposed occupancy-linked e-CPSs. After a validation experiment, the proposed model can accurately report occupant counts for system energy management. The AdaBoost method and type C occupancy report the highest detection accuracy of 72.7%. Type A occupancy has an absolute error and root mean squared error of 2.54 and 3.30, and both values for type B occupancy are 2.41 and 3.06, respectively. The energy simulation reports 24.7%, 26.4%, and 26.3% energy saving potentials by implementing these three types of occupancy information in e-CPSs, respectively.

This study contributes to the development of high-precision and large-scale human-centric services in e-CPSs. For future studies, it is suggested to investigate large-scale and more complicated system coordination and incorporate more information to bridge the energy system and CPSs, such as environmental conditions and occupants' feedback. In addition, the concept of occupancy-lined e-CPSs can be transplanted to smart grid management to optimize power supply across multiple buildings.

ACKNOWLEDGMENT

This work was financially supported by the Hong Kong General Research Fund

755(GRF) – Early Career Scheme, #21204816, and the National Natural Science
756Foundation of China (NSFC), #51508487. Any opinions, findings, conclusions, or
757recommendations expressed in this paper are those of the authors and do not
758necessarily reflect the views of GRF and NSFC.

759

760 REFERENCES

- 761[1] U.S. Energy Information Administration 2017.
762 https://www.eia.gov/energyexplained/index.cfm?page=us_energy_use.
- 763[2] Rajkumar R, Lee ILI, Sha LSL, Stankovic J. Cyber-physical systems: The next
764 computing revolution. Design Automation Conference (DAC), 2010 47th
765 ACM/IEEE 2010:0–5. doi:10.1145/1837274.1837461.
- 766[3] Gupta SKS, Mukherjee T, Varsamopoulos G, Banerjee A. Research directions
767 in energy-sustainable cyber-physical systems. Sustainable Computing:
768 Informatics and Systems 2011;1:57–74. doi:10.1016/j.suscom.2010.10.003.
- 769[4] Peng Rong, Pedram M. Power-aware scheduling and dynamic voltage setting
770 for tasks running on a hard real-time system. Asia and South Pacific
771 Conference on Design Automation, 2006., IEEE; 2006, p. 473–8.
772 doi:10.1109/ASPDAC.2006.1594730.
- 773[5] Dobson I. Energy Cyber-Physical Systems : Research Challenges and
774 Opportunities 2016.
- 775[6] Szilagyi I, Wira P. An Intelligent System for Smart Buildings using Machine
776 Learning and Semantic Technologies : A Hybrid Data-Knowledge Approach
777 2018:20–5.
- 778[7] Kleissl J, Agarwal Y. Cyber-physical energy systems: focus on smart buildings.
779 Proceedings of the 47th Design Automation Conference on - DAC '10
780 2010:749. doi:10.1145/1837274.1837464.
- 781[8] Balaji B, Abdullah M, Faruque A, Dutt N, Gupta R, Agarwal Y. Models ,
782 Abstractions , and Architectures : The Missing Links in Cyber-Physical
783 Systems n.d.
- 784[9] Zhao P, Simoes MG, Suryanarayanan S. A conceptual scheme for cyber-
785 physical systems based energy management in building structures. 2010 9th
786 IEEE/IAS International Conference on Industry Applications - INDUSCON
787 2010 2010:1–6. doi:10.1109/INDUSCON.2010.5739891.
- 788[10] Paridari K, El-Din Mady A, La Porta S, Chabukswar R, Blanco J, Teixeira A, et
789 al. Cyber-Physical-Security Framework for Building Energy Management
790 System 2016. doi:10.1109/ICCPS.2016.7479072.
- 791[11] Francisco A, Truong H, Khosrowpour A, Taylor JE, Mohammadi N. Occupant
792 perceptions of building information model-based energy visualizations in eco-
793 feedback systems. Applied Energy 2018;221:220–8.
794 doi:10.1016/j.apenergy.2018.03.132.
- 795[12] Hong T, Chou S., Bong T. Building simulation: an overview of developments
796 and information sources. Building and Environment 2000;35:347–61.

- doi:10.1016/S0360-1323(99)00023-2.
- 798[13] Modelica and the Modelica Association — Modelica Association n.d.
799 <https://www.modelica.org/>.
- 800[14] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO,
801 et al. EnergyPlus: creating a new-generation building energy simulation
802 program. *Energy and Buildings* 2001;33:319–31. doi:10.1016/S0378-
803 7788(00)00114-6.
- 804[15] Delwati M, Merema B, Breesch H, Helsen L, Sourbron M. Impact of demand
805 controlled ventilation on system performance and energy use. *Energy and*
806 *Buildings* 2018;174:111–23. doi:10.1016/j.enbuild.2018.06.015.
- 807[16] Hong T, Sun K, Zhang R, Hinokuma R, Kasahara S, Yura Y. Development and
808 validation of a new variable refrigerant flow system model in EnergyPlus.
809 *Energy and Buildings* 2016;117:399–411.
810 doi:10.1016/J.ENBUILD.2015.09.023.
- 811[17] Stamatescu G, Stamatescu I, Arghira N, Calofir V, Fagarasan I. Building
812 Cyber-Physical Energy Systems 2016.
- 813[18] Behl M, Jain A, Mangharam R. Data-Driven Modeling, Control and Tools for
814 Cyber-Physical Energy Systems. 2016 ACM/IEEE 7th International
815 Conference on Cyber-Physical Systems, ICCPS 2016 - Proceedings 2016.
816 doi:10.1109/ICCPS.2016.7479093.
- 817[19] Ferreira PM, Ruano AE, Silva S, Conceição EZE. Neural networks based
818 predictive control for thermal comfort and energy savings in public buildings.
819 *Energy and Buildings* 2012;55:238–51. doi:10.1016/J.ENBUILD.2012.08.002.
- 820[20] Costanzo GT, Iacovella S, Ruelens F, Leurs T, Claessens BJ. Experimental
821 analysis of data-driven control for a building heating system. *Sustainable*
822 *Energy, Grids and Networks* 2016;6:81–90.
823 doi:10.1016/J.SEGAN.2016.02.002.
- 824[21] Belafi Z, Hong T, Reith A. Smart building management vs. intuitive human
825 control—Lessons learnt from an office building in Hungary. *Building*
826 *Simulation* 2017;10:811–28. doi:10.1007/s12273-017-0361-4.
- 827[22] Wang Y, Shao L. Understanding occupancy pattern and improving building
828 energy efficiency through Wi-Fi based indoor positioning. *Building and*
829 *Environment* 2017;114:106–17. doi:10.1016/j.buildenv.2016.12.015.
- 830[23] Menezes AC, Cripps A, Bouchlaghem D, Buswell R. Predicted vs. actual
831 energy performance of non-domestic buildings: Using post-occupancy
832 evaluation data to reduce the performance gap. *Applied Energy* 2012;97:355–
833 64. doi:10.1016/j.apenergy.2011.11.075.
- 834[24] Liang X, Hong T, Shen GQ. Improving the accuracy of energy baseline models

- 835 for commercial buildings with occupancy data. *Applied Energy* 2016;179:247–
836 60. doi:10.1016/j.apenergy.2016.06.141.
- 837[25] Wang W, Chen J, Huang G, Lu Y. Energy efficient HVAC control for an IPS-
838 enabled large space in commercial buildings through dynamic spatial
839 occupancy distribution. *Applied Energy* 2017.
840 doi:10.1016/J.APENERGY.2017.06.060.
- 841[26] Barbeito I, Zaragoza S, Tarrío-Saavedra J, Naya S. Assessing thermal comfort
842 and energy efficiency in buildings by statistical quality control for
843 autocorrelated data. *Applied Energy* 2017;190:1–17.
844 doi:10.1016/J.APENERGY.2016.12.100.
- 845[27] Zhang S, Cheng Y, Fang Z, Huan C, Lin Z. Optimization of room air
846 temperature in stratum-ventilated rooms for both thermal comfort and energy
847 saving. *Applied Energy* 2017;204:420–31.
848 doi:10.1016/J.APENERGY.2017.07.064.
- 849[28] Korkas CD, Baldi S, Michailidis I, Kosmatopoulos EB. Occupancy-based
850 demand response and thermal comfort optimization in microgrids with
851 renewable energy sources and energy storage. *Applied Energy* 2016;163:93–
852 104. doi:10.1016/j.apenergy.2015.10.140.
- 853[29] Chen X, Wang Q, Srebric J. Occupant feedback based model predictive control
854 for thermal comfort and energy optimization: A chamber experimental
855 evaluation. *Applied Energy* 2016;164:341–51.
856 doi:10.1016/j.apenergy.2015.11.065.
- 857[30] Lim G-H, Keumala N, Ghafar NA. Energy saving potential and visual comfort
858 of task light usage for offices in Malaysia. *Energy and Buildings*
859 2017;147:166–75. doi:10.1016/J.ENBUILD.2017.05.004.
- 860[31] Shen E, Hu J, Patel M. Energy and visual comfort analysis of lighting and
861 daylight control strategies. *Building and Environment* 2014;78:155–70.
862 doi:10.1016/J.BUILDENV.2014.04.028.
- 863[32] Siano P. Demand response and smart grids—A survey. *Renewable and*
864 *Sustainable Energy Reviews* 2014;30:461–78. doi:10.1016/j.rser.2013.10.022.
- 865[33] Strbac G. Demand side management: Benefits and challenges. *Energy Policy*
866 2008;36:4419–26. doi:10.1016/j.enpol.2008.09.030.
- 867[34] Nguyen A-T, Reiter S, Rigo P. A review on simulation-based optimization
868 methods applied to building performance analysis. *Applied Energy*
869 2014;113:1043–58. doi:10.1016/j.apenergy.2013.08.061.
- 870[35] Díaz JA, Jiménez MJ. Experimental assessment of room occupancy patterns in
871 an office building. Comparison of different approaches based on CO₂
872 concentrations and computer power consumption. *Applied Energy*

- 873 2017;199:121–41. doi:10.1016/j.apenergy.2017.04.082.
- 874[36] Oldewurtel F, Sturzenegger D, Morari M. Importance of occupancy
875 information for building climate control. *Applied Energy* 2013;101:521–32.
876 doi:10.1016/j.apenergy.2012.06.014.
- 877[37] Hong T, Yan D, D ’oca S, Chen C-F. Ten questions concerning occupant
878 behavior in buildings: The big picture. *Building and Environment*
879 2017;114:518–30. doi:10.1016/j.buildenv.2016.12.006.
- 880[38] Yan D, Hong T, Dong B, Mahdavi A, D’Oca S, Gaetani I, et al. IEA EBC
881 Annex 66: Definition and simulation of occupant behavior in buildings. *Energy*
882 and Buildings 2017;156:258–70. doi:10.1016/J.ENBUILD.2017.09.084.
- 883[39] Hong T, Taylor-Lange SC, D ’oca S, Yan D, Corngati SP, D’Oca S, et al.
884 Advances in research and applications of energy-related occupant behavior in
885 buildings. *Energy and Buildings* 2016;116:694–702.
886 doi:10.1016/j.enbuild.2015.11.052.
- 887[40] Yan D, O’Brien W, Hong T, Feng X, Burak Gunay H, Tahmasebi F, et al.
888 Occupant behavior modeling for building performance simulation: Current
889 state and future challenges. *Energy and Buildings* 2015;107:264–78.
890 doi:10.1016/j.enbuild.2015.08.032.
- 891[41] Hong T, Yan D, D ’oca S, Chen C-F. Ten questions concerning occupant
892 behavior in buildings: The big picture. *Building and Environment*
893 2017;114:518–30. doi:10.1016/j.buildenv.2016.12.006.
- 894[42] Kim Y-S, Heidarinejad M, Dahlhausen M, Srebric J. Building energy model
895 calibration with schedules derived from electricity use data. *Applied Energy*
896 2017;190:997–1007. doi:10.1016/j.apenergy.2016.12.167.
- 897[43] Yang J, Santamouris M, Lee SE, Deb C. Energy performance model
898 development and occupancy number identification of institutional buildings.
899 *Energy and Buildings* 2016;123:192–204.
900 doi:10.1016/J.ENBUILD.2015.12.018.
- 901[44] Yang Z, Becerik-Gerber B. The coupled effects of personalized occupancy
902 profile based HVAC schedules and room reassignment on building energy use.
903 *Energy and Buildings* 2014;78:113–22. doi:10.1016/j.enbuild.2014.04.002.
- 904[45] Pisello AL, Asdrubali F. Human-based energy retrofits in residential buildings:
905 A cost-effective alternative to traditional physical strategies. *Applied Energy*
906 2014;133:224–35. doi:10.1016/j.apenergy.2014.07.049.
- 907[46] Chen Y, Liang X, Hong T, Luo X. Simulation and visualization of energy-
908 related occupant behavior in office buildings. *Building Simulation*
909 2017;10:785–98. doi:10.1007/s12273-017-0355-2.
- 910[47] Jin M, Bekiaris-Liberis N, Weekly K, Spanos CJ, Bayen AM. Occupancy

- 911 Detection via Environmental Sensing. IEEE Transactions on Automation
 912 Science and Engineering 2018;15:443–55. doi:10.1109/TASE.2016.2619720.
- 913[48] Jin M, Jia R, Spanos CJ. Virtual Occupancy Sensing: Using Smart Meters to
 914 Indicate Your Presence. IEEE Transactions on Mobile Computing
 915 2017;16:3264–77. doi:10.1109/TMC.2017.2684806.
- 916[49] Weekly K, Zou H, Xie L, Jia QS, Bayen AM. Indoor occupant positioning
 917 system using active rfid deployment and particle filters. Proceedings - IEEE
 918 International Conference on Distributed Computing in Sensor Systems,
 919 DCOSS 2014 2014:35–42. doi:10.1109/DCOSS.2014.53.
- 920[50] Wang W, Chen J, Lu Y, Wei H-H. Energy conservation through flexible HVAC
 921 management in large spaces: An IPS-based demand-driven control (IDC)
 922 system. Automation in Construction 2017;83:91–107.
 923 doi:10.1016/J.AUTCON.2017.08.021.
- 924[51] Wang W, Lin Z, Chen J. Promoting Energy Efficiency of HVAC Operation in
 925 Large Office Spaces with a Wi-Fi Probe enabled Markov Time Window
 926 Occupancy Detection Approach. World Engineers Summit – Applied Energy
 927 Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference,
 928 WES-CUE 2017, Singapore: 2017.
- 929[52] Chen J, Ahn C. Assessing occupants’ energy load variation through existing
 930 wireless network infrastructure in commercial and educational buildings.
 931 Energy and Buildings 2014;82:540–9. doi:10.1016/j.enbuild.2014.07.053.
- 932[53] Balaji B, Xu J, Nwokafor A, Gupta R, Agarwal Y. Sentinel: Occupancy Based
 933 HVAC Actuation using Existing WiFi Infrastructure within Commercial
 934 Buildings. Conference: Proceedings of the 11th ACM Conference on
 935 Embedded Networked Sensor Systems, 2013. doi:10.1145/2517351.2517370.
- 936[54] Jin M, Jia R, Kang Z, Konstantakopoulos IC, Spanos CJ. PresenceSense: zero-
 937 training algorithm for individual presence detection based on power
 938 monitoring. Proceedings of the 1st ACM Conference on Embedded Systems for
 939 Energy-Efficient Buildings - BuildSys ’14 2014:1–10.
 940 doi:10.1145/2674061.2674073.
- 941[55] Zou H, Jiang H, Yang J, Xie L, Spanos CJ. Non-intrusive occupancy sensing in
 942 commercial buildings. Energy and Buildings 2017;154:633–43.
 943 doi:10.1016/j.enbuild.2017.08.045.
- 944[56] Zou H, Zhou Y, Yang J, Spanos CJ. Device-free occupancy detection and
 945 crowd counting in smart buildings with WiFi-enabled IoT. Energy and
 946 Buildings 2018;174:309–22. doi:10.1016/j.enbuild.2018.06.040.
- 947[57] Zhao Y, Zeiler W, Boxem G, Labeodan T. Virtual occupancy sensors for real-
 948 time occupancy information in buildings. Building and Environment

949 2015;93:9–20. doi:10.1016/j.buildenv.2015.06.019.
 950[58] Wang W, Chen J, Hong T, Zhu N. Occupancy prediction through Markov based
 951 feedback recurrent neural network (M-FRNN) algorithm with WiFi probe
 952 technology. *Building and Environment* 2018;138:160–70.
 953 doi:10.1016/J.BUILDENV.2018.04.034.
 954[59] Wang W, Chen J, Song X. Modeling and predicting occupancy profile in office
 955 space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference
 956 approach. *Building and Environment* 2017;124:130–42.
 957 doi:10.1016/J.BUILDENV.2017.08.003.
 958[60] ANSI/ASHRAE Standard 62.1. 2013:Ventilation for Acceptable Indoor Air
 959 Quality. 2013.
 960[61] Eppy Tutorial — eppy 0.5.44 documentation n.d.
 961 https://pythonhosted.org/eppy/Main_Tutorial.html.
 962[62] Ang BW, Wang H, Ma X. Climatic influence on electricity consumption: The
 963 case of Singapore and Hong Kong. *Energy* 2017;127:534–43.
 964 doi:10.1016/j.energy.2017.04.005.
 965[63] Zou H, Zhou Y, Jiang H, Chien S-C, Xie L, Spanos CJ. WinLight: A WiFi-
 966 based occupancy-driven lighting control system for smart building. *Energy and*
 967 *Buildings* 2018;158:924–38. doi:10.1016/J.ENBUILD.2017.09.001.
 968[64] Boroojeni KG, Amini MH, Iyengar SS. *Smart Grids: Security and Privacy*
 969 *Issues*. Cham: Springer International Publishing; 2017. doi:10.1007/978-3-319-
 970 45050-6.
 971